Statistical Inference for Stochastic Volatility Models

Md. Nazmul Ahsan

Doctor of Philosophy

Department of Economics
McGill University
Montreal, Quebec, Canada

November 2020

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Doctor of Philosophy

Dedication

This thesis is dedicated to my parents, who raised me and taught me to think independently. They always respected and supported every decision I made. It is also dedicated to my wife Atia and my daughter Zara. This thesis could not be finished without their love and emotional support.

Contributions

This thesis contains collaborative works with my supervisor Jean-Marie Dufour (Professor of Economics at McGill University) as well as my individual research work.

The first essay is published as an article in *Advances in Econometrics* (2019), Volume 40A, 157-201, joint with Prof. Dufour. I initiated this research project under the direction of Prof. Dufour. The literature review, theoretical and empirical sections, simulations and the writing are my own work. A part of theoretical results and a proof in this chapter are the contributions of Prof. Dufour. He also contributed to revising the manuscript and discussions of the model.

The second essay is also based on joint work with Prof. Dufour. He initiated and directed the whole research project. The literature review, theoretical and empirical sections, simulations, applications and the writing are my own work. Prof. Dufour contributed to the writing and editing of the paper.

The third chapter is my independent research paper. I contributed to coming up with the research idea, working with the data, estimating the model and writing the manuscript.

All three studies are original scholarship and distinct contributions to Financial Econometrics and Volatility Modelling.

Acknowledgments

It is an absolute pleasure to thank all the people who made this thesis possible.

First and foremost, I would like to express my sincere gratitude to my supervisor, Professor Jean-Marie Dufour, for his continuous support of my Ph.D. study and research. His motivation, enthusiasm and immense knowledge and research experience were invaluable in the development of my research. His dedication to research and students has taught me a lot about both work and life.

I would also like to thank Professors John Galbraith and Victoria Zinde-Walsh for their encouragement and support, for providing countless valuable suggestions, and for many valuable conversations that clarified, for me, a number of issues related to this thesis.

Special thanks to Professors Christian Genest and Jean-Michel Zakoian for serving as my thesis examiners and for their valuable comments and suggestions.

I am also grateful for the assistance provided by other faculty and staff members of the McGill Economics department including Saraswata Chaudhuri, Russell Davidson, Kenneth MacKenzie, Markus Poschke, Ngo Van Long, Lisa Stevenson and Judy Dear. I benefited greatly from discussions with current and former fellow students at McGill — Byunguk Kang, Purevdorj Tuvaandorj, Masaya Takano, Tianyu He, Mathieu Blais, and Jinjing Liu. Also I thank my brother (Razib Ahsan) and sisters (Aurin and Nahida) for their deepest love and continuous support throughout my life.

Acknowledgements of chapters

For Chapter 2, the authors thank John Galbraith, Lynda Khalaf, Hashem Pesaran, Giovanni Urga, Pascale Valéry, Victoria Zinde-Walsh, an anonymous referee, the editor Justin Tobias, as well as the CIREQ-McGill Seminar participants for useful comments.

For Chapter 3, the authors thank Torben Andersen, Manabu Asai, Richard Baillie, Russell Davidson, Firmin Doko Tchatoka, John Galbraith, René Garcia, Byunguk Kang, Lynda Khalaf, Jinjing Liu, Richard Luger, Franck Moraux, Denis Pelletier, Hashem Pesaran, Eric Renault, Lars Stentoft, Masaya Takano, Purevdorj Tuvaandorj, Aman Ullah, Pascale Valéry, Victoria Zinde-Walsh, as well as the seminar and conference participants for useful comments and constructive discussions. Earlier versions of this paper were presented at 11^{th} French Econometrics Conference (Marseille, 2019), NBER-NSF Time Series Conference (The Chinese University of Hong Kong, 2019), Workshop on Econometrics and Statistics (City University of Hong Kong, 2019), Annual Conference of IAAE (Cyprus, 2019), 36^{th} International Conference of the French Finance Association, Université Laval (Quebec City, 2019), BU Pi-day Econometrics Conference (Boston, 2019), Canadian Econometrics Study Group Meeting (Ottawa, 2018), International conference on Time Series and Forecasting (Granada, 2018), Annual Conference of IAAE (Montreal, 2018), 52^{nd} Annual Conference of the CEA (Montreal, 2018), CIREQ Econometrics Conference (Recent Advances in the Method of Moments, Montreal, 2018), University of Bergamo (Bergamo, 2018), University of Southern California (Los Angeles, 2018), Camp Econometrics (New York, 2018), York University (Toronto, 2018), CIREQ-McGill Lunch Seminar (Montreal, 2016, 2017), 11^{th} World Congress of the Econometric Society (Montreal, 2015), and 11^{th} CIREQ PhD Students' Conference (Montreal, 2015).

For Chapter 4, I thank Bertille Antoine, Marine Carrasco, Firmin Doko Tchatoka, Peter Fuleky, Silvia Gonçalves, Joann Jasiak, Sayar Karmakar, Lynda Khalaf, Jans-Peter Kreiss, Alex

Maynard, Dan Nordman, John Galbraith, Ke-Li Xu, Victoria Zinde-Walsh for useful comments and constructive discussions. Earlier versions of this paper were presented at Bernoulli-IMS Symposium 2020, Joint Statistical Meetings 2020, Canadian Econometrics Study Group Meeting (Montreal, 2019), 53rd Annual Conference of the CEA (Banff, 2019), CIREQ Montreal Econometrics Conference on Recent Advances on Bootstrap Methods (Montreal, 2019), 14th CIREQ Ph.D. Students' Conference (Montreal, 2018), CIREQ Econometrics Conference (Montreal, 2016), 12th CIREQ Ph.D. Students' Conference (Montreal, 2016), CIREQ-McGill Lunch Seminar (2017, 2016).

Abstract

Although stochastic volatility (SV) models have many appealing features, estimation and inference on SV models are challenging problems due to the inherent difficulty of evaluating the likelihood function. The existing methods are either computationally costly and/or inefficient. This thesis studies and contributes to the SV literature from the estimation, inference, and volatility forecasting viewpoints. It consists of three essays, which include both theoretical and empirical contributions. On the whole, the thesis develops easily applicable statistical methods for stochastic volatility models.

The first essay proposes computationally simple moment-based estimators for the first-order SV model. In addition to confirming the enormous computational advantage of the proposed estimators, the results show that the proposed estimators match (or exceed) alternative estimators in terms of precision – including Bayesian estimators proposed in this context, which have the best performance among alternative estimators. Using this simple estimator, we study three crucial test problems (no persistence, no latent specification of volatility, and no stochastic volatility hypothesis), and evaluate these null hypotheses in three ways: asymptotic critical values, a parametric bootstrap procedure, and a maximized Monte Carlo procedure. The proposed methods are applied to daily observations on the returns for three major stock prices [Coca-Cola, Walmart, Ford], and the Standard and Poor's Composite Price Index. The results show the presence of stochastic volatility with strong persistence.

The second essay studies the problem of estimating higher-order stochastic volatility [SV(p)] models. The estimation of SV(p) models is even more challenging and rarely considered in the literature. We propose several estimators for higher-order stochastic volatility models. Among these, the simple winsorized ARMA-based estimator is uniformly superior in terms of bias and RMSE to other estimators, including the Bayesian MCMC estimator. The

proposed estimators are applied to stock return data, and the usefulness of the proposed estimators is assessed in two ways. First, using daily returns on the S&P 500 index from 1928 to 2016, we find that higher-order SV models – in particular an SV(3) model – are preferable to an SV(1), from the viewpoints of model fit and both asymptotic and finite-sample tests. Second, using different volatility proxies (squared return and realized volatility), we find that higher-order SV models are preferable for out-of-sample volatility forecasting, whether a high volatility period (such as financial crisis) is included in the estimation sample or the forecasted sample. Our results highlight the usefulness of higher-order SV models for volatility forecasting.

In the final essay, we introduce a novel class of generalized stochastic volatility (GSV) models which utilize high-frequency (HF) information (realized volatility (RV) measures). GSV models can accommodate nonstationary volatility process, various distributional assumptions, and exogenous regressors in the latent volatility equation. Instrumental variable methods are employed to provide a unified framework for the analysis (estimation and inference) of GSV models. We consider the parameter inference problem in GSV models with nonstationary volatility and develop identification-robust methods for joint hypotheses involving the volatility persistence parameter and the autocorrelation parameter of the composite error (or the noise ratio). For distributional theory, three different sets of assumptions are considered. In simulations, the proposed tests outperform the usual asymptotic test regarding size and exhibit excellent power. We apply our inference methods to IBM price and option data and identify several empirical relationships.

Résumé

Les modèles de volatilité stochastique (SV) ont plusieurs caractéristiques désirables, mais l'estimation et l'inférence pour ces modèles sont des défis de taille à cause de la difficulté d'évaluer la fonction de vraisemblance en présence d'une variable latente. Les méthodes existantes sont intensives en calcul ou inefficaces. Cette thèse contribue à l'application de tels modèles au niveau de l'estimation, de l'inférence et de la prévision. Elle consiste en trois essais, qui constituent des contributions tant théoriques qu'empiriques au développement de nouveaux outils statistiques facilement applicables.

Le premier essai propose une méthode de moments qui facilite l'estimation du modèle SV de premier ordre. En plus de confirmer un énorme avantage de calcul, les résultats démontrent que l'estimateur proposé surpasse en précision les autres estimateurs proposés précédemment – notamment un estimateur de type bayésien qui semble être le meilleur à ce jour. Grâce à cet estimateur simple, nous étudions trois problèmes de test importants dans ce cadre (l'absence de persistance, l'absence de volatilité latente, l'absence de volatilité stochastique). Nous comparons trois manières différentes d'effectuer ces tests: valeurs critiques asymptotiques, bootstrap paramétrique, et test de Monte Carlo maximisé. Les méthodes sont appliquées à des observations quotidiennes sur les rendements de trois compagnies majeures (Coca-Cola, Walmart et Ford), ainsi qu'à l'indice de prix Standard & Poor's. Les résultats démontrent une forte persistance de la volatilité stochastique.

Le deuxième essai se penche sur l'estimation de modèles de volatilité stochastique d'ordre supérieur [SV(p)]. L'estimation de tels modèles comporte des défis additionnels qui font qu'ils ne sont presque jamais utilisés. Nous proposons plusieurs estimateurs faciles à calculer pour ce type de modèle. Nous trouvons qu'une méthode de type ARMA est supérieure en termes de biais et d'erreur quadratique moyenne, incluant un estimateur bayésien basé sur l'algorithme

MCMC. Les estimateurs proposés sont appliqués à des données de rendement de marché (S&P 500). Nous démontrons les avantages du modèle SV(p) et des méthodes d'estimation proposées de deux manières différentes. Premièrement, en utilisant les rendements quotidiens du S&P 500 de 1928 à 2016, ainsi que des tests asymptotiques et exacts, nous trouvons qu'un modèle SV(3) est préférable au modèle SV(1). Deuxièmement, en utilisant des variables « proxy » pour la volatilité (rendement au carré et volatilité réalisée), nous obtenons que les modèles SV(p) sont préférables pour effectuer des prévisions hors échantillon. Ces résultats sont robustes à l'exclusion de la période de crise financière.

L'essai final introduit une nouvelle classe de modèles de volatilité stochastique généralisé (GSV) qui utilise des mesures de volatilité réalisée à haute fréquence. Les modèles GSV peuvent accommoder des processus non stationnaires, des hypothèses distributionnelles variées, et peuvent inclure des variables latentes. Nous proposons des méthodes de variables instrumentales afin d'obtenir un cadre unifié pour l'analyse du modèle GSV. Nous développons aussi une méthode robuste à la non-identification afin de tester une hypothèse jointe sur le paramètre de persistance et l'autocorrélation des termes d'erreurs. Au niveau de la théorie distributionnelle, nous proposons une théorie qui inclut trois ensembles différents d'hypothèses. Une expérience de simulation démontre que les tests proposés améliorent le contrôle du niveau des tests par rapport aux méthodes asymptotiques usuelles et possèdent une bonne puissance. Nous appliquons les méthodes proposées au prix de l'action IBM et à des données d'options. Plusieurs relations empiriques intéressantes émergent de ces résultats.

Contents

1 Introduction **1**

2 A simple efficient moment-based estimator for the stochastic volatility model **5**

2.1 Introduction 6
2.2 Framework 11
2.3 Simple ARMA-type estimators 13
2.4 Stationarity, ergodicity and mixing properties 16
2.5 Asymptotic distributional theory 18
2.6 Hypothesis testing 20
2.6.1 Asymptotic tests 21
2.6.2 Simulation-based finite-sample tests 22
2.6.3 Implicit standard error 27
2.7 Simulation study 28
2.7.1 Estimation 28
2.7.2 Testing 30
2.8 Applications to stock price volatilities 32
2.9 Conclusion 35
2.10 Appendix 36
2.10.1 Proofs 36
2.10.2 Figures 44
2.10.3 Tables 45

3 Simple estimators for higher-order stochastic volatility models and forecasting **52**

3.1 Introduction 53
3.2 Framework 59
3.3 Higher-order stochastic volatility 62

3.4 Stationarity, ergodicity and mixing properties . . . 66
3.5 Simple estimation methods . . . 68
3.5.1 Simple moment-based estimation . . . 68
3.5.2 ARMA-based estimation . . . 71
3.5.3 Restricted estimation . . . 73
3.5.4 ARMA-based winsorized estimation . . . 73
3.5.5 Recursive estimation for SV(p) models . . . 75
3.6 Andersen-Sørensen type GMM estimation . . . 77
3.7 Asymptotic distributional theory . . . 80
3.7.1 Moment-based estimators . . . 80
3.7.2 ARMA-based estimators . . . 82
3.8 Monte Carlo tests . . . 85
3.9 Simulation study . . . 90
3.10 Empirical applications . . . 94
3.10.1 Empirical evidence . . . 94
3.10.2 Volatility forecast performance . . . 96
3.10.2.1 Forecasting squared return . . . 100
3.10.2.2 Forecasting realized volatility . . . 103
3.11 Conclusion . . . 104
3.12 Appendix . . . 106
3.12.1 Proofs . . . 106
3.12.2 Tables . . . 127
3.12.3 Figures . . . 151
3.12.4 Forecasting with SV(p) models . . . 152
3.12.5 Forecasting with GARCH models . . . 154
3.12.6 Heterogenous Autoregressive model of Realized Volatility . . . 156

4 High-frequency instruments and identification-robust inference for stochastic volatility models 158
4.1 Introduction . . . 159
4.2 Framework . . . 165
4.3 Finite-sample procedures . . . 168
4.3.1 Anderson-Rubin-type procedure . . . 170
4.3.2 Anderson-Rubin-type point-optimal procedure (AR^*) . . . 173

4.3.3 Split-sample-type procedure . . . 175
4.3.4 Split-sample-type point-optimal procedure (SS^*) . . . 178
4.3.5 Inference on general transformations . . . 179
4.4 Finite-sample procedures with possibly non-Gaussian errors . . . 181
4.5 Asymptotic distributional theory . . . 186
4.6 Simulation study . . . 190
4.6.1 Test size of asymptotic t-test . . . 192
4.6.2 Performance of the proposed tests . . . 192
4.6.2.1 Test performance under $M1$ with weak instruments . . . 193
4.6.2.2 Test performance under $M2$ with weak instruments . . . 194
4.6.2.3 Test performance under $M3$ with low-frequency instruments . . . 195
4.6.2.4 Test performance under $M4$ with high-frequency instruments . . . 195
4.7 Application to stock prices . . . 198
4.7.1 Data description . . . 198
4.7.2 High-frequency instruments of asset price variability . . . 199
4.7.2.1 Classes of realized measures not robust to jumps . . . 200
4.7.2.2 Classes of realized measures robust to jumps . . . 202
4.7.2.3 Additional HF measures and jump variations . . . 203
4.7.3 Final instrument set, econometric model and test statistics . . . 204
4.7.4 Results . . . 204
4.7.4.1 Strength of IV's . . . 205
4.7.4.2 Projection-based confidence sets . . . 205
4.8 Conclusion . . . 210
4.9 Appendix . . . 212
4.9.1 Proofs . . . 212
4.9.2 Figures . . . 218
4.9.3 Tables . . . 220

5 Conclusion and future work **236**

List of Tables

2.1 Comparison of different winsorized versions of simple ARMA estimator 45
2.2 Comparison of different estimation methods with respect to bias and RMSE for the SV model using simulated data 46
2.3 Comparison of different estimation methods with respect to bias and RMSE for the SV model using simulated data 47
2.4 Comparison of different estimation methods with respect to relative time for the SV model using simulated data 48
2.5 Empirical levels of asymptotic, bootstrap and MMC tests based on LR-type statistic 48
2.6 Empirical powers of asymptotic, bootstrap and MMC tests based on LR-type statistic 48
2.7 Summary Statistics 49
2.8 ARMA based estimates of SV model 49
2.9 ARMA based estimates of SV model with ISE 50
2.10 Exact and asymptotic tests based on ARMA based estimators 50
2.11 Empirical estimates of S&P 500 index 51

3.1 Comparison of different winsorized ARMA-SV estimators 128
3.2 Comparison of different estimation methods with respect to bias and RMSE for an SV(2) model using simulated data. 129
3.3 Comparison of different estimation methods with respect to bias and RMSE for an SV(2) model using simulated data. 130
3.4 Comparison of different estimation methods with respect to bias and RMSE for an SV(2) model using simulated data. 131
3.5 Comparison of different estimation methods with respect to bias and RMSE for an SV(2) model using simulated data. 132

3.6 Comparison of different estimation methods with respect to inadmissible values for an SV(2) model using simulated data. 133
3.7 Comparison of different estimation methods with respect to relative time for an SV(2) model using simulated data 133
3.8 Summary statistics 134
3.9 ARMA-based empirical estimates of SV(p) models. 135
3.10 W-ARMA-SV-OLS estimates with different level of winsorization. 136
3.11 Summary statistics of full sample of experiment - 1 137
3.12 Summary statistics of full sample of experiment - 2 138
3.13 Forecasting squared return: relative MSE and associated MCS p-value during moderate volatility regimes 139
3.14 Forecasting squared return: relative MAE and associated MCS p-value during moderate volatility regimes 140
3.15 Forecasting squared return: relative R2LOG and associated MCS p-value during moderate volatility regimes 141
3.16 Forecasting squared return: relative MSE and associated MCS p-value during high volatility regimes 142
3.17 Forecasting squared return: relative MAE and associated MCS p-value during high volatility regimes 143
3.18 Forecasting squared return: relative R2LOG and associated MCS p-value during high volatility regimes 144
3.19 Forecasting realized volatility: relative MSE and associated MCS p-value during moderate volatility regimes 145
3.20 Forecasting realized volatility: relative MAE and associated MCS p-value during moderate volatility regimes 146
3.21 Forecasting realized volatility: relative R2LOG and associated MCS p-value during moderate volatility regimes 147
3.22 Forecasting realized volatility: relative MSE and associated MCS p-value during high volatility regimes 148
3.23 Forecasting realized volatility: relative MAE and associated MCS p-value during high volatility regimes 149
3.24 Forecasting realized volatility: relative R2LOG and associated MCS p-value during high volatility regimes 150

4.1 Corresponding values of λ_0 220
4.2 Size of asymptotic t-type test for $H_0 : \phi = \phi_0$ (nominal level: 5%) 220
4.3 Size comparison of joint tests ($H_0 : \phi = \phi_0, \rho = \rho_0$) under $M1$ with weak instruments, nominal level: 5% 221
4.4 Power comparison of joint tests ($H_0 : \phi = 0.50, \rho = 0.30$) under $M1$ with weak instruments, nominal level: 5% 222
4.5 Size comparison of joint tests ($H_0 : \phi = \phi_0, \rho = \rho_0$) under $M2$ with weak instruments, nominal level: 5% 223
4.6 Power comparison of joint tests ($H_0 : \phi = 0.50, \rho = 0.30$) under $M2$ with weak instruments, nominal level: 5% 224
4.7 Size and power comparison of joint tests under $M3$ with low-frequency (past lags) instruments, nominal level: 5% 225
4.8 Size and power comparison of joint tests under $M3$ with $\epsilon_t \sim N(0, \pi^2/2)$ and low-frequency (past lags) instruments, nominal level: 5% 226
4.9 Size comparison of tests ($H_0 : \phi_{lf} = \phi_0$) under $M4$ with high-frequency instruments, nominal level: 5% 227
4.10 Power comparison of tests ($H_0 : \phi_{lf} = 0$) under $M4$ with high-frequency instruments, nominal level: 5% 228
4.11 Description of instruments. 229
4.12 First-stage regression with daily instruments 230
4.13 First-stage regression with all IV's 231
4.14 Projection-based 90% confidence intervals for the volatility persistence parameter ϕ - Strong instruments 232
4.15 Projection-based 90% confidence intervals for the volatility persistence parameter ϕ - Weak instruments 233
4.16 Projection-based 90% confidence intervals for the volatility persistence parameter ϕ - Strong instruments (Several lags) 234
4.17 Projection-based 90% confidence intervals for the volatility persistence parameter ϕ - Different combinations of strong instruments 235

List of Figures

2.1 S&P 500: Four W-ARMA estimators of ϕ as a function of the number of lags 44

3.1 S&P 500 (1996-2016): W-ARMA-SV-OLS estimators of volatility persistence parameters (ϕ's) as a function of the number of lags (J) 151

3.2 S&P 500 (2006-2016): W-ARMA-SV-OLS estimators of volatility persistence parameters (ϕ's) as a function of the number of lags (J) 151

3.3 S&P 500 (2006-2010): W-ARMA-SV-OLS estimators of volatility persistence parameters (ϕ's) as a function of the number of lags (J) 152

4.1 IBM: 2009-2013: Precision of different classes of instruments. 218

4.2 IBM: 2009-2013: Precision of different subclasses of HF instruments. 219

Chapter 1

Introduction

Time-varying volatility of asset returns is a widespread feature of financial markets. This property has been known for a long time; early discussions include Mandelbrot (1963) and Fama (1965). Accurate characterization and prediction of the dynamic volatility are essential in many areas of financial decision making, such as asset pricing, portfolio selection, option pricing, and risk management. Volatility modelling was started in the early 1980s, when Rob Engle introduced autoregressive conditional heteroskedastic (ARCH) models. Volatility modeling is still, and will remain for long, one of the most active research topics of financial econometrics.

To deal with time-varying volatility, two main classes of parametric models have been proposed in the literature to estimate and forecast dynamic volatility: (1) GARCH-type models [Engle (1982), Bollerslev (1986)]; (2) stochastic volatility (SV) models [Taylor (1982, 1986)]. The main distinction between GARCH and SV models is that the variance process of the latter has an additional error term which captures the effect of any new information coming to the market, so conditional on the information set $\mathcal{F}_{t-1}$, volatility σ_t^2 is not known in SV models but rather an unobserved random variable. Several reviews of GARCH and SV literature are available; for GARCH, see Bollerslev (2010), and for SV, see Ghysels et al. (1996), Broto and Ruiz (2004), and Shephard (2005). SV models are also common in macroeconomic modelling; see Cogley and Sargent (2005), Primiceri (2005), Benati (2008), Koop et al. (2009), Koop and Korobilis (2013), and Liu and Morley (2014).

SV models may be preferable to GARCH-type models for several reasons. *First*, SV models are discrete-time formulations of continuous-time diffusion processes used in theoretical finance for derivative pricing and portfolio optimization; see Taylor (1994), Shephard and Andersen (2009). *Second*, SV models do not appear to require various *ad hoc* adjustments, like the addition of a random jump component or non-Gaussian innovations. These modifications improve the performance of the standard GARCH, but these are evidently unnecessary for SV models; see Carnero et al. (2004), Chan and Grant (2016). *Third*, SV models often provide more accurate volatility forecasts than GARCH models, indicating that the time-varying volatility is better modelled as a latent stochastic process; see Kim et al. (1998), Yu (2002), Poon and Granger (2003), Koopman et al. (2005). *Finally*, it is easy to derive the probabilistic properties (stationarity, ergodicity and mixing) of SV models than GARCH models; see Davis and Mikosch (2009). In contrast, the stationarity of a GARCH process is difficult to establish; see Nelson (1990), Bougerol and Picard (1992), Lindner (2009).

Although SV models have many appealing features, the estimation and inference are challenging due to the inherent problem of evaluating the likelihood function. The existing methods are either computationally costly and/or inefficient. This thesis studies and contributes to the SV literature from the estimation, inference, and volatility forecasting viewpoints. It consists of three essays, which include both theoretical and empirical contributions. On the whole, the thesis develops easily applicable statistical methods for stochastic volatility models. Chapter 2 proposes computationally simple and efficient estimators for the first-order SV model. The proposed class of estimators is based on a small number of moment equations derived from an ARMA representation associated with the SV model, along with the possibility of using "winsorization" to improve stability and efficiency. We call these ARMA-SV estimators. Closed-form expressions for ARMA-SV estimators are obtained, and no numerical optimization procedure or choice of initial parameter values is required. The asymptotic distributional theory of the proposed estimators is studied. Due to their computational simplicity, the ARMA-SV estimators allow one to make reliable – even exact – simulation-based inference, through the application of Monte Carlo (MC) test or bootstrap methods. We compare them in a simulation experiment with a wide array of alternative estimation methods, in terms of bias, root mean square error and computation time. In addition to confirming the enormous computational advantage of the proposed estimators, the results show that ARMA-SV estimators match (or exceed) alternative estimators in terms of precision – including Bayesian estimators proposed in this context, which have the best performance among alternative estimators. The proposed methods are applied to daily observations on the returns for three major stock prices [Coca-Cola, Walmart, Ford] over the period 1980-2015, and to the Standard and Poor's Composite Price Index over the 2000-2017 period. The results confirm the presence of stochastic volatility with strong persistence.

Chapter 3 studies the problem of estimating higher-order stochastic volatility [SV(p)] models. The estimation of SV(p) models is even more challenging and rarely considered in the literature. In this paper, we propose simple moment-based estimators for such models – in particular ARMA-type estimators – which are both computationally inexpensive and remarkably accurate. The proposed estimators do not require choosing a sampling algorithm, initial parameter values, or an auxiliary model. To reduce the risk of getting inadmissible (nonstationary) solutions, we suggest winsorized versions of the simple ARMA-SV estimators. We also show that a Durbin-Levinson-type updating algorithm can be applied to recursively estimate models of increasing order p. The asymptotic distribution of the estimators is established. We compare by simulation the proposed estimators to a Bayesian MCMC estimator. The results show that the simple winsorized ARMA-SV estimator is uniformly superior to other estimators in terms of bias and root mean square error. The proposed estimators are applied to stock return data, and the usefulness of the proposed estimators is assessed in two ways. First, using the daily return on the S&P 500 index from 1928 to 2016, we find that higher-order SV models – in particular an SV(3) model – are preferable to a SV(1), from the viewpoint model fit and both asymptotic and finite-sample tests. Second, using different volatility proxies (the squared return of S&P 500 index and the realized volatility of S&P 500, FTSE100, NASDAQ100, N225, SSMI20 indices), we conduct two out-of-sample forecast experiments: (1) we forecast a moderately volatile period after the late-2000s financial crisis; (2) we forecast a highly volatile period, *i.e.*, the core financial crisis. We compare the accuracy of volatility forecasts among SV(p) models, GARCH models, and Heterogenous Autoregressive model of Realized Volatility (HAR-RV) models. The results suggest that SV(p) models perform better than other models in most cases. This finding holds even if a high volatility period (such as financial crisis) is included in the estimation sample or the forecasted sample. Formal prediction tests, *i.e.*, model confidence set procedure, also support these inferences. Our findings highlight the usefulness of higher-order SV models for volatility forecasting.

In Chapter 4, we introduce a novel class of generalized stochastic volatility (GSV) models, which utilize high-frequency (HF) information (realized volatility (RV) measures). GSV models can accommodate nonstationary volatility, various distributional assumptions, and exogenous regressors in the latent volatility equation. Instrumental variable methods are employed to provide a unified framework for GSV models' analysis (estimation and inference). We consider the parameter inference problem in GSV models with nonstationary volatility and develop identification-robust methods for joint hypotheses involving the volatility persistence parameter and the autocorrelation parameter of the composite error (or the noise ratio). For inference about the volatility persistence parameter, projection techniques are applied. The proposed tests include Anderson-Rubin-type (AR) tests, a dynamic version of the split-sample (SS) procedure, and point-optimal versions of these tests (AR^* and SS^*). For distributional theory, three different sets of assumptions are considered: (1) for Gaussian errors, we provide exact tests and confidence sets; (2) for a wide class of parametric non-Gaussian errors (possibly heavy-tailed), we establish that exact Monte Carlo procedures can be applied using the statistics considered; (3) under weaker distributional assumptions, we show these tests are asymptotically valid. A comprehensive Monte Carlo study indicates that the proposed tests outperform the usual asymptotic test regarding size and exhibit excellent power in empirically realistic settings. We apply our inference methods to IBM's price and option data (2009-2013). We consider 175 different instruments (IV's) spanning 22 classes and analyze their ability to describe the low-frequency volatility. The IV's are compared based on the average length of confidence intervals, which are produced by the proposed tests. The superior instrument set mostly consists of 5-minute HF realized measures, and these IV's produce confidence sets where the volatility persistence parameter lies roughly between 0.85 and 1.0. This outcome suggests that the volatility process is highly persistent and close to unit-root. We find RVs with higher frequency produce wider confidence intervals compared to RVs with slightly lower frequency, showing that these confidence intervals adjust to absorb market microstructure noise or discretization error. Further, when we consider irrelevant or weak IV's (jumps and signed jumps), the proposed tests produce unbounded confidence intervals. Although jumps contain little information content regarding the low-frequency volatility, we find evidence that there may be a nonlinear relationship between jumps and the low-frequency volatility.

Chapter 2

A simple efficient moment-based estimator for the stochastic volatility model

Abstract

We study the problem of estimating the parameters of the stochastic volatility (SV) process [Taylor (1982, 1986)], a model of conditional heteroskedasticity with several attractive features, especially for financial applications. However, due to the presence of latent variables, likelihood-based methods are difficult to apply, and statistical inference (estimation and testing) for this model is challenging. The existing methods are either computationally costly and/or inefficient. In this paper, we propose computationally simple estimators for the SV model, which are at the same time highly efficient. The proposed class of estimators is based on a small number of moment equations derived from an ARMA representation associated with the SV model, along with the possibility of using "winsorization" to improve stability and efficiency. We call these ARMA-SV estimators. Closed-form expressions for ARMA-SV estimators are obtained, and no numerical optimization procedure or choice of initial parameter values is required. The asymptotic distributional theory of the proposed estimators is studied. Due to their computational simplicity, the ARMA-SV estimators allow one to make reliable – even exact – simulation-based inference, through the application of Monte Carlo (MC) test or bootstrap methods. We compare them in a simulation experiment with a wide array of alternative estimation methods, in terms of bias, root mean square error and computation time. In addition to confirming the enormous computational advantage of the proposed estimators, the results show that ARMA-SV estimators match (or exceed) alternative estimators in terms of precision – including Bayesian estimators proposed in this context, which have the best performance among alternative estimators. The proposed methods are applied to daily observations on the returns for three major stock prices [Coca-Cola, Walmart, Ford] over the period 1980-2015, and to the Standard and Poor's Composite Price Index over the 2000-2017 period. The results confirm the presence of stochastic volatility with strong persistence.

Key words: Stochastic volatility; Latent variable; ARCH; Moment estimator; Generalized method of moments; Quasi-maximum likelihood; Bayesian estimator; Markov Chain Monte Carlo; Asymptotic distribution; Monte Carlo test; Maximized Monte Carlo test; Stock returns.

Journal of Economic Literature classification: C11; C13; C15; C22; G1

2.1 Introduction

Modelling the time-varying variance or conditional heteroskedasticity of asset returns is one of the major problems of financial econometrics. To deal with such features, two main classes of parametric models have been proposed: (1) ARCH [Engle (1982)] and GARCH models [Bollerslev (1986)], where volatility is modelled as a deterministic function of past shocks; (2) stochastic volatility (SV) models [Taylor (1982, 1986)], where volatility is a latent stochastic process.

SV models may be preferable to GARCH-type models for several reasons. *First*, SV models constitute discrete versions of continuous-time diffusion processes, which are widely used in the option-pricing literature; see Hull and White (1987), Taylor (1994), Shephard and Andersen (2009). *Second*, SV models are flexible and relatively robust to model misspecification. GARCH models often require adding a random jump component or allowing for innovations with heavy-tailed distributions to tackle these problems. Such modifications substantially improve the performance of the standard GARCH, but do not appear to be required for SV models; see Carnero et al. (2004), Chan and Grant (2016). *Third*, the SV model performs better than GARCH-type models in volatility forecasting, which suggests that time-varying volatility is better modelled as a latent first-order autoregression; see Kim et al. (1998), Yu (2002), Poon and Granger (2003), Koopman et al. (2005). *Fourth*, one can easily derive the statistical properties (stationarity, ergodicity, mixing) of SV models, while this appears more difficult for GARCH models; see Davis and Mikosch (2009). In particular, conditions for the stationarity of GARCH models are relatively difficult to establish; see Nelson (1990), Bougerol and Picard (1992) and Lindner (2009). *Finally*, SV models belong to the analytically convenient class of state-space models [Harvey (1989)].

Several reviews of the literature on SV models are available; see Ghysels et al. (1996), Broto and Ruiz (2004), and Shephard (2005). SV models are also important in macroeconometric modelling following the seminal work of Cogley and Sargent (2005) and Primiceri (2005). Recent papers along these lines include Benati (2008), Koop et al. (2009), Koop and Korobilis (2013), Liu and Morley (2014).

Despite their appealing features, SV models are much less popular than GARCH-type models in the empirical literature. As pointed out by Bos (2012), this may be explained by two reasons. *First*, estimating the parameters of an SV model is much more complex than it is for a GARCH model, since SV models have no closed-form likelihood function. *Second*, many statistical packages (such as EVIEWS, GAUSS, MATLAB, R, S+, SAS, TSP, STATA, PYTHON, OX, etc.) have options for incorporating GARCH effects, while programs for estimating SV models appear to be much less widespread (some routines in R and MATLAB are available).

Proposed estimation methods for SV models include:

1. the generalized method of moments (GMM) [Melino and Turnbull (1990), Andersen and Sørensen (1996)];

2. quasi-maximum likelihood (QML) [Nelson (1988), Harvey et al. (1994), Ruiz (1994)];

3. the simulated method of moments (SMM) [Gallant and Tauchen (1996), Monfardini (1998), Andersen et al. (1999)];

4. Monte Carlo likelihood (MCL) [Sandmann and Koopman (1998)];

5. simulated maximum likelihood (SML) [Danielsson and Richard (1993), Danielsson (1994), Durham (2006, 2007), Richard and Zhang (2007)];

6. estimation based on linear representations (LiR) [Francq and Zakoïan (2006)];

7. methods based on Bayesian Markov Chain Monte Carlo (MCMC) [Jacquier et al. (1994), Kim et al. (1998), Chib et al. (2002), Flury and Shephard (2011)];

8. closed-form moment-based estimators (DV) [Dufour and Valéry (2006, 2009)].

The above estimation procedures are typically based on simulation techniques and/or require numerical optimization. The only exception is the closed-form estimator of Dufour and Valéry (2006, 2009). Methods such as SML, MCL, SMM, and Bayesian MCMC [through the Metropolis-Hastings algorithm or the Gibbs sampler] require the use of simulation techniques. These methods are computationally expensive, inflexible across models, and may converge quite slowly; see Broto and Ruiz (2004). Furthermore, some of these methods require one to choose a sampling algorithm, initial parameters, or an auxiliary model. The choice of initial values for QML, GMM or MCMC plays a vital role in convergence [a large number of non-converging GMM estimations is reported by Andersen et al. (1999)]. As usual, GMM can easily be adversely affected by an ill-conditioned weighting matrix.

Among these estimators, only the closed-form estimator of Dufour and Valéry (2006, 2009) is analytically tractable, computationally simple, and easy to implement. However, it tends to be less precise than some other estimators. In this paper, we propose improved simple moment-based estimators for the SV model, which retain the computational advantages of the method described by Dufour and Valéry (2006) and match (or exceed) the precision of alternative estimators. To do this, we exploit an ARMA representation which can be associated with the SV model, along with a "winsorization" technique originally proposed by Kristensen and Linton (2006) for GARCH models.[1] The proposed class of estimators can be viewed as an ARMA-type extension of the approach used in Dufour and Valéry (2006), which leads to a small but different set of moments. To be more specific, the contributions of the paper can be summarized as follows.

First, after spelling out the relevant ARMA-type equations, we show that these yield autocovariances which can be solved in closed form for the parameters of the SV model. As the moments involved can be easily estimated from the data, we obtain in this way computationally simple estimators, without the need to use numerical optimization or initial values. We call these *simple ARMA-SV* estimators. In particular, the persistence parameter of the process is estimated by a simple ratio of easily estimable sample autocovariances.

Second, as the proposed estimator may not satisfy stationarity restrictions and can be sensitive to outliers in small samples, we propose winsorized versions of the simple ARMA-SV estimators (*W-ARMA-SV* estimators), where the persistence parameter of the SV model is estimated using a combination of several ratios of sample autocovariances [such as weighted averages, the median, or an OLS-based weighting]. This modification remains computationally simple and improves the stability and precision of the estimators. Indeed, we show in simulations that W-ARMA-SV estimators improve (or match) alternative estimators in terms of precision – including Bayesian estimators proposed in this context, which have the best

[1] By exploiting the ARMA representation of GARCH processes, Kristensen and Linton (2006), Sbrana and Poloni (2013) and Hafner and Linton (2017) have proposed closed-form moment estimators for GARCH(1,1), multivariate GARCH(1,1) and exponential GARCH(1,1) models, respectively.

performance among alternative estimators. In particular, an OLS-based W-ARMA-SV appears to have the best performance.

Third, due to their computational simplicity, the proposed ARMA-SV estimators can be useful for several purposes.

1. Since they can be easily be simulated, the proposed estimators constitute ideal candidates for building simulation-based tests, even exact tests through the application of the Monte Carlo test method [see Dufour (2006)], as opposed to procedures based on establishing asymptotic distributions. Interestingly, exact tests obtained in this way do not depend on stationarity assumptions, and so may be especially useful when the latent volatility process has a unit root (or is close to this structure).

2. Methods which involve repeated estimation, such as out-of-sample forecasting based on a rolling window scheme, become easily applicable.

Fourth, we study the asymptotic properties of the proposed estimators under standard regularity assumptions. In particular, we show that the estimators are $\sqrt{T}$-consistent and asymptotically normal (when the fourth moment of the latent volatility process exists), at least with linear winsorization.

Fifth, we report Monte Carlo simulations comparing the performance of our simple estimators with alternative available estimators, in terms of bias, standard deviation, and mean square error. We make four important observations: (1) the OLS-type W-ARMA-SV estimator dominates the other winsorized estimators considered; (2) ARMA-SV estimators have an excellent overall performance: they clearly dominate other non-Bayesian estimators (QML, GMM, DV) and match the Bayesian estimator; (3) these results underscore that using too many moments can be very costly from an efficiency viewpoint: it is preferable to use a small number of well-chosen moments; (4) the proposed ARMA-SV estimators (simple and winsorized) are extremely efficient in terms of computation time, especially when compared with the Bayesian estimator.

Sixth, we present some simulation evidence on the performance of likelihood-ratio-type (LR-type) tests based on ARMA-SV estimators, for a number of basic hypotheses in this context (no volatility persistence, no random variation in volatility, fixed volatility). Three approaches for controlling the level of the tests are considered: (1) using a standard chi-square asymptotic approximation; (2) local Monte Carlo tests (or parametric bootstrapping); (3) maximized Monte Carlo (MMC) tests. We find that tests based on an asymptotic approximation can be quite unreliable, but simulation-based tests (bootstrapping, MMC) perform well in terms of level control and power.

Seventh, we present empirical applications to daily observations on the returns for three major stock prices [Coca-Cola, Walmart, and Ford] over the period 1980-2015, and to the Standard and Poor's Composite Price Index over the 2000-2017 period. In this study, using the proposed estimation method, we find evidence that the returns on these stocks exhibit stochastic volatility with strong persistence. We also implemented MC tests to construct more reliable finite-sample inference since the estimated volatility persistence parameter is close to the unit circle. Three crucial null hypotheses relevant to this context are considered: (a) no persistence in latent volatility; (b) no latent specification of the volatility process; (c) no stochastic volatility. These are tested following three approaches: asymptotic critical values, a local Monte Carlo (or parametric bootstrap) procedure, and a maximized Monte Carlo (MMC) procedure. All three hypotheses are decisively rejected, irrespective of the test approach used. For the S&P composite index, the results of our estimation method are compared with those obtained by a Bayesian MCMC method. The estimates based on the two methods are remarkably close, even though our technique requires much less computation time.

This paper is organized as follows. Section 2.2 specifies the model, assumptions, and motivation. Section 2.3 describes simple estimators for the SV model. Section 2.4 reviews the stationarity, ergodicity and mixing properties of the SV process. Section 2.5 develops the asymptotic distributional theory for the simple estimator. Section 2.6 discusses how finite-sample Monte Carlo tests can be applied using the proposed simple estimator. Section 2.7 presents the simulation study. The empirical application is presented in Section 2.8. We conclude in Section 2.9. The proofs, tables, and figures are available in the Appendix 2.10.

2.2 Framework

We consider a standard discrete-time stochastic volatility (SV) model of the type described by Taylor (1986) and Ghysels et al. (1996). Specifically, we say that a variable y_t follows a discrete-time SV process if it satisfies the following assumption, where $t \in \mathbb{N}_0$ and $\mathbb{N}_0$ represents the non-negative integers.

Assumption 2.2.1. STOCHASTIC VOLATILITY MODEL. *The process* $\{y_t : t \in \mathbb{N}_0\}$ *satisfies the equations*

$$y_t = \sigma_y \exp(\frac{w_t}{2}) z_t , \tag{2.2.1}$$

$$w_t = \phi w_{t-1} + \sigma_v v_t , \tag{2.2.2}$$

where the vectors $(z_t, v_t)'$ *are i.i.d. according to a* $\mathrm{N}[0, I_2]$ *distribution, while* ϕ, σ_y *and* σ_v *are fixed parameters.*

We also make a stationarity assumption as follows.

Assumption 2.2.2. STATIONARITY. *The process* $l_t := (y_t, w_t)'$ *is strictly stationary.*

The above stationarity condition entails $|\phi| < 1$ and $w_0 \sim \mathrm{N}[0, \sigma_v^2/(1-\phi^2)]$. The SV model involves two stochastic processes which describe the dynamics of y_t and the latent log-volatilities w_t. When y_t is an asset return, the latent process w_t in (2.2.2) can be interpreted as a random flow of uncertainty shocks or new information in financial markets, while ϕ represents volatility persistence. This type of volatility model naturally fits into the theoretical framework of modern financial theory.

Let us now transform y_t by taking the logarithm of its squared value. We get in this way the following *measurement equation*:

$$\begin{aligned}\log(y_t^2) &= \log(\sigma_y^2) + w_t + \log(z_t^2) = \{\log(\sigma_y^2) + \mathbb{E}[\log(z_t^2)]\} + w_t + \{\log(z_t^2) - \mathbb{E}[\log(z_t^2)]\} \\ &= \mu + w_t + \epsilon_t \end{aligned} \tag{2.2.3}$$

where

$$\mu := \mathbb{E}[\log(y_t^2)] = \log(\sigma_y^2) + \mathbb{E}[\log(z_t^2)] , \quad \epsilon_t := \log(z_t^2) - \mathbb{E}[\log(z_t^2)] . \tag{2.2.4}$$

Under the normality assumption for z_t, the errors ϵ_t are i.i.d. according to the distribution of a centered $\log(\chi_1^2)$ random variable [*i.e.*, ϵ_t has mean zero and variance $\mathbb{E}(\epsilon_t^2)$]. The cumulant generating function of the $\log(\chi_1^2)$ distribution is

$$\begin{aligned} M(s) &= \log\mathbb{E}\big[\exp\big(s\log(\chi_1^2)\big)\big] = \log\mathbb{E}\big[\big(\chi_1^2\big)^s\big] = \log\left[\frac{2^s\Gamma\big((1/2)+s\big)}{\Gamma(1/2)}\right] \\ &= s\log(2) + \log[\Gamma\big((1/2)+s\big)] - \log[\Gamma(1/2)]\,, \quad \text{for } s \geq 0\,, \end{aligned} \tag{2.2.5}$$

where $\Gamma(z) := \int_0^\infty x^{z-1}e^{-x}\,dx$ is the *gamma function*; see Wishart (1947). The m^{th} moment of a $\log(\chi_1^2)$ random variable is the m^{th} derivative of $M(s)$ evaluated at $s=0$, and the corresponding cumulants are:

$$\kappa_m = \begin{cases} \log(2) + \psi(\frac{1}{2}), & \text{if } m = 1, \\ \psi^{(m-1)}(\frac{1}{2}), & \text{if } m > 1, \end{cases} \tag{2.2.6}$$

where

$$\psi(z) := \frac{d}{dz}\ln[\Gamma(z)] = \frac{\Gamma'(z)}{\Gamma(z)} \tag{2.2.7}$$

is the *digamma function* and

$$\psi^{(m)}(z) := \frac{d^m}{dz^m}\psi(z) = \frac{d^{m+1}}{dz^{m+1}}\ln[\Gamma(z)] \tag{2.2.8}$$

is the polygamma function of order m [*i.e.*, the $(m+1)$-th order derivative of the logarithm of the gamma function].

Using the relationship between cumulants (κ_m) and central moments ($\tilde{\mu}_m$) given by

$$\tilde{\mu}_m = \begin{cases} 0, & \text{if } m = 1 \\ \kappa_m + \sum_{j=1}^{m-2}\binom{m-1}{j}\kappa_{m-j}\tilde{\mu}_j, & \text{if } m > 1 \end{cases},$$

and (2.2.6), we get:

$$\mathbb{E}[\log(z_t^2)] = \kappa_1 = \log(2) + \psi(1/2) \simeq -1.2704\,, \tag{2.2.9}$$

$$\sigma_\epsilon^2 := \mathbb{E}(\epsilon_t^2) = \mathrm{Var}[\log(z_t^2)] = \tilde{\mu}_2 = \kappa_2 = \psi^{(1)}(1/2) = \pi^2/2\,, \tag{2.2.10}$$

$$\mathbb{E}(\epsilon_t^3) = \tilde{\mu}_3 = \kappa_3 = \psi^{(2)}(1/2)\,, \quad \mathbb{E}(\epsilon_t^4) = \tilde{\mu}_4 = \kappa_4 + 3\kappa_2^2 = \psi^{(3)}(1/2) + 3\sigma_\epsilon^2 = \pi^4 + 3\sigma_\epsilon^2\,; \tag{2.2.11}$$

see Abramowitz and Stegun (1970, Chapter 6). The $\log(\chi_1^2)$ distribution is often approximated by a normal distribution with mean of -1.2704 and variance of $\pi^2/2$ [see Broto and Ruiz (2004)], or by a mixture distribution [Kim et al. (1998)].

On setting

$$y_t^* := \log(y_t^2) - \mu, \tag{2.2.12}$$

the SV model (2.2.3) can be written as

$$y_t^* = w_t + \epsilon_t. \tag{2.2.13}$$

By combining (2.2.2) and (2.2.13), we see that the SV model can be written in state-space form:

$$w_t = \phi w_{t-1} + v_t, \qquad \text{(State Transition Equation)} \tag{2.2.14}$$

$$y_t^* = w_t + \epsilon_t, \qquad \text{(Measurement Equation)} \tag{2.2.15}$$

where w_t is the logarithm of latent daily volatility, y_t^* is the logarithm of the daily squared return corrected by its mean, where the variables v_t are i.i.d. $\mathrm{N}(0, \sigma_v^2)$ and the ϵ_t's are i.i.d. $\log(\chi_1^2)$; for further discussion of this representation, see Nelson (1988), Harvey et al. (1994), Ruiz (1994), Shephard (1994), Breidt and Carriquiry (1996), Harvey and Shephard (1996), Kim et al. (1998), Sandmann and Koopman (1998), Steel (1998), Chib et al. (2002), Knight et al. (2002), Francq and Zakoïan (2006), Omori et al. (2007).

2.3 Simple ARMA-type estimators

In this section, we propose simple estimators for the SV model by exploiting the autocovariance structure of y_t^*. For this purpose, we consider moments and cross-moments of y_t^* which differ in a crucial way from those used by Dufour and Valéry (2006). In the latter paper, the parameters of the SV model are obtained through the following equations based on moments of y_t^2 [instead of $\log(y_t^2)$]:

$$\phi = \frac{\log[\mathbb{E}(y_t^2 y_{t-1}^2)/(\mathbb{E}(y_t^2))^2]}{\log[\mathbb{E}(y_t^4)/3(\mathbb{E}(y_t^2))^2]}, \quad \sigma_y = \frac{3^{1/4}\mathbb{E}(y_t^2)}{[\mathbb{E}(y_t^4)]^{1/4}}, \quad \sigma_v = \{(1-\phi^2)\log[\mathbb{E}(y_t^4)/\big(3E(y_t^2)\big)]\}^{1/2}. \tag{2.3.1}$$

In this paper, we derive a closed-form estimator of the SV model based on the ARMA representation of y_t^*.

The ARMA representation of the process y_t^* and its autocovariance structure are given in the following proposition.

Proposition 2.3.1. ARMA REPRESENTATION OF LOG-SV PROCESS. *Under the Assumptions 2.2.1 - 2.2.2, the process y_t^* defined in (2.2.12) has the following ARMA$(1,1)$ representation:*

$$y_t^* = \phi y_{t-1}^* + \eta_t - \theta\eta_{t-1} \tag{2.3.2}$$

with $\eta_t - \theta\eta_{t-1} = v_t + \epsilon_t - \phi\epsilon_{t-1}$, where the error processes $\{v_t\}$ and $\{\epsilon_t\}$ are mutually independent, the errors v_t are i.i.d. $\mathrm{N}(0, \sigma_v^2)$*, and the errors ϵ_t are i.i.d. according to the distribution of a* $\log(\chi_1^2)$ *random variable.*

The above proposition provides simple expressions for the autocovariances and parameters of the SV model. For future reference, we state these properties in two corollaries.

Corollary 2.3.2. AUTOCOVARIANCES OF LOG-SV PROCESS. *Under the assumptions of Proposition 2.3.1, the autocovariances of the process y_t^* defined in (2.2.12) satisfy the following equations:*

$$\operatorname{cov}(y_t^*, y_{t-k}^*) := \gamma_{y^*}(k) = \begin{cases} \phi\gamma_{y^*}(k-1) + \sigma_v^2 + \sigma_\epsilon^2, & \text{if } k = 0, \\ \phi\gamma_{y^*}(k-1) - \phi\sigma_\epsilon^2, & \text{if } k = 1, \\ \phi\gamma_{y^*}(k-1), & \text{if } k \geq 2. \end{cases} \tag{2.3.3}$$

Corollary 2.3.3. CLOSED-FORM EXPRESSIONS FOR SV PARAMETERS. *Under the assumptions of Proposition 2.3.1, we have:*

$$\phi = \frac{\gamma_{y^*}(k+1)}{\gamma_{y^*}(k)}, \quad \text{for } k \geq 1, \tag{2.3.4}$$

$$\sigma_y^2 = \exp[\mu - \mu_2], \tag{2.3.5}$$

$$\sigma_v^2 = (1-\phi^2)[\gamma_{y^*}(0) - (\pi^2/2)], \tag{2.3.6}$$

where $\gamma_{y^*}(k) = \operatorname{cov}(y_t^*, y_{t-k}^*)$, with y_t^* and μ defined in (2.2.12), and $\mu_2 := \mathbb{E}[\log(z_t^2)] \simeq -1.2704$.

From (2.3.4), we see that ϕ can be obtained from several autocovariance ratios, which can be easily estimated with the corresponding empirical moments:

$$\hat{\gamma}_{y^*}(k) = \frac{1}{T-k}\sum_{t=1}^{T-k}[\log(y_t^2)-\hat{\mu}][\log(y_{t+k}^2)-\hat{\mu}], \quad \hat{\mu} = \frac{1}{T}\sum_{t=1}^{T}\log(y_t^2). \tag{2.3.7}$$

Of these, the ratio $\gamma_{y^*}(2)/\gamma_{y^*}(1)$ is the one for which we can use the largest number of observations. This suggests the following parameter estimators:

$$\hat{\phi} = \frac{\hat{\gamma}_{y^*}(2)}{\hat{\gamma}_{y^*}(1)}, \quad \hat{\sigma}_y^2 = \exp(\hat{\mu}+1.2704), \quad \hat{\sigma}_v^2 = (1-\hat{\phi}^2)[\hat{\gamma}_{y^*}(0)-(\pi^2/2)]\,. \tag{2.3.8}$$

We call these the *simple ARMA-SV* estimators.

A shortcoming of the above simple ARMA-type estimator is that it can yield inadmissible parameter values, *e.g.* with $|\hat{\phi}| \geq 1$. This issue can arise especially in small samples or in the presence of outliers. To deal with a similar problem, Kristensen and Linton (2006) proposed to use "winsorization" which substantially increases the probability of getting admissible values.

From (2.3.4), it is easy to see that

$$\phi = \sum_{j=1}^{\infty} w_j \frac{\gamma_{y^*}(j+1)}{\gamma_{y^*}(j)} \tag{2.3.9}$$

for any sequence w_j such that $\sum_{j=1}^{\infty} w_j = 1$. This suggests a more general class of estimators for ϕ obtained by averaging several sample analogs of the ratios $\gamma_{y^*}(j+1)/\gamma_{y^*}(j)$:

$$\tilde{\phi} = \sum_{j=1}^{J} w_j \frac{\hat{\gamma}_{y^*}(j+1)}{\hat{\gamma}_{y^*}(j)} \tag{2.3.10}$$

where $1 \leq J \leq T-2$ with $\sum_{j=1}^{J} w_j = 1$, and T is the length of the time series. We call such estimators *winsorized ARMA-SV* estimators (or *W-ARMA-SV* estimators). Other (possibly nonlinear) averaging methods, such as the median, may also be used.

In the simulation section below, we consider four different winsorized ARMA-SV estimators based on (2.3.10). These estimators are also considered by Hafner and Linton (2017) in the context of closed-form estimation of the EGARCH(1,1) model. The first one ($\hat{\phi}_M$) is an arithmetic mean of sample covariance ratios (equal weights): we set

$$w_j = 1/J, \quad j = 1, \ldots, J, \tag{2.3.11}$$

in (2.3.10). The second one ($\hat{\phi}_{LD}$) has linearly declining weights: we set

$$w_j = (2/J)[1-(j/(J+1))], \quad j = 1, \ldots, J, \tag{2.3.12}$$

in (2.3.10). The third one ($\hat{\phi}_{MED}$) is the median of J autocovariance ratios:

$$\hat{\phi}_{MED} = \text{med}\{\hat{\gamma}_{y^*}(j+1)/\hat{\gamma}_{y^*}(j) : j = 1, \ldots, J\}. \tag{2.3.13}$$

The fourth one ($\hat{\phi}_{OLS}$) is based on the OLS regression of $\hat{\gamma}_{y^*}(j+1)\, w_j^{1/2}$ on $\hat{\gamma}_{y^*}(j)\, w_j^{1/2}$ without intercept: this suggests the estimate

$$\hat{\phi}_{OLS} = \left(\bar{a}'\bar{a}\right)^{-1}\bar{a}'\bar{e} \tag{2.3.14}$$

where $\bar{a} = [\hat{\gamma}_{y^*}(1)\, w_1^{1/2}, \ldots, \hat{\gamma}_{y^*}(J)\, w_J^{1/2}]'$ and $\bar{e} = [\hat{\gamma}_{y^*}(2)\, w_1^{1/2}, \ldots, \hat{\gamma}_{y^*}(J+1)\, w_J^{1/2}]'$. Clearly, different OLS-based W-ARMA-SV can be generated by considering different weights $w_1, \ldots, w_J$. In our simulations below as well as empirical applications, we focus on the case where the weights are equal [see (2.3.11)]: in this case,

$$\hat{\phi}_{OLS} = \frac{\sum_{j=1}^{J} \hat{\gamma}_{y^*}(j)\, \hat{\gamma}_{y^*}(j+1)}{\sum_{j=1}^{J} \hat{\gamma}_{y^*}(j)^2} \tag{2.3.15}$$

All these estimators depend on J. For $J = 1$, they all yield the simple ARMA-SV estimator $\hat{\phi} = \hat{\gamma}_{y^*}(2)/\hat{\gamma}_{y^*}(1)$.

2.4 Stationarity, ergodicity and mixing properties

In SV models, the independence between the noise (z_t) and the volatility variable (w_t) allows for a simpler probabilistic structure than for GARCH processes. This independence is one of the attractive features of SV models. Indeed, the problem of finding a necessary and sufficient condition for stationarity of GARCH processes was tackled fairly late [see Nelson (1990) and Bougerol and Picard (1992)]; for reviews, see Straumann (2005) and Francq and Zakoïan (2010). To establish the large-sample properties for our estimator, we consider the case where $(w_t, y_t)'$ is strictly stationary and ergodic. The following results ensure the stationarity, ergodicity and mixing for the standard SV model; see Carrasco and Chen (2002).

Result 2.4.1. STATIONARITY AND ERGODICITY. *Let $\{z_t\}$ and $\{v_t\}$ be two independent processes such that $\{z_t\}$ is a sequence of i.i.d. real-valued random variables, independent of w_0, with $\mathbb{E}(z_t) = 0$ and $\mathbb{E}(z_t^2) = 1$, and z_t has a continuous strictly positive density (with respect to the Lebesgue measure) on the real line. Suppose also that $|\phi| < 1$ and there is an integer $s \geq 1$ such that*

$$\mathbb{E}(|v_t|^s) < \infty. \tag{2.4.1}$$

Then the following properties hold.

(i) *$\mathbb{E}(|w_t|^s) < \infty$ and $\{w_t\}$ is Markov geometrically ergodic.*

(ii) *If $\{w_t\}$ is initialized from its stationary distribution, the processes $\{w_t\}$ and $\{y_t\}$ are strictly stationary and exponential β-mixing, and this property is preserved by any continuous transformation of $\{w_t\}$, such as $\{\exp(w_t/2)\}$.*

(iii) *If $\mathbb{E}\big(|\ln(|z_t|)|^s\big) < \infty$, then $\mathbb{E}\Big(\big|\ln(|y_t|)\big|^s\Big) < \infty$.*

Note that the latter part of the above result follows easily on observing that $y_t = \exp(w_t/2)\sigma_y z_t$ entails

$$\ln|y_t| = w_t/2 + \ln|\sigma_y| + \ln|z_t|\,. \tag{2.4.2}$$

The stochastic volatility process $\{y_t\}$ is a hidden Markov process since it includes a latent Markov chain $\{w_t\}$. Further, the process $\{w_t\}$ is independent of the i.i.d. noise process $\{z_t\}$. Proposition 2.1 of Genon-Catalot et al. (2000) show that a hidden Markov model y_t is ergodic and strong mixing if the hidden chain $\{w_t\}$ is ergodic and strong mixing. In the context of the SV model, (2.4.2) and Proposition 4 of Carrasco and Chen (2002) entail the following result.

Result 2.4.2. BETA MIXING. *Let $\{y_t\}$ be a generalized hidden Markov model with a hidden chain $\{w_t\}$.*

(i) *If $\{w_t\}$ is geometrically ergodic, the process $\{(w_t, \ln|y_t|)\}$ is Markov geometrically ergodic.*

(ii) *If $\{w_t\}$ is stationary β-mixing, $\{\ln|y_t|\}$ is stationary β-mixing with a decay rate at least as fast as that of $\{w_t\}$.*

We thus have the following basic property of the SV process: if $\{w_t\}$ is initialized from its stationary distribution, $\{\ln|y_t|\}$ is strictly stationary and exponential β-mixing, and so is the process $\{(y_t, w_t)'\}$.

2.5 Asymptotic distributional theory

We will now study the asymptotic distribution of the estimator $\hat{\theta} := (\hat{\phi}, \hat{\sigma}_y, \hat{\sigma}_v)'$ under the following set of assumptions.

Assumption 2.5.1. DISTRIBUTION OF THE ERROR PROCESSES. *The error processes z_t and v_t are mutually independent and $\{z_t\}$ is a sequence of i.i.d. real-valued random variables, independent of w_0. The probability distribution of z_t has a continuous density with respect to Lebesgue measure on the real line, and its density is positive on $(-\infty, +\infty)$. The transformed error ϵ_t satisfies $\mathbb{E}(|\epsilon_t|^s) < \infty$, where s is an integer such that $s \geq 1$.*

Assumption 2.5.2. STATIONARITY OF THE LATENT PROCESS. *The latent process $\{w_t\}$ is strictly stationary with $|\phi| < 1$, $\mathbb{E}(|w_t|^s) < \infty$, and $\mathbb{E}(|v_t|^s) < \infty$. where s is a positive integer.*

Under the Assumptions 2.5.1 and 2.5.2 with $s = 2$, the process $\{y_t^*\}$ is strictly stationarity and geometrically ergodic with exponential β-mixing (see results 2.4.1 and 2.4.2) with finite second moments, *i.e.*, $\mathbb{E}\left[(y_t^*)^2\right] < \infty$. In the following lemma, using ergodicity, we prove the consistency of the empirical moments in (2.3.7).

Lemma 2.5.1. CONSISTENCY OF EMPIRICAL MOMENTS. *Under the Assumptions 2.5.1 and 2.5.2 with $s = 2$, the estimators $\hat{\Gamma}(m) := [\hat{\gamma}_{y^*}(0), \hat{\gamma}_{y^*}(1), \ldots, \hat{\gamma}_{y^*}(m)]'$ and $\hat{\mu}$ defined by (2.3.7) satisfy: for any $m \geq 0$,*

$$\hat{\mu} \xrightarrow{p} \mu \quad \text{and} \quad \hat{\Gamma}(m) \xrightarrow{p} \Gamma(m) := [\gamma_{y^*}(0), \gamma_{y^*}(1), \ldots, \gamma_{y^*}(m)]' . \tag{2.5.1}$$

The Assumptions 2.5.1 and 2.5.2 with $s=4$ are sufficient for the SV model to have a strictly stationary solution with a finite fourth moment of y_t^*, *i.e.*, $\mathbb{E}\left[(y_t^*)^4\right]<\infty$. Note that the fourth moment of y_t^* translates into the eighth moment of y_t. This solution will be β-mixing with geometrically decreasing mixing coefficients. In the following lemma, using a Central Limit Theorem for stationary ergodic processes (Lindeberg-Levy theorem for dependent processes), we give the asymptotic distribution of the empirical moments in (2.3.7).

Lemma 2.5.2. ASYMPTOTIC DISTRIBUTION OF EMPIRICAL MOMENTS. *Under the assumptions 2.5.1, 2.5.2 with $s=4$, the estimators $\hat{\Gamma}(m)=[\hat{\gamma}_{y^*}(0),\hat{\gamma}_{y^*}(1),\ldots,\hat{\gamma}_{y^*}(m)]'$ and $\hat{\mu}$ defined by (2.3.7) satisfy:*

$$\sqrt{T}\begin{bmatrix}\hat{\mu}-\mu\\ \hat{\Gamma}(m)-\Gamma(m)\end{bmatrix}\xrightarrow{d}\mathrm{N}\left(0,\ \begin{bmatrix}V_\mu & C'_{\mu,\Gamma(m)}\\ C_{\mu,\Gamma(m)} & V_{\Gamma(m)}\end{bmatrix}\right) \tag{2.5.2}$$

where

$$V_\mu := \gamma_{y^*}(0)+2\sum_{\tau=1}^{\infty}\gamma_{y^*}(\tau)\,,\quad V_{\Gamma(m)}=\mathrm{Var}(\Lambda_t)+2\sum_{\tau=1}^{\infty}\mathrm{cov}(\Lambda_t,\,\Lambda_{t+\tau})\,,\quad C=(\bar{c},\ 0_{[1\times m]})'\,, \tag{2.5.3}$$

$$\Lambda_t := [\Lambda_{t,0},\Lambda_{t,1},\ldots,\Lambda_{t,m}]'\,, \tag{2.5.4}$$

$$\Lambda_{t,k} := y_t^* y_{t+k}^* - \gamma_{y^*}(k) = [\log(y_t^2)-\mu][\log(y_{t+k}^2)-\mu]-\gamma_{y^*}(k)\,,\quad k=0,\ldots,m\,, \tag{2.5.5}$$

$$\bar{c} := C_{\mu,\Gamma(0)} = 2\sum_{t=1}^{\infty}\mathbb{E}[y_t^{*3}] = 2\sum_{t=1}^{\infty}(\mathbb{E}[w_t^3]+\mathbb{E}[\epsilon_t^3]) = 2\sum_{t=1}^{\infty}\mathbb{E}[\epsilon_t^3]\,. \tag{2.5.6}$$

This in turn yields the asymptotic distribution of the simple ARMA-type estimator $(\hat{\phi},\hat{\sigma}_y,\hat{\sigma}_v)'$.

Theorem 2.5.3. ASYMPTOTIC DISTRIBUTION OF SIMPLE ARMA-TYPE ESTIMATOR. *Under the assumptions 2.5.1, 2.5.2 with $s=4$, the estimator $\hat{\theta}:=(\hat{\phi},\hat{\sigma}_y,\hat{\sigma}_v)'$ given in (2.3.8) is consistent,* i.e., $\hat{\theta}\xrightarrow{p}\theta$, *and*

$$\sqrt{T}(\hat{\theta}-\theta)\xrightarrow{d}\mathrm{N}[0,\,V] \tag{2.5.7}$$

where $\theta:=(\phi,\sigma_y,\sigma_v)'$,

$$V=G(\beta)\begin{bmatrix}V_\mu & C'_{\mu,\Gamma(3)}\\ C_{\mu,\Gamma(3)} & V_{\Gamma(3)}\end{bmatrix}G(\beta)'\,, \tag{2.5.8}$$

$$G(\beta) := \frac{\partial D}{\partial \beta'} = \begin{bmatrix} 0 & 0 & -\gamma_{y^*}(2)/\gamma_{y^*}(1)^2 & 1/\gamma_{y^*}(1) \\ \sigma_y/2 & 0 & 0 & 0 \\ 0 & \frac{1}{2}\sqrt{\frac{\kappa_1}{\kappa_2}} & \frac{\gamma_{y^*}(2)^2}{\gamma_{y^*}(1)^3}\sqrt{\frac{\kappa_2}{\kappa_1}} & -\frac{\gamma_{y^*}(2)}{\gamma_{y^*}(1)^2}\sqrt{\frac{\kappa_2}{\kappa_1}} \end{bmatrix}, \tag{2.5.9}$$

$$D := D(\beta) = (D_\phi, D_{\sigma_y}, D_{\sigma_v})', \quad \beta := [\mu, \gamma_{y^*}(0), \gamma_{y^*}(1), \gamma_{y^*}(2)]', \tag{2.5.10}$$

$$D_\phi := \gamma_{y^*}(2)/\gamma_{y^*}(1), \quad D_{\sigma_y} := \exp(\mu + 1.27)^{1/2}, \quad D_{\sigma_v} = \kappa_1 \kappa_2, \tag{2.5.11}$$

$$\sigma_y = \sqrt{\exp(\mu + 1.2704)}, \quad \kappa_1 = [1 - (\gamma_{y^*}(2)/\gamma_{y^*}(1))^2], \quad \kappa_2 = [\gamma_{y^*}(0) - \pi^2/2]. \tag{2.5.12}$$

An estimator of the covariance matrix V can be obtained by first estimating V_μ, $C_{\mu,\Gamma(3)}$ and $V_{\Gamma(3)}$ using heteroskedasticity and autocorrelation consistent (HAC) variance estimators [see Den Haan and Levin (1997) and Robinson and Velasco (1997)] and then substituting $\hat{\beta} = [\hat{\mu}, \hat{\gamma}_{y^*}(0), \hat{\gamma}_{y^*}(1), \hat{\gamma}_{y^*}(2)]'$ into $G(\beta)$. In our empirical applications, we use a Bartlett kernel estimator with the bandwidth varying with the sample size; see Newey and West (1994). One can alternatively use the analytic expressions of $\gamma_{y^*}(k)$ to obtain an estimator of V_μ. The ARMA-type estimator can be viewed as a GMM-type estimator, so one can also use GMM standard errors.

Theorem 2.5.3 covers the simplest ARMA-SV estimator. The asymptotic distribution of more general winsorized estimators can be derived in the same way upon using Lemmas 2.5.1 - 2.5.2.

2.6 Hypothesis testing

In this section, we discuss how to test a hypothesis on an SV model. First, we discuss asymptotic tests based on t-type and LR-type test statistics. Second, we show how to construct finite-sample tests using the Monte Carlo test technique.

2.6.1 Asymptotic tests

The SV model has three parameters, given by $\theta = (\phi, \sigma_y, \sigma_v)'$. To test the values of individual parameters, we can consider t-type statistics of the form:

$$T(\theta_1) = (\hat{\theta}_1 - \theta_1)/SE(\hat{\theta}_1)$$

where the standard error $SE(\hat{\theta}_1)$ is calculated from the asymptotic covariance matrix given in (2.5.7).

For testing a joint hypothesis, we consider GMM-based LR-type statistics, based on the following moment-based objective function:

$$M_T(\theta) := g_T(\theta)' A_T \, g_T(\theta) \tag{2.6.1}$$

where $\theta := (\phi, \sigma_y, \sigma_v)'$, $g_T(\theta)$ is 3×1 vector of moment functions, defined as

$$g_T(\theta) = \begin{bmatrix} \hat{\mu} + 1.2704 - \log(\sigma_y^2) \\ \hat{\gamma}_{y^*}(0) + \hat{\gamma}_{y^*}(1) - (\pi^2/2) - [\sigma_v^2/(1-\phi)] \\ \hat{\gamma}_{y^*}(2) - \phi\hat{\gamma}_{y^*}(1) \end{bmatrix} \tag{2.6.2}$$

and A_T is an appropriate weighting matrix. $M_T(\theta)$ is (up to an asymptotically negligible term) a GMM objective function.

The first moment function follows from (2.2.4). The second moment function follows from Corollary 2.3.2 on adding the equations for $k = 0$ and $k = 1$ [in (2.3.3)]: this yields

$$\hat{\gamma}_{y^*}(0) + \hat{\gamma}_{y^*}(1) - (\pi^2/2) - [\sigma_v^2/(1-\phi)] = 0. \tag{2.6.3}$$

The third moment condition corresponds to equation (2.3.3) with $k = 2$.

Since the number of moment functions in (2.6.2) is equal to the number of parameters, we take $A_T = I_3$ and consider the GMM-type objective function

$$M_T^*(\theta) = g_T(\theta)' g_T(\theta). \tag{2.6.4}$$

To test hypotheses on θ, the LR-type statistic is the difference between the restricted and unrestricted optimal values of the objective function:

$$LR_T = T[M_T^*(\hat{\theta}_0) - M_T^*(\hat{\theta})] \tag{2.6.5}$$

where $\hat{\theta}$ is the unrestricted estimator and $\hat{\theta}_0$ is the constrained estimator under the null hypothesis. Under standard regularity conditions, the asymptotic distribution of LR_T is χ_r^2 where r is the number of constraints; see Newey and West (1987), Newey and McFadden (1994), Dufour et al. (2017). Note however that usual regularity conditions may not be satisfied when some parameters are not identified or the null hypothesis involves the frontier of the parameter space.

2.6.2 Simulation-based finite-sample tests

We now discuss simulation-based inference procedures for the SV model. Simulation-based methods are tractable in the context of this study for two reasons: (1) the SV model is a parametric model, and we can easily simulate it; (2) the proposed test statistics for SV parameters are based on computationally inexpensive estimators and thus can also be easily simulated. Using our proposed computationally simple estimator, one can construct more reliable finite-sample inference using Monte Carlo tests.

The technique of Monte Carlo tests was originally proposed by Dwass (1957) for implementing permutation tests and by Barnard (1963) for continuous test statistics; for a review, see Dufour and Khalaf (2001), and for further generalizations and proofs, see Dufour (2006). It has the great attraction of providing exact (randomized) tests based on any statistic whose finite-sample distribution can be simulated, even though it may be analytically intractable. One can replace the unknown theoretical distribution $F(S|\theta)$, where $\theta = (\phi, \sigma_y, \sigma_v)$, by its sample analogue based on the statistics $S_1(\theta), \ldots, S_N(\theta)$ simulated under the null hypothesis.

Let us first consider the case of pivotal statistics, *i.e.* the case where the distribution of the test statistic under the null hypothesis does not depend on nuisance parameters. We can then proceed as follows to obtain an exact critical region for testing a null hypothesis H_0.

1. Compute the observed test statistic S_0 from the available data.

2. Generate by Monte Carlo methods a vector $S(N) = (S_1, \ldots, S_N)$ of N i.i.d. replications of S under H_0.

3. From the simulated samples, compute the MC p-value $\hat{p}_N[S] := p_N[S_0; S(N)]$ where

$$p_N[x, S(N)] := \frac{NG_N[x; S(N)] + 1}{N+1}, \tag{2.6.6}$$

$$G_N[x; S(N)] := \frac{1}{N}\sum_{i=1}^{N} I_{[0,\infty)}(S_i - x), \quad I_{[0,\infty)}(x) = \begin{cases} 1 & \text{if } x \in [0,\infty), \\ 0 & \text{if } x \notin [0,\infty). \end{cases} \tag{2.6.7}$$

In other words,

$$p_N[S_0; S(N)] = \frac{NG_N[S_0; S(N)] + 1}{N+1} \tag{2.6.8}$$

where $NG_N[S_0; S(N)]$ is the number of simulated values greater than or equal to S_0. When $S_0, S_1, \ldots, S_N$ are all distinct [an event with probability one when the vector $(S_0, S_1, \ldots, S_N)'$ has an absolutely continuous distribution], $\hat{R}_N(S_0) = N + 1 - NG_N[S_0; S(N)]$ is the rank of S_0 in the series $S_0, S_1, \ldots, S_N$.

4. The MC critical region for a test of level α $(0 < \alpha < 1)$ is

$$\hat{p}_N[S] \leq \alpha\,. \tag{2.6.9}$$

If $\alpha(N+1)$ is an integer and the distribution of S is continuous under the null hypothesis, then under H_0,

$$P[\hat{p}_N[S] \leq \alpha] = \alpha; \tag{2.6.10}$$

see Dufour (2006).

Consider now the case where the distribution of the test statistic depends on nuisance parameters. In other words, we consider a model $\{(\Xi, \mathbb{A}_\Xi, P_\theta) : \theta \in \Omega\}$ where we assume that the distribution of S is determined by $P_{\bar{\theta}}$, where $\bar{\theta}$ represents the true parameter vector. To deal with this complication, the MC test procedure can be modified as follows.

1. To test the null hypothesis

$$H_0 : \bar{\theta} \in \Omega_0 \tag{2.6.11}$$

where $\emptyset \neq \Omega_0 \subset \Omega$, compute the observed test statistic S_0 from the available data.

2. For each $\theta \in \Omega_0$, we can generate by Monte Carlo methods a vector $S(N,\theta) = [(S_1(\theta),\ldots,S_N(\theta)]$ of N i.i.d. replications of S.

3. The simulated test statistics define the MC p-value function $\hat{p}_N[S|\theta] := p_N[S_0; S(N,\theta)]$ where
$$p_N[x; S(N,\theta)] := \frac{NG_N[x; S(N,\theta)]+1}{N+1}. \tag{2.6.12}$$

4. The p-value function $\hat{p}_N[S \mid \theta]$ as a function of θ is maximized over the parameter values compatible with Ω_0, *i.e.*, under the null hypothesis, and H_0 is rejected if N
$$\sup\{\hat{p}_N[S|\theta] : \theta \in \Omega_0\} \leq \alpha. \tag{2.6.13}$$

   If the number N of simulated statistics is chosen so that $\alpha(N+1)$ is an integer, then we have under H_0:
$$P_{\bar{\theta}}[\sup\{\hat{p}_N[S|\theta] : \theta \in \Omega_0\} \leq \alpha] \leq \alpha\,. \tag{2.6.14}$$

   Consequently the critical region in (2.6.13) has *level* α for testing H_0; for a proof, see Dufour (2006).

Because of the maximization in the critical region (2.6.13), the above test is called a *maximized Monte Carlo* (MMC) test. MMC tests provide valid inference under general regularity conditions such as unidentified models or time series processes involving unit roots. In particular, even though the moment conditions defining the estimator are derived under the stationarity assumption, this does not question in any way the validity of maximized MC tests, unlike the parametric bootstrap whose distributional theory is based on strong regularity conditions. Only the power of MMC tests may be affected. However, the simulated p-value function is not continuous, so standard gradient-based algorithms and quasi-Newton methods cannot be used to maximize it. But search methods applicable to non-differentiable functions are applicable, *e.g.* simulated annealing or Particle Swarm Optimization.

A simplified approximate version of the MMC procedure can alleviate its computational load whenever a consistent point or set estimate of θ is available. To do this, we reformulate the setup in order to allow for an increasing sample size, *i.e.*, now the test statistic depends on a sample of size T, $S = S_T$.

1. Compute S_{T0} the observed test statistic (based on data). By assumption, the distribution of S involves nuisance parameters under the null hypothesis H_0 in (2.6.11).

2. We suppose we have a consistent set estimator C_T of $\bar{\theta}$ (under H_0), i.e. C_T satisfies

$$\lim_{T\to\infty} P_{\bar{\theta}}[\bar{\theta} \in C_T] = 1 \text{ under } H_0. \tag{2.6.15}$$

3. For each $\theta \in \Omega_0$, we can generate by Monte Carlo methods a vector $S(N,\theta) = [(S_1(\theta),\ldots,S_N(\theta)]$ of N i.i.d. replications of S.

4. he simulated test statistics define the MC p-value function $\hat{p}_{TN}[S_T\,|\,\theta] := p_{TN}[S_{T0}; S_T(N,\theta)]$, where

$$p_{TN}[x; S_T(N,\theta)] := \frac{NG_{TN}[x; S_T(N,\theta)] + 1}{N+1}. \tag{2.6.16}$$

5. The p-value function $\hat{p}_{TN}[S_T\,|\,\theta]$ as a function of θ is maximized with respect to θ in C_T, and H_0 is rejected if

$$\sup\{\hat{p}_{TN}[S_T\,|\,\theta] : \theta \in C_T\} \le \alpha. \tag{2.6.17}$$

   If the number of simulated statistics N is chosen so that $\alpha(N+1)$ is an integer, then under H_0,

$$\lim_{T\to\infty} P_{\bar{\theta}}[\sup\{\hat{p}_{TN}[S_T\,|\,\theta] : \theta \in C_T\} \le \alpha] \le \alpha\,. \tag{2.6.18}$$

   The critical region in (2.6.17) has level α asymptotically.

In practice, it is easy to find a consistent set estimate of $\bar{\theta}$, whenever a *consistent* point estimate $\hat{\theta}_T$ of $\bar{\theta}$ is available (*e.g.*, a GMM estimator). For instance, any set of the form

$$C_T = \{\theta \in \Omega_0 : \left\|\hat{\theta}_T - \theta\right\| < \varepsilon\} \tag{2.6.19}$$

with $\varepsilon > 0$ a fixed positive constant independent of T, satisfies (2.6.15). The consistent set estimate MMC (CSEMMC) method is especially useful when the distribution of the test statistic is highly sensitive to nuisance parameters. Here, possible discontinuities in the asymptotic distribution are automatically overcome through a numerical maximization over a set which contains the true value of the nuisance parameter with probability one asymptotically (while there is no guarantee for the point estimate to converge sufficiently fast to overcome the discontinuity). It is worth noting that there is no need to maximize the p-value function with respect to unidentified parameters under the null hypothesis. Thus, parameters which are unidentified under the null hypothesis can be set to any fixed value and the maximization be performed only over the remaining identified nuisance parameters. When there are several nuisance parameters, one can use simulated annealing, an optimization algorithm which does not require differentiability. Indeed the simulated p-value function is not continuous, so standard gradient based methods cannot be used to maximize it. For an example where this is done on a VAR model involving a large number of nuisance parameters, see Dufour and Jouini (2006).

The test based on simulations using a point nuisance parameter estimate is called a *local Monte Carlo* (LMC) test. The term local reflects the fact that the underlying MC p-value is based on a specific choice for the nuisance parameter. If the set C_T in (2.6.17) is reduced to a single point estimate $\hat{\theta}_T$, *i.e.* $C_T = \{\hat{\theta}_T\}$, we get a LMC test

$$\hat{p}_{TN}[S_T \mid \hat{\theta}_T] \leq \alpha \tag{2.6.20}$$

which can be interpreted as a parametric bootstrap test. Note that no asymptotic argument on the number N of MC replications is required to obtain this result, a fundamental difference between the latter procedure and the parametric bootstrap method.

Even if $\hat{\theta}_T$ is a consistent estimate of $\bar{\theta}$ (under the null hypothesis), the set $C_T = \{\hat{\theta}_T\}$ does not generally satisfy condition (2.6.15). Additional assumptions are needed to show that the parametric bootstrap procedure yields an asymptotically valid test. It is computationally less costly but clearly less robust to violations of regularity conditions than the MMC procedure; for further discussion, see Dufour (2006). Furthermore, the LMC non-rejections are *exactly* conclusive in the following sense: if $\hat{p}_N[S\,|\,\hat{\theta}_0] > \alpha$, then the exact MMC test is clearly not significant at level α.

2.6.3 Implicit standard error

In this subsection, we show that *implicit standard errors* (ISE) for the components of a parameter vector $\theta = (\theta_1, \ldots, \theta_m)'$ can be derived from simulation-based confidence intervals. The asymptotic standard error proposed in Section 2.5 can be markedly different and may be quite unreliable in finite samples. To construct a more reliable standard error, we derive the ISE in the following way.

1. Calculate the (typically restricted) estimate $\hat{\theta}_0 = (\hat{\theta}_{10}, \ldots, \hat{\theta}_{m0})'$ from observed data $(\mathbb{Y}_0)$.

2. Using $\hat{\theta}_0$ as parameter value, generate N i.i.d. replications $\mathbb{Y}(N) = (\mathbb{Y}_1, \ldots, \mathbb{Y}_N)$ of $\mathbb{Y}$, by Monte Carlo methods.

3. From $\mathbb{Y}(N) = (\mathbb{Y}_1, \ldots, \mathbb{Y}_N)$, compute the corresponding parameter estimates.

4. For each component θ_i of θ, the confidence interval $[C_i(\alpha_L), C_i(\alpha_H)]$, with coverage $\alpha = \alpha_L - \alpha_H$, is constructed using the empirical α_{iL} quantile and the empirical α_{iH} quantile of $\hat{\theta}_i(N) = (\hat{\theta}_{i0}, \hat{\theta}_{i1}, \ldots, \hat{\theta}_{iN})$.

5. By analogy with usual Gaussian-based confidence intervals, we set $C_i(\alpha_L) = \hat{\theta}_{i0} - z(\alpha/2)\,\hat{\sigma}_{iL}$ and $C_i(\alpha_H) = \hat{\theta}_{i0} + z(\alpha/2)\,\hat{\sigma}_{iH}$, where $z(\alpha/2)$ satisfies $P[Z \geq z(\alpha/2)] = \alpha/2$ and $Z \sim \mathrm{N}(0,1)$. This suggests that two numbers could play the role of “standard errors” here:
$$\hat{\sigma}_{iL} = \frac{\hat{\theta}_{i0} - C_i(\alpha_L)}{z(\alpha/2)} := ISE_{iL}\,, \quad \hat{\sigma}_{iU} = \frac{C_i(\alpha_H) - \hat{\theta}_{i0}}{z(\alpha/2)} := ISE_{iH}\,. \tag{2.6.21}$$

6. Finally, a conservative ISE for θ_i is given by $\min\{ISE_{iL},\ ISE_{iH}\}$ and a liberal ISE is given by the average or $\max\{ISE_{iL},\ ISE_{iH}\}$.

2.7 Simulation study

In this section, we study by simulation the properties of our proposed estimators in terms of bias and root mean square error (RMSE). We also present some simulation evidence on the finite-sample properties of the LR-type test described in Section 2.6.

2.7.1 Estimation

We first investigate the finite-sample properties of the proposed winsorized ARMA-SV estimators for the persistence parameter ϕ, namely $\hat{\phi}_M$, $\hat{\phi}_{LD}$, $\hat{\phi}_{MED}$, and $\hat{\phi}_{OLS}$ (with equal weights). We generate SV processes with $(\phi, \sigma_y, \sigma_v) = (0.95, 0.2, 0.9)$, which represent typical estimates of a financial time series. These values are also representative of those obtained in our empirical study. We consider four different sample sizes ($T = 200, 500, 1000, 10000$) and use 1000 replications. All four winsorized estimators depend on the truncation parameter J, so we consider different values of the truncation parameter ($J = 1, 5, 10, 20, 30, 40, 50, 100$). Note $J = 1$ represents the simple ARMA-SV estimator. The simulation results are reported in Table 2.1, where the average values of the parameter estimates are reported under the different parameter estimates ($\hat{\phi}_M$, $\hat{\phi}_{LD}$, $\hat{\phi}_{MED}$, $\hat{\phi}_{OLS}$), along with the estimated standard errors of the estimators (SD), and the frequencies of inadmissible parameter estimates (NIV).

Clearly, the different estimators perform best with values of J in the range of 5 to 10; low and large values of J produce inferior results. As expected, the performance improves with the sample size T. But the estimators based on relatively simple weighted averages produce inadmissible parameter values even in large samples for higher J. However, the number of unacceptable parameter values decline as the sample size increases. This fact also tells us that the variability of estimated ACF is also going down as the sample size increases. Median and OLS estimates are better than the weighted estimator while OLS based estimates are superior. OLS estimates outperformed other three estimators in terms of bias and standard error, across different sample sizes particularly in small samples. Further, it is also robust to different values of J. From the reported results, there may be a bias-variance trade-off for higher values of J. Finally, we suggest to use OLS for winsorizing and use small values of J for large samples or vice versa.

Second, we compare the statistical performance of the proposed estimators (ARMA-SV and W-ARMA-SV) with alternative estimators (QML, GMM, Bayesian-MCMC, DV). The W-ARMA-SV estimator considered is the OLS-based one with $J = 10$. Bayesian estimates are computed using the R package *stochvol* [Kastner (2016)]. We consider two SV models where parameter values of $(\phi, \sigma_y, \sigma_v)$ are $M_1 = (0.95, 0.2, 0.9)$ and $M_2 = (0.98, 0.025, 1)$. The parameters were selected to represent values often found in empirical applications of hourly or daily returns, where it is observed that ϕ is very close or exactly one and the estimated value of σ_v^2 between 0.01 and 2.77; see Ruiz (1994). The simulations use 1000 replications and we present results for two different sample sizes ($T = 500, 2000$).

In our simulations, we encountered frequent non-convergence problems with GMM estimation. These simulated samples had to be discarded. The DV and simple ARMA-SV estimators of ϕ also occasionally produced values outside the stationary region. These samples were also discarded. So the bias and RMSE obtained are thus conditional on the non-occurrence of non-convergence or inadmissible values. Such problems are practically non-existent with the OLS-based W-ARMA-SV estimator.

Table 2.2 reports the estimation results for model M_1. From this table, we see that the GMM estimator performs poorly in terms of bias and RMSE. The W-ARMA-SV and Bayesian estimators are almost unbiased. For the estimation of ϕ, the DV method yields the biggest RMSE, and the Bayesian method produces the smallest RMSE. The W-ARMA-SV method yields the smallest RMSE for σ_y and the Bayesian estimation yields the smallest RMSE for σ_v. The results for the two sample sizes are qualitatively similar ($T = 500, 2000$) and indicate that estimator precision increases with the sample size.

The results for the M_2 model are reported in Table 2.3. These are very similar to the results of the M_1 model. In this setting, compared to M_1, all methods (except the GMM estimator) have smaller bias and RMSE. These results may be because in this setting the ϕ is nearly unit root. Again the W-ARMA-SV estimator exhibits good statistical properties regarding bias and RMSE than several computationally expensive estimators. Furthermore, from Table 2.4, the winsorized estimator is highly time-efficient and the margin of this time efficiency is enormous compared to other estimators except for the DV estimator.

2.7.2 Testing

We now investigate the finite-sample level and power of the LR-type tests based on the statistic defined in equation (2.6.5). This LR-type test statistic corresponds to the difference between the restricted and the unrestricted optimal values of the objective function; see Section 2.6 for further details.

We consider the following three null hypotheses, which are of fundamental in our setup.

1. No persistence (no clustering) in volatility in the SV model [$\phi = 0$]. In this case, volatility is random but not persistent.

2. The volatility process is not a latent stochastic process [$\sigma_v = 0$]. This problem is an important pre-test before one tries to include a latent stochastic process to drive the dynamics of the log-volatility. If this null hypothesis holds, we have an EGARCH-type model where, given the past, the logarithm of the variance is modeled as a deterministic process.

3. No stochastic volatility [$\phi = 0$ and $\sigma_v = 0$] This is also a crucial pre-test before allowing for conditional heteroskedasticity.

Three ways of implementing the tests are considered: asymptotic critical values, parametric bootstrap, and MMC. Parametric bootstrap (or LMC) tests are performed by replacing the nuisance parameters by their corresponding point estimates. MMC tests involve maximizing the p-value function over the nuisance parameter space.

For MMC tests of the no-persistence hypothesis, we have two nuisance parameters (σ_y, σ_v) and the set $C_T(\phi)$ over which we maximize the simulated p-value is

$$C_T(\phi) = \{(\sigma_y, \sigma_v) : |\sigma_y - \hat{\sigma}_y^0| \leq 0.3,\ \ \sigma_y \geq 0.01, |\sigma_v - \hat{\sigma}_v^0| \leq 1.25,\ \ \sigma_v \geq 0.01\} \tag{2.7.1}$$

where $(\hat{\sigma}_y^0, \hat{\sigma}_v^0)$ are the restricted estimates of (σ_y, σ_v) subject to $\phi = 0$. Note that estimate of ϕ has no influence on the estimate of σ_y, so the restricted and unrestricted estimates are the same. The bounds of 0.30 and 1.25 for the scale parameters correspond to more than 10 standard errors for each and we also restrict them to be positive. Note that, any fixed bound associated with a consistent estimator will lead to an asymptotically valid test provided that the probability of covering the true parameter converges to one as the sample size goes to infinity, see Dufour (2006).

For the second hypothesis (no latent volatility, $\sigma_v = 0$), the nuisance parameters are (ϕ, σ_y) and we maximize the p-value over the set

$$C_T(\sigma_v) = \{(\phi, \sigma_y) : |\, \phi - \hat{\phi}^0 \,| \leq 0.3, \;\; |\, \phi \,| \leq 0.99, |\, \sigma_y - \hat{\sigma}_y^0 \,| \leq 0.3, \;\; \sigma_y \geq 0.01\} \tag{2.7.2}$$

where $(\phi^0, \hat{\sigma}_y^0)$ are the restricted estimates of (ϕ, σ_y) subject to $\sigma_v = 0$ and they are equivalent to the unrestricted estimates. The bounds for nuisance parameters satisfy more than 10 standard errors. Further, we put additional stationarity restriction on ϕ and positivity restriction on σ_y.

For the no-stochastic-volatility hypothesis, the nuisance parameter set is

$$C_T(\phi, \sigma_v) = \{\sigma_y : |\, \sigma_y - \hat{\sigma}_y^0 \,| \leq 0.3, \;\; \sigma_y \geq 0.01\} \tag{2.7.3}$$

where $\hat{\sigma}_y^0$ is the restricted estimate of σ_y under the null hypothesis.

The number of replications used for the Monte Carlo tests is $N = 99$. The nominal level is $\alpha = 0.05$ and the rejection frequencies are estimated from 1000 simulations. T is the sample size of the series y_t which follows the process defined in equations (2.2.1) - (2.2.2). In the power study, the asymptotic critical points are locally level-corrected, *i.e.* the critical points are modified to ensure that the rejection frequency under the null hypothesis (for the specific nuisance parameter values considered) is equal to 0.05. Note that we use the term “locally level-corrected” instead of “size-corrected” because a true size correction would require one to ensure that the probability of rejecting the null hypothesis under all distributions compatible with null hypothesis (*i.e.*, for all values of the nuisance parameters) be less than or equal to the level α. However, finding the appropriate size-corrected critical values requires a numerical search that was not performed in the experiments. The corrected critical value is obtained by simulating the test statistic under the null hypothesis with 10000 replications for asymptotic tests and 1000 replications for MC-type tests. The maximization of MMC tests was done using grid search method and calculations were performed with the R system [MaxMC package of

Dufour and Neves (2018)].

Table 2.5 reports the empirical levels of asymptotic, bootstrap and MMC tests based on LR-type test statistics. In panel A (no-persistence hypothesis), the asymptotic tests severely over-reject the null hypothesis. Level distortions increase with the sample size, indicating that standard critical values are not asymptotically valid. In this case, bootstrap tests control the level very well while MMC tests are conservative. Results on testing that volatility is not a latent process appear in panel B. In this case, all three types of tests control the level. However, asymptotic tests are undersized compared to bootstrap and MMC tests. In panel C (no stochastic volatility hypothesis), the asymptotic tests are not valid (large size distortion) while bootstrap and MMC tests control the type I error.

We report the empirical powers of asymptotic, bootstrap and MMC tests in Table 2.6. In panel A, we report empirical powers for tests of $H_0 : \phi = 0$. We can see from the results that the bootstrap and MMC tests have more power than the asymptotic tests. The bootstrap tests have good power properties even when $T = 250$. Further, although the MMC tests may be conservative in this case, their power is quite close to the bootstrap tests and even perform better when $T = 1000$. For $H_0 : \sigma_v = 0$, we can see from panel B that all three tests have similar power. However, for the no stochastic volatility hypothesis [$H_0 : \sigma_v = 0$ (panel C)], the asymptotic tests are less powerful compared to other tests. Bootstrap and MMC-based LR-type tests exhibit excellent power – which increases with the sample size – and are identical to each other.

2.8 Applications to stock price volatilities

In this section, we estimate the SV model for stock price data. First, we estimate the SV model using our simple estimators and discuss the fit of this model. We then present more reliable inference by exploiting MC tests, since our simple estimators are convenient in the context of simulation-based inference procedures.

The SV model is fitted to daily returns of Coca-Cola (KO), Walmart (WMT), and Ford (F). The price data come from Wharton Research Data Services (WRDS) and the sample period is 02/01/1980 to 31/12/2015, giving us 9081 observations. This period covers the Black

Monday (1987), the Asian financial crisis (1997), the early 2000s recession (Dot-com bubble), the late-2000s Financial Crisis (Subprime mortgage crisis / United States housing bubble) and the recent Russian financial crisis (2014). The prices p_t are transformed into returns $r_t = 100[\log(p_t) - \log(p_{t-1})]$. The returns are converted to residual returns by $y_t = r_t - \hat{\mu}_r$ where $\hat{\mu}_r$ is the sample average of returns [2]. Table 2.7 reports the summary statistics.

Table 2.8 shows the parameter estimates of the SV models obtained by our simple ARMA estimator. To estimate ϕ, we used (2.3.14) with $J = 10$. For all three stocks, these results show that there is strong persistence in the volatility process during the period 1980-2015, and this is statistically significant. Further, we can see that our simple estimator is remarkably time efficient. In Table 2.9, we report the SV parameter estimates with the implicit standard error that we discussed in Section 2.6.3. From Tables 2.8 and 2.9, we can see that the asymptotic standard errors are larger compared to the ISE and this implies that the t-type test based on these asymptotic standard errors under rejecting the null hypothesis.

However, the p-values tabulated in Table 2.8 are based on the usual large-sample approximation based on HAC estimator. The variance-covariance $\hat{V}$ is estimated by a Bartlett kernel estimator with the bandwidth varying with the sample size, *i.e.* $m = [0.159T^{1/3}]$, where $[\cdot]$ denotes the integer part of the enclosed number; see Newey and West (1994). Note that for all three stocks, the empirical estimate of ϕ is close to 1, implying that the volatility processes are highly persistent.

Now, we consider three important test problems: (1) no persistence (or no clustering) hypothesis in the SV model [$\phi = 0$]; (2) volatility process is not a latent stochastic process [$\sigma_v = 0$]; (3) no stochastic volatility hypothesis [$\phi = 0$ and $\sigma_v = 0$]. To construct more reliable inference for these tests, we implemented asymptotic, parametric bootstrap and max-

[2] Note that our simple estimator is based on the log-squared transformation of residual returns minus the mean, *i.e.*, $\log(y_t^2) - \mu$, so we do not need to adjust for the inlier problem. When residuals returns, y_t, are very close to zero, the log-squared transformation yields large negative numbers, and this is so called inlier problem. In the extreme case, if the return is equal to 0, the log-squared transformation is not defined. To solve this problem, Fuller (1996) proposed the following modification of the log-squared transformation:

$$log(y_t^2) \cong log(y_t^2 + cs_y^2) - cs_y^2/(y_t^2 + cs_y^2), \quad t = 1, ..., T,$$

where s_y^2 is the sample variance of y_t and c is a small constant. The effect of this transformation is to reduce the kurtosis in the transformed observations by cutting down the long tail made up of the negative values obtained by taking the logarithms. In other words, it is a form of trimming; see Ghysels et al. (1996).

imized Monte Carlo (MMC) tests based on the GMM-type LR statistic; discussed in Section 2.6.2. Bootstrap and MMC tests are performed using the same procedure described in Section 2.7. However, the maximization was done using Particle Swarm optimization [introduced by Eberhart and Kennedy (1995) and Shi and Eberhart (1998)]. This algorithm evaluates a set of candidate solutions (particles) with random initial positions and the particles are set to move around the search. Compare to simulated annealing or genetic algorithm, it provides a flexible set of controls and methods. The number of replications used for Monte Carlo tests is N = (19, 99, 999) and results are reported in Table 2.10. We can see from the results, that the three versions (asymptotic, bootstrap and MMC) of the LR-type test reject all null hypotheses. These results suggest that the volatility of financial returns is highly persistent, and driven by an additional noise process. Further, a latent stochastic process is a more appropriate specification for the volatility process.

In the above applications, we have a relatively large sample and focus on the OLS-based W-ARMA-SV estimator with $J = 10$. We will now study the daily returns on the S&P 500 index over the period 01/01/2000 – 31/12/2017 (4529 observations). The sample in this case is smaller, and we address two critical issues: (1) the selection of winsorized estimator of ϕ; (2) the choice of J. We then compare the four winsorized estimators of ϕ discussed in Section 2.3 along with different choices of J. In Figure 2.1, we draw the values of the estimated ϕ against the number of lags J for $J = 5, \ldots, 150$. The weighted estimators are usually outside the stationary bound, and the median estimator is oscillating around the MCMC estimate, while the no intercept regression estimator is much closer to the MCMC estimate than the others. We also looked at higher values of J, and a similar pattern is observed.

In Table 2.11, we compare our simple estimates with the Bayesian MCMC ones. To estimate ϕ, we used (2.3.14) with $J = 100$. Bayesian estimates are computed using the R package *stochvol* with 50000 draws after discarding 50000 burn-ins. From the results, we can see that the empirical estimates of these two methods are very similar. However, the difference in elapsed time between these estimations is huge, and the simple estimator is 8214 times faster than the Bayesian estimator. These results confirm that the simple W-ARMA-SV estimator not only produces accurate estimates, but is also vastly more efficient from a computational viewpoint.

2.9 Conclusion

In this paper, we have proposed computationally simple estimators for the SV model, which are based on a small number of moment functions and exploit a winsorization technique. Compared with alternative existing procedures for this model, the proposed class of estimators enjoys a considerable advantage in terms of computation time. Further, it dominates other non-Bayesian estimators and matches the standard Bayesian estimator, in terms of bias and RMSE. The asymptotic distribution of the proposed estimators was studied, and testing procedures based on the new estimators were proposed. Due to its computational simplicity, the proposed class of estimators allow one to build reliable (even exact) simulation-based tests for the SV model.

The results in this paper also underscore the pitfalls of using too many moments in the context of moment-based (or GMM) inference, an important side observation made in other contexts; see Dufour and Taamouti (2003), Dufour and Valéry (2006), and Chao and Swanson (2007). Using many moments can entail large efficiency losses, just like putting too many irrelevant regressors in a linear regression can blow up estimator variances and mean square errors.

We fitted the SV model using our simple estimator to stock return time series. We found that the volatility process highly persistent and near the unit root. We also implemented MC tests to construct more reliable finite-sample inference. We considered three important testing problems (no persistence, latent specification of volatility process, and no stochastic volatility), which are decisively rejected by both asymptotic and finite-sample tests.

One can exploit computationally simple estimators in the context of out-of-sample forecasting. This requires extensive investigation where one may conduct various out-of-sample experiments using different volatility proxies [the squared return and the realized volatility] across different models. This is a potential extension of this paper and we investigate this venture in the following chapter.

2.10 Appendix

2.10.1 Proofs

PROOF OF PROPOSITION 2.3.1 From (2.2.14) - (2.2.15), we have

$$y_t^* = w_t + \epsilon_t, \quad \phi(B)w_t = v_t, \tag{2.10.1}$$

where $\phi(B) = (1 - \phi B)$, and the error processes v_t and ϵ_t are i.i.d. $N(0, \sigma_v^2)$ and i.i.d. $\log(\chi_1^2)$ random variables, respectively. Furthermore, Assumption 2.2.1 implies that v_t's and ϵ_t's are independent. Now, applying $\phi(B)$ to both sides of (2.2.15) yields

$$\phi(B)y_t^* = \phi(B)w_t + \phi(B)\epsilon_t = v_t + \phi(B)\epsilon_t. \tag{2.10.2}$$

Consider the right hand side of (2.10.2). This is clearly a covariance stationary process. By the Wold decomposition theorem it must have a moving average representation. Since the autocovariance function cuts off for lags $k > 1$ it must be an $MA(1)$ process, say $\theta(B)\eta_t = (1 - \theta B)\eta_t$. Hence, y_t^* must be an $ARMA(1,1)$ process; see Granger and Morris (1976). The moving average parameter θ and the white noise variance σ_η^2 of this $ARMA(1,1)$ process can be found by equating the autocovariance function of the right hand side of (2.10.2) with that of $\theta(B)\eta_t$ for lags $k = 0, 1$ and solving the following non-linear equations

$$(1+\theta^2)\sigma_\eta^2 = \sigma_v^2 + (1+\phi^2)\sigma_\epsilon^2, \qquad -\theta\sigma_\eta^2 = -\phi\sigma_\epsilon^2. \tag{2.10.3}$$

Note that there may be multiple solutions, only some of which result in an invertible process.

□

PROOF OF COROLLARY 2.3.2 From Proposition 2.3.1, the process y_t^* satisfies the following equation:

$$y_t^* = \phi y_{t-1}^* + v_t + \epsilon_t - \phi\epsilon_{t-1}. \tag{2.10.4}$$

On multiplying both sides of (2.10.4) by y^*_{t-k} and taking expectation, we have:

$$\gamma_{y^*}(k) = \phi\gamma_{y^*}(k-1) + \mathbb{E}[v_t y^*_{t-k}] + \mathbb{E}[\epsilon_t y^*_{t-k}] - \phi\mathbb{E}[\epsilon_{t-1} y^*_{t-k}]. \tag{2.10.5}$$

We then get:

$$\begin{aligned} \gamma_{y^*}(0) &= \phi\gamma_{y^*}(1) + \mathbb{E}[v_t y^*_t] + \mathbb{E}[\epsilon_t y^*_t] - \phi\mathbb{E}[\epsilon_{t-1} y^*_t] \\ &= \phi\gamma_{y^*}(1) + \sigma^2_v + \sigma^2_\epsilon - \phi\mathbb{E}[\epsilon_{t-1}(\phi y^*_{t-1} - \phi\epsilon_{t-1})] \\ &= \phi\gamma_{y^*}(1) + \sigma^2_v + \sigma^2_\epsilon + \phi^2\sigma^2_\epsilon - \phi^2\sigma^2_\epsilon \\ &= \phi\gamma_{y^*}(1) + \sigma^2_v + \sigma^2_\epsilon, \end{aligned} \tag{2.10.6}$$

$$\begin{aligned} \gamma_{y^*}(1) &= \phi\gamma_{y^*}(0) + \mathbb{E}[v_t y^*_{t-1}] + \mathbb{E}[\epsilon_t y^*_{t-1}] - \phi\mathbb{E}[\epsilon_{t-1} y^*_{t-1}] \\ &= \phi\gamma_{y^*}(0) + 0 + 0 - \phi\sigma^2_\epsilon = \phi\gamma_{y^*}(0) - \phi\sigma^2_\epsilon, \end{aligned} \tag{2.10.7}$$

and, for $k \geq 2$,

$$\begin{aligned} \gamma_{y^*}(k) &= \phi\gamma_{y^*}(k-1) + \mathbb{E}[v_t y^*_{t-k}] + \mathbb{E}[\epsilon_t y^*_{t-k}] - \phi\mathbb{E}[\epsilon_{t-1} y^*_{t-k}] \\ &= \phi\gamma_{y^*}(k-1) + 0 + 0 - 0 = \phi\gamma_{y^*}(k-1). \end{aligned} \tag{2.10.8}$$

Combining (2.10.6), (2.10.7), and (2.10.8), we obtain the autocovariance structure of the process y^*_t stated in the Corollary. □

PROOF OF COROLLARY 2.3.3 The estimator of ϕ is based on the autocovariance structure of the process y^*_t. By (2.3.3), we have for $k \geq 2$:

$$\phi = \frac{\gamma_{y^*}(k)}{\gamma_{y^*}(k-1)} \tag{2.10.9}$$

which yields (2.3.4). From (2.3.3) with $k = 1$ we get

$$\gamma_{y^*}(1) = \phi\gamma_{y^*}(0) - \phi\sigma^2_\epsilon \tag{2.10.10}$$

and, substituting $\gamma_{y^*}(1)$ into (2.3.3) with $k = 0$,

$$\gamma_{y^*}(0) = \phi[\phi\gamma_{y^*}(0) - \phi\sigma_\epsilon^2] + \sigma_v^2 + \sigma_\epsilon^2 . \tag{2.10.11}$$

hence, using $\sigma_\epsilon^2 = \pi^2/2$ [see (2.2.10)],

$$\sigma_v^2 = (1-\phi^2)[\gamma_{y^*}(0) - \sigma_\epsilon^2] = (1-\phi^2)[\gamma_{y^*}(0) - \pi^2/2] . \tag{2.10.12}$$

Finally, by definition,

$$\mu = \mathbb{E}[\log(y_t^2)] = \log(\sigma_y^2) + \mathbb{E}[\log(z_t^2)] = \log(\sigma_y^2) - 1.2704 \tag{2.10.13}$$

or, equivalently

$$\sigma_y^2 = \exp(\mu + 1.2704) . \tag{2.10.14}$$

$\square$

PROOF OF LEMMA 2.5.1 Under the assumptions 2.5.1 and 2.5.2 with $s = 2$, the process $\{y_t^*\}$ is strictly stationarity and geometrically ergodic with $\mathbb{E}[y_t^*] < \infty$ and $\mathbb{E}[y_t^* y_{t+k}^*] < \infty$. So the consistency property follows by the application of the Law of Large Numbers for stationary ergodic processes (*i.e.*, the Ergodic Theorem); see Davidson (1994, Theorem 13.12 and Corollary 13.14). $\square$

PROOF OF LEMMA 2.5.2 To establish the asymptotic normality of empirical moments, we shall use a central limit theorem (CLT) for dependent processes [see Davidson (1994, Theorem 24.5, p. 385)]. For that purpose, we first check the conditions under which this CLT holds. Set

$$X_t := \begin{bmatrix} \Psi_t \\ \Lambda_t \end{bmatrix}, \quad \Psi_t := \log(y_t^2) - \mu, \quad \Lambda_t := [\Lambda_{t,0}, \Lambda_{t,1}, \ldots, \Lambda_{t,m}]', \tag{2.10.15}$$

$$\Lambda_{t,k} := y_t^* y_{t+k}^* - \gamma_{y^*}(k) = [\log(y_t^2) - \mu][\log(y_{t+k}^2) - \mu] - \gamma_{y^*}(k) , \quad k = 0, 1, \ldots, m , \tag{2.10.16}$$

$$S_T := \sum_{t=1}^{T} X_t = \begin{bmatrix} \sum_{t=1}^{T} \Psi_t \\ \sum_{t=1}^{T} \Lambda_t \end{bmatrix}, \tag{2.10.17}$$

and consider the subfields $\mathcal{F}_t = \sigma(s_t, s_{t-1}, \ldots)$ where $s_t = (y_t, w_t)'$. We will now show that

$$T^{-1/2} S_T \xrightarrow{d} \mathrm{N}\left(0, \begin{bmatrix} V_\mu & C'_{\mu,\Gamma(m)} \\ C_{\mu,\Gamma(m)} & V_{\Gamma(m)} \end{bmatrix}\right), \tag{2.10.18}$$

which in turn yields (2.5.2). To do this, we will check the following conditions:

(i) $\{X_t, \mathcal{F}_t\}$ is stationary and ergodic;

(ii) $\{X_t, \mathcal{F}_t\}$ is a L_1-mixingale of size -1;

(iii) $\limsup_{T\to\infty} T^{-1/2}\mathbb{E}\,\|S_T\| < \infty$, where $\|\cdot\|$ is the Euclidean norm.

(i) The fact that $\{X_t, \mathcal{F}_t\}$ is stationary and ergodic follows from results 2.4.1 and 2.4.2.

(ii) - (1) A mixing zero-mean process is an adapted L_1-mixingale with respect to the sub-fields $\mathcal{F}_t$ provided it is bounded in the L_1-norm [see Davidson (1994, Theorem 14.2, p. 211)]. To see that $\{X_t\}$ is bounded in the L_1-norm, we note that:

$$\mathbb{E}|\log(y_t^2) - \mu| = \mathbb{E}|y_t^*| \le (\mathbb{E}|y_t^*|^2)^{1/2} = (\mathbb{E}[y_t^{*2}])^{1/2} = \sqrt{\gamma_{y^*}(0)} < \infty, \tag{2.10.19}$$

$$\begin{aligned} \mathbb{E}|y_t^* y_{t+k}^* - \gamma_{y^*}(k)| &= \mathbb{E}|y_t^* y_{t+k}^*| - |\gamma_{y^*}(k)| \\ &\le \mathbb{E}|y_t^* y_{t+k}^*| \\ &\le (\mathbb{E}|y_t^*|^2)^{1/2} (\mathbb{E}|y_{t+k}^*|^2)^{1/2} \\ &= (\mathbb{E}[y_t^{*2}])^{1/2} (\mathbb{E}[y_{t+k}^{*2}])^{1/2} \\ &= \mathbb{E}[y_t^{*2}] = \gamma_{y^*}(0) < \infty, \quad \text{for } k = 0, 1, \ldots, m, \end{aligned} \tag{2.10.20}$$

where the inequality in (2.10.19) is the application of Lyapunov's inequality and the second inequality in (2.10.20) follows from the Hölder's inequality.

(ii) - (2) We now show that $\{X_t, \mathcal{F}_t\}$ is a L_1-mixingale of size -1. From the discussion in Section 2.4, we know that X_t is β-mixing, so it has mixing coefficients of the type $\beta_T = \psi \rho^T$,

$\psi > 0$, $0 < \rho < 1$. To show that $\{X_t\}$ is of size -1, its mixing coefficients β_T must be $O(T^{-\varphi})$, with $\varphi > 1$ [see Davidson (1994, Definition 16.1, p. 247)]. Indeed,

$$\frac{\rho^T}{T^{-\varphi}} = T^{\varphi}\exp(T\log\rho) = \exp(\varphi\log T)\exp(T\log\rho) = \exp[\varphi(\log T) + T(\log\rho)]. \tag{2.10.21}$$

Since $\lim_{T\to\infty}[\varphi(\log T) + T(\log\rho)] = -\infty$, we get

$$\lim_{T\to\infty}\exp[\varphi(\log T) + T(\log\rho)] = 0. \tag{2.10.22}$$

This holds in particular for $\varphi > 1$; see Rudin (1976, Theorem 3.20(d), p. 57).

(iii) To show that $\limsup_{T\to\infty} T^{-1/2}\mathbb{E}\|S_T\| < \infty$, we first observe that $\mathbb{E}(S_T) = 0$ and, using the Cauchy-Schwarz inequality,

$$\begin{aligned}(T^{-1/2}\mathbb{E}\|S_T\|)^2 &\le \frac{1}{T}\mathbb{E}(\|S_T\|^2) = \frac{1}{T}\mathbb{E}(S_T'S_T) = \frac{1}{T}\mathrm{tr}[\mathbb{E}(S_TS_T')] = \frac{1}{T}\mathrm{tr}[\mathrm{Var}(S_T)]\\ &= \mathrm{tr}[\mathrm{Var}(T^{-1/2}S_T)].\end{aligned} \tag{2.10.23}$$

It is thus sufficient to show that

$$\limsup_{T\to\infty}\mathrm{tr}[\mathrm{Var}(T^{-1/2}S_T)] < \infty. \tag{2.10.24}$$

We now consider separately the components Ψ_t and Λ_t of X_t.

(iii) - (1) Set

$$S_{T1} := \sum_{t=1}^{T}\Psi_t, \quad \zeta_\Psi(\tau) := \mathrm{cov}(\Psi_t, \Psi_{t+\tau}). \tag{2.10.25}$$

Then

$$\zeta_\Psi(\tau) = \mathbb{E}[(\log(y_t^2) - \mu)(\log(y_{t+\tau}^2) - \mu)] = \mathbb{E}[y_t^* y_{t+\tau}^*] = \gamma_{y^*}(\tau), \tag{2.10.26}$$

$$\begin{aligned}\mathrm{Var}(T^{-1/2}S_{T1}) &= \frac{1}{T}\left[\sum_{t=1}^{T}\mathrm{Var}(\Psi_t) + \sum_{t\neq s}\mathrm{cov}(\Psi_t, \Psi_s)\right] = \frac{1}{T}\left[T\zeta_\Psi(0) + 2\sum_{\tau=1}^{T}(T-\tau)\zeta_\Psi(\tau)\right]\\ &= \zeta_\Psi(0) + 2\sum_{\tau=1}^{T}(1-\frac{\tau}{T})\zeta_\Psi(\tau) = \gamma_{y^*}(0) + 2\sum_{\tau=1}^{T}(1-\frac{\tau}{T})\gamma_{y^*}(\tau),\end{aligned} \tag{2.10.27}$$

hence

$$\begin{aligned}
\limsup_{T\to\infty} \mathrm{Var}(T^{-1/2}S_{T1}) &= \limsup_{T\to\infty} [\gamma_{y^*}(0) + 2\sum_{\tau=1}^{T}(1-\frac{\tau}{T})\gamma_{y^*}(\tau)] \\
&= \gamma_{y^*}(0) + 2\sum_{\tau=1}^{\infty}\gamma_{y^*}(\tau) = \sum_{\tau=-\infty}^{\infty}\gamma_{y^*}(\tau) \\
&\leq \sum_{\tau=-\infty}^{\infty}|\gamma_{y^*}(\tau)| < \infty. \qquad (2.10.28)
\end{aligned}$$

This convergence is due to the fact that y_t^* follows an ARMA$(1,1)$ process with $|\phi| < 1$. So y_t^* can be viewed as an MA(∞) process with absolutely summable coefficients, which implies the absolute summability of autocovariances [see Hamilton (1994, chapter 3, page 52)]. This entails

$$\limsup_{T\to\infty} T^{-1/2}\mathbb{E}|S_{T1}| < \infty. \qquad (2.10.29)$$

(iii) - (2) Set

$$S_{T2} := \sum_{t=1}^{T} \Lambda_t = [S_{T2,0}, S_{T2,1}, \ldots, S_{T2,m}]', \qquad (2.10.30)$$

$$S_{T2,k} := \sum_{t=1}^{T} \Lambda_{t,k}, \quad \zeta_{\Lambda_k}(\tau) := \mathrm{cov}(\Lambda_{t,k}, \Lambda_{t+\tau,k}), \quad k = 0, 1, \ldots, m. \qquad (2.10.31)$$

Then, for $k = 0, 1, \ldots, m$,

$$\begin{aligned}
\zeta_{\Lambda_k}(\tau) &= \mathbb{E}[\big(y_t^* y_{t+k}^* - \gamma_{y^*}(k)\big)\big(y_{t+\tau}^* y_{t+\tau+k}^* - \gamma_{y^*}(k)\big)] = \mathbb{E}[y_t^* y_{t+k}^* y_{t+\tau}^* y_{t+\tau+k}^*] - \gamma_{y^*}(k)^2 \\
&= \mathbb{E}[y_t^* y_{t+k}^*]\mathbb{E}[y_{t+\tau}^* y_{t+\tau+k}^*] + \mathrm{cov}(y_t^*, y_{t+\tau}^*)\mathrm{cov}(y_{t+k}^*, y_{t+\tau+k}^*) \\
&\quad + \mathrm{cov}(y_t^*, y_{t+\tau+k}^*)\mathrm{cov}(y_{t+k}^*, y_{t+\tau}^*) - \gamma_{y^*}(k)^2 \\
&= \gamma_{y^*}(k)^2 + \gamma_{y^*}(\tau)^2 + \gamma_{y^*}(\tau+k)\gamma_{y^*}(\tau-k) - \gamma_{y^*}(k)^2 \\
&= \gamma_{y^*}(\tau)^2 + \gamma_{y^*}(\tau+k)\gamma_{y^*}(\tau-k), \qquad (2.10.32)
\end{aligned}$$

hence

$$\mathrm{Var}(T^{-1/2}S_{T2,k}) = \frac{1}{T}\left[\sum_{t=1}^{T}\mathrm{Var}(\Lambda_{t,k}) + \sum_{t\neq s}\mathrm{cov}(\Lambda_{t,k}, \Lambda_{s,k})\right] = \frac{1}{T}\left[T\zeta_{\Lambda_k}(0) + 2\sum_{\tau=1}^{T}(T-\tau)\zeta_{\Lambda_k}(\tau)\right]$$

$$
\begin{aligned}
&= \zeta_{\Lambda_k}(0) + 2\sum_{\tau=1}^{T}(1-\frac{\tau}{T})\zeta_{\Lambda_k}(\tau) \\
&= \gamma_{y^*}(0)^2 + \gamma_{y^*}(k)\gamma_{y^*}(-k) \\
&\quad +2\sum_{\tau=1}^{T}(1-\frac{\tau}{T})[\gamma_{y^*}(\tau)^2 + \gamma_{y^*}(\tau+k)\gamma_{y^*}(\tau-k)]\,,
\end{aligned}
\tag{2.10.33}
$$

and

$$
\begin{aligned}
\limsup_{T\to\infty} \mathrm{Var}(T^{-1/2}S_{T2,k}) &= \gamma_{y^*}(0)^2 + \gamma_{y^*}(k)\gamma_{y^*}(-k) \\
&\quad + \limsup_{T\to\infty}[2\sum_{\tau=1}^{T}(1-\frac{\tau}{T})[\gamma_{y^*}(\tau)^2 + \gamma_{y^*}(\tau+k)\gamma_{y^*}(\tau-k)]] \\
&= \sum_{\tau=-\infty}^{\infty}[\gamma_{y^*}(\tau)^2 + \gamma_{y^*}(\tau+k)\gamma_{y^*}(\tau-k)] \\
&= \sum_{\tau=-\infty}^{\infty}\gamma_{y^*}(\tau)^2 + \sum_{\tau=-\infty}^{\infty}\gamma_{y^*}(\tau+k)\gamma_{y^*}(\tau-k) \\
&= \sum_{\tau=-\infty}^{\infty}\gamma_{y^*}(\tau)^2 + \sum_{\tau=-\infty}^{\infty}\gamma_{y^*}^2(\tau+k) < \infty.
\end{aligned}
\tag{2.10.34}
$$

This convergence is due to the fact that absolute summability implies square-summability. We deduce that

$$
\limsup_{T\to\infty} T^{-1/2}\mathbb{E}\left|S_{T2,k}\right| < \infty, \quad k = 0, 1, \ldots, m. \tag{2.10.35}
$$

Combining (2.10.29) and (2.10.35), we get, for any $(m+2)\times 1$ fixed real vector $a \neq 0$,

$$
\limsup_{T\to\infty} T^{-1/2}\mathbb{E}\left|a' S_T\right| < \infty. \tag{2.10.36}
$$

It is also clear properties (i) and (ii) also hold if we replace S_T by $a'S_T$. Thus we can apply Theorem 24.5 of Davidson (1994) to $a'S_T$ to state that $T^{-1/2}(a'S_T)$ is asymptotically normal. Since this holds for any $a \neq 0$, it follows from the Cramér-Wold theorem that $T^{-1/2}\sum_{t=1}^{T} X_t$ is asymptotically multinormal:

$$
T^{-1/2}S_T = T^{-1/2}\sum_{t=1}^{T} X_t = \sqrt{T}\begin{bmatrix} \hat{\mu} - \mu \\ \hat{\Gamma}(m) - \Gamma(m) \end{bmatrix} \xrightarrow{d} \mathrm{N}[0, V], \tag{2.10.37}
$$

where

$$V = \lim_{T\to\infty} \mathbb{E}\{[T^{-1/2}S_T][T^{-1/2}S_T]'\}, \tag{2.10.38}$$

$$V = \begin{bmatrix} V_\mu & C'_{\mu,\Gamma(m)} \\ C_{\mu,\Gamma(m)} & V_{\Gamma(m)} \end{bmatrix}. \tag{2.10.39}$$

Using (2.10.28) and (2.10.33), we have:

$$V_\mu = \gamma_{y^*}(0) + 2\sum_{\tau=1}^{\infty} \gamma_{y^*}(\tau), \tag{2.10.40}$$

$$V_{\Gamma(m)} = \mathrm{Var}(\Lambda_t) + 2\sum_{\tau=1}^{\infty} \mathrm{cov}(\Lambda_t, \Lambda_{t+\tau}), \tag{2.10.41}$$

$$C_{\mu,\Gamma(m)} = [C_{\mu 0}, C_{\mu 1}, \ldots, C_{\mu m}]', \tag{2.10.42}$$

$$\begin{aligned} C_{\mu k} &= \sum_t \mathrm{cov}(\Psi_t, \Lambda_{t,k}) = 2\sum_{t=1}^{\infty} \mathbb{E}[\Psi_t\big(y_t^* y_{t+k}^* - \gamma_{y^*}(k)\big)] \\ &= 2\sum_{t=1}^{\infty} \mathbb{E}[y_t^*\big(y_t^* y_{t+k}^* - \gamma_{y^*}(k)\big)] = 2\sum_{t=1}^{\infty} [\mathbb{E}(y_t^{*2} y_{t+k}^*) - \mathbb{E}(y_t^*)\gamma_{y^*}(k)] \\ &= 2\sum_{t=1}^{\infty} \mathbb{E}(y_t^{*2} y_{t+k}^*), \quad k = 0, 1, 2, \ldots, m. \end{aligned} \tag{2.10.43}$$

Further, for $k = 0$, we substitute $y_t^* = w_t + \epsilon_t$ to get

$$\bar{c} := C_{\mu 0} = 2\sum_{t=1}^{\infty} \mathbb{E}(y_t^{*3}) = 2\sum_{t=1}^{\infty} [\mathbb{E}(w_t^3) + \mathbb{E}(\epsilon_t^3)] = 2\sum_{t=1}^{\infty} \mathbb{E}(\epsilon_t^3). \tag{2.10.44}$$

Since $\{z_t\}$ is a sequence of i.i.d. N[0, 1] random variables, we have $\mathbb{E}(\epsilon_t^3) = \psi^{(2)}(\frac{1}{2})$ [see (2.2.11)], which is equal to $-14\mathcal{Z}(3)$ where $\mathcal{Z}(\cdot)$ is Riemann's Zeta function with $\mathcal{Z}(3) = 1.20205$.[3] For $k = 1, \ldots, m$, it is easily seen that $C_{\mu k} = 0$ from the MA(∞) representation of w_t. So $C_{\mu,\Gamma(m)}$ is a $(m+1)\times 1$ vector given by $(\bar{c}, 0_{[m\times 1]})'$, with $\bar{c}$ is defined in (2.10.44). Finally, (2.5.2) follows on observing that

$$\sqrt{T}\begin{bmatrix} \hat{\mu} - \mu \\ \hat{\Gamma}(m) - \Gamma(m) \end{bmatrix} - T^{-1/2}S_T \xrightarrow[T\to\infty]{p} 0. \tag{2.10.45}$$

[3] The Riemann Zeta function for $s \in \mathbb{C}$ with $\mathrm{Re}(s) > 1$ is defined as $\mathcal{Z}(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}$.

□

PROOF OF THEOREM 2.5.3 It is easily seen that D is a continuously differentiable mapping of $(\mu, \gamma_{y^*}(0), \gamma_{y^*}(1), \gamma_{y^*}(2))$. The convergence result stated in (2.5.7) follows from the standard result for differentiable transformations of asymptotically normally distributed variables together with the application of multivariate delta method. □

2.10.2 Figures

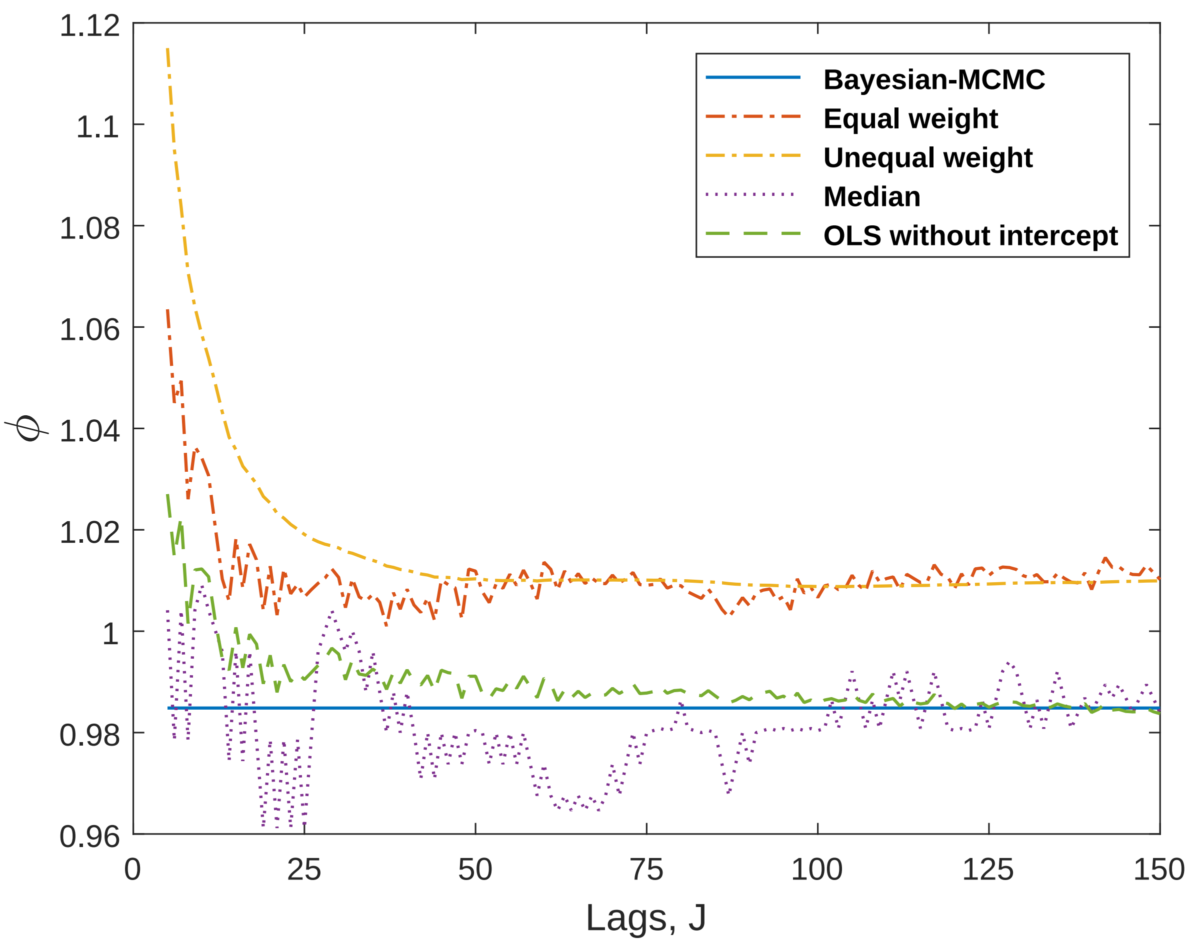

Figure 2.1. S&P 500: Four W-ARMA-SV estimators of ϕ as a function of the number of lags. The solid reference line is the Bayesian estimator based on the MCMC method.

2.10.3 Tables

Table 2.1. Comparison of different winsorized ARMA-SV estimators (W-ARMA-SV)
SV model with parameters: $(\phi, \sigma_y, \sigma_v) = (0.95, 0.2, 0.9)$

J	$\hat{\phi}_M$	SD	NIV	$\hat{\phi}_{LD}$	SD	NIV	$\hat{\phi}_{MED}$	SD	NIV	$\hat{\phi}_{OLS}$	SD	NIV
						$T = 200$						
1	0.89	0.083	218	0.89	0.083	218	0.89	0.083	218	0.89	0.083	218
5	0.93	0.051	36	0.93	0.045	33	0.91	0.064	73	**0.91**	**0.055**	0
10	0.91	0.122	64	0.93	0.064	35	0.91	0.074	29	0.91	0.058	0
20	0.82	0.310	237	0.89	0.179	146	0.89	0.094	13	0.90	0.064	0
30	0.74	0.359	375	0.81	0.290	261	0.89	0.105	15	0.89	0.068	0
40	0.66	0.410	447	0.75	0.348	350	0.89	0.110	13	0.89	0.072	0
50	0.61	0.409	492	0.71	0.370	406	0.89	0.107	19	0.89	0.075	0
100	0.61	0.421	520	0.63	0.401	485	0.90	0.109	25	0.90	0.081	0
						$T = 500$						
1	0.93	0.042	123	0.93	0.042	123	0.93	0.042	123	0.93	0.042	123
5	0.94	0.023	1	0.94	0.023	1	0.94	0.028	13	0.94	0.023	0
10	0.94	0.024	0	**0.94**	**0.021**	0	0.94	0.026	1	0.94	0.023	0
20	0.93	0.119	33	0.94	0.053	7	0.94	0.034	0	0.94	0.024	0
30	0.89	0.199	134	0.92	0.140	67	0.93	0.036	0	0.94	0.025	0
40	0.85	0.252	254	0.89	0.190	150	0.93	0.039	2	0.94	0.025	0
50	0.81	0.307	335	0.87	0.235	232	0.93	0.042	3	0.94	0.025	0
100	0.72	0.355	478	0.78	0.308	407	0.94	0.044	5	0.94	0.026	0
						$T = 1000$						
1	0.94	0.030	61	0.94	0.030	61	0.94	0.030	61	0.94	0.030	61
5	0.95	0.015	0	0.95	0.015	0	0.95	0.019	2	0.95	0.015	0
10	0.95	0.015	0	**0.95**	**0.014**	0	0.95	0.017	0	**0.95**	**0.014**	0
20	0.94	0.027	1	0.95	0.016	0	0.94	0.019	0	0.94	0.015	0
30	0.93	0.076	47	0.94	0.065	11	0.94	0.022	0	0.94	0.015	0
40	0.89	0.216	146	0.92	0.145	65	0.94	0.023	0	0.94	0.015	0
50	0.87	0.232	244	0.90	0.170	132	0.94	0.024	1	0.94	0.015	0
100	0.74	0.359	458	0.82	0.278	352	0.95	0.027	1	0.95	0.016	0
						$T = 10000$						
1	0.95	0.009	0	0.95	0.009	0	0.95	0.009	0	0.95	0.009	0
5	**0.95**	**0.004**	0	**0.95**	**0.004**	0	0.95	0.005	0	**0.95**	**0.004**	0
10	**0.95**	**0.004**	0	**0.95**	**0.004**	0	0.95	0.005	0	**0.95**	**0.004**	0
20	0.95	0.005	0	**0.95**	**0.004**	0	0.95	0.005	0	**0.95**	**0.004**	0
30	0.95	0.007	0	0.95	0.005	0	0.95	0.006	0	0.95	0.005	0
40	0.95	0.010	1	0.95	0.020	0	0.95	0.007	0	0.95	0.005	0
50	0.94	0.086	18	0.95	0.035	4	0.95	0.007	0	0.95	0.005	0
100	0.84	0.279	314	0.89	0.185	170	0.95	0.009	0	0.95	0.005	0

Note – The estimators compared are the simple and winsorized estimators defined in Section 2.3. The estimated means of the different estimators are reported below the corresponding columns ($\hat{\phi}_M$, $\hat{\phi}_{LD}$, $\hat{\phi}_{MED}$, $\hat{\phi}_{OLS}$). For $J = 1$, all the estimators reduce to the simple ARMA-SV estimator. SD is the estimated standard error based on the simulation. NIV stands for the number of inadmissible parameter values produced by the estimators (over 1000).

Table 2.2. Comparison of different estimation methods for the SV model: bias and RMSE
Model: $M_1 = (0.95, 0.2, 0.9)$

	$T = 500$			$T = 2000$		
	ϕ	σ_y	σ_v	ϕ	σ_y	σ_v
True value	0.95	0.2	0.9	0.95	0.2	0.9
				Bias		
QML	-0.0135	0.0174	0.0285	-0.0031	0.0053	0.0070
GMM	0.1392	0.0367	-0.5435	0.0923	0.0487	-0.1032
Bayesian-MCMC	-0.0097	0.0167	**0.0221**	**-0.0025**	0.0052	0.0065
DV*	-0.1764	0.3804	0.0160	-0.1143	0.3379	0.0334
ARMA-SV**	-0.0152	0.0138	0.0148	-0.0037	0.0014	-0.0070
W-ARMA-SV($J = 10$)	**-0.0109**	**0.0135**	0.0257	-0.0027	**0.0016**	**0.0027**
				RMSE		
QML	0.0279	0.0957	0.1312	0.0104	0.0441	0.0609
GMM	0.2470	0.0861	0.6381	0.1308	0.0682	0.3802
Bayesian-MCMC	**0.0211**	0.0882	**0.0828**	**0.0088**	0.0428	**0.0411**
DV*	0.2526	0.5214	0.4032	0.1727	0.3860	0.4284
ARMA-SV**	0.0450	0.0842	0.2934	0.0222	0.0415	0.1822
W-ARMA-SV($J = 10$)	0.0254	0.**0838**	0.1213	0.0103	**0.0415**	0.0602

Notes:
1. GMM is the generalized method of moment estimator of Andersen and Sørensen (1996) with 24 moments.
2. QML is the quasi-maximum likelihood estimator of Ruiz (1994).
3. Bayesian-MCMC is the Bayesian estimator based on Markov Chain Monte Carlo methods proposed by Jacquier et al. (1994).
4. We used R package *stochvol* of Kastner (2016) for the Bayesian estimation.
5. DV is the simple moment estimator of Dufour and Valéry (2006).
6. ARMA-SV is the simple ARMA based estimator proposed in Section 2.3 with no winsorizing.
7. W-ARMA-SV is the winsorized ARMA estimator based on OLS with $J = 10$ that proposed in Section 2.3.
8. *DV produces 356 and 362 inadmissible values, out of 1000 simulations, of ϕ when $T = 500$ and $T = 2000$, respectively.
9. **ARMA-SV produces 123 and 14 inadmissible values, out of 1000 simulations, of ϕ when $T = 500$ and $T = 2000$, respectively.
10. Bias and RMSE of DV/ARMA-SV method are calculated from the acceptable values of $\hat{\phi}$ only. These RMSEs are not comparable to other rows.

Table 2.3. Comparison of different estimation methods for the SV model: bias and RMSE
Model: $M_2 = (0.98, 0.025, 1)$

	$T = 500$			$T = 2000$		
	ϕ	σ_y	σ_v	ϕ	σ_y	σ_v
True value	0.98	0.025	1	0.98	0.025	1
			Bias			
QML	-0.0109	0.0224	0.0231	-0.0024	0.0048	0.0054
GMM	-0.1429	-0.0830	-0.3149	-0.0754	-0.0739	-0.1009
Bayesian-MCMC	**-0.0089**	0.0188	**0.0207**	**-0.0020**	0.0047	0.0058
DV*	-0.2016	0.7273	-0.0135	-0.1334	0.8745	-0.0091
ARMA-SV**	-0.0096	0.0149	-0.0066	-0.0024	0.0031	-0.0065
W-ARMA-SV($J = 10$)	-0.0093	**0.0152**	0.0315	-0.0022	**0.0032**	**0.0018**
			RMSE			
QML	0.0185	0.1294	0.1199	0.0061	0.0193	0.0555
GMM	0.6571	0.1306	0.6899	0.5415	0.0952	0.6842
Bayesian-MCMC	**0.0155**	0.0644	**0.0817**	**0.0056**	0.0192	**0.0409**
DV*	0.2725	2.0592	0.4407	0.1823	1.7896	0.4524
ARMA-SV**	0.0228	0.0511	0.3103	0.0093	0.0178	0.1757
W-ARMA-SV($J = 10$)	0.0167	**0.0514**	0.1264	0.0060	**0.0178**	0.0593

Notes:
1. GMM is the generalized method of moment estimator of Andersen and Sørensen (1996) with 24 moments.
2. QML is the quasi-maximum likelihood estimator of Ruiz (1994).
3. Bayesian-MCMC is the Bayesian estimator based on Markov Chain Monte Carlo methods proposed by Jacquier et al. (1994).
4. We used R package *stochvol* of Kastner (2016) for the Bayesian estimation.
5. DV is the simple moment estimator of Dufour and Valéry (2006).
6. ARMA-SV is the simple ARMA based estimator proposed in Section 2.3 with no winsorizing.
7. W-ARMA-SV is the winsorized ARMA estimator based on OLS with $J = 10$ that proposed in Section 2.3.
8. *DV produces 365 and 361 inadmissible values, out of 1000 simulations, of ϕ when $T = 500$ and $T = 2000$, respectively.
9. **ARMA-SV produces 70 and 5 inadmissible values, out of 1000 simulations, of ϕ when $T = 500$ and $T = 2000$, respectively.
10. Bias and RMSE of DV/ARMA-SV method are calculated from the acceptable values of $\hat{\phi}$ only. These RMSEs are not comparable to other rows.

Table 2.4. Comparison of different estimation methods with respect to relative time for the SV model using simulated data

Relative computing time with respect to W-ARMA-SV estimator			
	$T = 500$	$T = 1000$	$T = 2000$
QML	67.21	76.14	86.14
GMM	225.23	234.22	245.57
Bayesian-MCMC	1055.92	1099.87	1133.98
DV	1.03	1.03	1.06
W-ARMA-SV($J = 10$)	1.00	1.00	1.00

Table 2.5. Empirical levels of asymptotic, bootstrap and MMC tests based on LR-type statistic, (nominal size: α = 5%).

	(A) No persistence			(B) No latent process			(C) No stochastic volatility		
	$H_0 : \phi = 0$			$H_0 : \sigma_v = 0$			$H_0 : \phi = \sigma_v = 0$		
	$\sigma_y = 1, \sigma_v = 2$			$\phi = 0.5, \sigma_y = 1$			$\sigma_y = 1$		
T	Asy	Bootstrap	MMC	Asy	Bootstrap	MMC	Asy	Bootstrap	MMC
250	23.8	5.7	1.2	2.2	4.7	4.6	27.1	4.2	4.2
500	22.0	4.4	0.4	2.7	4.9	4.9	25.5	5.7	5.7
750	24.3	4.5	1.3	2.9	5.2	5.2	25.1	5.0	5.0
1000	24.5	5.9	1.3	2.1	5.4	5.4	24.2	6.3	6.3

Notes: Rejection frequencies are reported in percentages. Simulations are computed on 1000 replications.

Table 2.6. Empirical powers of asymptotic, bootstrap and MMC tests based on LR-type statistic
(nominal size: α = 5%).

	(A) No persistence			(B) No latent process			(C) No stochastic volatility		
	$H_0 : \phi = 0$			$H_0 : \sigma_v = 0$			$H_0 : \phi = \sigma_v = 0$		
	$H_1 : \phi = 0.5$			$H_1 : \sigma_v = 1$			$H_1 : \phi = 0.95, \sigma_v = 2$		
	$\sigma_y = 1, \sigma_v = 2$			$\phi = 0.5, \sigma_y = 1$			$\sigma_y = 1$		
T	Asy	Bootstrap	MMC	Asy	Bootstrap	MMC	Asy	Bootstrap	MMC
250	64.5	80.1	66.4	24.4	24.4	24.4	16.7	39.4	39.4
500	74.7	93.4	88.4	50.3	50.1	50.1	47.3	67.8	67.8
750	78.9	97.7	96.0	71.6	71.7	71.7	61.6	82.2	82.2
1000	80.2	98.2	98.7	82.4	81.6	81.6	69.4	88.7	88.7

Notes: Rejection frequencies are reported in percentages. Simulations are computed on 1000 replications. Asymptotic tests and MC tests are locally level-corrected when the probability of type I error exceeds 0.05.

Table 2.7. Summary Statistics

	Series	Mean	SD	Kurtosis	SK	Range	Max	Min	LB(10)
Coca Cola (KO)	y_t	0.00	1.04	762.73	-20.69	54.57	7.80	-46.77	16.72
	y_t^2	1.07	29.57	3645.53	56.29	2187.61	2187.61	0.00	0.07
	y_t^*	0.00	3.05	7.59	-1.91	21.22	10.48	-10.74	423.11
Walmart (WMT)	y_t	0.00	1.15	353.76	-13.93	35.41	5.09	-30.32	25.95
	y_t^2	1.33	24.93	1156.64	33.83	919.59	919.59	0.00	0.27
	y_t^*	0.00	2.96	5.97	-1.60	18.77	9.26	-9.52	523.48
Ford (F)	y_t	0.00	1.18	128.97	-5.28	41.68	11.24	-30.44	19.31
	y_t^2	1.40	15.84	2575.61	47.31	926.74	926.74	0.00	15.88
	y_t^*	0.00	2.86	6.32	-1.65	17.93	8.72	-9.21	507.36

Notes:

1. $y_t = r_t - \hat{\mu}_r$ is the residual return, y_t^2 is the square of residual return and y_t^* is the residual of log square of residual return.
2. LB(10) is the heteroskedasticity-corrected Ljung - Box statistics with 10 lags. The critical values for LB(10) are: 15.99 (10%), 18.31 (5%), and 23.21 (1%).

Table 2.8. ARMA-type estimates of the SV model

		Coca Cola (KO)	Walmart (WMT)	Ford (F)
ϕ		0.9119 (0.0353)	0.9218 (0.0339)	0.9402 (0.0356)
σ_y		0.4668 (0.0054)	0.5589 (0.0065)	0.7347 (0.0090)
σ_v		0.8589 (0.1269)	0.7570 (0.1209)	0.6137 (0.1344)
Time (in seconds)	0.03			
Sample Size	9081			

Notes:

1. Standard errors are in parenthesis.
2. ARMA estimates are based on OLS with $J = 10$ that proposed in Section 2.3.

Table 2.9. ARMA-type estimates of the SV model with ISE

		Coca Cola (KO)	Walmart (WMT)	Ford (F)
ϕ		0.9119 (0.0065)	0.9218 (0.0063)	0.9402 (0.0054)
σ_y		0.4668 (0.0200)	0.5589 (0.0237)	0.7347 (0.0333)
σ_v		0.8589 (0.0231)	0.7570 (0.0223)	0.6137 (0.0197)
Time (in seconds)	0.61			
Sample Size	9081			

Notes:

1. Implicit standard errors are in parenthesis and calculated at $\alpha = 0.05$ with N=99.
2. ARMA estimates are based on OLS with $J = 10$ that proposed in Section 2.3.

Table 2.10. Exact and asymptotic tests based on ARMA-type estimators

(A) Test of no-persistence – $H_0: \phi = 0$								
	Asymptotic tests		Bootstrap tests			MMC tests		
	S_0	p-value	$N = 19$	$N = 99$	$N = 999$	$N = 19$	$N = 99$	$N = 999$
KO	2153.66	0	0.05	0.01	0.001	0.05	0.01	0.001
WMT	1521.80	0	0.05	0.01	0.001	0.05	0.01	0.001
F	3296.17	0	0.05	0.01	0.001	0.05	0.01	0.001
(B) Test of volatility is not a latent process – $H_0: \sigma_v = 0$								
	Asymptotic tests		Bootstrap tests			MMC tests		
	S_0	p-value	$N = 19$	$N = 99$	$N = 999$	$N = 19$	$N = 99$	$N = 999$
KO	245901	0	0.05	0.01	0.001	0.05	0.01	0.001
WMT	193164	0	0.05	0.01	0.001	0.05	0.01	0.001
F	143214	0	0.05	0.01	0.001	0.05	0.01	0.001
(C) Test of no stochastic volatility – $H_0: \phi = \sigma_v = 0$								
	Asymptotic tests		Bootstrap tests			MMC tests		
	S_0	p-value	$N = 19$	$N = 99$	$N = 999$	$N = 19$	$N = 99$	$N = 999$
KO	248055	0	0.05	0.01	0.001	0.05	0.01	0.001
WMT	194685	0	0.05	0.01	0.001	0.05	0.01	0.001
F	146510	0	0.05	0.01	0.001	0.05	0.01	0.001

Notes:

1. Null hypotheses are tested against right-sided alternatives.
2. ARMA estimates are based on OLS with $J = 10$ that proposed in Section 2.3.
3. S_0 is the GMM based LR statistic.

Table 2.11. Empirical estimates of S&P 500 index

2000 - 2017, Number of observations: 4529				
	ϕ	σ_y	σ_v	Time (in seconds)
Bayesian-MCMC	0.9850 (0.0035)	0.3863 (0.0369)	0.1808 (0.0152)	354.3
W-ARMA-SV	0.9856 (0.0069)	0.3590 (0.0364)	0.2185 (0.0175)	0.043
Relative Time Efficiency: 8214				

Notes:

1. Implicit standard errors are in parenthesis (calculated from 99 simulations).
2. Bayesian-MCMC estimates are computed using R package *stochvol* with 50000 draws and 50000 burn-in.

Chapter 3

Simple estimators for higher-order stochastic volatility models and forecasting

Abstract

We study the problem of estimating higher-order stochastic volatility [SV(p)] models. Due to the inherent difficulty of evaluating the likelihood function – a general feature of non-linear latent variable models – estimation and inference for SV models constitute challenging problems. Most of the existing estimation methods are confined to the SV(1) model and are computationally expensive, inflexible [not easy to generalize for SV(p) models], difficult to implement, and typically inefficient. The estimation of SV(p) models is even more challenging and rarely considered in the literature. In this paper, we propose simple moment-based estimators for such models – in particular ARMA-type estimators – which are both computationally inexpensive and remarkably accurate. The proposed estimators do not require choosing a sampling algorithm, initial parameter values, or an auxiliary model. To reduce the risk of getting inadmissible (nonstationary) solutions, we suggest winsorized versions of the simple ARMA-SV estimators. We also show that a Durbin-Levinson-type updating algorithm can be applied to recursively estimate models of increasing order p. The asymptotic distribution of the estimators is established. Due to their computational simplicity, the proposed estimators allow one to make finite-sample inference through the technique of Monte Carlo (MC) tests. We compare by simulation the proposed estimators to a Bayesian MCMC estimator. The results show that the simple winsorized ARMA-SV estimator is uniformly superior to other estimators in terms of bias and root mean square error. The proposed estimators are applied to stock return data, and the usefulness of the proposed estimators is assessed in two ways. First, using the daily return on the S&P 500 index from 1928 to 2016, we find that higher-order SV models – in particular an SV(3) model – are preferable to an SV(1), from the viewpoints model fit and both asymptotic and finite-sample tests. Second, using different volatility proxies (the squared return of S&P 500 index and the realized volatility of S&P 500, FTSE100, NASDAQ100, N225, SSMI20 indices), we conduct two out-of-sample forecast experiments: (1) we forecast a moderately volatile period after the late-2000s financial crisis; (2) we forecast a highly volatile period, *i.e.*, the core financial crisis. We compare the accuracy of volatility forecasts among SV(p) models, GARCH models, and Heterogenous Autoregressive model of Realized Volatility (HAR-RV) models. The results suggest that SV(p) models perform better than other models in most cases. This finding holds even if a high volatility period (such as financial crisis) is included in the estimation sample or the forecasted sample. Formal prediction tests, *i.e.*, model confidence set procedure, also support these inferences. Our findings highlight the usefulness of higher-order SV models for volatility forecasting.

Key words: generalized method of moments, Markov Chain Monte Carlo, Monte Carlo tests, stochastic volatility, asymptotic distribution, stock returns, realized variance, volatility forecasting, high-frequency data.
Journal of Economic Literature Classification: C15, C22, C53, C58.

3.1 Introduction

Time-varying volatility of asset returns is a widespread feature of financial markets. This property has been known for a long time; early discussions include Mandelbrot (1963) and Fama (1965). Two main classes of parametric models have been proposed in the literature to estimate and forecast dynamic volatility: (1) GARCH-type models [Engle (1982), Bollerslev (1986)]; (2) stochastic volatility (SV) models [Taylor (1982, 1986)]. The main distinction between GARCH and SV models is that the variance process of the latter has an additional error term which captures the effect of any new information coming to the market, so conditional on the information set $\mathcal{F}_{t-1}$, volatility σ_t^2 is not known in SV models but rather an unobserved random variable. Several reviews of GARCH and SV literature are available; for GARCH, see Bollerslev (2010), and for SV, see Ghysels et al. (1996), Broto and Ruiz (2004), and Shephard (2005). SV models are also common in macroeconomic modelling; see Cogley and Sargent (2005), Primiceri (2005), Benati (2008), Koop et al. (2009), Koop and Korobilis (2013), and Liu and Morley (2014).

SV models may be preferable to GARCH-type models for several reasons. *First*, SV models are discrete-time formulations of continuous-time diffusion processes used in theoretical finance for derivative pricing and portfolio optimization; see Taylor (1994), Shephard and Andersen (2009). *Second*, SV models do not appear to require various *ad hoc* adjustments, like the addition of a random jump component or non-Gaussian innovations. These modifications improve the performance of the standard GARCH, but these are evidently unnecessary for SV models; see Carnero et al. (2004), Chan and Grant (2016). *Third*, SV models often provide more accurate volatility forecasts than GARCH models, indicating that the time-varying volatility is better modelled as a latent stochastic process; see Kim et al. (1998), Yu (2002), Poon and Granger (2003), Koopman et al. (2005). *Finally*, it is easy to derive the probabilistic properties (stationarity, ergodicity and mixing) of SV models than GARCH models; see Davis and Mikosch (2009). In contrast, the stationarity of a GARCH process is difficult to establish; see Nelson (1990), Bougerol and Picard (1992), Lindner (2009).

Despite these attractive features, SV models are clearly less popular than GARCH models. The main reasons for this appear to be the following. *First*, the estimation of SV models is much more complicated than it is for GARCH-type models. In particular, due to the presence of latent variables, likelihood-based methods are difficult to apply, and statistical inference (estimation and testing) for SV models is quite challenging. Consequently, a variety of methods have been proposed to estimate SV models. The vast majority of these are either computer-intensive and/or inefficient. *Second*, many statistical packages (such as EVIEWS, GAUSS, MATLAB, R, S+, SAS, TSP, STATA, PYTHON, OX, etc.) have many options for incorporating GARCH effects, whereas SV models lack statistical packages. Nevertheless, some routines in R and MATLAB for SV models are available.

Earlier work on the estimation of SV models has focused on the first-order SV model, where the latent volatility process is modelled as a first-order autoregression. These include: quasi-maximum likelihood (QML) [Nelson (1988), Harvey et al. (1994), Ruiz (1994)], the generalized method of moments (GMM) [Melino and Turnbull (1990), Andersen and Sørensen (1996)], the simulated method of moments (SMM) [Gallant and Tauchen (1996), Monfardini (1998), Andersen et al. (1999)], Monte Carlo likelihood (MCL) [Sandmann and Koopman (1998)], simulated maximum likelihood (SML) [Danielsson and Richard (1993), Danielsson (1994), Durham (2006, 2007), Richard and Zhang (2007)], the method based on linear representation [Francq and Zakoïan (2006)], closed-form moment-based estimators [Dufour and Valéry (2006, 2009), Ahsan and Dufour (2019)], and Bayesian techniques based on Markov Chain Monte Carlo (MCMC) methods [Jacquier et al. (1994), Kim et al. (1998), Chib et al. (2002), Fiorentini et al. (2004), Flury and Shephard (2011)].

Apart from the closed-form estimators, the above estimation methods are based on simulation techniques and/or numerical optimization. Simulation-based methods such as SML, MCL, SMM, and Bayesian MCMC methods [via the Metropolis-Hastings algorithm or the Gibbs sampler] are computer-intensive, inflexible across models, hard to implement in practice, and may converge very slowly; see Broto and Ruiz (2004). Implementing these methods requires one to choose a sampling scheme, initial parameters, and an auxiliary model (which is largely conventional). The choice of initial parameter values for QML, GMM or MCMC plays a pivotal role in convergence. In particular, a poorly assigned prior may lead to a fragile Bayesian inference. In the context of GMM estimation, Broto and Ruiz (2004) pointed out that the criterion surface is highly irregular, so optimization often fails to converge in small samples [Andersen and Sørensen (1996) documented a large number of non-converging GMM estimations]. Further, GMM usually produces imprecise estimates due to an ill-conditioned weighting matrix. By contrast, the closed-form moment-based estimators are analytically tractable, computationally simple, and very easy to implement.

In this paper, we study higher-order stochastic volatility [SV(p)] models where the underlying volatility process follows an autoregressive process of order p. In particular, we focus on the estimation and forecasting issues of SV(p) models. The estimation of SV(p) models is even more challenging than it is for an SV(1) model. Consequently, SV(p) models are rarely estimated in financial econometrics literature; exceptions are SMM of Gallant et al. (1997), MCL of Asai (2008), Bayesian MCMC of Chan and Grant (2016). However, in line with these studies, motivations for SV(p) models are as follows:

1. It is a natural extension of the basic SV(1) model, which can only generate geometrically decaying autocovariance function, whereas volatility process generically features persistent memory.

2. As pointed out by Asai (2008) and Meddahi (2003), the latent volatility process of a multi-factor stochastic volatility (MFSV) model can be interpreted as a linear combination of latent and independent AR(1) processes which aggregate to an ARMA(p, q) process. So, the higher-order autoregressive terms in SV models naturally emerge from the aggregation process.

3. The empirical results of these studies suggest that higher-order models provide more flexibility to represent volatility persistence, heavy tails and may capture the effects of jumps as well.

4. Empirical evidence in this paper suggests that higher-order SV models may be preferable for both in-sample model fitting and out-of-sample volatility forecasting.

In this paper, we study the problem of estimating the parameters of an SV(p) model. Due to its intrinsic complexity, the work on this problem remains scarce, and the proposed ones are inflexible, computationally costly, and limited to low orders [see Gallant et al. (1997), Asai (2008), Chan and Grant (2016)]. Instead, we propose here two simple moment-based estimation methods for SV(p) models.

1. We extend the closed-form estimators of Dufour and Valéry (2009) and develop a class of simple estimators for SV(p) models based on the moment structure of returns. We call this approach, the *simple EDV* method.

2. We exploit the non-Gaussian ARMA representation of SV(p) models and derive an estimator which we call the *simple ARMA-SV* estimators.[1] The ARMA-SV method uses the moment structure of the logarithm of squared residual returns.

These estimators are analytically tractable and computationally inexpensive. In particular, they can be readily implemented without using any numerical optimization, and they do not require one to choose an arbitrary initial parameter or an auxiliary model. Further, we also suggest GMM-type estimators for SV(p) models. These GMM estimators are extensions of Andersen and Sørensen (1996).

The proposed moment-based estimators (simple estimators and GMM estimators) may violate stationarity conditions in the presence of outliers or in small samples. To circumvent this problem, we suggest restricted estimation where the estimates are restrained on the space of acceptable parameter solutions by adjusting the eigenvalues that lie on or outside the unit circle.

Further, in the case of ARMA-SV method, we suggest winsorized versions of the ARMA-SV estimator (*W-ARMA-SV* estimators), which substantially increases the probability of getting acceptable values and also improves efficiency [Hafner and Linton (2017)]. In proposing winsorized methods, autoregressive parameters of the latent volatility process [these parameters capture the volatility clustering of a financial time series] are estimated using a combination of several ratios of sample autocovariance matrices, including weighted averages, the median,

[1] In the context of continuous-time stochastic volatility models, Meddahi (2003) derives the ARMA representation of integrated and realized variances when the spot variance depends linearly on two autoregressive factors.

or an OLS-based weighting. This computationally simple adjustment improves the stability and accuracy of the estimators. Indeed, we show in simulations that W-ARMA-SV estimators improve the precision. Especially, an OLS-based W-ARMA-SV estimator uniformly outperforms all other estimators in terms of bias and RMSE by a significant margin — including the Bayesian estimator proposed in this context.

Using these simple estimators, we develop recursive estimation procedures for SV(p) models by exploiting Durbin-Levinson-type (DL) algorithms. We discuss two algorithms, which allow recursive-in-order calculation of the parameters of higher-order SV processes. The proposed procedures generalize the recursion of Durbin (1960) [which pertains to pure autoregressive models] and of Tsay and Tiao (1984) [which applies to autoregressive-moving average models].

The proposed computationally inexpensive estimators can be useful in several contexts. Since SV models are parametric models involving only a finite number of unknown parameters, using these proposed estimators, one can construct simulation-based tests, even exact tests based on the Monte Carlo (MC) test technique [see Dufour (2006)], as opposed to procedures based on establishing asymptotic distributions. In particular, exact tests obtained in this way do not depend on stationarity assumptions, and consequently are useful when the latent volatility process has a unit root (or is close to this structure). Furthermore, proposed estimators are helpful for estimation schemes which require repeated estimation based on a rolling window method, for example, Backtesting of risk measures (such as Value-at-Risk or Expected Shortfall) in the context of risk management.

We derive the asymptotic properties of the proposed simple estimators under standard regularity assumptions, showing consistency and asymptotic normality when the fourth moment of the latent volatility process exists. Due to the $\sqrt{T}$-consistency, our simple estimators can be effortlessly applied to very large samples, which are not rare in empirical finance. In these situations, estimators based on simulation technique and/or numerical optimization often require substantial computational effort to achieve convergence. So instead of using computationally costly estimators, one may prefer to use estimators that are available in analytical form.

Using Monte Carlo simulations, we study the statistical properties of our estimators and compare them with the Bayesian MCMC method. The simulation results confirm that the W-ARMA-SV estimator that based on OLS (W-ARMA-SV-OLS) has excellent statistical properties in terms of bias and RMSE. It uniformly outperforms all other estimators, including the Bayesian estimator regarding bias and RMSE. This result holds across different simulation designs and for all individual parameters. Furthermore, the simple estimators are highly efficient in terms of computation time, compared to other estimators.

We present empirical applications related to SV(p) models and the ARMA-SV estimator. *First*, using the daily return on the S&P 500 index from 1928 to 2016, we find that an SV(3) model is the best one from the viewpoint of in-sample fit, using both asymptotic and finite-sample tests. *Second*, using different volatility proxies [the squared return of S&P 500 index and the realized volatility of S&P 500, FTSE100, NASDAQ100, N225, SSMI20 indices], we conduct two out-of-sample forecast experiments: (1) a moderately volatile period after the late-2000s financial crisis; (2) a highly volatile period, *i.e.*, the core financial crisis. We compare the accuracy of volatility forecasts among SV(p) models, GARCH models, and Heterogenous Autoregressive model of Realized Volatility (HAR-RV) models. The results suggest that SV(p) models perform better than other models in most cases. This finding holds even if a high volatility period (such as financial crisis) is included in the estimation sample or the forecasted sample. These inferences are not only based on a standard forecasting precision assessment [such as using MSE and MAE statistics] but also on formal prediction tests, using the MCS procedure of Hansen et al. (2011). Our findings highlight the usefulness of higher-order SV models for volatility forecasting.

The W-ARMA-SV-OLS estimator proposed in this paper can be interpreted as a *parsimonious moment-based* estimator where only a few (well chosen) moments are used. In a moment-based (or GMM) inference, using too many moments can be very costly from an estimation efficiency viewpoint as well as forecasting. Indeed, we show in our simulations and empirical applications that the W-ARMA-SV-OLS estimator exhibits the best performance in both estimation and forecasting, as well as numerical efficiency.

The paper proceeds as follows. Section 3.2 specifies the model and its assumptions. Section 3.4 discusses the stationarity, ergodicity and mixing properties of SV(p) models. Section 3.5 proposes simple estimators and their recursive prediction algorithms. Section 3.6 proposes

GMM type estimators for SV(p) models, and Section 3.8 discusses the MC test technique. Section 3.7 develops asymptotic theories for simple estimators. Section 3.9 presents the simulation study, and Section 3.10 presents the empirical applications. We conclude in Section 3.11. The proofs, tables, and figures are provided in the Appendix 3.12.

3.2 Framework

We consider a standard discrete-time SV process of order p, which is described below following Taylor (1986), Ghysels et al. (1996) and Gallant et al. (1997). Specifically, we say that a variable y_t follows a discrete-time SV(p) process if it satisfies the following assumption, where $t \in \mathbb{N}_0$, and $\mathbb{N}_0$ represents the non-negative integers.

Assumption 3.2.1. STOCHASTIC VOLATILITY OF ORDER p. *The process* $\{y_t : t \in \mathbb{N}_0\}$ *satisfies the equations*

$$y_t = \sigma_y \exp(w_t/2) z_t \,, \tag{3.2.1}$$

$$w_t = \sum_{j=1}^{p} \phi_j w_{t-j} + \sigma_v v_t \,, \tag{3.2.2}$$

where the vectors $(z_t, v_t)'$ *are i.i.d. according to a* $\mathrm{N}[0, I_2]$ *distribution, while* $(\phi_1, \ldots, \phi_p, \sigma_y, \sigma_v)'$ *are fixed parameters.*

We also make a stationarity assumption as follows.

Assumption 3.2.2. STATIONARITY. *The process* $l_t = (y_t, w_t)'$ *is strictly stationary.*

The last assumption entails that all the roots of the characteristic equation of the volatility process $[\phi(B) = 0]$ lie outside the unit circle [*i.e.*, $\phi(z) \neq 0$ for $|z| \leq 1$], and $w_0 \sim N[0, \sigma_v^2/(1 - \sum_{j=1}^{p} \phi_j^2)]$.

The SV(p) model consists of two stochastic processes, where y_t describes the dynamics of asset returns and $w_t := \log(\sigma_t^2)$ captures the dynamics of latent log volatilities. Usually the y_t's are residual returns, such that

$$y_t := r_t - \mu_r \,, \quad r_t := 100[\log(p_t) - \log(p_{t-1})] \,,$$

where μ_r is the mean of returns (r_t) and p_t is the raw prices of an asset.[2] The latent process w_t can be interpreted as a random flow of uncertainty shocks or new information in financial markets, while ϕ_j's capture the volatility persistence. This type of volatility model naturally fits into the theoretical framework of modern financial theory.

Let us now transform y_t by taking the logarithm of its squared value. We get in this way the following *measurement equation*:

$$\begin{aligned}\log(y_t^2) &= \log(\sigma_y^2) + w_t + \log(z_t^2) = \{\log(\sigma_y^2) + \mathbb{E}[\log(z_t^2)]\} + w_t + \{\log(z_t^2) - \mathbb{E}[\log(z_t^2)]\} \\ &= \mu + w_t + \epsilon_t \end{aligned} \tag{3.2.3}$$

where

$$\mu := \mathbb{E}[\log(y_t^2)] = \log(\sigma_y^2) + \mathbb{E}[\log(z_t^2)]\,, \quad \epsilon_t := \log(z_t^2) - \mathbb{E}[\log(z_t^2)]\,. \tag{3.2.4}$$

Note that this logarithmic transformation entails no information loss since the distribution of z_t is symmetric (see Remark 1 of Francq and Zakoïan (2006)). Furthermore, even if v_t and z_t are not mutually independent, they are uncorrelated if the joint distribution of v_t and z_t is symmetric, that is $f(v_t, z_t) = f(-v_t, -z_t)$; see Harvey et al. (1994).

Under the normality assumption for z_t, the errors ϵ_t are i.i.d. according to the distribution of a centered $\log(\chi_1^2)$ random variable [*i.e.*, ϵ_t has mean zero and variance $\mathbb{E}(\epsilon_t^2)$]. The cumulant generating function of $\log(\chi_1^2)$ distribution is:

$$\begin{aligned} M(s) &= \log \mathbb{E}\big[\exp\big(s\log(\chi_1^2)\big)\big] = \log\Big[\mathbb{E}\big(\chi_1^2\big)^s\Big] = \log\left[\frac{2^s\,\Gamma\big((1/2)+s\big)}{\Gamma(1/2)}\right] \\ &= s\log(2) + \log[\Gamma\big((1/2)+s\big)] - \log[\Gamma(1/2)]\,, \quad \text{for } s \geq 0\,, \end{aligned} \tag{3.2.5}$$

where $\Gamma(z) := \int_0^\infty x^{z-1} e^{-x}\, dx$ is the *gamma function*; see Wishart (1947). The m^{th} cumulant of the $\log(\chi_1^2)$ random variable is the m^{th} derivative of $M(s)$ evaluated at $s = 0$. Thus, the

[2] It is noteworthy to mention that y_t is ordinarily the error term of any time series regression model, see for example Jurado et al. (2015).

corresponding cumulants (κ_m) and central moments ($\tilde{\mu}_m$) are:

$$\kappa_m = \begin{cases} \log(2) + \psi\left(\frac{1}{2}\right), & \text{if } m = 1 \\ \psi^{(m-1)}\left(\frac{1}{2}\right), & \text{if } m > 1 \end{cases}, \qquad \tilde{\mu}_m = \begin{cases} 0, & \text{if } m = 1 \\ \kappa_m + \sum_{j=1}^{m-2} \binom{m-1}{j} \kappa_{m-j} \tilde{\mu}_j, & \text{if } m > 1 \end{cases}, \tag{3.2.6}$$

where

$$\psi(z) := \frac{d}{dz} \log\left[\Gamma(z)\right] = \frac{\Gamma'(z)}{\Gamma(z)} \tag{3.2.7}$$

is the *digamma function* and

$$\psi^{(m)}(z) := \frac{d^m}{dz^m} \psi(z) = \frac{d^{m+1}}{dz^{m+1}} \log[\Gamma(z)] \tag{3.2.8}$$

is the *polygamma function* of order m [*i.e.*, the $(m+1)$-th order derivative of the logarithm of the *gamma function*].

From (3.2.6), we get:

$$\mathbb{E}[\log(z_t^2)] = \kappa_1 = \log(2) + \psi(1/2) \simeq -1.2704, \tag{3.2.9}$$

$$\sigma_\epsilon^2 := \mathbb{E}(\epsilon_t^2) = \operatorname{Var}[\log(z_t^2)] = \tilde{\mu}_2 = \kappa_2 = \psi^{(1)}(1/2) = \pi^2/2, \tag{3.2.10}$$

$$\mathbb{E}(\epsilon_t^3) = \tilde{\mu}_3 = \kappa_3 = \psi^{(2)}(1/2), \quad \mathbb{E}(\epsilon_t^4) = \tilde{\mu}_4 = \kappa_4 + 3\kappa_2^2 = \psi^{(3)}(1/2) + 3\sigma_\epsilon^2 = \pi^4 + 3\sigma_\epsilon^2; \tag{3.2.11}$$

see Abramowitz and Stegun (1970, Chapter 6). The $\log(\chi_1^2)$ distribution is often approximated by a normal distribution with mean of -1.2704 and variance of $\pi^2/2$ [see Broto and Ruiz (2004)], or by a mixture distribution [Kim et al. (1998)].

On setting

$$y_t^* := \log(y_t^2) - \mu, \tag{3.2.12}$$

the SV model (3.2.3) can be written as

$$y_t^* = w_t + \epsilon_t. \tag{3.2.13}$$

By combining (3.2.2) and (3.2.13), we see that the SV(p) model can be written in state-space form:

$$\text{State Transition Equation:} \quad w_t = \sum_{j=1}^{p} \phi_j w_{t-j} + v_t\,, \tag{3.2.14}$$

$$\text{Measurement Equation:} \quad y_t^* = w_t + \epsilon_t\,, \tag{3.2.15}$$

where w_t is a logarithm of latent daily volatility, y_t^* is a logarithm of the daily squared return corrected by its mean, where the variables v_t are i.i.d. $\mathrm{N}(0,\sigma_v^2)$, and the ϵ_t's are i.i.d. $\log(\chi_1^2)$; for further discussion of this representation, see Nelson (1988), Harvey et al. (1994), Ruiz (1994), Shephard (1994), Breidt and Carriquiry (1996), Harvey and Shephard (1996), Kim et al. (1998), Sandmann and Koopman (1998), Steel (1998), Chib et al. (2002), Knight et al. (2002), Francq and Zakoïan (2006), Omori et al. (2007).

3.3 Higher-order stochastic volatility

In this section, we discuss the econometric motivation for SV(p) models. It has been well documented that the volatility process is driven by at least two factors: one factor captures the salient properties of volatility, such as randomness and persistence, and a second one to deal with the shape of the conditional distribution of financial returns such as fat-tails; examples of these studies include Gallant et al. (1999), Meddahi (2001), Alizadeh et al. (2002), Barndorff-Nielsen et al. (2002), Bollerslev and Zhou (2002), Chernov et al. (2003), and Durham (2006, 2007). Some of these studies also considered more than two factors and tried to fit the volatility process of asset returns. This type of factor model is important for capturing nonlinearities in financial returns and improves the fit dramatically. However, these proposed models need highly complex numerical optimization techniques, and they are not tractable analytically. It is worth noting that we can always transform MFSV models to SV models which have ARMA representation in the log volatility process [SV(p,q)]. This transformation is perfectly acceptable, since we can recuperate the MFSV parameters form the estimates of the transformed model parameters. Further, instead of an SV(p,q) model, we can estimate an SV(p) model and recuperate the SV(p,q) parameters from there.

Assumption 3.3.1. MULTI-FACTOR STOCHASTIC VOLATILITY MODEL OF ORDER M. *The process*

$\{y_t : t \in \mathbb{N}_0\}$ *satisfies the equations*

$$y_t = \sigma_y \exp(\sum_{i=1}^{m} \frac{w_{it}}{2}) z_t ,$$

$$w_{it} = \phi_{if} w_{it-1} + \sigma_{iv} v_{it} , \quad \left|\phi_{if}\right| < 1 , \quad i = 1, \ldots, m,$$

where $\theta_m^{MFSV} := (\sigma_y, (\phi_{if})_{i=1}^m, (\sigma_{iv})_{i=1}^m)'$ *are fixed parameters, and* $(z_t, v_{it})'$ *are i.i.d. Gaussian such that* z_t *is* $\mathrm{N}[0, 1]$ *and* v_{it}*'s are* $\mathrm{N}[0, I_m]$ *and* $\mathbb{E}[v_{it} z_t] = 0 \ \forall i$.

Lemma 3.3.1. SV(p, q) REPRESENTATION OF MFSV MODEL. *The model MFSV*(m) *defined by Assumption 3.3.1 has the following SV*$(m, m-1)$ *representation:*

$$y_t = \sigma_y \exp(\frac{w_t}{2}) z_t , \tag{3.3.1}$$

$$w_t = \sum_{j=1}^{m} \alpha_j w_{t-j} + \sigma_v v_t - \sigma_v \sum_{j=1}^{m-1} \beta_j v_{t-j} , \tag{3.3.2}$$

where $\theta_{m,m-1} := (\sigma_y, (\alpha_j)_{j=1}^m, (\beta_j)_{j=1}^{m-1}, \sigma_v)'$ *are fixed parameters and* $(z_t, v_t)'$, $t \in \mathbb{N}_0$, *are i.i.d. according to a* $\mathrm{N}[0, I_2]$ *distribution. All the roots of characteristic equations* $[(1 - \alpha_1 B - \cdots - \alpha_m B^m) = 0$ *and* $(1 - \beta_1 B - \cdots - \beta_{m-1} B^{m-1}) = 0]$ *lie outside the unit circle.*

To understand the Lemma 3.3.1, we consider the following example.

Example 1. SV(2,1) REPRESENTATION OF THE MFSV(2) MODEL. *Under Assumption 3.3.1, the volatility process of an MFSV*(2) *model, which is driven by the sum of two independent AR*(1) *process,* i.e., $w_t = w_{1t} + w_{2t}$, *where*

$$w_{1t} - \phi_{1f} w_{1t-1} = (1 - \phi_{1f} B) w_{1t} = \sigma_{1v} v_{1t} ,$$

$$w_{2t} - \phi_{2f} w_{2t-1} = (1 - \phi_{2f} B) w_{2t} = \sigma_{2v} v_{2t} .$$

Using the aggregation principle of autoregressive processes [see Granger and Morris (1976)], we know that AR(p) + *AR*(q) = *ARMA*$(p + q, max(p, q))$ *and in particular, if we add two independent AR*(1) *processes, then we can get an ARMA*$(2, 1)$ *process. This could be achieved as follows.*

Since $w_t = w_{1t} + w_{2t}$, and v_{1t} and v_{2t} are two independent white noise, it follows that

$$(1-\phi_{1f}B)(1-\phi_{2f}B)w_t = (1-\phi_{2f}B)\sigma_{1v}v_{1t} + (1-\phi_{1f}B)\sigma_{2v}v_{2t},$$

or

$$(1-\alpha_1 B - \alpha_2 B^2)w_t = (1-\beta_1 B)\sigma_v v_t. \tag{3.3.3}$$

This is an ARMA$(2,1)$ process with AR parameters, $\alpha_1 = \phi_{1f}+\phi_{2f}$ and $\alpha_2 = -\phi_{1f}\phi_{2f}$, hence

$$\phi_{1f} = \frac{\alpha_1 - \sqrt{\alpha_1^2 + 4\alpha_2}}{2}, \quad \phi_{2f} = \frac{\alpha_1 + \sqrt{\alpha_1^2 + 4\alpha_2}}{2}. \tag{3.3.4}$$

The RHS of 3.3.3 is an invertible $MA(1)$ process with variance $[(1+\phi_{2f}^2)\sigma_{1v}^2 + (1+\phi_{1f}^2)\sigma_{2v}^2]$ and autocovariance at lag 1 $[-(\phi_{2f}\sigma_{1v}^2 + \phi_{1f}\sigma_{2v}^2)]$. Thus we get an SV$(2,1)$ with

$$\alpha_1 = \phi_{1f}+\phi_{2f}, \alpha_2 = -\phi_{1f}\phi_{2f}, \tag{3.3.5}$$

$$[(1+\phi_{2f}^2)\sigma_{1v}^2 + (1+\phi_{1f}^2)\sigma_{2v}^2] = (1+\beta_1^2)\sigma_v^2, \tag{3.3.6}$$

$$[-(\phi_{2f}\sigma_{1v}^2 + \phi_{1f}\sigma_{2v}^2)] = -\beta_1\sigma_v^2. \tag{3.3.7}$$

SV(p,q) models are more parsimonious but difficult to estimate, whereas SV(p) models are not parsimonious but easy to estimate and use. We can estimate an SV(p) model instead of an SV(p,q) model and recuperate the parameters of SV(p,q) model from the estimates of SV(p) parameters. This process is based on the AR approximation of ARMA-type latent volatility process. Some estimators, based on autoregressive approximation, have been proposed for general ARMA models. These methods derive ARMA estimates from an approximate AR process, where a linear regression or other technique is used to extract information from the full set of AR coefficients; see for example Hannan and Rissanen (1982), Saikkonen (1986), Koreisha and Pukkila (1990) and Galbraith and Zinde-Walsh (1997). In the context of VARMA models, this type of method is used by Dufour and Pelletier (2005) and Dufour and Jouini (2014).

Lemma 3.3.2. INFINITE-ORDER SV REPRESENTATION OF SV(P,Q) PROCESS. *The model SV(p,q)*

defined by lemma 3.3.1 [*where* $p = m$ *and* $q = m-1$] *has the following SV*(∞) *representation:*

$$y_t = \sigma_y \exp(\frac{w_t}{2}) z_t, \tag{3.3.8}$$

$$\sum_{j=0}^{\infty} (-\phi_j) w_{t-j} = \sigma_v v_t, \tag{3.3.9}$$

where $\theta_{\infty}^{SV} := (\sigma_y, \left\{\phi_j\right\}_{j=1}^{\infty}, \sigma_v)$ *are fixed parameters and* $(z_t, v_t)'$, $t \in \mathbb{N}_0$, *are i.i.d. according to a* $\mathrm{N}[0, I_2]$ *distribution.*

The SV(∞) model, given in Lemma 3.3.2, can be replaced by a truncated SV(p) model and we can recuperate the SV(p, q) parameters from it. Using standard results on the representation of an ARMA(p, q) process [see Fuller (1996), Ch. 2, page 74], we have the following expression that relate the parameters of SV(∞) and SV(p, q) model:

$$\sum_{j=0}^{\infty} (-\phi_j) w_{t-j} = \sigma_v v_t, \tag{3.3.10}$$

where

$$\begin{aligned}
&\phi_0 = -1, \\
&\phi_1 = -\beta_1 + \alpha_1, \\
&\phi_2 = -\beta_1 \phi_1 + \beta_2 + \alpha_2, \\
&\vdots \\
&\phi_j = -\sum_{i=1}^{min(j,q)} \beta_i \phi_{j-i} + \alpha_j, \quad (j \le p) \\
&\phi_l = -\sum_{i=1}^{min(l,q)} \beta_i \phi_{l-i}, \quad (l > p).
\end{aligned}$$

Given the above equations, we can identify the parameters of an SV(p, q) model from the parameters of an SV(k) model. The identification requires $k \ge p + q$. To understand the whole identification process, we illustrate the following example.

Example 2. SV(2,1) PARAMETERS FROM AN SV(3) MODEL PARAMETERS. *Under the lemma 3.3.1, we have SV*(2,1) *model where the volatility process is driven by an ARMA*(2,1) *process. To identify an SV*(2,1) *model from an SV*(3) *model, we use the following equations:*

$$\phi_1 = -\beta_1 + \alpha_1, \quad \phi_2 = -\beta_1 \phi_1 + \alpha_2, \quad \phi_3 = -\beta_1 \phi_2.$$

Solving the above equations yields the parameters of SV$(2,1)$ model in terms of SV(3) parameter:

$$\alpha_1 = \phi_1 + \frac{\phi_3}{\phi_2}, \quad \alpha_2 = \phi_2 + \frac{\phi_1\phi_3}{\phi_2}, \quad \beta_1 = -\frac{\phi_3}{\phi_2}.$$

The whole identification process is as follows:

1. *The MFSV(2) model, where the volatility process is driven by two independent AR(1) process, has an SV$(2,1)$ representation by aggregation.*

2. *The SV(∞) representation follows from the invertibility of the MA part of the SV$(2,1)$ model.*

3. *Estimate an SV(3) model [instead of an SV(∞)] and recuperate the SV$(2,1)$ parameters.*

4. *Further, from the SV$(2,1)$ parameters, we can identify the AR factor polynomials of the MFSV(2) model by using* (3.3.4).

From most of the empirical studies, it is prominent that researchers try to fit a distribution that provides best fits for the volatility of asset return. In this section, we point out that an SV(p) model may be served better in that respect since it is not only a natural extension SV(1) model but also an approximated representation of the MFSV or the SV(p,q) model.

3.4 Stationarity, ergodicity and mixing properties

The mutual independence of the noise (z_t) and the volatility sequence (w_t) is one of the attractive probabilistic features of SV models. This statistical property of SV models allows for a much simpler probabilistic structure than that of GARCH-type models. It is difficult to establish a necessary and sufficient condition for stationarity of a GARCH process. Nelson (1990) established a solution for the GARCH$(1,1)$ case and Bougerol and Picard (1992) for the general GARCH(p,q) case. For a review of the stationarity of GARCH processes, one may refer to Straumann (2005) or Francq and Zakoïan (2010). From Carrasco and Chen (2002), the following results ensure the stationarity, ergodicity and mixing condition of SV(p) models. These probabilistic conditions are useful in order to establish the large sample properties of the estimators of SV(p) models. Indeed, we need $\{(y_t, w_t)'\}$ to be strictly stationary and ergodic with appropriate mixing condition.

Result 3.4.1. STATIONARITY AND ERGODICITY. *Let $\{z_t\}$ and $\{v_t\}$ be two independent processes such that $\{z_t\}$ is a sequence of i.i.d. real-valued random variables, independent of w_0, with $\mathbb{E}(z_t) = 0$ and $\mathbb{E}(z_t^2) = 1$, and z_t has a continuous positive density with respect to Lebesgue measure on real line. Also, assume that all the roots of the characteristic equation of the volatility process* $[\phi(z) := 1 - \phi_1 z - \cdots - \phi_p z^p = 0]$ *lie outside the unit circle* [i.e., $\phi(z) \neq 0$ *for* $|z| \leq 1$] *and there is an integer $s \geq 1$ such that*

$$\mathbb{E}\big(|v_t|^s\big) < \infty. \tag{3.4.1}$$

Then the following properties hold.

(*i*) $\mathbb{E}[|w_t|^s] < \infty$ *and $\{w_t\}$ is Markov geometrically ergodic.*

(*ii*) *If $\{w_t\}$ is initialized from its stationary distribution, then $\{w_t\}$ and $\{y_t\}$ are strictly stationary and exponential β - mixing and this property is preserved by any continuous transformation of $\{w_t\}$, such as* $\exp(w_t/2)$.

(*iii*) *If* $\mathbb{E}\big(\big|\log(|z_t|)\big|^s\big) < \infty$, *then* $\mathbb{E}\big(\big|\log(|y_t|)\big|^s\big) < \infty$.

Note that the latter part of the above result follows from $y_t = \exp(w_t/2)\sigma_y z_t$, which implies

$$\log|y_t| = (w_t/2) + \log|\sigma_y| + \log|z_t|. \tag{3.4.2}$$

The stochastic volatility model $\{y_t\}$ is a hidden Markov model since it includes a latent Markov chain $\{w_t\}$ and $\{w_t\}$ is independent of the i.i.d. noise process $\{z_t\}$. Proposition 2.1 of Genon-Catalot et al. (2000) show that a hidden Markov model $\{y_t\}$ is ergodic and strong mixing if the hidden chain $\{w_t\}$ is ergodic and strong mixing. We can get a similar result in the context of SV models by using the Proposition 4 of Carrasco and Chen (2002) and (3.4.2).

Result 3.4.2. BETA MIXING. *Let $\{y_t\}$ be a generalized hidden Markov model with a hidden chain $\{w_t\}$. Then,* (*i*) *if* $\{w_t\}$ *is geometrically ergodic, then the process* $\{(w_t, \log|y_t|)\}$ *is Markov geometrically ergodic;*

(*ii*) *if* $\{w_t\}$ *is stationary* β*-mixing, then* $\{\log|y_t|\}$ *is stationary* β*-mixing with a decay rate at least as fast as that of* $\{w_t\}$.

We thus have the following basic property of the SV(p) process: if $\{w_t\}$ is initialized from its stationary distribution, $\log|y_t|$ is strictly stationary and exponential β-mixing, and so is the process $(y_t, w_t)'$.

3.5 Simple estimation methods

In this section, we propose simple estimators for SV(p) models, including the corresponding recursive procedures. Besides, we also suggest alternative methods to improve the performance of these simple estimators.

3.5.1 Simple moment-based estimation

This moment-based estimator is the extension of Dufour and Valéry (2006, 2009) and it is based on the moments of the following identity that can be obtained from substituting (3.2.2) into (3.2.1):

$$y_t := \sigma_y \exp\Big[\Big(\sum_{j=1}^{p} \phi_j w_{t-j} + \sigma_v v_t\Big)/2\Big] z_t, \quad \forall t. \tag{3.5.1}$$

The moments and cross-moments of y_t [$y_t := y_t(\theta)$ where $\theta := (\phi_1, \ldots, \phi_p, \sigma_y, \sigma_v)'$] are given in the following Lemma which is a generalization of Lemma 3.1 of Dufour and Valéry (2006).

Lemma 3.5.1. MOMENTS AND CROSS-MOMENTS OF THE VOLATILITY PROCESS. *Under the assumptions 3.2.1 - 3.2.2, and if* $U \sim \mathrm{N}[0, 1]$, *then* $\mathbb{E}(U^{2p+1}) = 0, \ \ \forall p \in \mathbb{N}$ *and* $\mathbb{E}(U^{2p}) = \frac{2p!}{2^p p!}, \ \ \forall p \in \mathbb{N}$*; then the moments and cross-moments of* $y_t = \sigma_y \exp((\sum_{j=1}^{p} \phi_j w_{t-j} + \sigma_v v_t)/2) z_t$ *are given by* *the following formulas: For k, l and $m \in \mathbb{N}$, we have:*

$$\mu_k(\theta) := \mathbb{E}(y_t^k) = \begin{cases} \sigma_y^k \frac{k!}{2^{k/2}(k/2)!} \exp\left[\frac{k^2}{8} \frac{\sigma_v^2}{1-\sum_{j=1}^p \phi_j \rho_j}\right] & \text{if } k \text{ is even,} \\ 0 & \text{if } k \text{ is odd,} \end{cases} \tag{3.5.2}$$

$$\mu_{k,l}(m|\theta) := \mathbb{E}(y_t^k y_{t+m}^l) = \begin{cases} \sigma_y^{k+l} \frac{k!}{2^{k/2}(k/2)!} \frac{l!}{2^{l/2}(l/2)!} \exp\left[\frac{1}{8} \frac{\sigma_v^2(k^2+l^2+2kl\rho_m)}{1-\sum_{j=1}^p \phi_j \rho_j}\right] & \text{if } k \text{ and } l \text{ are even} \\ 0 & \text{otherwise} \end{cases} \tag{3.5.3}$$

where $\rho_j := \text{corr}(w_t, w_{t+j})$.

Dufour and Valéry (2006) derived a closed-form solution for an SV(1) model by exploiting Lemma 3.5.1. We now derive a closed-form solution for the higher-order SV process by using Lemma 3.5.1. In following Lemma, we show it for an SV(p) model where $p = 2$:

Lemma 3.5.2. CLOSED-FORM MOMENT EQUATIONS SOLUTION FOR THE SV(2) MODEL. *Using Lemma 3.5.1, we have following moment equations solution:*

$$\phi_1 = \frac{-\left[\log\left(\mu_{2,2}(1)/\mu_2^2\right)\right]\left[\log\left(3\mu_{2,2}(2)/\mu_4\right)\right]}{\left[\log\left(\mu_4/(3\mu_2^2)\right)\right]^2 - \left[\log\left(\mu_{2,2}(1)/\mu_2^2\right)\right]^2}, \tag{3.5.4}$$

$$\phi_2 = \frac{-\left[\log\left(\mu_{2,2}(1)/\mu_2^2\right)\right]^2 + \left[\log\left(\mu_{2,2}(2)/\mu_2^2\right)\right]\left[\log\left(\mu_4/(3\mu_2^2)\right)\right]}{\left[\log\left(\mu_4/(3\mu_2^2)\right)\right]^2 - \left[\log\left(\mu_{2,2}(1)/\mu_2^2\right)\right]^2}, \tag{3.5.5}$$

$$\sigma_y = 3^{1/4}\mu_2/\mu_4^{1/4}, \tag{3.5.6}$$

$$\sigma_v = \left[\log\left(\mu_4/(3\mu_2^2)\right) - \phi_1 \log\left(\mu_{2,2}(1)/\mu_2^2\right) - \phi_2 \log\left(\mu_{2,2}(2)/\mu_2^2\right)\right]^{1/2}. \tag{3.5.7}$$

where $\mu_k := \mu_k(\theta)$ *and* $\mu_{k,l}(m) := \mu_{k,l}(m|\theta)$.

Using Lemma 3.5.1, we derive higher-order autocovariance functions of y_t^2, y_t^4, $y_t^2 y_{t-1}^2$ and $y_t^2 y_{t-2}^2$ given in the following Lemma. These autocovariance functions are useful for the derivation of asymptotic properties of the SV(2) estimator defined in Lemma 3.5.2.

Lemma 3.5.3. HIGHER-ORDER AUTOCOVARIANCE FUNCTIONS. *Under the assumptions of*

Lemma 3.5.1, let $X_t = (X_{1t}, X_{2t}, X_{3t}, X_{4t})'$ *with*

$$X_{1t} = y_t^2 - \mu_2(\theta), \quad X_{2t} = y_t^4 - \mu_4(\theta), \quad X_{3t} = y_t^2 y_{t-1}^2 - \mu_{2,2}(1\,|\theta), \quad X_{4t} = y_t^2 y_{t-2}^2 - \mu_{2,2}(2\,|\theta).$$

Then the auto-covariances $\zeta_i(\tau) := \mathrm{cov}(X_{i,t}, X_{i,t+\tau})$, $i = 1,2,3,4$ *are given by:*

$$\zeta_1(\tau) = \mu_2^2(\theta)[\exp(\gamma_\tau) - 1], \tag{3.5.8}$$

$$\zeta_2(\tau) = \mu_4^2(\theta)[\exp(4\gamma_\tau) - 1], \quad \forall\tau \geq 1, \tag{3.5.9}$$

$$\zeta_3(\tau) = \mu_{2,2}^2(1\,|\theta)[\exp(\gamma_{\tau-1} + 2\gamma_\tau + \gamma_{\tau+1}) - 1], \quad \forall\tau \geq 2, \tag{3.5.10}$$

$$\zeta_4(\tau) = \mu_{2,2}^2(2\,|\theta)[\exp(\gamma_{\tau-2} + 2\gamma_\tau + \gamma_{\tau+2}) - 1], \quad \forall\tau \geq 3, \tag{3.5.11}$$

where $\gamma_j := \mathrm{cov}(w_t, w_{t+j})$.

Now it is natural to estimate $\mu_2(\theta)$, $\mu_4(\theta)$, $\mu_{2,2}(1\,|\theta)$, and $\mu_{2,2}(2\,|\theta)$ by the corresponding empirical moments:

$$\hat{\mu}_2 = \frac{1}{T}\sum_{t=1}^{T} y_t^2, \quad \hat{\mu}_4 = \frac{1}{T}\sum_{t=1}^{T} y_t^4, \quad \hat{\mu}_{2,2}(1) = \frac{1}{T}\sum_{t=1}^{T} y_t^2 y_{t-1}^2, \quad \hat{\mu}_{2,2}(2) = \frac{1}{T}\sum_{t=1}^{T} y_t^2 y_{t-2}^2. \tag{3.5.12}$$

This yields the following estimators of the SV coefficients:

$$\hat{\phi}_1 = \frac{-\left[\log\left(\hat{\mu}_{2,2}(1)/\hat{\mu}_2^2\right)\right]\left[\log\left(3\hat{\mu}_{2,2}(2)/\hat{\mu}_4\right)\right]}{\left[\log\left(\hat{\mu}_4/(3\hat{\mu}_2^2)\right)\right]^2 - \left[\log\left(\hat{\mu}_{2,2}(1)/\hat{\mu}_2^2\right)\right]^2}, \tag{3.5.13}$$

$$\hat{\phi}_2 = \frac{-\left[\log\left(\hat{\mu}_{2,2}(1)/\hat{\mu}_2^2\right)\right]^2 + \left[\log\left(\hat{\mu}_{2,2}(2)/\hat{\mu}_2^2\right)\right]\left[\log\left(\hat{\mu}_4/(3\hat{\mu}_2^2)\right)\right]}{\left[\log\left(\hat{\mu}_4/(3\mu_2^2)\right)\right]^2 - \left[\log\left(\hat{\mu}_{2,2}(1)/\hat{\mu}_2^2\right)\right]^2}, \tag{3.5.14}$$

$$\hat{\sigma}_y = 3^{1/4}\hat{\mu}_2/\hat{\mu}_4^{1/4}, \tag{3.5.15}$$

$$\hat{\sigma}_v = \left[\log\left(\hat{\mu}_4/(3\hat{\mu}_2^2)\right) - \hat{\phi}_1\log\left(\hat{\mu}_{2,2}(1)/\hat{\mu}_2^2\right) - \hat{\phi}_2\log\left(\hat{\mu}_{2,2}(2)/\hat{\mu}_2^2\right)\right]^{1/2}. \tag{3.5.16}$$

From the above analysis, it is clear that the procedure of Dufour and Valéry (2006) can be easily extended to an SV(2) process. We refer this method as the *simple EDV* estimator. This estimator is computationally much simpler than those based on numerical optimization techniques. Similarly, one can compute other higher-order SV models. The expressions of SV(3) or SV(4) estimators are lengthier, so we do not include those equations in the text. However, using this moment-based estimator, we propose a recursive estimation algorithm for SV(p) models in Section 3.5.5.

3.5.2 ARMA-based estimation

In this subsection, we propose another simple estimator for SV(p) models by exploiting the autocovariance structure of y_t^*. We consider a set of moments which are based on $y_t^* = (\log(y_t^2) - \mu)$. The ARMA representation of the observed process $\{y_t^*\}$ is given in the following proposition.

Proposition 3.5.4. ARMA REPRESENTATION OF SV(P) MODELS. *Under the assumptions 3.2.1 - 3.2.2, the process y_t^* defined in* (3.2.12) *has the following ARMA*(p, p) *representation:*

$$y_t^* = \sum_{j=1}^{p} \phi_j y_{t-j}^* + \eta_t - \sum_{j=1}^{p} \theta_j \eta_{t-j} \tag{3.5.17}$$

with $\eta_t - \sum_{j=1}^{p} \theta_j \eta_{t-j} = v_t + \epsilon_t - \sum_{j=1}^{p} \phi_j \epsilon_{t-j}$, where the error processes $\{v_t\}$ and $\{\epsilon_t\}$ are mutually independent, the errors v_t are i.i.d. $N(0, \sigma_v^2)$, and the errors ϵ_t are i.i.d. according to the distribution of a $\log(\chi_1^2)$ *random variable.*

From the above proposition, we have simple expressions for the autocovariances and parameters of the SV(p) model, and these are given in following corollaries.

Corollary 3.5.5. AUTOCOVARIANCES OF THE OBSERVED PROCESS. *Under the assumptions of Proposition 3.5.4, the autocovariances of the observed process y_t^* defined in* (3.2.12) *satisfy the following equations:*

$$\operatorname{cov}(y_t^*, y_{t-k}^*) := \gamma_{y^*}(k) = \begin{cases} \phi_1 \gamma_{y^*}(k-1) + \cdots + \phi_p \gamma_{y^*}(k-p) + \sigma_v^2 + \sigma_\epsilon^2; & \text{if } k = 0, \\ \phi_1 \gamma_{y^*}(k-1) + \cdots + \phi_p \gamma_{y^*}(k-p) - \phi_k \sigma_\epsilon^2; & \text{if } 1 \le k \le p, \\ \phi_1 \gamma_{y^*}(k-1) + \cdots + \phi_p \gamma_{y^*}(k-p); & \text{if } k > p. \end{cases} \tag{3.5.18}$$

Corollary 3.5.6. CLOSED-FORM EXPRESSIONS FOR SV(P) PARAMETERS. *Under the assumptions of Proposition 3.5.4, we have:*

$$\phi_{\boldsymbol{p}} = \boldsymbol{\Gamma}^{-1}_{(\boldsymbol{p}+\boldsymbol{j}-\boldsymbol{1})}\gamma_{(\boldsymbol{p}+\boldsymbol{j})}, \quad j \geq 1 \tag{3.5.19}$$

$$\sigma_y = [\exp(\mu + 1.27)]^{1/2}, \tag{3.5.20}$$

$$\sigma_v = [\gamma_{y^*}(0) - \phi_{\boldsymbol{p}}^{'}\gamma_{(\boldsymbol{1})} - \pi^2/2]^{1/2}, \tag{3.5.21}$$

where $\phi_{\boldsymbol{p}} := (\phi_1, \ldots, \phi_p)^{'}$, $\gamma_{(\boldsymbol{p}+\boldsymbol{j})} := [\gamma_{y^*}(p+j), \ldots, \gamma_{y^*}(2p+j-1)]^{'}$ *are vectors and* $\boldsymbol{\Gamma}_{(\boldsymbol{p}+\boldsymbol{j}-\boldsymbol{1})}$ *is a p-dimensional Toeplitz matrices such that*

$$\boldsymbol{\Gamma}_{(\boldsymbol{p}+\boldsymbol{j}-\boldsymbol{1})} := \begin{bmatrix} \gamma_{y^*}(p+j-1) & \gamma_{y^*}(p+j-2) & \cdots & \gamma_{y^*}(j) \\ \gamma_{y^*}(p+j) & \gamma_{y^*}(p+j-1) & \cdots & \gamma_{y^*}(j+1) \\ \vdots & \vdots & & \vdots \\ \gamma_{y^*}(2p+j-2) & \gamma_{y^*}(2p+j-3) & \cdots & \gamma_{y^*}(p+j-1) \end{bmatrix}.$$

where p *is the SV order,* $\gamma_{y^*}(k) = \text{cov}(y_t^*, y_{t-k}^*)$, *with* y_t^* *and* μ *defined in* (3.2.12).

Now, it is natural to estimate $\gamma_{y^*}(k)$ and μ by the corresponding empirical moments:

$$\hat{\gamma}_{y^*}(k) = \frac{1}{T-k}\sum_{t=1}^{T-k} y_t^* y_{t+k}^*, \qquad \hat{\mu} = \frac{1}{T}\sum_{t=1}^{T}\log(y_t^2), \tag{3.5.22}$$

where by construction y_t^* is a mean corrected process. Setting $j = 1$ in (3.5.19) and replacing theoretical moments by their corresponding empirical moments yield the following *simple ARMA-SV* estimator of the SV(p) coefficients:

$$\hat{\phi}_{\boldsymbol{p}} = \hat{\boldsymbol{\Gamma}}^{-1}_{(\boldsymbol{k},\boldsymbol{p})}\hat{\gamma}_{(\boldsymbol{k},\boldsymbol{p})}, \tag{3.5.23}$$

$$\hat{\sigma}_y = [exp(\hat{\mu} + 1.27)]^{1/2}, \tag{3.5.24}$$

$$\hat{\sigma}_v = [\hat{\gamma}_{y^*}(0) - \hat{\phi}_{\boldsymbol{p}}^{'}\hat{\gamma}_{(\boldsymbol{k},\boldsymbol{p})} - \pi^2/2]^{1/2}. \tag{3.5.25}$$

3.5.3 Restricted estimation

These simple estimators may yield a solution outside the admissible area, *i.e.*, some of the eigenvalues of the latent volatility process [it is an AR(p) process] may lie outside the unit circle or equal to unity. This issue can arise especially in small samples or in the presence of outliers. When this happens, a simple fix is projecting the estimate on the space of acceptable parameter solutions by altering the eigenvalues that lie on or outside the unit circle. The characteristic equation of the latent AR(p) process is given by $C(\lambda) = \lambda^p - \phi_1 \lambda^{p-1} - \cdots - \phi_p = 0$, and the stationary condition requires all roots lie inside the unit circle, *i.e.*, $|\lambda_i| < 1,\ i = 1, \cdots, p$. If the estimated parameters fail to satisfy this condition, then the restricted estimation can be done in the following two steps:

1. Given the estimated unstable parameters, we calculate the roots of the characteristic equation and restrict their absolute values to less than unity.

2. Given these restricted roots, we calculate the constrained parameters which ensure stationarity.

For example, in case of an SV(2) model, the characteristic equation of the latent volatility process is $C(\lambda) = \lambda^2 - \phi_1 \lambda - \phi_2 = 0$. It may have two types of roots: (i) if $\phi_1^2 + 4\phi_2 \geq 0$, then $C(\lambda)$ has two real roots, and these are given by $\lambda_{1,2} = \frac{\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}}{2}$ and (ii) if $\phi_1^2 + 4\phi_2 < 0$ then $C(\lambda)$ has two complex roots, and these are given by $\lambda_{1,2} = \frac{\phi_1}{2} \pm i \frac{\sqrt{-(\phi_1^2 + 4\phi_2)}}{2}$. When the estimated polynomial coefficients produce an unstable solution, then we restrict the absolute value of the roots less than unity, *i.e.* $|\lambda_{1,2}| < 1$ or $|\lambda_{1,2}| = 1 - \Delta$ where Δ is a very small number. Given these restricted roots, we solve for restricted parameters which ensure the stationarity condition. These steps can be done very easily in MATLAB. In MATLAB, the **roots** function calculates the roots given the parameters, and the **poly** function calculates the parameters given the roots.

3.5.4 ARMA-based winsorized estimation

We can achieve better stability and efficiency of ARMA-SV estimator by using "winsorization" which exploits (3.5.19). Winsorization (censoring) substantially increases the probability of getting admissible values. From (3.5.19), it is easy to see that:

$$\phi_{\boldsymbol{p}} = \sum_{j=1}^{\infty} \omega_j \boldsymbol{\Gamma}_{(\boldsymbol{p}+\boldsymbol{j}-\mathbf{1})}^{-\mathbf{1}} \gamma_{(\boldsymbol{p}+\boldsymbol{j})} \tag{3.5.26}$$

for any ω_j sequence with $\sum_{j=1}^{\infty} \omega_j = 1$. Using (3.5.26), we can define a more general class of estimators for $\phi_{\boldsymbol{p}}$ by taking a weighted average of several sample analogs of the ratio $\boldsymbol{\Gamma}_{(\boldsymbol{p}+\boldsymbol{j}-\mathbf{1})}^{-\mathbf{1}} \gamma_{(\boldsymbol{p}+\boldsymbol{j})}$:

$$\tilde{\phi}_{\boldsymbol{p}} = \sum_{j=1}^{J} \omega_j \hat{\boldsymbol{\Gamma}}_{(\boldsymbol{p}+\boldsymbol{j}-\mathbf{1})}^{-\mathbf{1}} \hat{\gamma}_{(\boldsymbol{p}+\boldsymbol{j})}, \tag{3.5.27}$$

where $1 \le J \le T - p$ with $\sum_{j=1}^{J} \omega_j = 1$ and T is the length of time series. We can expect that a sufficiently general class of weights may improve the efficiency of the ARMA-SV estimators.

Using (3.5.27), we can propose alternative estimators of $\phi_{\boldsymbol{p}}$ and we call these estimators *winsorized ARMA-SV* estimators (or *W-ARMA-SV* estimators). Other (possibly nonlinear) averaging methods, such as the median, may also be used. We consider four types of winsorized estimators based on the expression given by (3.5.27) in the simulation section. These estimators are also considered by Hafner and Linton (2017) in the context of closed-form estimation of the EGARCH(1, 1) model.

1. The first, $\hat{\phi}_{\boldsymbol{p}}^{m}$, is an arithmetic mean of sample ratios (equal weights) where we set

$$\omega_j = 1/J, \quad j = 1, \ldots, J, \tag{3.5.28}$$

in (3.5.27). This type of winsorization is also considered by Kristensen and Linton (2006) in the context of the GARCH(1, 1) model estimation.

2. The second, $\hat{\phi}_{\boldsymbol{p}}^{ld}$, is a mean of ratios with linearly declining weights, *i.e.*, it is the estimator in

$$\omega_j = (2/J)[1 - (j/(J+1))], \quad j = 1, \ldots, J. \tag{3.5.29}$$

3. The third is the median: $\hat{\phi}_{\boldsymbol{p}}^{\text{med}} = \text{med}\{\hat{\boldsymbol{\Gamma}}_{(\boldsymbol{p}+\boldsymbol{j}-\mathbf{1})}^{-\mathbf{1}} \hat{\gamma}_{(\boldsymbol{p}+\boldsymbol{j})}\} \quad j = 1, \ldots, J$, where

$$\hat{\phi}_i^{\text{med}} = \left(\text{med}\{\hat{\boldsymbol{\Gamma}}_{(\boldsymbol{p}+\boldsymbol{j}-\mathbf{1})}^{-\mathbf{1}} \hat{\gamma}_{(\boldsymbol{p}+\boldsymbol{j})}\}\right)_{(i,1)}. \tag{3.5.30}$$

4. The fourth is the OLS without intercept regression estimator, given by

$$\hat{\phi}_{\boldsymbol{p}}^{ols} = (\bar{a}'\bar{a})^{-1}\bar{a}'\bar{e}, \tag{3.5.31}$$

where $\bar{a} = (\hat{\Gamma}_{(p)}\omega_1^{1/2},\ldots,\hat{\Gamma}_{(p+J-1)}\omega_J^{1/2})'$ and $\bar{e} = (\hat{\gamma}_{(\boldsymbol{p+1})}\omega_1^{1/2},\ldots,\hat{\gamma}_{(\boldsymbol{p+J})}\omega_J^{1/2})'$. Clearly, different OLS-based W-ARMA-SV can be generated by considering different weights $w_1,\ldots,w_J$. In our simulations below as well as empirical applications, we focus on the case where the weights are equal [see (3.5.28)]. Note that, in case of an SV(2), the W-ARMA-SV-OLS (with equal weights) yields:

$$\hat{\phi}_1^{ols} = \frac{\sum_{j=1}^{J}[\hat{\gamma}_{y^*}(j+1)\hat{\gamma}_{y^*}(j+2)-\hat{\gamma}_{y^*}(j)\hat{\gamma}_{y^*}(j+3)][\hat{\gamma}_{y^*}(j+1)^2-\hat{\gamma}_{y^*}(j)\hat{\gamma}_{y^*}(j+2)]}{\sum_{j=1}^{J}[\hat{\gamma}_{y^*}(j+1)^2-\hat{\gamma}_{y^*}(j)\hat{\gamma}_{y^*}(j+2)]^2} \tag{3.5.32}$$

$$\hat{\phi}_2^{ols} = \frac{\sum_{j=1}^{J}[\hat{\gamma}_{y^*}(j+1)\hat{\gamma}_{y^*}(j+3)-\hat{\gamma}_{y^*}(j+2)^2][\hat{\gamma}_{y^*}(j+1)^2-\hat{\gamma}_{y^*}(j)\hat{\gamma}_{y^*}(j+2)]}{\sum_{j=1}^{J}[\hat{\gamma}_{y^*}(j+1)^2-\hat{\gamma}_{y^*}(j)\hat{\gamma}_{y^*}(j+2)]^2}. \tag{3.5.33}$$

The above simplification [simple regressions] follows from (3.5.19) with $p=2$, which can be written as following:

$$\begin{bmatrix} \phi_1 \\ \phi_2 \end{bmatrix} = \begin{bmatrix} \gamma_{y^*}(j+1) & \gamma_{y^*}(j) \\ \gamma_{y^*}(j+2) & \gamma_{y^*}(j+1) \end{bmatrix}^{-1} \begin{bmatrix} \gamma_{y^*}(j+2) \\ \gamma_{y^*}(j+3) \end{bmatrix} = \begin{bmatrix} \frac{\gamma_{y^*}(j+1)\gamma_{y^*}(j+2)-\gamma_{y^*}(j)\gamma_{y^*}(j+3)}{\gamma_{y^*}(j+1)^2-\gamma_{y^*}(j)\gamma_{y^*}(j+2)} \\ \frac{\gamma_{y^*}(j+1)\gamma_{y^*}(j+3)-\gamma_{y^*}(j+2)^2}{\gamma_{y^*}(j+1)^2-\gamma_{y^*}(j)\gamma_{y^*}(j+2)} \end{bmatrix}. \tag{3.5.34}$$

All these estimators are depend on J and for $J=1$, they are equivalent to the simple ARMA-SV estimator which is given by (3.5.23).

3.5.5 Recursive estimation for SV(p) models

Previously we had shown that it is possible to derive higher-order closed-form solution for SV(p) models. In this section, we propose recursive estimation algorithms for SV(p) models. We use an alternative method provided by Durbin (1960) that avoids the matrix inversion in the Yule-Walker equations. This method is called the Durbin-Levinson (DL) Algorithm, and it is a prediction algorithm. One central feature of DL Algorithm is that we will automatically get partial autocorrelations and mean-squared errors associated with our predictions. For notational convenience, we use a different indexation for the autoregressive parameters of the volatility process [only for this section]. For example, the SV(p) parameters are now denoted by $\Theta_p^{SV} := (\left\{\phi_{p,j}\right\}_{j=1}^{p}, \sigma_{pv}, \sigma_y)'$.

Under the assumptions 3.2.1-3.2.2, the latent volatility process is a stationary AR(p) process that satisfies the Yule-Walker equations. Thus we can apply DL algorithm that is designed for a pure autoregressive process, and we obtain parameters of higher-order SV models recursively. We obtain the extension of the Dufour and Valéry (2006) estimator for SV(p) models by using the following recursive formulae:

$$\phi_{p,p} = \frac{\rho_p - \sum_{j=1}^{p-1} \phi_{p-1,j}\ \rho_{p-j}}{1 - \sum_{j=1}^{p-1} a_{p-1,j}\ \rho_j}, \tag{3.5.35}$$

$$\phi_{p,j} = \phi_{p-1,j} - \phi_{p,p}\phi_{p-1,p-j}, \quad \forall j = 1,2,\ldots,p-1, \tag{3.5.36}$$

where ρ_j is the auto-correlation of the autoregressive process at lag j. Using the following Lemma, we can get the solution of an SV(p) model from an SV($p-1$) model:

Lemma 3.5.7. Recursive moment equation solution. *Under the assumptions 3.2.1 - 3.2.2, the parameters of the SV(p) model,* i.e., $\Theta_p^{SV} := (\left\{\phi_{p,j}\right\}_{j=1}^{p}, \sigma_{pv}, \sigma_y)'$, *can be recursively estimated from the SV($p-1$) model by the following algorithm:*

$$\hat{\sigma}_y = \frac{3^{1/4}\hat{\mu}_2}{\hat{\mu}_4^{1/4}}, \tag{3.5.37}$$

$$\hat{\phi}_{p,p} = \frac{\hat{\rho}_p - \sum_{j=1}^{p-1} \hat{\phi}_{p-1,j}\hat{\rho}_{p-j}}{1 - \sum_{j=1}^{p-1} \hat{\phi}_{p-1,j}\hat{\rho}_j}, \tag{3.5.38}$$

$$\hat{\phi}_{p,j} = \hat{\phi}_{p-1,j} - \hat{\phi}_{p,p}\hat{\phi}_{p-1,p-j} \quad , j = 1,2,\ldots,p-1, \tag{3.5.39}$$

$$\hat{\sigma}_v = \left[(1 - \sum_{j=1}^{k} \hat{\phi}_{p-1,j}\hat{\rho}_j)\log\left(\hat{\mu}_4/(3\hat{\mu}_2^2)\right)\right]^{1/2}, \tag{3.5.40}$$

where

$$\hat{\rho}_j := \frac{\log\left(\hat{\mu}_{2,2}(j)/\hat{\mu}_2^2\right)}{\log\left(\hat{\mu}_4/(3\hat{\mu}_2^2)\right)}. \tag{3.5.41}$$

Note that after calculating the sample auto-correlations, we can estimate the parameters of the model in the second stage with the help of a DL algorithm. The recursive estimation of the ARMA-SV estimator exploits extended Yule-Walker (EYW) equations of the observed process. When the MA order is fixed, the system of the EYW equations constitutes a nested Toeplitz system. A *Generalized Durbin-Levinson* algorithm for the ARMA-SV estimator for SV(p) model is useful when neither the AR order nor the MA order is known. We consider the case $i = p$, *i.e.*, the MA order is p, which also implies that the AR order is p.

For $i = 0$, use the Durbin-Levinson algorithm to calculate

$$\{\hat{\phi}^{(0)}_{p,j} \mid p \geq 1, \quad j = 1, \ldots, p\}.$$

For $i \geq 1$, calculate

$$\hat{\phi}^{(i-1)}_{p,0} = -1,$$

and

$$\hat{\phi}^{(i)}_{p,j} = \hat{\phi}^{(i-1)}_{p+1,j} - \frac{\hat{\phi}^{(i-1)}_{p+1,p+1}}{\hat{\phi}^{(i-1)}_{p,p}} \hat{\phi}^{(i-1)}_{p,j-1}, \quad \text{where } p \geq 1, \quad j = 1, \ldots, p,$$

$$\hat{\sigma}_y = [\exp(\hat{\mu} + 1.27)]^{1/2},$$

$$\hat{\sigma}_{pv} = [\hat{\gamma}_{y^*}(0) - \sum_{j=1}^{p} \hat{\phi}_{p,j} \hat{\gamma}_{y^*}(j) - \pi^2/2]^{1/2}.$$

This algorithm is the same as Tsay and Tiao (1984) algorithm [except for equations involving $\hat{\sigma}_y$ and $\hat{\sigma}_{pv}$] for calculating the extended sample autocorrelation function under the stationarity assumption.

3.6 Andersen-Sørensen type GMM estimation

The simple estimators proposed in the previous section can be viewed as specific cases of GMM estimators where we used a few number of moments. In this section, we propose GMM estimators for SV(p) models with many moments in line with Andersen and Sørensen (1996). In literature, Andersen and Sørensen (1996) proposed GMM estimators for the SV(1) model but the GMM estimator for SV(p) models remains to be discussed. The GMM estimation was formalized by Hansen (1982), and since then it has become one of the most popular methods of estimation for many models in economics and finance. Unlike the MLE, GMM does not require complete knowledge of the distribution of the data. The GMM estimator of SV(p) model is a natural extension of our simple closed-form moment-based estimator. Following the general methodology of GMM, our goal is to minimize the quadratic form with respect to the parameter vector:

$$M_T = [\bar{g}_T(Y_T) - \mu(\theta)]^{'} \hat{\Omega}_T [\bar{g}_T(Y_T) - \mu(\theta)] \tag{3.6.1}$$

where $\mu(\theta)$ is a vector of moments, $\bar{g}_T(Y_T)$ the corresponding vector of empirical moments based on the vector $Y_T = (y_1, \ldots, y_T)^{'}$, and $\hat{\Omega}_T$ a positive-definite (possibly random) matrix. We compute the corresponding sample averages such that

$$g_T(\theta) := \bar{g}_T(Y_T) - \mu(\theta) = \sum_{t=1}^{T} [\bar{g}_t(Y_t) - \mu(\theta)].$$

Now under the standard regularity assumptions,

$$\sqrt{T}[\hat{\theta}_T(\Omega) - \theta_0] \xrightarrow{d} N[0, V(\theta_0 | \Omega)], \tag{3.6.2}$$

where

$$V(\theta_0 | \Omega) = [J(\theta)\Omega J(\theta)^{'}]^{-1} J(\theta)\Omega\Omega_* \Omega J(\theta)^{'} [J(\theta)\Omega J(\theta)^{'}]^{-1}, \tag{3.6.3}$$

and $J(\theta) = \frac{\partial \mu^{'}}{\partial \theta}$. Furthermore, if (i) $J(\theta)$ is a square matrix, or (ii) Ω_* is non-singular and $\Omega = \Omega_*^{-1}$, then

$$V(\theta_0 | \Omega) = [J(\theta)\Omega_*^{-1} J(\theta)^{'}]^{-1} := V_*(\theta). \tag{3.6.4}$$

The $V_*(\theta_0)$ is the smallest possible asymptotic covariance matrix for a method-of-moments estimator based on $M_T(\theta)$. The latter, in particular, is reached when the dimensions of μ and θ are the same, in which case the estimator is obtained by solving the equation

$$\bar{g}_T(Y_T) = \mu(\hat{\theta}_T)$$

Consistent estimators $V(\theta_0 | \Omega)$ and $V_0(\theta_0)$ can be obtained by replacing θ_0 and Ω_* with their consistent estimators. The sample analog of Ω_* is given by

$$\hat{\Omega}_* = \hat{\mathbb{S}}_0 + \sum_{i=1}^{\infty}(\hat{\mathbb{S}}_i + \hat{\mathbb{S}}_i^{'}). \tag{3.6.5}$$

Given this structure, it is natural to estimate $\hat{\Omega}_*$ by truncating this infinite sum and using the sample auto-covariances, where

$$\hat{\mathbb{S}}_j = \frac{1}{T}\sum_{t=j+1}^{T}[g_{t-j}(\hat{y}) - \mu(\theta)][g_{t-j}(\hat{y}) - \mu(\theta)]^{'}$$

with θ replaced by a consistent estimator of it.[3] However for Ω_*, we need to use the heteroskedasticity and autocorrelation covariance (HAC) matrices to avoid any potential inconsistency caused by inappropriate assumptions about the dynamic specification of $[g_t(\hat{y}) - \mu(\theta)]$. This estimator is consistent under relatively weak assumptions on the dependence structure of the process, and this class consists of estimators of the form:

$$\hat{\Omega}_{*,HAC} = \hat{\mathbb{S}}_0 + \sum_{i=1}^{T-1}\omega_{i,T}(\hat{\mathbb{S}}_i + \hat{\mathbb{S}}_i^{'}) \tag{3.6.6}$$

where $\omega_{i,T}$ is known as the kernel (or weight), and it must be chosen to ensure: (i) consistency and (ii) positive semi-definiteness of $\hat{\Omega}_*$. In the literature, there have been few proposed kernel functions that can fit into the above equation.[4] Thus a consistent estimator of $\hat{V}_*(\theta_0)$ is given by

$$\hat{V}_* = [J(\hat{\theta}_T)\hat{\Omega}_*^{-1}J(\hat{\theta}_T)^{'}]^{-1}.$$

Andersen and Sørensen (1996), based on a Monte Carlo simulations study, address several issues related to GMM estimation of SV(1) model. One issue of GMM estimation is the choice of the number of moment conditions. If weighted appropriately, by increasing the number of moment conditions (using additional information), one cannot make the parameter estimates worse. However, the weighting matrix, Ω, must itself be estimated, and with q moment conditions, we need to estimate $q(q+1)/2$ elements of Ω, and a larger number of moment

[3] The truncation parameter l_T is allowed to grow with the sample size such that: $l_T \to \infty$ as $T \to \infty$ and $l_T = 0(T^{1/3})$ see White and Domowitz (1984).

[4] (i) Bartlett kernel by Newey and West (1987), (ii) Parzen kernel by Gallant (1987), and (iii) Quadratic spectral kernel by Andrews (1991).

conditions could lead to poorer estimates of Ω and worse estimates of the parameters. Another issue with GMM is that there is not much guidance on which moment conditions to use. For SV models, one can construct moment conditions based on infinitely many functions of returns; see Melino and Turnbull (1990).

3.7 Asymptotic distributional theory

In this section, we derive asymptotic properties for our simple estimators. For the asymptotic distribution of GMM-type estimators; see Hansen (1982).

3.7.1 Moment-based estimators

Dufour and Valéry (2006) derived an asymptotic distributional theory for the SV(1) estimator. In line with that we establish the asymptotic distributional theory for the moment-based SV(p) estimators. Our approach for constructing an MM estimator is to minimize the quadratic form:

$$M_T = [\bar{g}_T(Y_T) - \mu(\theta)]^{'} \hat{\Omega}_T [\bar{g}_T(Y_T) - \mu(\theta)] \tag{3.7.1}$$

where $\mu(\theta)$ is a vector of moments, $\bar{g}_T(y_T)$ the corresponding vector of empirical moments based on the vector $Y_T = (y_1, \ldots, y_T)^{'}$, and $\hat{\Omega}_T$ a positive-definite (possibly random) matrix. Of course, this estimator belongs to the general family of moment estimators, for which a number of general asymptotic results do exist; see Hansen (1982), Gouriéroux and Monfort (1995) (Volume 1, Chapter 9) and Newey and McFadden (1994).

It is worth noting at this stage that Andersen and Sørensen (1996) did refer to the asymptotic distribution of the usual GMM estimator as derived in Hansen (1982) for the SV(1) model, but without checking the suitable regularity conditions. We want to find the estimator $\hat{\theta}_T(\hat{\Omega}_T)$ by minimizing $M_T(\theta)$, and for that we will consider the following assumptions, where θ_0 denotes the "true" value of the parameter vector θ.

Assumption 3.7.1. ASYMPTOTIC NORMALITY OF EMPIRICAL MOMENTS.

$$\sqrt{T}[\bar{g}_T(Y_T) - \mu(\theta_0)] \xrightarrow{d} N[0, \Omega_*]$$

where $Y_T := (y_1, \ldots, y_T)^{'}$ *and*

$$\Omega_* = \lim_{T\to\infty} \mathbb{E}\{T[\bar{g}_T(Y_T) - \mu(\theta_0)][\bar{g}_T(Y_T) - \mu(\theta_0)]^{'}\}$$

Assumption 3.7.2. ASYMPTOTIC NON-SINGULARITY OF WEIGHT MATRIX.

$$\operatorname*{plim}_{T\to\infty}(\hat{\Omega}_T) = \Omega,$$

where $det(\Omega) \neq 0$.

Assumption 3.7.3. DIFFERENTIABILITY OF WEIGHT MATRIX. $\mu(\theta_0)$ *is twice continuously differentiable in an open neighborhood of* θ_0 *and the Jacobian matrix* $J(\theta_0)$ *has full rank, where* $J(\theta) = \frac{\partial \mu^{'}}{\partial \theta}$.

Given these assumptions, the asymptotic distribution of $\hat{\theta}_T$ is determined by a standard argument on method-of-moments estimation.

Lemma 3.7.1. ASYMPTOTIC DISTRIBUTION OF METHOD-OF-MOMENTS ESTIMATOR. *Under the assumptions 3.2.1 - 3.2.2 and 3.7.1 - 3.7.3,*

$$\sqrt{T}[\hat{\theta}_T - \theta_0] \xrightarrow{d} N[0, V(\theta_0|\Omega)] \tag{3.7.2}$$

where

$$V(\theta_0|\Omega) = [J(\theta)\Omega J(\theta)^{'}]^{-1} J(\theta)\Omega\Omega_*\Omega J(\theta)^{'}[J(\theta)\Omega J(\theta)^{'}]^{-1} \tag{3.7.3}$$

where $J(\theta) = \frac{\partial \mu^{'}}{\partial \theta}$. *If, furthermore, (i)* $J(\theta)$ *is a square matrix, or (ii)* Ω_* *is non-singular and* $\Omega = \Omega_*^{-1}$, *then*

$$V(\theta_0|\Omega) = [J(\theta)\Omega_*^{-1} J(\theta)^{'}]^{-1} := V_*(\theta). \tag{3.7.4}$$

Here, $V_*(\theta_0)$ is the smallest possible asymptotic covariance matrix for a method-of-moments estimator based on $M_T(\theta)$, and a consistent estimator of $\hat{V}_*(\theta_0)$ is given by

$$\hat{V}_* = [J(\hat{\theta}_T)\hat{\Omega}_*^{-1} J(\hat{\theta}_T)^{'}]^{-1}.$$

Since we are using a number of moments equal to the number of parameters, the moment estimator can be obtained by taking $\hat{\Omega}_T$ equal to an identity matrix so that Assumption 3.7.2 automatically holds. Thus we only need to show that the Assumption 3.7.1 holds.

Lemma 3.7.2. ASYMPTOTIC DISTRIBUTION FOR EMPIRICAL MOMENTS. *Under the assumptions 3.2.1 - 3.2.2 with $p > 1$, we have:*

$$\sqrt{T}[\bar{g}_T(Y_T) - \mu(\theta_0)] \xrightarrow{d} N[0, \Omega_*] \tag{3.7.5}$$

where $\bar{g}_T(Y_T) = \sum_{t=1}^{T} g_t$, $g_t = [y_t^2, y_t^4, y_t^2 y_{t-1}^2, \ldots, y_t^2 y_{t-p}^2]'$, and

$$\Omega_* = V[g_t] = \mathbb{E}[g_t g_t'] - \mu(\theta_0)\mu(\theta_0)'.$$

3.7.2 ARMA-based estimators

We derive the asymptotic properties of the ARMA-SV estimator $\hat{\theta} := (\hat{\phi}_1, \ldots, \hat{\phi}_p, \hat{\sigma}_y, \hat{\sigma}_v)'$ under the following set of assumptions.

Assumption 3.7.4. DISTRIBUTION OF THE ERROR PROCESSES. *The error processes z_t and v_t are mutually independent and $\{z_t\}$ is a sequence of i.i.d. real-valued random variables, independent of w_0. The probability distribution of z_t has a continuous density with respect to Lebesgue measure on real line, and its density is positive on $(-\infty, +\infty)$. The transformed error ϵ_t satisfies $\mathbb{E}(|\epsilon_t|)^s < \infty$, where s is an positive integer.*

Assumption 3.7.5. STATIONARITY OF THE LATENT PROCESS. *The latent process $\{w_t\}$ is strictly stationary with $\mathbb{E}(|w_t|)^s < \infty$ and there is an integer $s \geq 1$ such that*

$$\mathbb{E}(|v_t|)^s < \infty, \quad \sum_{j=1}^{\infty} \left|\psi_j\right|^s < \infty, \tag{3.7.6}$$

where $w_t = \phi^{-1}(B) v_t = \psi(B) v_t$ which follows from $\phi(z) \neq 0$ for $|z| \leq 1]$ where the characteristic equation of the volatility process $\phi(z) := 1 - \phi_1 z - \cdots - \phi_p z^p = 0$.

Under the assumptions 3.7.4 and 3.7.5 with $s = 2$, the observed process $\{y_t^*\}$ is strictly stationarity and geometrically ergodic with exponential β-mixing (see results 3.4.1 and 3.4.2) with finite second moment, *i.e.*, $\mathbb{E}\left[(y_t^*)^2\right] < \infty$. In the following Lemma, using Ergodic theorem, we prove the consistency of the empirical moments in (3.5.22).

Lemma 3.7.3. CONSISTENCY OF EMPIRICAL MOMENTS. *Under the assumptions 3.7.4 and 3.7.5 with $s = 2$, the estimators $\hat{\Gamma}(m) := [\hat{\gamma}_{y^*}(0), \hat{\gamma}_{y^*}(1), \ldots, \hat{\gamma}_{y^*}(m)]'$ and $\hat{\mu}$ defined by (3.5.22) satisfy:*

$$\hat{\mu} \xrightarrow{p} \mu \quad \text{and} \quad \hat{\Gamma}(m) \xrightarrow{p} \Gamma(m) := [\gamma_{y^*}(0), \gamma_{y^*}(1), \ldots, \gamma_{y^*}(m)]'. \tag{3.7.7}$$

The assumptions 3.7.4 and 3.7.5 with $s = 4$ are sufficient for the SV model to have a strictly stationary solution with a finite fourth moment of y_t^*, *i.e.*, $\mathbb{E}\left[(y_t^*)^4\right] < \infty$. Note that the fourth moment of y_t^* translates into the eighth moment of y_t. This solution will be β-mixing with geometrically decreasing mixing coefficients. In the following Lemma, using a Central Limit theorem for the stationary and ergodic process (Lindeberg-Levy theorem for the dependent process), we present the asymptotic distribution of the empirical moments in (3.5.22).

Lemma 3.7.4. ASYMPTOTIC DISTRIBUTION OF EMPIRICAL MOMENTS. *Under the assumptions 3.7.4 and 3.7.5 with $s = 4$, the estimators $\hat{\Gamma}(m) = [\hat{\gamma}_{y^*}(0), \hat{\gamma}_{y^*}(1), \ldots, \hat{\gamma}_{y^*}(m)]'$ and $\hat{\mu}$ defined by (3.5.22) satisfy:*

$$\sqrt{T}\begin{bmatrix} \hat{\mu} - \mu \\ \hat{\Gamma}(m) - \Gamma(m) \end{bmatrix} \xrightarrow{d} N\left(0, \begin{bmatrix} V_\mu & C'_{\mu,\Gamma(m)} \\ C_{\mu,\Gamma(m)} & V_{\Gamma(m)} \end{bmatrix}\right), \tag{3.7.8}$$

where

$$V_\mu = \gamma_{y^*}(0) + 2\sum_{\tau=1}^{\infty} \gamma_{y^*}(\tau), \quad V_{\Gamma(m)} = \mathrm{Var}(\Lambda_t) + 2\sum_{\tau=1}^{\infty} \mathrm{cov}(\Lambda_t, \Lambda_{t+\tau}), \quad C_{\mu,\Gamma(m)} = (\bar{c},\, 0_{[1\times m]})', \tag{3.7.9}$$

$$\Lambda_t := [\Lambda_{t,0}, \Lambda_{t,1}, \ldots, \Lambda_{t,m}]', \tag{3.7.10}$$

$$\Lambda_{t,k} := y_t^* y_{t+k}^* - \gamma_{y^*}(k) = [\log(y_t^2) - \mu][\log(y_{t+k}^2) - \mu] - \gamma_{y^*}(k), \quad k = 0, \ldots, m, \tag{3.7.11}$$

$$\bar{c} := C_{\mu,\Gamma(0)} = 2\sum_{t=1}^{\infty} \mathbb{E}[y_t^{*3}] = 2\sum_{t=1}^{\infty} (\mathbb{E}[w_t^3] + \mathbb{E}[\epsilon_t^3]) = 2\sum_{t=1}^{\infty} \mathbb{E}[\epsilon_t^3]. \tag{3.7.12}$$

This in turn yields the asymptotic distribution of the simple ARMA-type estimator $(\hat{\phi}_1, \ldots, \hat{\phi}_p, \hat{\sigma}_y, \hat{\sigma}_v)'$.

Theorem 3.7.5. ASYMPTOTIC DISTRIBUTION OF SIMPLE ARMA-SV ESTIMATOR. *Under the assumptions 3.7.4 and 3.7.5 with* $s = 4$, *the estimator* $\hat{\theta} := (\hat{\phi}_1, \ldots, \hat{\phi}_p, \hat{\sigma}_y, \hat{\sigma}_v)'$ *given in* (3.5.23) - (3.5.25) *is consistent,* i.e., $\hat{\theta} \xrightarrow{p} \theta$, *and*

$$\sqrt{T}(\hat{\theta} - \theta) \xrightarrow{d} \mathrm{N}[0, V], \tag{3.7.13}$$

where $\theta := (\phi_1, \ldots, \phi_p, \sigma_y, \sigma_v)'$ *and*

$$V = G(\beta) \begin{bmatrix} V_\mu & C'_{\mu,\Gamma(2p)} \\ C_{\mu,\Gamma(2p)} & V_{\Gamma(2p)} \end{bmatrix} G(\beta)', \tag{3.7.14}$$

$$G(\beta) := \frac{\partial D_p}{\partial \beta'}, \quad D_p := D_p(\beta) = (D_{\phi_{\boldsymbol{p}}}, D_{\sigma_y}, D_{\sigma_v})', \quad \beta := [\mu, \gamma_{y^*}(0), \gamma_{y^*}(1), \ldots, \gamma_{y^*}(2p)]', \tag{3.7.15}$$

$$D_{\phi_{\boldsymbol{p}}} := \boldsymbol{\Gamma}_{(\boldsymbol{p})}{}^{-1}\gamma_{(\boldsymbol{p+1})}, \quad D_{\sigma_y} := \exp(\mu + 1.27)^{1/2}, \quad D_{\sigma_v} = [\gamma_{y^*}(0) - \phi'_{\boldsymbol{p}}\gamma_{(\boldsymbol{1})} - \pi^2/2]^{1/2}, \tag{3.7.16}$$

$$\phi_{\boldsymbol{p}} := (\phi_1, \ldots, \phi_p)', \quad \gamma_{(\boldsymbol{p+1})} = [\gamma_{y^*}(p+1), \ldots, \gamma_{y^*}(2p)]', \tag{3.7.17}$$

$$\boldsymbol{\Gamma}_{(\boldsymbol{p})_{[p\times p]}} = \begin{bmatrix} \gamma_{y^*}(p) & \gamma_{y^*}(p-1) & \cdots & \gamma_{y^*}(1) \\ \gamma_{y^*}(p+1) & \gamma_{y^*}(p) & \cdots & \gamma_{y^*}(2) \\ \vdots & \vdots & & \vdots \\ \gamma_{y^*}(2p-1) & \gamma_{y^*}(2p-2) & \cdots & \gamma_{y^*}(p) \end{bmatrix}, \tag{3.7.18}$$

$$\gamma_{y^*}(k) = \mathrm{cov}(y^*_t, y^*_{t-k}), \quad y^*_t = (\log y_t^2 - \mu), \quad \mu := \mathbb{E}[\log(y_t^2)]. \tag{3.7.19}$$

The explicit form of the analytical moment derivative, $G(\beta)$, is given in the proof. An estimator of the covariance matrix V can be obtained by first estimating V_μ, $C_{\mu,\Gamma(p)}$ and $V_\Gamma(2p)$ using heteroskedasticity and autocorrelation consistent (HAC) variance estimators [see Den Haan and Levin (1997) and Robinson and Velasco (1997)] and then substituting $\hat{\beta} = [\hat{\mu},\ \hat{\gamma}_{y^*}(0),\ \hat{\gamma}_{y^*}(1),\ \hat{\gamma}_{y^*}(2), \ldots,\ \hat{\gamma}_{y^*}(2p)]'$ into $G(\beta)$. In our empirical applications, we use a Bartlett kernel estimator with the bandwidth varying with the sample size; see Newey and West (1994). One can alternatively use the analytic expressions of $\gamma_{y^*}(k)$ to obtain an estimator of V_μ. The ARMA-type estimator can be viewed as a GMM-type estimator, so one can also use GMM standard errors.

Theorem 3.7.5 covers the simplest ARMA-SV estimator. It is easy to see that the asymptotic distribution of more general winsorized estimators can be derived in the same way upon using Lemmas 3.7.3 - 3.7.4.

3.8 Monte Carlo tests

In this section, we discuss simulation-based inference procedures for SV(p) models. The simulation-based methods are more attainable in the context of this study for two reasons:

1. the SV(p) model is a parametric model with a finite number of parameters, and we can effortlessly simulate this model;

2. we can simulate the test statistic of SV(p) parameters, which is based on a computationally inexpensive estimator. So, using our proposed computationally simple estimators, one can easily construct more reliable finite-sample inference.

It should be noted that the simulation-based procedure may not be attainable when the estimator is computationally expensive, so that we cannot simulate the test statistic easily.

We now examine the usefulness of our simple estimators in the context of simulation-based inference, *i.e.*, Monte Carlo test technique. The technique of Monte Carlo tests was originally proposed by Dwass (1957) for implementing permutation tests and did not involve nuisance parameters. This technique was also independently proposed by Barnard (1963); for a review, see Dufour and Khalaf (2001) and for a general discussion and proofs, see Dufour (2006). It has the great attraction of providing exact (randomized) tests based on any statistic whose finite-sample distribution may be intractable but can be simulated. One can replace the unknown or intractable theoretical distribution $F(S|\theta)$, where $\theta := (\phi_1, \ldots, \phi_p, \sigma_y, \sigma_v)'$, by its sample analog based on the statistics $S_1(\theta), \ldots, S_N(\theta)$ simulated under the null hypothesis.

Let us first consider the pivotal statistics case, *i.e.*, the case where the distribution of the test statistic under the null hypothesis does not depend on nuisance parameters. We can then proceed as follows to obtain an exact critical region.

1. Let S_0 be the observed test statistic (calculated from data).

2. By Monte Carlo methods, draw N i.i.d. replications of S, denoted by $S(N) = (S_1, \ldots, S_N)$ under H_0.

3. From the simulated samples compute the MC p-value $\hat{p}_N[S] := p_N[S_0; S(N)]$, where

$$p_N[x, S(N)] := \frac{NG_N[x; S(N)] + 1}{N+1} \tag{3.8.1}$$

$$G_N[x; S(N)] := \frac{1}{N}\sum_{i=1}^{N} I_{[0,\infty)}(S_i - x), \qquad I_{[0,\infty)}(x) = \begin{cases} 1 & \text{if } x \in [0, \infty), \\ 0 & \text{if } x \notin [0, \infty). \end{cases} \tag{3.8.2}$$

   In other words, $p_N[S_0; S(N)] = (NG_N[S_0; S(N)] + 1)/(N+1)$ where $NG_N[S_0; S(N)]$ is the number of simulated values which are greater than or equal to S_0 . When $S_0, S_1, \ldots, S_N$ are all distinct [an event with probability one when the vector $(S_0, S_1, \ldots, S_N)'$ has an absolutely continuous distribution], $\hat{R}_N(S_0) = N + 1 - NG_N[S_0; S(N)]$ is the rank of S_0 in the series $S_0, S_1, \ldots, S_N$.

4. The MC critical region is: $\hat{p}_N[S] \leq \alpha$, $0 < \alpha < 1$. If $\alpha(N+1)$ is an integer and the distribution of S is continuous under the null hypothesis, then under null,

$$P[\hat{p}_N[S] \leq \alpha] = \alpha; \tag{3.8.3}$$

   see Dufour (2006).

We will now study the case where the distribution of the test statistic depends on nuisance parameters. In other words, we consider a model $\{(\Xi, \mathbb{A}_{\Xi}, P_\theta) : \theta \in \Omega\}$ where we assume that the distribution of S is determined by $P_{\bar{\theta}}$, where $\bar{\theta}$ represents the true parameter vector. To deal with this complication, the MC test procedure can be modified as follows.

1. To test the null hypothesis

$$H_0 : \bar{\theta} \in \Omega_0,$$

   where $\Omega_0 \subset \Omega$, we calculate the relevant test statisticS_0 based on data.

2. For each $\theta \in \theta_0$, by Monte Carlo methods, we generate N i.i.d. replications of S : $S(N, \theta) = [(S_1(\theta), \ldots, S_N(\theta)]$.

3. Using these simulated test statistics, we compute the MC p-value $\hat{p}_N[S|\theta] := p_N[S_0; S(N,\theta)]$, where

$$p_N[x; S(N,\theta)] := \frac{NG_N[x; S(N,\theta)] + 1}{N+1}. \tag{3.8.4}$$

4. The p-value function $\hat{p}_N[S|\theta]$ as a function of θ is maximized over the parameter values compatible with the Ω_0, *i.e.*, the null hypothesis, and H_0 is rejected if

$$\sup_{\theta\in\Omega_0} \hat{p}_N[S|\theta] \leq \alpha. \tag{3.8.5}$$

   If the number of simulated statistics N is chosen so that $\alpha(N+1)$ is an integer, then we have under H_0:

$$P[\sup_{\theta\in\Omega_0}\{\hat{p}_N[S|\theta]\} \leq \alpha] \leq \alpha. \tag{3.8.6}$$

   The test defined by $\hat{p}_N[S|\theta] \leq \alpha$ has size α for known θ . Treating θ as a nuisance parameter and Ω_0 is a nuisance parameter set consistent with null, the test is *exact at level* α; for a proof, see Dufour (2006).

Because of the maximization in the critical region (3.8.5) the test is called a *maximized Monte Carlo* (MMC) test. MMC tests provide valid inference under general regularity conditions such as almost-unidentified models or time series processes involving unit roots. In particular, even though the moment conditions defining the estimator are derived under the stationarity assumption, this does not question in any way the validity of maximized MC tests, unlike the parametric bootstrap whose distributional theory is based on strong regularity conditions. Only the power of MMC tests may be affected. However, the simulated p-value function is not continuous, so standard gradient-based methods cannot be used to maximize it. But search methods applicable to non-differentiable functions are applicable, e.g. simulated annealing [see Goffe et al. (1994)].

A simplified approximate version of the MMC procedure can alleviate its computational load whenever a consistent point or set estimate of θ is available. To do this, we reformulate the setup in order to allow for an increasing sample size, *i.e.*, now the test statistic depends on a sample of size T, $S = S_T$.

1. Let S_{T0} be the observed test statistic (based on data) and the distribution of S involves nuisance parameters under the null and $\bar{\theta} \in \Omega_0$ with $\Omega_0 \subset \Omega$ and $\Omega_0 \neq \emptyset$.
2. we have a consistent set estimator C_T of θ (under H_0) such that
$$\lim_{T\to\infty} P[\bar{\theta} \in C_T] = 1 \text{ under } H_0. \tag{3.8.7}$$
3. For each $\theta \in C_T$, by Monte Carlo methods, we generate N i.i.d. replications of S : $S_T(N,\theta) = [(S_{T1}(\theta),\ldots,S_{TN}(\theta)]$.
4. Using these simulations we compute the MC p-value $\hat{p}_{TN}[S_T|\theta] := p_{TN}[S_{T0}; S_T(N,\theta)]$, where
$$p_{TN}[x; S_T(N,\theta)] := \frac{NG_{TN}[x; S_T(N,\theta)] + 1}{N+1}. \tag{3.8.8}$$
5. The p-value function $\hat{p}_{TN}[S_T|\theta]$ as a function of θ is maximized with respect to θ in C_T, and H_0 is rejected if
$$\sup\{\hat{p}_{TN}[S_T|\theta] : \theta \in C_T\} \leq \alpha. \tag{3.8.9}$$
If the number of simulated statistics N is chosen so that $\alpha(N+1)$ is an integer, then we have under H_0:
$$\lim_{T\to\infty} P[\sup\{\hat{p}_{TN}[S_T|\theta] : \theta \in C_T\} \leq \alpha] \leq \alpha, \tag{3.8.10}$$
i.e., we control for the level asymptotically.

In practice, it is easy to find a consistent set estimate of $\bar{\theta}$, whenever a *consistent* point estimate $\hat{\theta}_T$ of $\bar{\theta}$ available (e.g. a GMM estimator). For instance, any set of the form

$$C_T = \{\theta : \left\|\hat{\theta}_T - \theta\right\| < d\} \tag{3.8.11}$$

with d a fixed positive constant independent of T, satisfies (3.8.7). The consistent set estimate MMC (CSEMMC) method is especially useful when the distribution of the test statistic is highly sensitive to nuisance parameters. Here, possible discontinuities in the asymptotic distribution are automatically overcome through a numerical maximization over a set that contains the true value of the nuisance parameter with probability one asymptotically (while there is no guarantee for the point estimate to converge sufficiently fast to overcome the discontinuity). It is worth noting that there is no need to maximize the p-value function with respect to unidentified parameters under the null hypothesis. Thus, parameters which are unidentified under the null hypothesis can be set to any fixed value and the maximization be performed only over the remaining identified nuisance parameters. When there are several nuisance parameters, one can use simulated annealing, an optimization algorithm which does not require differentiability. Indeed, the simulated p-value function is not continuous, so standard gradient based methods cannot be used to maximize it. An example where this is done on a VAR model involving a large number of nuisance parameters, see Dufour and Jouini (2006).

In Dufour and Khalaf (2002), they call the test based on simulations using a point nuisance parameter estimate a *local* MC (LMC) test. The term local reflects the fact that the underlying MC p-value is based on a specific choice for the nuisance parameter. Here if the set C_T in (3.8.9) is reduced to a single point estimate $\hat{\theta}_T$, *i.e.* $C_T = \{\hat{\theta}_T\}$, we get a LMC test

$$\hat{p}_{TN}[S_T|\hat{\theta}_T] \leq \alpha, \tag{3.8.12}$$

which can be interpreted as a parametric bootstrap test. Note that no asymptotic argument on the number N of MC replications is required to obtain this result; this is the fundamental difference between the latter procedure and the parametric bootstrap method.

Even if $\hat{\theta}_T$ is a consistent estimate of θ (under the null hypothesis), the condition (3.8.7) is not usually satisfied in this case, so additional assumptions are needed to show that the parametric bootstrap procedure yields an asymptotically valid test. It is computationally less costly but clearly less robust to violations of regularity conditions than the MMC procedure; for further discussion, see Dufour (2006). Furthermore, the LMC non-rejections are *exactly* conclusive in the following sense: if $\hat{p}_N[S|\hat{\theta}_0] > \alpha$, then the exact *Maximized Monte Carlo* (MMC) test is clearly not significant at level α.

3.9 Simulation study

In this section, we investigate the properties of the proposed estimators in terms of bias and root mean square error (RMSE) through simulation.

First, we study the finite-sample properties of the winsorized ARMA-SV estimators discussed in Section 3.5.4. We generate an SV(2) processes with $(\phi_1, \phi_2, \sigma_y, \sigma_v) = (0.30, 0.60, 0.025, 2.5)$. We consider different sample sizes $T = (500, 1000, 2000)$ and use 1000 replications. All four censored estimators depend on the truncation parameter J, so we use different values of $J = (1, 5, 10, 20, 30, 40, 50, 100)$. Simulation results are reported in Table 3.1. It is striking that weighted [equal weights and linearly declining weights] and median estimators perform very poorly. These estimators produce a large number of inadmissible values (NIV) for the parameter estimates. Their inferior performance may be due to the high variability of estimated ACF. Further, these estimators give inadmissible parameter values even in large samples and also in different values of J. On the other hand, the simple ARMA-SV estimator outperforms these above mention estimators and produces few inadmissible parameter values. However, the number of impermissible parameter values decline as the sample size increases, and it should be emphasized that the ARMA-SV estimator did not produce any unbound solutions when $T = 2000$. This fact also implies that the variability of estimated ACF goes down as the sample size increases.

However, it is remarkable that the OLS estimates are highly robust, and it outperforms the other three winsorized estimators as well as the ARMA-SV estimator in terms of bias and RMSE, across different sample sizes, particularly in small samples. Further, it is also robust to different values of J. From the reported results; there may be a bias-variance trade-off for higher values of J. Finally, we suggest to use OLS for winsorizing and use small values of J for large samples (or the other way round).

Now we explore the statistical performance of our proposed estimators, these include the moment estimator, the simple ARMA-SV estimator, the winsorized ARMA-SV (W-ARMA-SV) estimator [it is the no intercept regression with $J = 10$] and GMM estimators, in terms of bias and RMSE. Globally, there is no uniform ranking between the different estimators, but the performance of the Bayesian estimator remains superior among the competing methods in the context of SV(1) model. Under the following simulation designs, we compare our proposed estimators to the Bayesian estimator. We use the MATLAB code of Chan and Grant (2016) for the Bayesian estimation with their specified prior.

We simulate four SV(2) models where parameter values of $(\phi_1, \phi_2, \sigma_y, \sigma_v)$ are $M_1 = (0.30, 0.60, 0.025, 2.5)$, $M_2 = (0.95, -0.85, 0.5, 2.5)$, $M_3 = (0.45, 0.45, 0.25, 2.5)$ and $M_4 = (0.0, 0.90, 0.025, 2.5)$. The parameters have been selected arbitrarily since empirical applications of SV(p) models are rare in the literature. The variance of the returns process determined by σ_y and the magnitude of σ_y tends to vary depending on the measurement frequency [intraday, daily, or monthly]. The estimation of σ_y will typically not have much of an effect on the estimation of the other parameters $(\phi_1, \phi_2, \sigma_v)$. The simulations use 1000 replications, and we consider two different sample sizes, $T = (500, 2000)$. These sample sizes are adequate in the sense that in case of low-frequency financial data, the sample size of $T = 1200$ observations corresponds with roughly five years of daily returns, whereas for high-frequency financial data, the sample size of $T = 1200$ observations corresponds with fifteen days of five-minute intraday returns [one trading day is equal to 78 five-minute intraday returns].

In our GMM setting, we consider two sets of moments. One set contains the 24 moment conditions similar to the set of moments that are consider by Andersen and Sørensen (1996) in the context of SV(1) estimation. They recommend using moment conditions for GMM estimation based on lower-order moments, since higher-order moments tend to exhibit erratic finite-sample behavior. The other set considers 6 moment conditions. The large and small sets are denoted by M_L and M_S and given by

$$M_L = \begin{pmatrix} |y_t|^j - \mu_j(\theta) \text{ for } \ j = 1, \ldots, 4 \\ |y_t||y_{t-j}| - \mu_{1,1}(j\,|\theta) \text{ for } \ j = 1, \ldots, 10 \\ y_t^2 y_{t-j}^2 - \mu_{2,2}(j\,|\theta) \text{ for } \ j = 1, \ldots, 10 \end{pmatrix} \quad \text{and} \quad M_S = \begin{pmatrix} |y_t|^j - \mu_j(\theta) \text{ for } \ j = 1, \ldots, 4 \\ y_t^2 y_{t-j}^2 - \mu_{2,2}(j\,|\theta) \text{ for } \ j = 1, 2 \end{pmatrix},$$

where

$$\mu_1(\theta) := \mathbb{E}(|y_t|) = \sigma_y (2/\pi)^{1/2} \exp\left[\gamma_0/8\right], \quad \mu_2(\theta) := \mathbb{E}(y_t^2) = \sigma_y^2 \exp\left[\gamma_0/2\right],$$

$$\mu_3(\theta) := \mathbb{E}(|y_t|^3) = 2\sigma_y^3 (2/\pi)^{1/2} \exp\left[9\gamma_0/8\right], \quad \mu_4(\theta) := \mathbb{E}(y_t^4) = 3\sigma_y^4 \exp\left[2\gamma_0\right],$$

$$\mu_{1,1}(j\,|\theta) := \mathbb{E}(|y_t||y_{t-j}|) = \sigma_y^2 (2/\pi) \exp\left[\gamma_0(1+\rho_j)/4\right], \quad j = 1, \ldots, 10,$$

$$\mu_{2,2}(j|\theta) := \mathbb{E}(y_t^2 y_{t-j}^2) = \sigma_y^4 \exp\left[\gamma_0(1+\rho_j)\right], \quad j = 1, \ldots, 10, \quad \gamma_0 = \sigma_v^2/(1-\phi_1\rho_1-\phi_2\rho_2).$$

Also, we employ two types of GMM estimators based on the choice of weighting matrix (inverse of the asymptotic covariance matrix and HAC covariance matrix using Bartlett Kernel).

Tables 3.2 - 3.5 report the estimation results for model $M_1 - M_4$. These Tables also include the corresponding restricted estimation. In case of an unstable solution, we employ the restricted estimation [as discussed in Section 3.5.3] where the absolute values of the roots are adjusted to less than unity , *i.e.*, $|\lambda_i| = 1-\Delta$ with $\Delta = 0.0001$. We also report the number of inadmissible values (over 1000) for each model in Table 3.6.

From Table 3.6, we can see that GMM, EDV, and ARMA-SV estimators produce several unacceptable parameter values, and NIV goes down as sample size increases. We have similar results for all simulated models. In these cases, we discard those simulations from the calculation. The simple ARMA-SV provides few NIV, and it gives impermissible values in only 0.1% of all simulations when $T = 2000$. The EDV method and the efficient GMM with 24 moments produce a substantial NIV while other GMM estimators produce several NIV. The W-ARMA-SV estimator and Bayesian estimator give no NIV. It should be noted that the Bayesian algorithm draws values under stationarity restriction.

We report the results of M_1 in Table 3.2. The results suggest that W-ARMA-SV estimates are superior, while GMM, EDV and MCMC estimators are inferior in terms of bias and RMSE. For each parameter, the W-ARMA-SV and ARMA-SV estimators produce the smallest and the second smallest bias and RMSE, respectively. The ARMA-SV produces 17 inadmissible values when $T = 500$ and none when $T = 2000$. RMSE associated with the restricted ARMA-SV (R-ARMA-SV) estimates are close to ARMA-SV estimates. In most cases, restricted estimation corresponding to other methods [EDV, GMM] improves the RMSE but produces more bias, showing that it reduces the variance of the estimates. We also find that GMM, EDV and MCMC methods are biased, and the size of these biases is substantial. For the larger samples, $T = 2000$, we have almost identical results for all the parameter estimates as with $T = 500$. Again W-ARMA-SV outperforms all other estimators in terms of bias and RMSE. The RMSE of ARMA-SV and W-ARMA-SV estimates decreases as the sample size increases, shows the consistency of these estimators.

The results of M_2, M_3 and M_4 models are reported in Tables 3.3-3.5, and results are comparable to the ones reported in Table 3.2. Again, in these designs, the performance of ARMA-SV and W-ARMA-SV estimators stands out. However, there are some important findings. In M_2, EDV and 24 moments efficient GMM produce numerous inadmissible parameter values, almost 95% when $T = 500$ and 99% when $T = 2000$. In M_3, we find that some of GMM, EDV and MCMC estimates are exploded with substantial bias and RMSE and this problem persists even when $T = 2000$. All the GMM estimators perform very badly and this is also true for the Bayesian estimator. The performance of GMM estimators cast doubt on the advice that one should use many moments, thereby the chance of including irrelevant ones goes up. This assertion is documented in the literature on asymptotic theory; see Buse (1992), Chao and Swanson (2007), and Dufour and Valéry (2006). Note that, overidentification leads to biased GMM estimators in finite-samples. Concurring evidence, based on finite-sample optimality results and Monte Carlo simulations, is also available in Dufour and Taamouti (2003).

We also encounter the non-convergence problem with the Bayesian estimation. The EDV estimator of σ_v^2 sometimes produces a negative value. Note that in each simulation, the W-ARMA-SV estimator yields a solution. We can draw several conclusions from these simulation results for the W-ARMA-SV and ARMA-SV estimator. First, when $T = 500$, these estimators provide accurate estimates since it outperforms all other estimators in terms of bias and RMSE. Second, the W-ARMA-SV is more efficient than other estimators in terms of RMSE for all simulation design. These results show that W-ARMA-SV not only improves stability but also increases efficiency. From Table 3.7, the simple estimators are highly time efficient, and the margin of time efficiency is huge compared to other estimators.

Simulation results also show that the Bayesian method is very fragile, and the convergence of this method depends on the choice of the prior distribution. The specified prior [for the SV(2) model] of Chan and Grant (2016) produces a substantial bias for all four parameter estimates, indicating that their choice of the prior distribution is terrible. A better chosen prior could conceivably have a better performance, but the result will always depend on the true (unknown) parameter values whose domain gets larger as the order of the process increases. The Bayesian method requires different prior distributions as the order of SV(p) model changes, and the computational cost increases as well with p.

3.10 Empirical applications

In this section, we demonstrate two empirical applications of SV(p) models. *First*, we examine the fit of SV(p) models with real data to see the empirical evidence of this type of parametric models. *Second*, we further extend our analysis and compare the forecast performance of three typical volatility models in out-of-sample forecasting experiments, models include the GARCH-type, SV-type and high-frequency realized volatility based models.

3.10.1 Empirical evidence

The SV(p) models are fitted to daily observations of the Standard and Poor's (S&P) Composite Price Index. The raw series p_t is converted to returns by the transformation $r_t := 100[\log(p_t) - \log(p_{t-1})]$ and the returns are converted to residual returns by $y_t := r_t - \hat{\mu}_r$, where $\hat{\mu}_r$ is the sample average of returns. The sample period is from January 3, 1928 to September 27, 2016 and the number of observations is $T = 23,372$. This data is obtained from Wharton Research Data Services (WRDS). The sample includes many volatile periods that cover the Great Depression (1929), the Second World War (1937-45), the OPEC oil price shock (1973), the Black Monday (1987), the Asian financial crisis (1997), the early 2000s recession (Dot-com bubble), the late-2000s financial crisis (subprime mortgage crisis / United States housing bubble) and the recent Russian financial crisis (2014).

Table 3.8 reports summary statistics of the daily residual returns (y_t) and its several transformed series ($y_t^2, \log|y_t|, y_t^*$). We observe that the skewness and kurtosis of y_t and y_t^2 show the evidence of non-normal distribution, while the distributions of log-transformed residual returns are close to normal. This result is consistent with most empirical studies. Table 3.9 shows the parameter estimates of the SV(p) models (where $p = 1, 2, 3, 4$) using our W-ARMA-SV estimator. We use (3.5.31) with $J = 7$ with equal weights to estimate ϕ_p. Since we have a long financial time series, small values of $J \in (1:10)$ is more appropriate (according to simulation results), and $J = 7$ provides us stable solutions across the estimated models, *i.e.*, roots of the estimated autoregressive parameters of the latent volatility process are inside the unit circle.

Our results show that there is some persistence in the volatility process during the period 1928-2016, and this is statistically significant. We have also found that parameters of SV(p) models, where $p = 1,2,3$, are statistically significant. This finding suggests that the latent volatility process can be treated as an autoregressive process of order more than one. It is also in line with Asai (2008). However, the estimated SV(4) model is insignificant. Again from Table 3.9, we can see that the W-ARMA-SV estimator is extremely efficient from the viewpoint of computation time.

Table 3.9 reports p-values based on the usual large-sample approximation, *i.e.*, the HAC covariance estimator. The variance-covariance $\hat{V}$ is estimated by a Bartlett kernel estimator with the bandwidth varying with the sample size, *i.e.* $m = [1.14T^{1/3}]$, where $[\cdot]$ denotes the integer part of the enclosed number; see Newey and West (1994). Note that the asymptotic standard error can be markedly different and may be quite unreliable in finite-samples. To construct a more reliable finite-sample inference, we can compute the Monte Carlo tests (discussed thoroughly in section 3.8) since our estimator is convenient for use in the context of computationally costly inference techniques. We implemented parametric Bootstrap or LMC tests as discussed in section 3.8 where we replace the nuisance parameters by corresponding point estimates and simulate the test statistic under the null hypothesis. Except for ϕ_1 and ϕ_2 parameters of the SV(3) model and σ_y in all models, we test each coefficient equals zero against a right-sided alternative employing a t-type test statistic. Note that, we cannot test $\phi_1 = 0$ and $\phi_2 = 0$ in the SV(3) model because each of these restrictions leads to the latent volatility process non-stationary. In this situation, we cannot have a stationary SV(3) model, which makes ARMA-based estimation is infeasible. Thus we test $\phi_1 = 0.2$ and $\phi_2 = -0.4$ against a right-sided and a left-sided alternative, respectively. Further, we test $\sigma_y = 0.01$ against a right-sided alternative since when $\sigma_y = 0$, SV models are unidentified. Results of LMC tests are also reported in Table 3.9. From the results, we can see that the SV(4) model is not statistically significant in both asymptotic and finite-sample parametric Bootstrap tests, and this entails that an SV(3) model could be more suitable for the volatility dynamics of this sample periods. Note that, one can easily exploit these simple estimators and construct exact tests based on MMC procedure as discussed in Section 3.8. To summarize, the results presented in Table 3.9 indicate that SV models with additional lag terms in the volatility process may be appropriate to model the S&P 500 index.

The choice of J plays a very crucial role in ARMA-SV estimation; therefore, we scrutinize the possible choices of J with small samples. We consider three sample periods:

1. The first sample period is from January 2, 1996 to September 27, 2016. This is roughly 20 years and the number of daily observations $T = 5222$.

2. In the second sample, we consider approximately 10 years of daily observations from January 3, 2006 to September 27, 2016, where $T = 2703$.

3. The third sample includes the 2008 financial crisis, from January 3, 2006 to December 31, 2010, which gives $T = 1259$.

For each of these samples, we consider SV(p) models with $p = [1, \ldots, 4]$ and plot the W-ARMA-SV-OLS estimates of the volatility persistence parameters for $J = [1, \ldots, 100]$. We also report the corresponding SV(p) estimates for $J = [10, 20, 30, 40, 50, 60, 70, 80, 90, 100]$ in Table 3.10.

From Figure 3.1-3.3 and Table 3.10, we can see the evolution of ϕ's as a function of J. At $J = 25$, almost all models provide stationary solutions while when $J = 50$, we get very stable solutions across different samples and models. It should be noted that the truncation parameter J plays a smoothing role and simulation results suggest that there is a bias-variance trade-off for ϕ estimators as J increases. Therefore, we desire a moderate level of winsorizing, and in the volatility forecasting application below we consider $J = 50$.

3.10.2 Volatility forecast performance

We evaluate the volatility forecast performance amongst GARCH, SV, and realized volatility based models.[5] Volatility has long been modeled and forecasted using GARCH models because of the earlier discussed complexity of SV models. We considered several popular GARCH-type models in our experiments, these include: GARCH models of Bollerslev (1986),

[5] Realized volatility (RV), is a model free volatility, received much attention among the financial economists and econometricians as an accurate measure of the true latent volatility under the ideal market assumption [Andersen and Bollerslev (1998), Barndorff-Nielsen and Shephard (2001)] and it can be used as a proxy for true latent volatility (for details about RV and related measure, see Section 4.7.2).

Exponential GARCH (EGARCH) models of Nelson (1991) and GJR models of Glosten et al. (1993). For the details of these models and their forecast equations, see appendix 3.12.5.

We also consider high-frequency based Heterogenous Autoregressive model of Realized Volatility (HAR-RV) models of Corsi (2009). In this study, we use a logarithmic version of HAR-RV model since the logarithmic transformation of RV appears approximately Gaussian; see Andersen, Bollerslev, Diebold and Ebens (2001), Andersen, Bollerslev, Diebold and Labys (2001). The HAR-RV model takes into account the long memory feature of realized volatility, and among the models proposed to forecast realized volatility, it stands out because of its simplicity (for details, see appendix 3.12.6).

For SV(p) models, we exploit the state-space representation in (3.2.14-3.2.15) and calculate forecasts based on the Kalman filter. The SV(p) parameters are computed using our simple method, where we used (3.5.31) with $J = 50$ and fixed the value of J before any estimations. Given the simple estimates, we calculate the forecasts of SV(p) models through the Kalman filter. For the details of this forecasting procedure and out-of-sample forecasting equations, see appendix 3.12.4.

We use three loss measures to evaluate the forecast accuracy. These include MSE, MAE, and R2LOG. MSE and MAE are the mean squared error and mean absolute error, respectively, and R2LOG is the logarithmic loss function proposed by Pagan and Schwert (1990a) and can penalize volatility forecast asymmetry in high and low level of volatility. These loss measures are defined as follows:

$$MSE: l_t = (\hat{\sigma}_t^2 - h_{t|t-k}^2)^2, \quad MAE: l_t = |\hat{\sigma}_t^2 - h_{t|t-k}^2|, \quad R2LOG: l_t = (\log \hat{\sigma}_t^2 - \log h_{t|t-k})^2,$$

where $\hat{\sigma}_t^2$ is an unbiased *ex-post* proxy of conditional variance (such as squared return or realized volatility) and $h_{t|t-k}$ is a volatility forecast based on $t-k$ information set where $k > 0$.

Using the above loss functions, we also compute the model confidence set (MCS) procedure proposed by Hansen et al. (2011). The model confidence set involves a sequence of tests for equal predictive ability (EPA) hypothesis. Given a model set $\mathcal{M}_0$, which contains m competing forecast models, the null hypothesis is that all models in $\mathcal{M}_0$ have equal predictive accuracy. If the null hypothesis is rejected at a given confidence level α, then the worst performing model in $\mathcal{M}_0$ is eliminated. After that, the EPA test is repeated until the null hypothesis is accepted. When the null hypothesis is accepted, the remainder composes $1-\alpha$ confidence set, $\hat{\mathcal{M}}^*_{1-\alpha}$.

We now briefly discuss how it is implemented. Define the relative loss differential between models by

$$d_{i,j,t} = l_{i,t} - l_{j,t}, \quad \text{for all} \;\; i, j \in \mathcal{M}, \quad t = 1, \ldots, T,$$

be the simple loss of model i relative to any other model j at time t. Using the loss differential between competing models, the MCS procedure tests the EPA hypothesis in two alternative ways.

$$H_0 : \mu_{ij} = 0 \quad \text{for all} \;\; i, j \in \mathcal{M} \quad \text{and} \quad H_A : \mu_{ij} \neq 0 \quad \text{for some} \;\; i, j \in \mathcal{M} \tag{3.10.1}$$

or

$$H_0 : \mu_{i,\cdot} = 0 \quad \text{for all} \;\; i \in \mathcal{M} \quad \text{and} \quad H_A : \mu_{i,\cdot} \neq 0 \quad \text{for some} \;\; i \in \mathcal{M} \tag{3.10.2}$$

where $\mu_{ij} = \mathbb{E}(d_{i,j})$ and $\mu_{i,\cdot} = \mathbb{E}(d_{i,\cdot})$. The two statistics, used in the model confidence set test, are expressed as follows:

$$MCS\_T_{R,\mathcal{M}} = \max_{i,j \in \mathcal{M}} |\, t_{i,j} \,| \quad \text{and} \quad MCS\_T_{\max,\mathcal{M}} = \max_{i \in \mathcal{M}} t_{i,\cdot}, \tag{3.10.3}$$

where

$$t_{i,j} = \frac{\bar{d}_{i,j}}{\sqrt{\widehat{\text{Var}}(\bar{d}_{i,j})}}, \quad t_{i,\cdot} = \frac{\bar{d}_{i,\cdot}}{\sqrt{\widehat{\text{Var}}(\bar{d}_{i,\cdot})}},$$

$$\bar{d}_{i,\cdot} = m^{-1} \sum_{j \in \mathcal{M}} \bar{d}_{i,j}, \quad \bar{d}_{i,j} = T^{-1} \sum_{t=1}^{T} d_{i,j,t} \quad \text{for} \;\; i, j \in \mathcal{M},$$

while $\widehat{\text{Var}}(\bar{d}_{i,\cdot})$ and $\widehat{\text{Var}}(\bar{d}_{i,j})$ are bootstrap estimates of $\text{Var}(\bar{d}_{i,\cdot})$ and $\text{Var}(\bar{d}_{i,j})$, respectively. In our calculations, we perform a block-bootstrap using a block length of 12 days and 10000 bootstrap replications. The first statistic, $t_{i,j}$, is used in the well-known test for comparing two forecasts; see Diebold and Mariano (2002) and West (1996), while the second one, $t_{i,\cdot}$, is used in Hansen et al. (2003), Hansen (2005), and Hansen et al. (2011).

We conduct two out-of-sample forecast experiments using different volatility proxy:

1. **Design 1 (Moderate volatility regimes)**: In this setting, we consider a sample period from September 01, 2005 to August 31, 2010. The in-sample is from September 01, 2005 to August 31, 2008 and the out-of-sample is from September 01, 2008 to August 31, 2010. We forecast a moderately volatile period but the in-sample contains the most volatile part of the late-2000s financial crisis.

2. **Design 2 (High volatility regimes)**: In this design, we consider a sample period, from January 01, 2005 to December 31, 2009. The in-sample is from January 01, 2005 to December 31, 2007 and the out-of-sample is from January 01, 2008 to December 31, 2009. The out-of-sample includes a highly volatile period, *i.e.*, the late-2000s financial crisis (Subprime mortgage crisis / United States housing bubble).

In both designs, we consider a sample of five years that split into three years span of in-sample and two years span of out-of-sample. Three years span for the in-sample window is adequate for finding the most accurate volatility forecasts; see Kambouroudis and McMillan (2015).

Within the SV and GARCH framework, the key element is the specification for conditional variance. Parametric SV and GARCH models utilize daily returns (typically squared returns) to extract information about the current level of volatility, and this information is used to form expectations about the next period's volatility. Although the squared return is a noisy measure, it is a conditionally unbiased estimator of the daily conditional variance. In contrast, Andersen and Bollerslev (1998) suggest that realized volatility (which is based on cumulative intraday squared returns) is a more accurate proxy for true latent volatility. Therefore, we examine out-of-sample volatility forecasts across competing models using different loss functions as well as the MCS procedure with squared return and realized volatility proxies.

We computed out-of-sample forecasts using rolling (moving) window method and computed for a range of forecast horizons which are 1-day, 2-day, 1-week, 2-week, 3-week and 1-month. In this rolling forecasts setup, an initial sample using data from $t = 1, \ldots, T$ is used to determine a window width T, to estimate the models, and to form $h-$step ahead out-of-sample forecasts starting at time T. Then the window is moved ahead one time period, the models are re-estimated using data from $t = 2, \ldots, T+1$, and $h-$step ahead out-of-sample forecasts are produced starting at time $T+1$. This process is repeated until no more $h-$step ahead forecasts can be computed.

3.10.2.1 Forecasting squared return

Now we consider the daily squared return as volatility proxy and evaluate the volatility forecast performance among GARCH, SV and HAR-RV models using the S&P 500 index. The high-frequency RV estimates and prices of S&P 500 index are obtained from the Oxford-Man Institute's Realized Library; see Heber et al. (2009). The raw prices p_t are converted to returns by the transformation $r_t = 100[\log(p_t) - \log(p_{t-1})]$. The returns are converted to residual returns by $y_t = r_t - \hat{\mu}_r$ where $\hat{\mu}_r$ is the sample average of returns. Note that, y_t^2 is the volatility proxy at time t.

We consider a modified version of the HAR-RV model [that defined in (3.12.100)], where the daily squared return is the dependent variable and realized volatilities are independent variables. We are using additional information from the high-frequency data to forecast the squared return. However, there is a problem in measuring the realized volatility from high-frequency data. The high-frequency estimate of realized volatility may be very unstable because of the market microstructure noise, which captures a mixture of frictions inherent in the trading mechanism: bid-ask bounces, discreteness of price changes, different price impacts due to differences in trade sizes, slow response of prices to a block trade, strategic component of the order flow, inventory control effects, etc. The choice of RV estimator is important. We consider other RV estimates, including: realized bi-power variation (BV) [Barndorff-Nielsen and Shephard (2006)], realized semi-variance (RSV) [Barndorff-Nielsen et al. (2010)], realized kernel (RK) [Barndorff-Nielsen et al. (2008, 2011)], median realized volatility (MedRV) [Andersen et al. (2012)], two-scale realized kernels (TSRK) [Ikeda (2013)]. We also consider subsampled RV, BV, and RSV. Subsampling, introduced by Zhang et al. (2005), is a simple way to improve the efficiency of sparse-sampled estimators.[6]

For each forecast experiment, we compute forecasts from three SV models, eleven

[6]Subsampling involves using a variety of "grids"of prices sampled at a given frequency to obtain a collection of realized measures, which are then averaged to yield the "subsampled"version of the estimator. For example, 5-minute RV can be computed using prices sampled at 10:30, 10:35, etc. and can also be computed using prices sampled at 10:31, 10:36, etc.

GARCH-type models, and nine HAR-RV type models. The eight GARCH-type models are: GARCH$(1,1)$, GARCH$(1,2)$, GARCH$(2,1)$, GARCH$(2,2)$, GARCH$(3,3)$, EGARCH$(1,1)$, EGARCH$(2,2)$, EGARCH$(3,3)$, GJR$(1,1)$, GJR$(2,2)$ and GJR$(3,3)$.

For S&P 500, in design 1, the sample period is from September 01, 2005 to August 31, 2010 and the number of observations is $T = 1258$. The in-sample is from September 01, 2005 to August 31, 2008 ($T = 753$) and the out-of-sample is from September 01, 2008 to August 31, 2010 ($T = 505$). In design 2, the sample period is from January 01, 2005 to December 31, 2009 and the number of observations is $T = 1259$. The in-sample is from January 01, 2005 to December 31, 2007 ($T = 754$) and the out-of-sample is from January 01, 2008 to December 31, 2009 ($T = 505$).

Tables 3.11-3.12 report summary statistics of daily variables, high-frequency RV estimates, and their logarithms. Using our out-of-sample forecasts, we calculate forecast evaluation measures, *i.e.*, MSE, MAE and R2LOG. Tables 3.13-3.18 presents the main results of our forecasting experiments. For easy comparison, we report the relative MSE, the relative MAE and the relative R2LOG of forecast error. These are relative to the reference model HAR-RV5, and hence, values smaller than unity indicate better forecast performance than the HAR-RV5 model. We also report the MCS p-value for the corresponding model.

We check the stationarity of SV(p) estimates and find that the W-ARMA-SV-OLS ($J = 50$) estimator produces a few unstable solutions for the SV(3) model. In these cases, the unstable estimates are modified by applying the restricted estimation (proposed in Section 3.5.3). We report both the unrestricted and restricted SV(3) forecasts in Table 3.13-3.18. However, these forecasts are qualitatively similar. Note that, in the realized volatility forecasting application below, we do not find any unstable parameter estimates of SV models.

In design 1, Tables 3.13-3.15, when we forecast a moderately unstable period after the core financial crisis, the forecasting performance of higher-order SV models [especially, the SV(3) model (unrestricted or restricted)] are superior to all other volatility models. This result holds across different forecast horizons, different evaluation measures and based on MCS. According to MCS, the SV(3) model dominates all other competing models, except for 1- and 2-weeks horizon as per MSE loss function. Several HAR-RV models performed well according to MSE, but these results are undermined by their performance in terms of MAE and R2LOG. Except for one- and two-day horizons, the forecasting performance of GARCH-type models is poor for all models, according to all loss measures and across different forecast horizons.

In design 2, Tables 3.16-3.18, when we forecast a highly volatile period, *i.e.*, the core financial crisis, SV(p) models perform better than other competing models in most cases (this holds across different forecast horizons and different evaluation measure) while HAR-RV models perform better than GARCH models. The SV(3) model produces the superior forecast in terms of MSE criteria in horizon 1-day, 2-day, 3-weeks, and 1-month. SV(p) models [SV(3) or SV(2)] are the top forecasting models based on MCS p-value when using MAE and R2LOG. Performances of GARCH and HAR-RV models are poor according to R2LOG. These models may produce asymmetric forecast errors because R2LOG heavily penalized asymmetry in a high and low level of volatility. However, HAR-RV models perform better than GARCH models according to RMSE, while GARCH models outperform HAR-RV models according to MAE and R2LOG. This implies that GARCH models produce large forecast errors because MSE heavily penalized any outlier.

In both settings, among HAR models, those that based on the subsampled version of RV estimators produces identical forecasts, implies that subsampling is not improving any forecast performance. Further, the performance of HAR models is inferior among all models in long-horizon. Note that in both designs, the financial crisis is included either in-sample or out-of-sample. During this time, the financial market is unstable, and the high-frequency RV estimators are affected by large market microstructure noise. The forecasting performance of HAR-RV models may be affected by these noisy RV estimators.

From Tables 3.13-3.18, we can see that except for a few instances, higher-order SV models perform better than GARCH-type and HAR-RV type models not only in all evaluation measures but also across different volatility regimes and horizons. In both of our out-of-sample experiments, higher-order SV models also outperform the first-order SV model. This finding suggests that additional lag terms in the latent volatility equation are essential for forecasting volatility.

3.10.2.2 Forecasting realized volatility

In previous forecast experiments, we forecast the daily squared return, which is a noisy proxy for the true latent volatility. Now, we consider realized volatility as a volatility proxy since it is an accurate measure of the true latent volatility; see Andersen and Bollerslev (1998), Barndorff-Nielsen and Shephard (2001). In this section, we compare the performance of three SV models to the HAR-RV model. Several estimators have been proposed in realized volatility literature, but following the results of Liu et al. (2015), we use the 5-minute RV which is constructed from five-minute intraday returns. In the case of SV(p) models, we replace the squared return by realized volatility (squared return is considered as the observed process in SV models) and then estimate the models by our W-ARMA-SV estimator.

We consider five assets: S&P 500, FTSE100, NASDAQ100, N225, SSMI20 indices. Their 5-minute realized volatilities are sourced from the Oxford-Man Institute's Realized Library. The main results of these forecast experiments are reported in Tables 3.19-3.24. For easy comparison, we report the relative MSE, the relative MAE, and the relative R2LOG of forecast error. These are relative to the HAR-RV model, and hence, values smaller than unity indicate better forecast performance than the HAR-RV model. We also report the MCS p-value for the corresponding model and highlight the best model by boldface color font.

In design 1, Tables 3.19-3.21, when we forecast a moderately volatile period after the financial crisis, in most cases, higher-order SV models [SV(2) or SV(3)] provide superior forecasts. This finding is consistent across different evaluation measures. Out of 30 cases (across five assets and six forecast horizons), SV models delivered the best forecast performance in almost all cases (according to MSE, MAE, R2LOG), except for the 1-week ahead forecasting of SSMI20 volatility in terms of MSE. The forecasting performances of higher-order SV models are ranked top in 80%, 97%, and 97% of cases according to MSE, MAE, R2LOG, respectively. If we consider only longer horizons (2-week, 3-week, 1-month) then these winning percentages of SV(p) models are increased to 87%, 100% and 100%, while for short horizons (1-day, 2-day, 1-week) these percentages are 73%, 93%, and 93%. One out of ninety cases (across five assets, six forecast horizons, and three loss measures), the HAR-RV model produces the best forecasting performance.

Between the SV(2) and SV(3) model, the performance of the SV(2) model is better in short forecast horizons but SV(3) is better in long horizons. This finding tells us that additional lag term is essential for forecasting realized volatility in long horizons. Compared to all other models, the forecasting performance of the SV(3) model is getting better as the forecast horizon increases. Note that the MCS p-values of other competing models declined significantly in long horizons. The performance of HAR-RV is clearly poor compared to SV(p) models in long horizons since the relative loss measures of SV(p) models are now lower.

In design 2, Tables 3.22-3.24, when we forecast highly volatile periods such as a crisis or expansion, the ranking of models are similar to design 1. Out of ninety cases (across five assets, six forecast horizons, and three loss measures), HAR-RV model produced the best forecasts in 3% of cases, whereas SV(p) models delivered best forecasts in 86% of cases.

In both designs, our findings suggest that SV(p) models are better in forecasting realized volatility. So fitting non-parametric volatility measures in traditional parametric models can provide better forecasting performance. This also tells us that the HAR-RV model is not capturing the proper mean dynamics that comes from the moving average part of the market microstructure noise during the financial crisis. As pointed out by Meddahi (2003), if several factors influence the dynamics of RV, then RV follows an ARMA-type process. In this study, within a parametric SV framework, we model realized volatility as a non-Gaussian ARMA process [see Lemma 3.5.4].

3.11 Conclusion

In this paper, we propose several estimators for higher-order SV models, and these include computationally simple estimators and GMM-type estimators. The motivation, as well as the stationarity, ergodicity and mixing properties of SV(p) models, are thoroughly discussed. This study also develops recursive estimation procedures for SV(p) models using simple estimators and derives asymptotic distributions of these simple estimators. We show that simple estimators are especially convenient for use in the context of simulation-based inference techniques, *i.e.*, Bootstrap or Monte Carlo tests.

In simulations, we compare our proposed estimators to the Bayesian estimator. The simple winsorized ARMA-SV estimator uniformly outperforms all other estimators in terms of bias and statistical efficiency. This conclusion holds across different simulation designs. Furthermore, proposed simple estimators are highly time efficient compared to other estimators.

Our results cast doubt on the use of a large number of moments. In this respect, one should not include too many instruments since it can increase the chance of including irrelevant ones in the estimation procedure. In particular, over-identification increases the bias of GMM estimators in finite-samples. Concurring evidence based on finite-sample optimality results and Monte Carlo simulations is also available in Dufour and Taamouti (2003). In an optimal GMM setting, the number of moment equations should be equal to the number of parameters, provided that these moments are well selected. These types of GMM estimators are efficient and good at forecasting. In that sense, the ARMA-SV estimator, based on a few moments, is a nearly optimal and parsimonious moment-based (or GMM) estimator.

In empirical illustrations, we find that asset returns can be better modeled as an SV(p) model, an observation confirmed by both asymptotic and finite-sample tests. We also find that the forecasting performance of higher-order SV models is superior to the one of GARCH and HAR-RV models. This finding holds even if a high volatility period (such as financial crisis) is included in the estimation sample or the forecasted sample. These inferences are not only based on a standard forecasting precision assessment but also on formal prediction tests. These findings highlight the usefulness of higher-order SV models for volatility forecasting.

3.12 Appendix

3.12.1 Proofs

PROOF OF LEMMA 3.3.1 Under the assumption 3.3.1, we have an MFSV model where the volatility process is driven by the sum of m independent AR(1) process. Granger and Morris (1976) shown that the sum of m independent AR(1) processes is an ARMA($m, m-1$) process. The proof follows from there. Note that Meddahi (2003) derived ARMA representation of integrated and realized variances when the spot variance depends linearly on two autoregressive factors. This class of processes includes affine, GARCH diffusion, as well as the eigenfunction stochastic volatility and the positive Ornstein-Uhlenbeck models. □

PROOF OF LEMMA 3.3.2 We consider an SV(p, q) model defined by Lemma 3.3.1 with the latent volatility process driven by an ARMA(p, q) such that

$$\alpha(B)w_t = \beta(B)\sigma_v v_t$$

where

$$\alpha(B) = 1 - \alpha_1 B - \alpha_2 B^2 - \ldots - \alpha_p B^p, \quad \beta(B) = 1 - \beta_1 B - \beta_2 B^2 - \ldots - \beta_q B^q$$

and where the innovations $\{v_t\}$ form a stationary, ergodic sequence such that, for the $\sigma-$algebra $\mathcal{F}_{t-1}$ generated by $\{v_\tau, \tau \leq t-1\}$, $\mathbb{E}(v_t|\mathcal{F}_{t-1}) = 0$ almost surely, $\mathbb{E}(v_t^2|\mathcal{F}_{t-1}) = 1$ almost surely, and $\mathbb{E}(v_t^4) < \infty$. Since the roots of the moving average polynomial lie outside the unit circle [the model is invertible], and there exists an infinite-order autoregressive representation of the latent volatility process such that

$$\sum_{j=0}^{\infty}(-\phi_j)w_{t-j} = \sigma_v v_t. \tag{3.12.1}$$

□

PROOF OF LEMMA 3.5.1 Under the assumptions 3.2.1 - 3.2.2, and if $U \sim \mathrm{N}[0, 1]$, then

$\mathbb{E}(U^{2p+1}) = 0, \quad \forall p \in \mathbb{N}$ and $\mathbb{E}(U^{2p}) = \frac{2p!}{2^p p!}, \quad \forall p \in \mathbb{N}$. Set $\mu_k(\theta) := \mathbb{E}(y_t^k)$ and $\mu_{k,l}(m|\theta) := \mathbb{E}(y_t^k y_{t+m}^l)$. Then, if k is even,

$$\begin{aligned}\mu_k(\theta) &= \mathbb{E}(y_t^k) = \sigma_y^k \mathbb{E}(z_t^k)\mathbb{E}\left[\exp(\frac{k w_t}{2})\right] = \sigma_y^k \frac{k!}{2^{k/2}(k/2)!}\exp\left[\frac{k^2}{8}\mathrm{Var}(w_t)\right] \\ &= \sigma_y^k \frac{k!}{2^{k/2}(k/2)!}\exp\left[\frac{k^2}{8}\frac{\sigma_v^2}{1-\sum_{j=1}^p \phi_j \rho_j}\right] \end{aligned} \tag{3.12.2}$$

and $\mu_k(\theta) = 0$, if k is odd, where $\rho_j := \mathrm{corr}(w_t, w_{t+j})$. For the cross-moments, we have: if k and l are even,

$$\begin{aligned}\mu_{k,l}(m|\theta) &= \mathbb{E}(y_t^k y_{t+m}^l) = \sigma_y^{k+l}\mathbb{E}(z_t^k)\mathbb{E}(z_{t+m}^l)\mathbb{E}\left[\exp(\frac{k w_t}{2} + \frac{l w_{t+m}}{2})\right] \\ &= \sigma_y^{k+l}\frac{k!}{2^{k/2}(k/2)!}\frac{l!}{2^{l/2}(l/2)!}\left[\exp(\frac{k^2}{8}\mathrm{Var}(w_t) + \frac{l^2}{8}\mathrm{Var}(w_{t+m}) + \frac{2kl}{8}\mathrm{cov}(w_t, w_{t+m}))\right] \\ &= \sigma_y^{k+l}\frac{k!}{2^{k/2}(k/2)!}\frac{l!}{2^{l/2}(l/2)!}\left[\exp(\frac{k^2}{8}\gamma_0 + \frac{l^2}{8}\gamma_0 + \frac{2kl}{8}\gamma_0\rho_m)\right] \\ &= \sigma_y^{k+l}\frac{k!}{2^{k/2}(k/2)!}\frac{l!}{2^{l/2}(l/2)!}\exp\left[\frac{1}{8}\gamma_0(k^2 + l^2 + 2kl\rho_m)\right] \\ &= \sigma_y^{k+l}\frac{k!}{2^{k/2}(k/2)!}\frac{l!}{2^{l/2}(l/2)!}\exp\left[\frac{1}{8}\frac{\sigma_v^2}{1-\sum_{j=1}^p \phi_j\rho_j}(k^2 + l^2 + 2kl\rho_m)\right], \end{aligned} \tag{3.12.3}$$

and $\mu_{k,l}(m|\theta) = 0$, otherwise. □

PROOF OF LEMMA 3.5.2 Using Lemma 3.5.1 and considering $k = 2$, $k = 4$, $k = l = 2$ & $m = 1$ and $k = l = 2$ & $m = 2$, we get:

$$\mu_2(\theta) := \mathbb{E}(y_t^2) = \sigma_y^2 \exp\left[\frac{1}{2}\gamma_0\right] \tag{3.12.4}$$

$$\mu_4(\theta) := \mathbb{E}(y_t^4) = 3\sigma_y^4 \exp\left[2\gamma_0\right] \tag{3.12.5}$$

$$\mu_{2,2}(1|\theta) := \mathbb{E}(y_t^2 y_{t-1}^2) = \sigma_y^4 \exp\left[\gamma_0(1+\rho_1)\right] \tag{3.12.6}$$

$$\mu_{2,2}(2|\theta) := \mathbb{E}(y_t^2 y_{t-2}^2) = \sigma_y^4 \exp\left[\gamma_0(1+\rho_2)\right] \tag{3.12.7}$$

where $\gamma_0 = \sigma_v^2/(1-\sum_{j=1}^{2}\phi_j\rho_j)$. From (3.12.4) and (3.12.5), we get:

$$\frac{\mathbb{E}(y_t^4)}{\left[\mathbb{E}(y_t^2)\right]^2} = 3\exp(\gamma_0),$$

hence

$$\gamma_0 = \log\left(\frac{\mathbb{E}(y_t^4)}{3\left[\mathbb{E}(y_t^2)\right]^2}\right).$$

From (3.12.4), we can write

$$\sigma_y^2 = \frac{\mathbb{E}(y_t^2)}{\exp(\gamma_0/2)}$$

or,

$$\sigma_y = \frac{\left[\mathbb{E}(y_t^2)\right]^{1/2}}{\left(\frac{\mathbb{E}(y_t^4)}{3\left[\mathbb{E}(y_t^2)\right]^2}\right)^{1/4}} = \frac{3^{1/4}\mathbb{E}(y_t^2)}{\left[\mathbb{E}(y_t^4)\right]^{1/4}}. \tag{3.12.8}$$

From (3.12.4) and (3.12.6), we get:

$$\gamma_0\rho_1 = \log\left(\frac{\mathbb{E}(y_t^2 y_{t-1}^2)}{\left[\mathbb{E}(y_t^2)\right]^2}\right)$$

or,

$$\gamma_1 = \log\left(\frac{\mathbb{E}(y_t^2 y_{t-1}^2)}{\left[\mathbb{E}(y_t^2)\right]^2}\right) = \log\left(\mathbb{E}(y_t^2 y_{t-1}^2)\right) - 2\log\left(\mathbb{E}(y_t^2)\right).$$

Similarly from (3.12.4) and (3.12.7), we get:

$$\gamma_2 = \log\left(\frac{\mathbb{E}(y_t^2 y_{t-2}^2)}{\left[\mathbb{E}(y_t^2)\right]^2}\right) = \log\left(\mathbb{E}(y_t^2 y_{t-2}^2)\right) - 2\log\left(\mathbb{E}(y_t^2)\right).$$

Now under the assumptions 3.2.2, the latent volatility process satisfy the Yule-Walker equations, see Fuller (1996). Thus, autocovariances of the volatility process satisfy the following equations:

$$\gamma_0 = \phi_1\gamma_1 + \phi_2\gamma_2 + (\sigma_v)^2, \quad \gamma_1 = \phi_1\gamma_0 + \phi_2\gamma_1, \quad \gamma_2 = \phi_1\gamma_1 + \phi_2\gamma_0.$$

Solving for ϕ_1 and ϕ_2 as functions of autocovariances, we get:

$$\phi_1 = \frac{-\gamma_1(\gamma_2-\gamma_0)}{(\gamma_0)^2-(\gamma_1)^2}, \quad \phi_2 = \frac{-(\gamma_1)^2+\gamma_2\gamma_0}{(\gamma_0)^2-(\gamma_1)^2}.$$

Substitute the values of γ_0, γ_1, and γ_2 into the Yule-Walker equations together with (3.12.8), we have following moment equations solution for the SV(2) parameters:

$$\phi_1 = \frac{-\left[\log\left(\mu_{2,2}(1)/\mu_2^2\right)\right]\left[\log\left(3\mu_{2,2}(2)/\mu_4\right)\right]}{\left[\log\left(\mu_4/(3\mu_2^2)\right)\right]^2 - \left[\log\left(\mu_{2,2}(1)/\mu_2^2\right)\right]^2}, \tag{3.12.9}$$

$$\phi_2 = \frac{-\left[\log\left(\mu_{2,2}(1)/\mu_2^2\right)\right]^2 + \left[\log\left(\mu_{2,2}(2)/\mu_2^2\right)\right]\left[\log\left(\mu_4/(3\mu_2^2)\right)\right]}{\left[\log\left(\mu_4/(3\mu_2^2)\right)\right]^2 - \left[\log\left(\mu_{2,2}(1)/\mu_2^2\right)\right]^2}, \tag{3.12.10}$$

$$\sigma_y = 3^{1/4}\mu_2/\mu_4^{1/4}, \tag{3.12.11}$$

$$\sigma_v = \left[\log\left(\mu_4/(3\mu_2^2)\right) - \phi_1\log\left(\mu_{2,2}(1)/\mu_2^2\right) - \phi_2\log\left(\mu_{2,2}(2)/\mu_2^2\right)\right]^{1/2}. \tag{3.12.12}$$

where $\mu_k := \mu_k(\theta) = \mathbb{E}(y_t^k)$ and $\mu_{k,l}(m) := \mu_{k,l}(m|\theta) = \mathbb{E}(y_t^k y_{t-m}^l)$.

□

Proof of Lemma 3.5.3 The autocovariance function of the first component of X_t is given by

$$\begin{aligned}\zeta_1(\tau) &:= \mathrm{cov}(X_{1,t}, X_{1,t+\tau}) = \mathbb{E}([y_t^2 - \mu_2(\theta)][y_{t+\tau}^2 - \mu_2(\theta)]) \\ &= \mathbb{E}(y_t^2 y_{t+\tau}^2) - \mu_2^2(\theta) = \sigma_y^4 \exp[\gamma_0(1+\rho_\tau)] - \mu_2^2(\theta) = \mu_2^2(\theta)[\exp(\gamma_\tau) - 1],\end{aligned} \tag{3.12.13}$$

where $\gamma_j := \mathrm{cov}(w_t, w_{t+j})$. Similarly,

$$\begin{aligned}\zeta_2(\tau) &:= \mathrm{cov}(X_{2,t}, X_{2,t+\tau}) = \mathbb{E}[y_t^4 - \mu_4(\theta)][y_{t+\tau}^4 - \mu_4(\theta)] = \mathbb{E}(y_t^4 y_{t+\tau}^4) - \mu_4^2(\theta) \\ &= 9\sigma_y^8 \exp[4\gamma_0(1+\rho_\tau)] - \mu_4^2(\theta) = \mu_4^2(\theta)[\exp(4\gamma_\tau) - 1], \quad \forall \tau \geq 1.\end{aligned} \tag{3.12.14}$$

$$\begin{aligned}\zeta_3(\tau) &:= \mathrm{cov}(X_{3,t}, X_{3,t+\tau}) = \mathbb{E}([y_t^2 y_{t-1}^2 - \mu_{2,2}(1|\theta)][y_{t+\tau}^2 y_{t+\tau-1}^2 - \mu_{2,2}(1|\theta)]) \\ &= \mathbb{E}[y_t^2 y_{t-1}^2 y_{t+\tau}^2 y_{t+\tau-1}^2] - \mu_{2,2}^2(1|\theta) \\ &= \sigma_y^8 \mathbb{E}[\exp(w_{t-1} + w_{t+\tau} + w_t + w_{t+\tau-1})] - \mu_{2,2}^2(1|\theta) \\ &= \sigma_y^8 \exp[2(\gamma_0 + \gamma_1) + \gamma_{\tau-1} + 2\gamma_\tau + \gamma_{\tau+1}] - \mu_{2,2}^2(1|\theta) \\ &= \sigma_y^8 \exp[2(\gamma_0 + \gamma_1)]\exp[\gamma_{\tau-1} + 2\gamma_\tau + \gamma_{\tau+1}] - \mu_{2,2}^2(1|\theta) \\ &= \mu_{2,2}^2(1|\theta)[\exp(\gamma_{\tau-1} + 2\gamma_\tau + \gamma_{\tau+1}) - 1], \quad \forall \tau \geq 2.\end{aligned} \tag{3.12.15}$$

$$
\begin{aligned}
\zeta_4(\tau) &= \mathrm{cov}(X_{4,t}, X_{4,t+\tau}) = \mathbb{E}([y_t^2 y_{t-2}^2 - \mu_{2,2}(2|\theta)][y_{t+\tau}^2 y_{t+\tau-2}^2 - \mu_{2,2}(2|\theta)]) \\
&= \mathbb{E}[y_t^2 y_{t-2}^2 y_{t+\tau}^2 y_{t+\tau-2}^2] - \mu_{2,2}^2(2|\theta) \\
&= \sigma_y^8 \mathbb{E}[\exp(w_{t-2} + w_{t+\tau} + w_t + w_{t+\tau-2})] - \mu_{2,2}^2(2|\theta) \\
&= \sigma_y^8 \exp[2(\gamma_0 + \gamma_2) + \gamma_{\tau-2} + 2\gamma_\tau + \gamma_{\tau+2}] - \mu_{2,2}^2(2|\theta) \\
&= \sigma_y^8 \exp[2(\gamma_0 + \gamma_2)] \exp[\gamma_{\tau-2} + 2\gamma_\tau + \gamma_{\tau+2}] - \mu_{2,2}^2(2|\theta) \\
&= \mu_{2,2}^2(2|\theta)[\exp(\gamma_{\tau-2} + 2\gamma_\tau + \gamma_{\tau+2}) - 1], \quad \forall \tau \geq 3.
\end{aligned}
\tag{3.12.16}
$$

□

PROOF OF PROPOSITION 3.5.4 From (3.2.14) - (3.2.15), we have

$$\phi(B) w_t = v_t,$$

and

$$y_t^* = w_t + \epsilon_t,$$

where $\phi(B) = (1 - \phi_1 B - \cdots - \phi_p B^p)$. The error processes v_t's and ϵ_t's are i.i.d. N$[0, \sigma_v^2]$ and i.i.d. $\log(\chi_1^2)$ random variables, respectively. Furthermore, assumption 3.2.1 implies that v_t's and ϵ_t's are independent. on applying $\phi(B)$ to both sides of (3.2.15) yields

$$\phi(B) y_t^* = \phi(B) w_t + \phi(B)\epsilon_t = v_t + \phi(B)\epsilon_t. \tag{3.12.17}$$

The right hand side of (3.12.17) is clearly a covariance stationary process. By the Wold decomposition theorem it must have a moving average representation. Since the autocovariance function cuts off for lags $k > p$ it must be an $MA(p)$ process, say $\theta(B)\eta_t = (1 - \theta_1 B - \cdots - \theta_p B^p)\eta_t$. Hence, y_t^* must be an ARMA(p,p) process [see equation (2.1) of Granger and Morris (1976)].

The moving average parameters $\theta_1, \theta_2, \ldots, \theta_p$ and the white noise variance σ_η^2 of this ARMA(p,p) process can be found by equating the autocovariance function of the right hand side of (3.12.17) with that of $\theta(B)\eta_t$ for lags $k = 0, 1, \ldots, p$ and solving the $p+1$ resulting nonlinear equations

$$
\begin{aligned}
(1+\theta_1^2+\cdots+\theta_p^2)\sigma_\eta^2 &= \sigma_v^2+(1+\phi_1^2+\cdots+\phi_p^2)\sigma_\epsilon^2,\\
(-\theta_1+\theta_1\theta_2+\cdots+\theta_{p-1}\theta_p)\sigma_\eta^2 &= (-\phi_1+\phi_1\phi_2+\cdots+\phi_{p-1}\phi_p)\sigma_\epsilon^2,\\
&\vdots\\
(-\theta_{p-1}+\theta_1\theta_p)\sigma_\eta^2 &= (-\phi_{p-1}+\phi_1\phi_p)\sigma_\epsilon^2,\\
-\theta_p\sigma_\eta^2 &= -\phi_p\sigma_\epsilon^2.
\end{aligned}
$$

Note that there may be multiple solutions, only some of which result in an invertible process.

□

PROOF OF COROLLARY 3.5.5 From Proposition 3.5.4, the observed process y_t^* satisfies the following equation:

$$y_t^* = \sum_{j=1}^{p}\phi_j y_{t-j}^* + \eta_t - \sum_{j=1}^{p}\theta_j\eta_{t-j} \tag{3.12.18}$$

or

$$y_t^* = \sum_{j=1}^{p}\phi_j y_{t-j}^* + v_t + \epsilon_t - \sum_{j=1}^{p}\phi_j\epsilon_{t-j}. \tag{3.12.19}$$

Multiply both sides of (3.12.19) by y_{t-k}^*, and taking expectation, we get:

$$\gamma_{y^*}(k) = \sum_{j=1}^{p}\phi_j\gamma_{y^*}(k-1) + \mathbb{E}[v_t y_{t-k}^*] + \mathbb{E}[\epsilon_t y_{t-k}^*] - \sum_{j=1}^{p}\phi_j\mathbb{E}[\epsilon_{t-j}y_{t-k}^*].$$

For $k=0$, we get

$$\begin{aligned}
\gamma_{y^*}(k) &= \sum_{j=1}^{p}\phi_j\gamma_{y^*}(k-j)+\mathbb{E}[v_t y_t^*]+\mathbb{E}[\epsilon_t y_t^*]-\sum_{j=1}^{p}\phi_j\mathbb{E}[\epsilon_{t-j}y_t^*] \\
&= \sum_{j=1}^{p}\phi_j\gamma_{y^*}(k-j)+\sigma_v^2+\sigma_\epsilon^2-\sum_{j=1}^{p}\phi_j\mathbb{E}[\epsilon_{t-j}(\phi_j y_{t-j}^*-\phi_j\epsilon_{t-j})] \\
&= \sum_{j=1}^{p}\phi_j\gamma_{y^*}(k-j)+\sigma_v^2+\sigma_\epsilon^2-\sum_{j=1}^{p}\phi_j^2\mathbb{E}[\epsilon_{t-j}y_{t-j}^*-\epsilon_{t-j}^2] \\
&= \sum_{j=1}^{p}\phi_j\gamma_{y^*}(k-j)+\sigma_v^2+\sigma_\epsilon^2-\sum_{j=1}^{p}\phi_j^2[\sigma_\epsilon^2-\sigma_\epsilon^2] \\
&= \sum_{j=1}^{p}\phi_j\gamma_{y^*}(k-j)+\sigma_v^2+\sigma_\epsilon^2.
\end{aligned} \tag{3.12.20}$$

Setting $1\le k\le p$, we get

$$\begin{aligned}
\gamma_{y^*}(k) &= \sum_{j=1}^{p}\phi_j\gamma_{y^*}(k-j)+\mathbb{E}[v_t y_{t-k}^*]+\mathbb{E}[\epsilon_t y_{t-k}^*]-\sum_{j=1}^{p}\phi_j\mathbb{E}[\epsilon_{t-j}y_{t-k}^*] \\
&= \sum_{j=1}^{p}\phi_j\gamma_{y^*}(k-j)+0+0-\phi_k\mathbb{E}[\epsilon_{t-k}y_{t-k}^*] = \sum_{j=1}^{p}\phi_j\gamma_{y^*}(k-j)-\phi_k\sigma_\epsilon^2.
\end{aligned} \tag{3.12.21}$$

Setting $k>p$, we get

$$\begin{aligned}
\gamma_{y^*}(k) &= \sum_{j=1}^{p}\phi_j\gamma_{y^*}(k-j)+\mathbb{E}[v_t y_{t-k}^*]+\mathbb{E}[\epsilon_t y_{t-k}^*]-\sum_{j=1}^{p}\phi_j\mathbb{E}[\epsilon_{t-1}y_{t-k}^*] \\
&= \sum_{j=1}^{p}\phi_j\gamma_{y^*}(k-j)+0+0-0 = \sum_{j=1}^{p}\phi_j\gamma_{y^*}(k-j).
\end{aligned} \tag{3.12.22}$$

Combining (3.12.20), (3.12.21), and (3.12.22), we get the autocovariance structure of the observed process that stated in the Corollary. □

PROOF OF COROLLARY 3.5.6 The estimator of ϕ_p is based on the autocovariance structure of the process y_t^*. This is the solution of p-system of equations from (3.5.18) with $k=p+$

$1,\dots,2p$. So

$$\begin{bmatrix} \gamma_{y^*}(p) & \gamma_{y^*}(p-1) & \cdots & \gamma_{y^*}(1) \\ \gamma_{y^*}(p+1) & \gamma_{y^*}(p) & \cdots & \gamma_{y^*}(2) \\ \vdots & \vdots & & \vdots \\ \gamma_{y^*}(2p-1) & \gamma_{y^*}(2p-2) & \cdots & \gamma_{y^*}(p) \end{bmatrix} \cdot \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix} = \begin{bmatrix} \gamma_{y^*}(p+1) \\ \gamma_{y^*}(p+2) \\ \vdots \\ \gamma_{y^*}(2p) \end{bmatrix},$$

or

$$\boldsymbol{\Gamma}_{(\boldsymbol{p})}\boldsymbol{\phi}_{\boldsymbol{p}} = \gamma_{(\boldsymbol{p+1})}$$

hence

$$\boldsymbol{\phi}_{\boldsymbol{p}} = \boldsymbol{\Gamma}_{(\boldsymbol{p})}^{\boldsymbol{-1}}\gamma_{(\boldsymbol{p+1})}, \tag{3.12.23}$$

where $\boldsymbol{\phi}_{\boldsymbol{p}} := (\phi_1,\dots,\phi_p)^{'}$, $\gamma_{(\boldsymbol{p+1})} = [\gamma_{y^*}(p+1),\dots,\gamma_{y^*}(2p)]^{'}$ are vectors and $\boldsymbol{\Gamma}_{(\boldsymbol{p})}$ is a p-dimensional Toeplitz matrices such that

$$\boldsymbol{\Gamma}_{(\boldsymbol{p})} := \begin{bmatrix} \gamma_{y^*}(p) & \gamma_{y^*}(p-1) & \cdots & \gamma_{y^*}(1) \\ \gamma_{y^*}(p+1) & \gamma_{y^*}(p) & \cdots & \gamma_{y^*}(2) \\ \vdots & \vdots & & \vdots \\ \gamma_{y^*}(2p-1) & \gamma_{y^*}(2p-2) & \cdots & \gamma_{y^*}(p) \end{bmatrix}.$$

Note that (3.12.23) is also valid for any $j \geq 1$ such that

$$\boldsymbol{\phi}_{\boldsymbol{p}} = \boldsymbol{\Gamma}_{(\boldsymbol{p+j-1})}^{\boldsymbol{-1}}\gamma_{(\boldsymbol{p+j})}, \tag{3.12.24}$$

where $\gamma_{(\boldsymbol{p+j})} := [\gamma_{y^*}(p+j),\dots,\gamma_{y^*}(2p+j-1)]^{'}$ and $\boldsymbol{\Gamma}_{(\boldsymbol{p+j-1})}$ is a p-dimensional Toeplitz matrices such that

$$\boldsymbol{\Gamma}_{(\boldsymbol{p+j-1})} := \begin{bmatrix} \gamma_{y^*}(p+j-1) & \gamma_{y^*}(p+j-2) & \cdots & \gamma_{y^*}(j) \\ \gamma_{y^*}(p+j) & \gamma_{y^*}(p+j-1) & \cdots & \gamma_{y^*}(j+1) \\ \vdots & \vdots & & \vdots \\ \gamma_{y^*}(2p+j-2) & \gamma_{y^*}(2p+j-3) & \cdots & \gamma_{y^*}(p+j-1) \end{bmatrix}.$$

Now from (3.5.18) with $k = 0$ we have:

$$\gamma_{y^*}(0) = \phi_1 \gamma_{y^*}(1) + \cdots + \phi_p \gamma_{y^*}(p) + \sigma_v^2 + \sigma_\epsilon^2,$$

hence

$$\sigma_v = [\gamma_{y^*}(0) - \sum_{j=1}^{p} \phi_j \gamma_{y^*}(j) - \pi^2/2]^{1/2} = [\gamma_{y^*}(0) - \boldsymbol{\phi}_{\boldsymbol{p}}' \boldsymbol{\gamma}_{(\mathbf{1})} - \pi^2/2]^{1/2}, \tag{3.12.25}$$

where $\boldsymbol{\phi}_{\boldsymbol{p}} := (\phi_1, \ldots, \phi_p)'$, $\boldsymbol{\gamma}_{(\mathbf{1})} := [\gamma_{y^*}(1), \ldots, \gamma_{y^*}(p)]'$ and $\sigma_\epsilon^2 = \psi^{(1)}(1/2) = \pi^2/2$. Now by construction,

$$\mu = \mathbb{E}[\log(y_t^2)] = \log(\sigma_y^2) + \mathbb{E}[\log(z_t^2)] = \log(\sigma_y^2) - 1.27, \tag{3.12.26}$$

or, equivalently

$$\sigma_y^2 = \exp(\mu + 1.27). \tag{3.12.27}$$

$\square$

Proof of Lemma 3.5.7 We are using an alternative method provided by Durbin (1960) that avoids the matrix inversion in the Yule-Walker equations. We derive the solution set of SV(2) parameters recursively from the solution set of SV(1) parameters, thus the results of Lemma 3.5.7 can easily identify from there.

We wish to find closed-form moment equations solution for $\Theta_1^{SV} := (\phi_{11}, \sigma_{1v}, \sigma_y)'$. Using Lemma 3.5.1, and considering $k = 2$, $k = 4$, $k = l = 2$ & $m = 1$, we get:

$$\mu_2(\theta) := \mathbb{E}(y_t^2) = \sigma_y^2 \exp\left[\frac{1}{2}\gamma_0\right] \tag{3.12.28}$$

$$\mu_4(\theta) := \mathbb{E}(y_t^4) = 3\sigma_y^4 \exp\left[2\gamma_0\right] \tag{3.12.29}$$

$$\mu_{2,2}(1|\theta) := \mathbb{E}(y_t^2 y_{t-1}^2) = \sigma_y^4 \exp\left[\gamma_0(1 + \rho_1)\right] \tag{3.12.30}$$

where $\gamma_0 = \sigma_{1v}^2/(1 - \phi_{11}\rho_1)$. Solving the above equations yields:

$$\gamma_0 = \log\left(\frac{\mathbb{E}(y_t^4)}{3\left[\mathbb{E}(y_t^2)\right]^2}\right), \tag{3.12.31}$$

$$\sigma_y = \frac{\left[\mathbb{E}(y_t^2)\right]^{1/2}}{\left(\frac{\mathbb{E}(y_t^4)}{3\left[\mathbb{E}(y_t^2)\right]^2}\right)^{1/4}} = \frac{3^{1/4}\mathbb{E}(y_t^2)}{\left[\mathbb{E}(y_t^4)\right]^{1/4}}, \tag{3.12.32}$$

$$\gamma_0\rho_1 = \log\left(\frac{\mathbb{E}(y_t^2 y_{t-1}^2)}{\left[\mathbb{E}(y_t^2)\right]^2}\right), \tag{3.12.33}$$

or,

$$\rho_1 = \log\left(\frac{\mathbb{E}(y_t^2 y_{t-1}^2)}{\left[\mathbb{E}(y_t^2)\right]^2}\right) / \log\left(\frac{\mathbb{E}(y_t^4)}{3\left[\mathbb{E}(y_t^2)\right]^2}\right). \tag{3.12.34}$$

As a result, we have:

$$\phi_{11} = \gamma_1/\gamma_0 = \rho_1 = \log\left(\frac{\mathbb{E}(y_t^2 y_{t-1}^2)}{\left[\mathbb{E}(y_t^2)\right]^2}\right) / \log\left(\frac{\mathbb{E}(y_t^4)}{3\left[\mathbb{E}(y_t^2)\right]^2}\right). \tag{3.12.35}$$

Now using the Yule-Walker equations, we get:

$$\gamma_0 = \phi_{11}\gamma_1 + \sigma_v^2 \Leftrightarrow 1 = \phi_{11}\rho_1 + \sigma_v^2/\gamma_0 = \phi_{11}^2 + \sigma_v^2/\gamma_0. \tag{3.12.36}$$

Hence

$$\sigma_v = \left[(1-\phi_{11}^2)\gamma_0\right]^{1/2} = \left[\left(1 - \left[\log\left(\frac{\mathbb{E}(y_t^2 y_{t-1}^2)}{\left[\mathbb{E}(y_t^2)\right]^2}\right) / \log\left(\frac{\mathbb{E}(y_t^4)}{3\left[\mathbb{E}(y_t^2)\right]^2}\right)\right]^2\right) \log\left(\frac{\mathbb{E}(y_t^4)}{3\left[\mathbb{E}(y_t^2)\right]^2}\right)\right]^{1/2}. \tag{3.12.37}$$

Thus, closed-form SV(1) parameter solutions are given in (3.12.32), (3.12.35) and (3.12.37).

Let us now consider the SV(2) model. We wish to find closed-form estimator for $\Theta_2^{SV} := (\phi_{21}, \phi_{22}, \sigma_{2v}, \sigma_y)'$. Using Lemma 3.5.1, and considering $k = 2$, $k = 4$, $k = l = 2$ & $m = 1$ and $k = l = 2$ & $m = 2$, we get:

$$\mu_2(\theta) := \mathbb{E}(y_t^2) = \sigma_y^2 \exp\left[\frac{1}{2}\gamma_0\right] \tag{3.12.38}$$

$$\mu_4(\theta) := \mathbb{E}(y_t^4) = 3\sigma_y^4 \exp\left[2\gamma_0\right] \tag{3.12.39}$$

$$\mu_{2,2}(1|\theta) := \mathbb{E}(y_t^2 y_{t-1}^2) = \sigma_y^4 \exp\left[\gamma_0(1+\rho_1)\right] \tag{3.12.40}$$

$$\mu_{2,2}(2|\theta) := \mathbb{E}(y_t^2 y_{t-2}^2) = \sigma_y^4 \exp\left[\gamma_0(1+\rho_2)\right] \tag{3.12.41}$$

where $\gamma_0 = \sigma^2_{2v}/(1-\sum_{j=1}^{2}\phi_{2j}\rho_j)$. Solving the above equations yields:

$$\gamma_0 = \log\left(\frac{\mathbb{E}(y_t^4)}{3\left[\mathbb{E}(y_t^2)\right]^2}\right), \tag{3.12.42}$$

$$\sigma_y = \frac{\left[\mathbb{E}(y_t^2)\right]^{1/2}}{\left(\frac{\mathbb{E}(y_t^4)}{3\left[\mathbb{E}(y_t^2)\right]^2}\right)^{1/4}} = \frac{3^{1/4}\mathbb{E}(y_t^2)}{\left[\mathbb{E}(y_t^4)\right]^{1/4}}, \tag{3.12.43}$$

$$\rho_1 = \log\left(\frac{\mathbb{E}(y_t^2 y_{t-1}^2)}{\left[\mathbb{E}(y_t^2)\right]^2}\right)/\log\left(\frac{\mathbb{E}(y_t^4)}{3\left[\mathbb{E}(y_t^2)\right]^2}\right), \tag{3.12.44}$$

$$\rho_2 = \log\left(\frac{\mathbb{E}(y_t^2 y_{t-2}^2)}{\left[\mathbb{E}(y_t^2)\right]^2}\right)/\log\left(\frac{\mathbb{E}(y_t^4)}{3\left[\mathbb{E}(y_t^2)\right]^2}\right). \tag{3.12.45}$$

From (3.12.44) and (3.12.45), we observe that:

$$\rho_j = \log\left(\frac{\mathbb{E}(y_t^2 y_{t-j}^2)}{\left[\mathbb{E}(y_t^2)\right]^2}\right)/\log\left(\frac{\mathbb{E}(y_t^4)}{3\left[\mathbb{E}(y_t^2)\right]^2}\right), \text{ where } \ j = 1,2. \tag{3.12.46}$$

Now from Durbin-Levinson recurrence formula, we have:

$$\phi_{22} = \frac{\rho_2 - \phi_{11}\rho_1}{1-\phi_{11}\rho_1} = \frac{\rho_2 - \rho_1^2}{1-\rho_1^2} = \frac{\rho_2 - \phi_{11}^2}{1-\phi_{11}^2} \tag{3.12.47}$$

and

$$\phi_{21} = \phi_{11} - \phi_{22}\phi_{11} = \frac{\rho_1 - \rho_1\rho_2}{1-\rho_1^2}. \tag{3.12.48}$$

On substituting $\phi_{11}, \rho_1, \rho_2$ [given in (3.12.35), (3.12.44), (3.12.45)], we get:

$$\begin{aligned}
\phi_{22} &= \frac{\left[\log\left(\frac{\mathbb{E}(y_t^2 y_{t-2}^2)}{\left[\mathbb{E}(y_t^2)\right]^2}\right)/\log\left(\frac{\mathbb{E}(y_t^4)}{3\left[\mathbb{E}(y_t^2)\right]^2}\right)\right] - \left[\log\left(\frac{\mathbb{E}(y_t^2 y_{t-1}^2)}{\left[\mathbb{E}(y_t^2)\right]^2}\right)/\log\left(\frac{\mathbb{E}(y_t^4)}{3\left[\mathbb{E}(y_t^2)\right]^2}\right)\right]^2}{1-\left[\log\left(\frac{\mathbb{E}(y_t^2 y_{t-1}^2)}{\left[\mathbb{E}(y_t^2)\right]^2}\right)/\log\left(\frac{\mathbb{E}(y_t^4)}{3\left[\mathbb{E}(y_t^2)\right]^2}\right)\right]^2} \\
&= \frac{-\left[\log\left(\mu_{2,2}(1)/\mu_2^2\right)\right]^2 + \left[\log\left(\mu_{2,2}(2)/\mu_2^2\right)\right]\left[\log\left(\mu_4/(3\mu_2^2)\right)\right]}{\left[\log\left(\mu_4/(3\mu_2^2)\right)\right]^2 - \left[\log\left(\mu_{2,2}(1)/\mu_2^2\right)\right]^2},
\end{aligned} \tag{3.12.49}$$

$$\begin{aligned}\phi_{21} &= \frac{\left[\log\left(\frac{\mathbb{E}(y_t^2 y_{t-1}^2)}{\left[\mathbb{E}(y_t^2)\right]^2}\right)/\log\left(\frac{\mathbb{E}(y_t^4)}{3\left[\mathbb{E}(y_t^2)\right]^2}\right)\right] - \left[\log\left(\frac{\mathbb{E}(y_t^2 y_{t-1}^2)}{\left[\mathbb{E}(y_t^2)\right]^2}\right)/\log\left(\frac{\mathbb{E}(y_t^4)}{3\left[\mathbb{E}(y_t^2)\right]^2}\right)\right]\left[\log\left(\frac{\mathbb{E}(y_t^2 y_{t-2}^2)}{\left[\mathbb{E}(y_t^2)\right]^2}\right)/\log\left(\frac{\mathbb{E}(y_t^4)}{3\left[\mathbb{E}(y_t^2)\right]^2}\right)\right]}{1-\left[\log\left(\frac{\mathbb{E}(y_t^2 y_{t-1}^2)}{\left[\mathbb{E}(y_t^2)\right]^2}\right)/\log\left(\frac{\mathbb{E}(y_t^4)}{3\left[\mathbb{E}(y_t^2)\right]^2}\right)\right]^2} \\ &= \frac{-\left[\log\left(\mu_{2,2}(1)/\mu_2^2\right)\right]\left[\log\left(3\mu_{2,2}(2)/\mu_4\right)\right]}{\left[\log\left(\mu_4/(3\mu_2^2)\right)\right]^2-\left[\log\left(\mu_{2,2}(1)/\mu_2^2\right)\right]^2}, \end{aligned} \tag{3.12.50}$$

and

$$\begin{aligned}\sigma_{2v} &= [(1-\phi_{21}\rho_1-\phi_{22}\rho_2)\gamma_0]^{1/2} \\ &= \left[\log\left(\mu_4/(3\mu_2^2)\right)-\phi_1\log\left(\mu_{2,2}(1)/\mu_2^2\right)-\phi_2\log\left(\mu_{2,2}(2)/\mu_2^2\right)\right]^{1/2}, \end{aligned} \tag{3.12.51}$$

where $\mu_k := \mu_k(\theta) = \mathbb{E}(y_t^k)$ and $\mu_{k,l}(m) := \mu_{k,l}(m|\theta) = \mathbb{E}(y_t^k y_{t-m}^l)$. Thus, using the Durbin-Levinson algorithm, we obtain the same results which require a matrix inversion. □

Proof of Lemma 3.7.1 The method-of-moments estimator $\hat{\theta}_T$ is solution of the following optimization problem:

$$\min_{\theta} M_T = [\bar{g}_T(Y_T)-\mu(\theta)]'\hat{\Omega}_T[\bar{g}_T(Y_T)-\mu(\theta)].$$

Under the assumption 3.7.2, the score condition associated with this problem is:

$$J(\theta)\hat{\Omega}_T[\mu(\theta)-\bar{g}_T(Y_T)] = 0.$$

A Taylor series expansion of the score condition around the true value of θ yields

$$0 = J(\theta)\hat{\Omega}_T[\mu(\theta)-\bar{g}_T(Y_T)] = J(\theta)\hat{\Omega}_T[\mu(\theta)-\bar{g}_T(Y_T)] + J(\theta)\hat{\Omega}_T J(\theta)'(\hat{\theta}_T-\theta) = O_p(T^{-1}).$$

After rearranging the equation and using assumption 3.7.2, we have

$$\sqrt{T}[\hat{\theta}_T-\theta] = [J(\theta)\Omega J(\theta)']^{-1}J(\theta)\Omega\sqrt{T}[\bar{g}_T(Y_T)-\mu(\theta)] + O_p(T^{-1/2}).$$

Now using assumptions 3.7.1, we get the asymptotic normality of $\hat{\theta}_T(\Omega)$ with asymptotic covariance matrix $V(\Omega)$ as specified in the Lemma. □

PROOF OF LEMMA 3.7.2 To establish the asymptotic normality of $[\bar{g}_T(Y_T) - \mu(\theta)]$; we need to use a central limit theorem (CLT) for dependent processes (see Davidson (1994), Theorem 24.5, p. 385). For that purpose, we first check the conditions under which this CLT holds. We workout this proof where $p = 2$. Setting

$$X_t := \begin{pmatrix} y_t^2 - \mu_2(\theta) \\ y_t^4 - \mu_4(\theta) \\ y_t^2 y_{t-1}^2 - \mu_{2,2}(1\,|\theta) \\ y_t^2 y_{t-2}^2 - \mu_{2,2}(2\,|\theta) \end{pmatrix} = g_t(\theta) - \mu(\theta)$$

$$S_T = \sum_{t=1}^{T} X_t = \sum_{t=1}^{T} [g_t(\theta) - \mu(\theta)]$$

and the subfields $\mathcal{F}_t = \sigma(s_t, s_{t-1}, \ldots)$ where $s_t = (y_t, w_t)^{'}$, we need to check the following conditions in order to get that $T^{-1/2} S_T = \sqrt{T}[\bar{g}_T(Y_T) - \mu(\theta_0)] \xrightarrow{d} \mathrm{N}[0, \Omega_*]$.

(i) $\{X_t, \mathcal{F}_t\}$ is stationary and ergodic,

(ii) $\{X_t, \mathcal{F}_t\}$ is a L_1-mixingale of size -1, and

(iii) $\limsup_{T\to\infty} T^{-1/2}\mathbb{E}|S_T| < \infty$.

(i) This follows from results 3.4.1 and 3.4.2.

(ii)-(1) A mixing zero-mean process is an adapted L_1- mixingale with respect to the sub-fields $\mathcal{F}_t$ provided it is bounded in the L_1-norm. To see that $\{X_t\}$ is bounded in the L_1-norm, using Theorem 14.2 of Davidson (1994), we have

$$\mathbb{E}\left|y_t^{2k} - \mu_{2k}(\theta)\right| \le \mathbb{E}(\left|y_t^{2k}\right| + \left|\mu_{2k}(\theta)\right|) = 2\mu_{2k}(\theta) < \infty, \quad for \quad k = 1, 2, \ldots,$$

$$\mathbb{E}\left|y_t^2 y_{t-k}^2 - \mu_{2,2}(k\,|\theta)\right| \le \mathbb{E}(\left|y_t^2 y_{t-k}^2\right| + \left|\mu_{2,2}(k\,|\theta)\right|) = 2\mu_{2,2}(k\,|\theta) < \infty, \quad for \quad k = 1, 2, \ldots \ .$$

(ii)-(2) We now need to show that the $\{X_t, \mathcal{F}_t\}$ is a L_1-mixingale of size -1. From the discussion in section 3.4, we know that X_t is β-mixing, so it has mixing coefficients of the type $\beta_T = \psi\rho^T$, $\psi > 0$, $0 < \rho < 1$. To show that $\{X_t\}$ is of size -1, its mixing coefficients β_T must be $O(T^{-\varphi})$, with $\varphi > 1$ (see Davidson (1994), Definition 16.1, p. 247). To see that

$$\lim_{T\to\infty} \frac{\rho^T}{T^{-\varphi}} = \lim_{T\to\infty} T^{\varphi}\exp(T\log\rho) = \lim_{T\to\infty} \exp(\varphi\log T)\exp(T\log\rho) = \lim_{T\to\infty} \exp(\varphi\log T + T\log\rho) = 0.$$

This holds in particular for $\varphi > 1$; see Rudin (1976) [Theorem 3.20(d), page 57].

(iii) We need to show that $\limsup_{T\to\infty} T^{-1/2}\mathbb{E}|S_T| < \infty$ and using Cauchy-Schwarz inequality, we have $\mathbb{E}\left|T^{-1/2}S_T\right| \leq T^{-1/2}\|S_T\|_2$. Now we can prove it by showing that

$$\limsup_{T\to\infty} T^{-1}\mathbb{E}(S_T S_T^{'}) < \infty \Leftrightarrow \limsup_{T\to\infty} \mathrm{Var}(T^{-1/2}S_T) < \infty$$

(iii)-(1) First and second components of S_T. Define $S_{T1} = \sum_{t=1}^{T} X_{1,t}$ where $X_{1,t} := y_t^2 - \mu_2(\theta)$ and compute:

$$\begin{aligned}\mathrm{Var}(T^{-1/2}S_{T1}) =& \frac{1}{T}\left[\sum_{t=1}^{T}\mathrm{Var}(X_{1,t}) + \sum_{t\neq s}\mathrm{cov}(X_{1,s}, X_{1,t})\right] = \frac{1}{T}\left[T\zeta_1(0) + 2\sum_{\tau=1}^{T}(T-\tau)\zeta_1(\tau)\right] \\ =&\zeta_1(0) + 2\sum_{\tau=1}^{T}(1-\frac{\tau}{T})\zeta_1(\tau).\end{aligned} \tag{3.12.52}$$

Now we must prove that $\sum_{\tau=1}^{T}(1-\frac{\tau}{T})\zeta_1(\tau)$ converges as $T\to\infty$. By Lemma 3.1.5 in Fuller (1976, p. 112), it is sufficient to show that $\sum_{\tau=1}^{\infty}\zeta_1(\tau)$ converge. Using Lemma 3.5.3, we have

$$\begin{aligned}\zeta_1(\tau) =& \mu_2^2(\theta)[\exp(\gamma_\tau) - 1] = \mu_2^2(\theta)\left[1 + \sum_{k=1}^{\infty}\frac{\gamma_\tau^k}{k!} - 1\right] = \mu_2^2(\theta)\left[\gamma_\tau\sum_{k=1}^{\infty}\frac{\gamma_\tau^{k-1}}{k!}\right] \\ =&\mu_2^2(\theta)\left[\gamma_\tau\sum_{k=0}^{\infty}\frac{\gamma_\tau^k}{(k+1)!}\right] \leq \mu_2^2(\theta)\left[\gamma_\tau\sum_{k=0}^{\infty}\frac{\gamma_\tau^k}{(k)!}\right] = \mu_2^2(\theta)\gamma_\tau\exp(\gamma_\tau).\end{aligned} \tag{3.12.53}$$

Therefore, the series

$$\sum_{\tau=1}^{\infty}\zeta_1(\tau) \leq \mu_2^2(\theta)\sum_{\tau=1}^{\infty}\gamma_\tau\exp(\gamma_\tau) \leq \mu_2^2(\theta)\exp(\gamma_1)\sum_{\tau=1}^{\infty}\gamma_\tau \leq \mu_2^2(\theta)\exp(\gamma_1)\underbrace{\sum_{\tau=1}^{\infty}\left|\gamma_\tau\right|}_{<\infty} < \infty \tag{3.12.54}$$

converges and by the Cauchy-Schwarz inequality we deduce that $\limsup_{T\to\infty} T^{-1/2}\mathbb{E}|S_{T1}| < \infty$. The proof is very similar for S_{T2}.

(iii)-(2) Third and fourth components of S_T. we just have to show that $\sum_{\tau=1}^{\infty}\zeta_3(\tau) < \infty$. By Lemma 3.5.3, we have for all $\tau \geq 2$:

$$
\begin{aligned}
\zeta_3(\tau) &= \mu_{2,2}^2(1|\theta)[\exp(\gamma_{\tau-1}+2\gamma_\tau+\gamma_{\tau+1})-1] = \mu_{2,2}^2(1|\theta)\left[1+\sum_{k=1}^{\infty}\frac{(\gamma_{\tau-1}+2\gamma_\tau+\gamma_{\tau+1})^k}{k!}-1\right] \\
&= \mu_{2,2}^2(1|\theta)\left[(\gamma_{\tau-1}+2\gamma_\tau+\gamma_{\tau+1})\sum_{k=1}^{\infty}\frac{(\gamma_{\tau-1}+2\gamma_\tau+\gamma_{\tau+1})^{k-1}}{k!}\right] \\
&= \mu_{2,2}^2(1|\theta)\left[(\gamma_{\tau-1}+2\gamma_\tau+\gamma_{\tau+1})\sum_{k=0}^{\infty}\frac{(\gamma_{\tau-1}+2\gamma_\tau+\gamma_{\tau+1})^{k}}{(k+1)!}\right] \\
&\leq \mu_{2,2}^2(1|\theta)\left[(\gamma_{\tau-1}+2\gamma_\tau+\gamma_{\tau+1})\sum_{k=0}^{\infty}\frac{(\gamma_{\tau-1}+2\gamma_\tau+\gamma_{\tau+1})^{k}}{(k)!}\right] \\
&= \mu_{2,2}^2(1|\theta)(\gamma_{\tau-1}+2\gamma_\tau+\gamma_{\tau+1})\exp(\gamma_{\tau-1}+2\gamma_\tau+\gamma_{\tau+1}).
\end{aligned}
\tag{3.12.55}
$$

Therefore, the series

$$
\begin{aligned}
\sum_{\tau=1}^{\infty}\zeta_3(\tau) &\leq \zeta_3(1)+\mu_{2,2}^2(1|\theta)\sum_{\tau=2}^{\infty}(\gamma_{\tau-1}+2\gamma_\tau+\gamma_{\tau+1})\exp(\gamma_{\tau-1}+2\gamma_\tau+\gamma_{\tau+1}) \\
&\leq \zeta_3(1)+\mu_{2,2}^2(1|\theta)\exp(\gamma_1+2\gamma_2+\gamma_3)\sum_{\tau=2}^{\infty}(\gamma_{\tau-1}+2\gamma_\tau+\gamma_{\tau+1}) \\
&\leq \zeta_3(1)+\mu_{2,2}^2(1|\theta)\exp(\gamma_1+2\gamma_2+\gamma_3)\left[\underbrace{\sum_{\tau=2}^{\infty}|\gamma_{\tau-1}|}_{<\infty}+2\underbrace{\sum_{\tau=2}^{\infty}|\gamma_{\tau}|}_{<\infty}+\underbrace{\sum_{\tau=2}^{\infty}|\gamma_{\tau+1}|}_{<\infty}\right] < \infty
\end{aligned}
\tag{3.12.56}
$$

converges and by the Cauchy-Schwarz inequality we deduce that $\limsup_{T\to\infty} T^{-1/2}\mathbb{E}|S_{T3}| < \infty$. The proof for S_{T4} is similar. Thus we can apply Theorem 24.5 of Davidson (1994) to each component S_{Ti}, $i = 1,2,3,4$ of S_T to state that: $T^{-1/2}S_{Ti} \xrightarrow{d} \mathrm{N}[0,\lambda_i]$ and then by Cramér-Wold theorem we can establish the limiting result that stated in Lemma 3.7.2 . □

PROOF OF LEMMA 3.7.3 Under the assumptions 3.7.4 and 3.7.5 with $s = 2$, the observed process $\{y_t^*\}$ is strictly stationarity and geometrically ergodic with $\mathbb{E}[y_t^*] < \infty$ and $\mathbb{E}[y_t^* y_{t+k}^*] < \infty$. So the consistency is a simple application of the Law of Large Numbers for stationary and ergodic processes, *i.e.*, the Ergodic theorem; see Theorem 13.12 and Corollary 13.14 of Davidson (1994). □

PROOF OF LEMMA 3.7.4 To establish the asymptotic normality of empirical moments; we shall use a CLT for dependent processes (see Davidson (1994), Theorem 24.5, p. 385). For that purpose, we first check the conditions under which this CLT holds. We set

$$X_t := \begin{bmatrix} \Psi_t \\ \Lambda_t \end{bmatrix}, \quad \Psi_t := \log(y_t^2) - \mu, \quad \Lambda_t := [\Lambda_{t,0}, \Lambda_{t,1}, \dots, \Lambda_{t,m}]', \tag{3.12.57}$$

$$\Lambda_{t,k} := y_t^* y_{t+k}^* - \gamma_{y^*}(k) = [\log(y_t^2) - \mu][\log(y_{t+k}^2) - \mu] - \gamma_{y^*}(k), \quad k = 0, 1, \dots, m, \tag{3.12.58}$$

$$S_T := \sum_{t=1}^{T} X_t = \begin{bmatrix} \sum_{t=1}^{T} \Psi_t \\ \sum_{t=1}^{T} \Lambda_t \end{bmatrix}, \tag{3.12.59}$$

and consider the subfields $\mathcal{F}_t = \sigma(s_t, s_{t-1}, \dots)$ where $s_t = (y_t, w_t)'$. We will now show that

$$T^{-1/2} S_T \xrightarrow{d} \mathrm{N}\left(0, \begin{bmatrix} V_\mu & C'_{\mu,\Gamma(m)} \\ C_{\mu,\Gamma(m)} & V_{\Gamma(m)} \end{bmatrix}\right), \tag{3.12.60}$$

which in turn yields (3.7.8). To do this, we will check the following conditions:

(i) $\{X_t, \mathcal{F}_t\}$ is stationary and ergodic;

(ii) $\{X_t, \mathcal{F}_t\}$ is a L_1-mixingale of size -1;

(iii) $\limsup_{T\to\infty} T^{-1/2}\mathbb{E}\|S_T\| < \infty$, where $\|\cdot\|$ is the Euclidean norm.

(i) The fact that $\{X_t, \mathcal{F}_t\}$ is stationary and ergodic follows from results 3.4.1 and 3.4.2.

(ii) - (1) A mixing zero-mean process is an adapted L_1-mixingale with respect to the sub-fields $\mathcal{F}_t$ provided it is bounded in the L_1-norm [see Davidson (1994, Theorem 14.2, p. 211)]. To see that $\{X_t\}$ is bounded in the L_1-norm, we note that:

$$\mathbb{E}|\log(y_t^2) - \mu| = \mathbb{E}|y_t^*| \le (\mathbb{E}|y_t^*|^2)^{1/2} = (\mathbb{E}[y_t^{*2}])^{1/2} = \sqrt{\gamma_{y^*}(0)} < \infty, \tag{3.12.61}$$

$$
\begin{aligned}
\mathbb{E}|y_t^* y_{t+k}^* - \gamma_{y^*}(k)| &= \mathbb{E}|y_t^* y_{t+k}^*| - |\gamma_{y^*}(k)| \\
&\leq \mathbb{E}|y_t^* y_{t+k}^*| \\
&\leq (\mathbb{E}|y_t^*|^2)^{1/2}(\mathbb{E}|y_{t+k}^*|^2)^{1/2} \\
&= (\mathbb{E}[y_t^{*2}])^{1/2}(\mathbb{E}[y_{t+k}^{*2}])^{1/2} \\
&= \mathbb{E}[y_t^{*2}] = \gamma_{y^*}(0) < \infty, \quad \text{for } \ k = 0, 1, \ldots, m,
\end{aligned} \tag{3.12.62}
$$

where the inequality in (3.12.61) is the application of Lyapunov's inequality and the second inequality in (3.12.62) follows from the Hölder's inequality.

(ii) - (2) We now show that $\{X_t, \mathcal{F}_t\}$ is a L_1−mixingale of size −1. From the discussion in Section 3.4 , we know that X_t is β-mixing, so it has mixing coefficients of the type $\beta_T = \psi\rho^T$, $\psi > 0$, $0 < \rho < 1$. To show that $\{X_t\}$ is of size -1, its mixing coefficients β_T must be $O(T^{-\varphi})$, with $\varphi > 1$ [see Davidson (1994, Definition 16.1, p. 247)]. Indeed,

$$
\frac{\rho^T}{T^{-\varphi}} = T^{\varphi}\exp(T\log\rho) = \exp(\varphi\log T)\exp(T\log\rho) = \exp[\varphi(\log T) + T(\log\rho)]. \tag{3.12.63}
$$

Since $\lim\limits_{T\to\infty} [\varphi(\log T) + T(\log\rho)] = -\infty$, we get

$$
\lim_{T\to\infty} \exp[\varphi(\log T) + T(\log\rho)] = 0. \tag{3.12.64}
$$

This holds in particular for $\varphi > 1$; see Rudin (1976, Theorem 3.20(d), p. 57).

(iii) To show that $\limsup\limits_{T\to\infty} T^{-1/2}\mathbb{E}\|S_T\| < \infty$, we first observe that $\mathbb{E}(S_T) = 0$ and, using the Cauchy-Schwarz inequality,

$$
\begin{aligned}
(T^{-1/2}\mathbb{E}\|S_T\|)^2 &\leq \frac{1}{T}\mathbb{E}(\|S_T\|^2) = \frac{1}{T}\mathbb{E}(S_T'S_T) = \frac{1}{T}\text{tr}[\mathbb{E}(S_T S_T')] = \frac{1}{T}\text{tr}[\text{Var}(S_T)] \\
&= \text{tr}[\text{Var}(T^{-1/2}S_T)].
\end{aligned} \tag{3.12.65}
$$

It is thus sufficient to show that

$$
\limsup_{T\to\infty} \text{tr}[\text{Var}(T^{-1/2}S_T)] < \infty. \tag{3.12.66}
$$

We now consider separately the components Ψ_t and Λ_t of X_t.

(iii) - (1) Set

$$S_{T1} := \sum_{t=1}^{T} \Psi_t\,, \quad \zeta_{\Psi}(\tau) := \mathrm{cov}(\Psi_t, \Psi_{t+\tau})\,. \tag{3.12.67}$$

Then

$$\zeta_{\Psi}(\tau) = \mathbb{E}[(\log(y_t^2) - \mu)(\log(y_{t+\tau}^2) - \mu)] = \mathbb{E}[y_t^* y_{t+\tau}^*] = \gamma_{y^*}(\tau)\,, \tag{3.12.68}$$

$$\begin{aligned} \mathrm{Var}(T^{-1/2} S_{T1}) &= \frac{1}{T}\left[\sum_{t=1}^{T} \mathrm{Var}(\Psi_t) + \sum_{t \neq s} \mathrm{cov}(\Psi_t, \Psi_s)\right] = \frac{1}{T}\left[T\zeta_{\Psi}(0) + 2\sum_{\tau=1}^{T}(T-\tau)\zeta_{\Psi}(\tau)\right] \\ &= \zeta_{\Psi}(0) + 2\sum_{\tau=1}^{T}(1-\frac{\tau}{T})\zeta_{\Psi}(\tau) = \gamma_{y^*}(0) + 2\sum_{\tau=1}^{T}(1-\frac{\tau}{T})\gamma_{y^*}(\tau) \end{aligned} \tag{3.12.69}$$

hence

$$\begin{aligned} \limsup_{T\to\infty} \mathrm{Var}(T^{-1/2} S_{T1}) &= \limsup_{T\to\infty} [\gamma_{y^*}(0) + 2\sum_{\tau=1}^{T}(1-\frac{\tau}{T})\gamma_{y^*}(\tau)] \\ &= \gamma_{y^*}(0) + 2\sum_{\tau=1}^{\infty}\gamma_{y^*}(\tau) = \sum_{\tau=-\infty}^{\infty}\gamma_{y^*}(\tau) \\ &\leq \sum_{\tau=-\infty}^{\infty}|\gamma_{y^*}(\tau)| < \infty. \end{aligned} \tag{3.12.70}$$

This convergence is due to the fact that y_t^* follows a stationary ARMA(p, p) process. So y_t^* can be viewed as an MA(∞) process with absolutely summable coefficients, which implies the absolute summability of autocovariances [see Hamilton (1994, chapter 3, page 52)]. By the Cauchy-Schwarz inequality this entails

$$\limsup_{T\to\infty} T^{-1/2}\mathbb{E}|S_{T1}| < \infty. \tag{3.12.71}$$

(iii) - (2) Set

$$S_{T2} := \sum_{t=1}^{T} \Lambda_t = [S_{T2,0}, S_{T2,1}, \ldots, S_{T2,m}]'\,, \tag{3.12.72}$$

$$S_{T2,k} := \sum_{t=1}^{T} \Lambda_{t,k}, \quad \zeta_{\Lambda_k}(\tau) := \mathrm{cov}(\Lambda_{t,k}, \Lambda_{t+\tau,k})\,, \quad k = 0, 1, \ldots, m\,. \tag{3.12.73}$$

Then, for $k = 0, 1, \ldots, m$,

$$\begin{aligned}
\zeta_{\Lambda_k}(\tau) &= \mathbb{E}[\big(y_t^* y_{t+k}^* - \gamma_{y^*}(k)\big)\big(y_{t+\tau}^* y_{t+\tau+k}^* - \gamma_{y^*}(k)\big)] = \mathbb{E}[y_t^* y_{t+k}^* y_{t+\tau}^* y_{t+\tau+k}^*] - \gamma_{y^*}(k)^2 \\
&= \mathbb{E}[y_t^* y_{t+k}^*]\mathbb{E}[y_{t+\tau}^* y_{t+\tau+k}^*] + \mathrm{cov}(y_t^*, y_{t+\tau}^*)\mathrm{cov}(y_{t+k}^*, y_{t+\tau+k}^*) \\
&\quad + \mathrm{cov}(y_t^*, y_{t+\tau+k}^*)\mathrm{cov}(y_{t+k}^*, y_{t+\tau}^*) - \gamma_{y^*}(k)^2 \\
&= \gamma_{y^*}(k)^2 + \gamma_{y^*}(\tau)^2 + \gamma_{y^*}(\tau+k)\gamma_{y^*}(\tau-k) - \gamma_{y^*}(k)^2 \\
&= \gamma_{y^*}(\tau)^2 + \gamma_{y^*}(\tau+k)\gamma_{y^*}(\tau-k), \qquad (3.12.74)
\end{aligned}$$

hence

$$\begin{aligned}
\mathrm{Var}(T^{-1/2} S_{T2,k}) &= \frac{1}{T}\left[\sum_{t=1}^{T}\mathrm{Var}(\Lambda_{t,k}) + \sum_{t\neq s}\mathrm{cov}(\Lambda_{t,k}, \Lambda_{s,k})\right] = \frac{1}{T}\left[T\zeta_{\Lambda_k}(0) + 2\sum_{\tau=1}^{T}(T-\tau)\zeta_{\Lambda_k}(\tau)\right] \\
&= \zeta_{\Lambda_k}(0) + 2\sum_{\tau=1}^{T}(1-\frac{\tau}{T})\zeta_{\Lambda_k}(\tau) \\
&= \gamma_{y^*}(0)^2 + \gamma_{y^*}(k)\gamma_{y^*}(-k) + 2\sum_{\tau=1}^{T}(1-\frac{\tau}{T})[\gamma_{y^*}(\tau)^2 \\
&\quad + \gamma_{y^*}(\tau+k)\gamma_{y^*}(\tau-k)], \qquad (3.12.75)
\end{aligned}$$

and

$$\begin{aligned}
\limsup_{T\to\infty}\mathrm{Var}(T^{-1/2} S_{T2,k}) &= \gamma_{y^*}(0)^2 + \gamma_{y^*}(k)\gamma_{y^*}(-k) \\
&\quad + \limsup_{T\to\infty}[2\sum_{\tau=1}^{T}(1-\frac{\tau}{T})[\gamma_{y^*}(\tau)^2 + \gamma_{y^*}(\tau+k)\gamma_{y^*}(\tau-k)]] \\
&= \sum_{\tau=-\infty}^{\infty}[\gamma_{y^*}(\tau)^2 + \gamma_{y^*}(\tau+k)\gamma_{y^*}(\tau-k)] \\
&= \sum_{\tau=-\infty}^{\infty}\gamma_{y^*}(\tau)^2 + \sum_{\tau=-\infty}^{\infty}\gamma_{y^*}(\tau+k)\gamma_{y^*}(\tau-k) \\
&= \sum_{\tau=-\infty}^{\infty}\gamma_{y^*}(\tau)^2 + \sum_{\tau=-\infty}^{\infty}\gamma_{y^*}^2(\tau+k) < \infty. \qquad (3.12.76)
\end{aligned}$$

This convergence is due to the fact that absolute summability implies square-summability. We deduce that

$$\limsup_{T\to\infty} T^{-1/2}\mathbb{E}\left|S_{T2,k}\right| < \infty, \quad k = 0, 1, \ldots, m. \qquad (3.12.77)$$

Combining (3.12.71) and (3.12.77), we get, for any $(m+2)\times 1$ fixed real vector $a \neq 0$,

$$\limsup_{T\to\infty} T^{-1/2}\mathbb{E}\left|a' S_T\right| < \infty. \tag{3.12.78}$$

It is also clear properties (i) and (ii) also hold if we replace S_T by $a'S_T$. Thus we can apply Theorem 24.5 of Davidson (1994) to $a'S_T$ to state that $T^{-1/2}(a'S_T)$ is asymptotically normal. Since this holds for any $a \neq 0$, it follows from the Cramér-Wold theorem that $T^{-1/2}\sum_{t=1}^{T} X_t$ is asymptotically multinormal:

$$T^{-1/2}S_T = T^{-1/2}\sum_{t=1}^{T} X_t = \sqrt{T}\begin{bmatrix} \hat{\mu}-\mu \\ \hat{\Gamma}(m)-\Gamma(m)\end{bmatrix} \xrightarrow{d} \mathrm{N}[0,\, V], \tag{3.12.79}$$

where

$$V = \lim_{T\to\infty} \mathbb{E}\{[T^{-1/2}S_T][T^{-1/2}S_T]'\} = \begin{bmatrix} V_\mu & C'_{\mu,\Gamma(m)} \\ C_{\mu,\Gamma(m)} & V_{\Gamma(m)} \end{bmatrix}. \tag{3.12.80}$$

Using (3.12.70) and (3.12.75), we have:

$$V_\mu = \gamma_{y^*}(0) + 2\sum_{\tau=1}^{\infty} \gamma_{y^*}(\tau), \tag{3.12.81}$$

$$V_{\Gamma(m)} = \mathrm{Var}(\Lambda_t) + 2\sum_{\tau=1}^{\infty} \mathrm{cov}(\Lambda_t,\, \Lambda_{t+\tau}), \tag{3.12.82}$$

$$C_{\mu,\Gamma(m)} = [C_{\mu 0},\, C_{\mu 1}, \ldots,\, C_{\mu m}]', \tag{3.12.83}$$

$$\begin{aligned} C_{\mu k} &= \sum_t \mathrm{cov}(\Psi_t,\, \Lambda_{t,k}) = 2\sum_{t=1}^{\infty} \mathbb{E}[\Psi_t\big(y_t^* y_{t+k}^* - \gamma_{y^*}(k)\big)] \\ &= 2\sum_{t=1}^{\infty} \mathbb{E}[y_t^*\big(y_t^* y_{t+k}^* - \gamma_{y^*}(k)\big)] = 2\sum_{t=1}^{\infty} [\mathbb{E}(y_t^{*2} y_{t+k}^*) - \mathbb{E}(y_t^*)\gamma_{y^*}(k)] \\ &= 2\sum_{t=1}^{\infty} \mathbb{E}(y_t^{*2} y_{t+k}^*), \quad k = 0, 1, 2, \ldots, m. \end{aligned} \tag{3.12.84}$$

Further, for $k=0$, we substitute $y_t^* = w_t + \epsilon_t$ to get

$$\bar{c} := C_{\mu 0} = 2\sum_{t=1}^{\infty} \mathbb{E}(y_t^{*3}) = 2\sum_{t=1}^{\infty} [\mathbb{E}(w_t^3) + \mathbb{E}(\epsilon_t^3)] = 2\sum_{t=1}^{\infty} \mathbb{E}(\epsilon_t^3). \tag{3.12.85}$$

Since $\{z_t\}$ is a sequence of i.i.d. N[0,1] random variables, we have $\mathbb{E}(\epsilon_t^3) = \psi^{(2)}(\frac{1}{2})$ [see (3.2.6)], which is equal to $-14\mathcal{Z}(3)$ where $\mathcal{Z}(\cdot)$ is Riemann's Zeta function with $\mathcal{Z}(3) = 1.20205$.[7] For $k = 1, \ldots, m$, it is easily seen that $C_{\mu k} = 0$ from the MA(∞) representation of w_t. So $C_{\mu,\Gamma(m)}$ is a $(m+1)\times 1$ vector given by $(\bar{c}, 0_{[m\times 1]})'$, with $\bar{c}$ is defined in (3.12.85). Finally, (3.7.8) follows on observing that

$$\sqrt{T}\begin{bmatrix} \hat{\mu}-\mu \\ \hat{\Gamma}(m)-\Gamma(m) \end{bmatrix} - T^{-1/2}S_T \xrightarrow[T\to\infty]{p} 0. \tag{3.12.86}$$

□

Proof of Theorem 3.7.5 It is easily seen that D_p is a continuously differentiable mapping of $[\mu, \gamma_{y^*}(0), \gamma_{y^*}(1), \ldots, \gamma_{y^*}(2p)]'$. The convergence result stated in (3.7.13) follows from the standard result for differentiable transformations of asymptotically normally distributed variables together with the application of the multivariate delta method.

In case of an SV(1) model,

$$D_1 := D_1(\beta) = (D_{\phi_1}, D_{\sigma_y}, D_{\sigma_v})', \quad \beta := [\mu, \gamma_{y^*}(0), \gamma_{y^*}(1), \gamma_{y^*}(2)]', \tag{3.12.87}$$

$$D_{\phi_1} = \gamma_{y^*}(2)/\gamma_{y^*}(1), \quad D_{\sigma_y} = \exp(\mu+1.27)^{1/2}, \quad D_{\sigma_v} = \bar{\kappa}_1\bar{\kappa}_2, \tag{3.12.88}$$

$$G(\beta) := \frac{\partial D_1}{\partial \beta'} = \begin{pmatrix} 0 & 0 & -\gamma_{y^*}(2)/\gamma_{y^*}(1)^2 & 1/\gamma_{y^*}(1) \\ \sigma_y/2 & 0 & 0 & 0 \\ 0 & \frac{1}{2}\sqrt{\frac{\bar{\kappa}_1}{\bar{\kappa}_2}} & \frac{\gamma_{y^*}(2)^2}{\gamma_{y^*}(1)^3}\sqrt{\frac{\bar{\kappa}_2}{\bar{\kappa}_1}} & -\frac{\gamma_{y^*}(2)}{\gamma_{y^*}(1)^2}\sqrt{\frac{\bar{\kappa}_2}{\bar{\kappa}_1}} \end{pmatrix} \tag{3.12.89}$$

where $\sigma_y := \sqrt{\exp(\mu+1.27)}$, $\bar{\kappa}_1 := [1-(\gamma_{y^*}(2)/\gamma_{y^*}(1))^2]$, $\bar{\kappa}_2 := [\gamma_{y^*}(0) - \pi^2/2]$.

Similarly, for an SV(2) model, we have

$$D_2 := D_2(\beta) = (D_{\phi_2}, D_{\sigma_y}, D_{\sigma_v})', \quad \beta := [\mu, \gamma_{y^*}(0), \gamma_{y^*}(1), \ldots \gamma_{y^*}(4)]', \quad D_{\phi_2} = [D_{\phi_1} \quad D_{\phi_2}]', \tag{3.12.90}$$

$$D_{\phi_1} := \frac{\gamma_{y^*}(2)\gamma_{y^*}(3) - \gamma_{y^*}(1)\gamma_{y^*}(4)}{\gamma_{y^*}^2(2) - \gamma_{y^*}(1)\gamma_{y^*}(3)}, \quad D_{\phi_2} := \frac{\gamma_{y^*}(2)\gamma_{y^*}(4) - \gamma_{y^*}^2(3)}{\gamma_{y^*}^2(2) - \gamma_{y^*}(1)\gamma_{y^*}(3)}, \tag{3.12.91}$$

$$D_{\sigma_y} = \exp(\mu+1.27)^{1/2}, \quad D_{\sigma_v} = [\gamma_{y^*}(0) - \pi^2/2 - \phi_1\gamma_{y^*}(1) - \phi_2\gamma_{y^*}(2)]^{1/2}, \tag{3.12.92}$$

[7] The Riemann Zeta function for $s \in \mathbb{C}$ with $\mathrm{Re}(s) > 1$ is defined as $\mathcal{Z}(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}$.

$$G(\beta) := \frac{\partial D_2}{\partial \beta'} = \begin{pmatrix} 0 & 0 & \bar{\kappa}_3(\gamma_{y^*}(4) - \phi_1\gamma_{y^*}(3)) & \bar{\kappa}_3(2\phi_1\gamma_{y^*}(2) - \gamma_{y^*}(3)) & -\bar{\kappa}_3(\phi_1\gamma_{y^*}(1) + \gamma_{y^*}(2)) & \bar{\kappa}_3\gamma_{y^*}(1) \\ 0 & 0 & \bar{\kappa}_3\phi_2\gamma_{y^*}(3) & -\bar{\kappa}_3(2\phi_2\gamma_{y^*}(2) + \gamma_{y^*}(4)) & -\bar{\kappa}_3(\phi_2\gamma_{y^*}(1) - 2\gamma_{y^*}(3)) & -\bar{\kappa}_3\gamma_{y^*}(2) \\ \frac{\sigma_y}{2} & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2\sigma_v} & \frac{\bar{\kappa}_3\bar{\kappa}_4 - \phi_1}{2\sigma_v} & \frac{\bar{\kappa}_3\bar{\kappa}_5 - \phi_2}{2\sigma_v} & \frac{\bar{\kappa}_3\bar{\kappa}_6}{2\sigma_v} & \frac{\bar{\kappa}_3(\gamma^2_{y^*}(2) - \gamma^2_{y^*}(1))}{2\sigma_v} \end{pmatrix}$$

where

$$\phi_1 = D_{\phi_1}, \quad \phi_2 = D_{\phi_2}, \quad \sigma_y = D_{\sigma_y}, \quad \sigma_v = D_{\sigma_v},$$

$$\bar{\kappa}_3 := [\gamma^2_{y^*}(2) - \gamma_{y^*}(1)\gamma_{y^*}(3)]^{-1},$$

$$\bar{\kappa}_4 := [\phi_1\gamma_{y^*}(1)\gamma_{y^*}(3) + \phi_2\gamma_{y^*}(2)\gamma_{y^*}(3) - \gamma_{y^*}(1)\gamma_{y^*}(4)],$$

$$\bar{\kappa}_5 := [\gamma_{y^*}(1)\gamma_{y^*}(2) + \gamma_{y^*}(2)\gamma_{y^*}(4) - 2\phi_1\gamma_{y^*}(1)\gamma_{y^*}(2) - 2\phi_2\gamma^2_{y^*}(2)],$$

$$\bar{\kappa}_6 := [\phi_1\gamma^2_{y^*}(1) + (1 + \phi_2)\gamma_{y^*}(1)\gamma_{y^*}(2) - 2\gamma_{y^*}(2)\gamma_{y^*}(3)].$$

□

3.12.2 Tables

Table 3.1. Comparison of different winsorized ARMA-SV estimators (W-ARMA-SV)
An SV(2) model with parameters: $(\phi_1, \phi_2, \sigma_y, \sigma_\nu) = (0.30, 0.60, 0.025, 2.5)$

	$T=500$							$T=1000$							$T=2000$						
	ϕ_1		ϕ_2		σ_ν			ϕ_1		ϕ_2		σ_ν			ϕ_1		ϕ_2		σ_ν		
J	Bias	RMSE	Bias	RMSE	Bias	RMSE	NIV	Bias	RMSE	Bias	RMSE	Bias	RMSE	NIV	Bias	RMSE	Bias	RMSE	Bias	RMSE	NIV
Arithmatic Mean																					
1	0.016	0.186	-0.032	0.179	0.001	0.179	17	0.012	0.125	-0.020	0.121	0.011	0.129	4	0.002	0.087	-0.006	0.084	0.005	0.092	0
5	0.032	0.371	-0.049	0.345	0.015	0.197	204	0.036	0.337	-0.043	0.316	0.013	0.135	141	0.024	0.256	-0.027	0.240	0.007	0.097	65
10	0.077	0.415	-0.094	0.387	0.045	0.214	444	0.065	0.411	-0.072	0.385	0.015	0.151	352	0.100	0.422	-0.099	0.396	0.018	0.109	322
20	0.132	0.438	-0.197	0.410	0.162	0.410	501	0.164	0.451	-0.175	0.425	0.068	0.215	495	0.143	0.446	-0.146	0.418	0.047	0.158	422
30	0.123	0.432	-0.224	0.431	0.233	0.527	565	0.123	0.439	-0.179	0.414	0.176	0.466	511	0.149	0.440	-0.178	0.422	0.111	0.344	475
40	0.111	0.428	-0.232	0.426	0.281	0.588	594	0.127	0.414	-0.223	0.419	0.260	0.631	593	0.132	0.416	-0.167	0.402	0.128	0.381	523
50	0.098	0.436	-0.258	0.456	0.374	0.744	651	0.093	0.428	-0.213	0.427	0.309	0.684	606	0.146	0.444	-0.209	0.448	0.206	0.516	602
100	0.106	0.470	-0.332	0.470	0.559	0.844	678	0.114	0.433	-0.250	0.449	0.344	0.720	647	0.130	0.428	-0.258	0.436	0.353	0.716	697
Weighted Mean with Linearly Declining Weights																					
1	0.016	0.186	-0.032	0.179	0.001	0.179	17	0.012	0.125	-0.020	0.121	0.011	0.129	4	0.002	0.087	-0.006	0.084	0.005	0.092	0
5	0.023	0.322	-0.040	0.300	0.014	0.187	127	0.025	0.259	-0.032	0.243	0.007	0.128	76	0.012	0.180	-0.015	0.169	0.004	0.092	26
10	0.048	0.396	-0.066	0.369	0.036	0.200	321	0.016	0.337	-0.025	0.315	0.005	0.139	210	0.052	0.355	-0.053	0.334	0.009	0.100	172
20	0.108	0.416	-0.141	0.392	0.087	0.266	436	0.124	0.426	-0.132	0.401	0.046	0.179	416	0.082	0.395	-0.084	0.370	0.025	0.118	340
30	0.140	0.437	-0.194	0.416	0.137	0.374	496	0.126	0.420	-0.153	0.390	0.095	0.287	445	0.115	0.415	-0.122	0.392	0.052	0.155	418
40	0.082	0.397	-0.191	0.401	0.256	0.598	540	0.139	0.446	-0.183	0.418	0.139	0.371	516	0.135	0.420	-0.149	0.399	0.079	0.272	469
50	0.127	0.437	-0.234	0.437	0.274	0.578	584	0.105	0.416	-0.176	0.421	0.204	0.488	538	0.150	0.450	-0.176	0.429	0.109	0.346	543
100	0.145	0.452	-0.313	0.465	0.441	0.732	623	0.137	0.427	-0.252	0.445	0.284	0.647	631	0.143	0.464	-0.244	0.467	0.276	0.688	640
Median																					
1	0.016	0.186	-0.032	0.179	0.001	0.179	17	0.012	0.125	-0.020	0.121	0.011	0.129	4	0.002	0.087	-0.006	0.084	0.005	0.092	0
5	0.011	0.220	-0.027	0.206	0.009	0.192	28	0.017	0.171	-0.025	0.159	0.010	0.132	13	0.005	0.124	-0.008	0.118	0.004	0.099	5
10	-0.044	0.193	0.017	0.182	0.036	0.200	37	-0.040	0.157	0.027	0.147	0.001	0.138	17	-0.019	0.142	0.013	0.133	0.005	0.099	9
20	-0.104	0.166	0.051	0.158	0.078	0.273	34	-0.094	0.140	0.069	0.130	0.031	0.186	8	-0.074	0.128	0.058	0.118	0.020	0.136	7
30	-0.110	0.153	0.063	0.153	0.057	0.333	57	-0.112	0.131	0.075	0.129	0.054	0.246	16	-0.104	0.120	0.077	0.112	0.044	0.184	12
40	-0.112	0.132	0.068	0.136	0.051	0.327	68	-0.114	0.124	0.083	0.122	0.040	0.286	27	-0.113	0.113	0.090	0.105	0.031	0.215	25
50	-0.111	0.125	0.080	0.129	0.012	0.319	125	-0.118	0.111	0.092	0.113	0.021	0.284	56	-0.120	0.108	0.100	0.105	0.010	0.248	30
100	-0.120	0.096	0.113	0.103	-0.070	0.295	128	-0.111	0.096	0.110	0.101	-0.076	0.264	96	-0.107	0.086	0.110	0.091	-0.069	0.226	83
OLS without Intercept																					
1	0.016	0.186	-0.032	0.179	0.001	0.179	17	0.012	0.125	-0.020	0.121	0.011	0.129	4	0.002	0.087	-0.006	0.084	0.005	0.092	0
5	0.013	0.164	-0.031	0.158	0.017	0.177	0	0.013	0.113	-0.021	0.109	0.008	0.125	0	0.001	0.078	-0.005	0.076	0.003	0.090	0
10	-0.002	0.145	-0.017	0.143	0.030	0.175	0	-0.002	0.109	-0.006	0.104	0.002	0.123	0	0.001	0.082	-0.005	0.080	0.005	0.088	0
20	-0.015	0.132	-0.004	0.129	0.017	0.173	0	-0.010	0.103	0.001	0.101	0.011	0.127	0	-0.003	0.076	-0.002	0.074	0.009	0.086	0
30	-0.019	0.132	-0.003	0.130	0.020	0.187	0	-0.018	0.100	0.008	0.097	0.002	0.126	0	-0.006	0.079	0.001	0.077	0.007	0.090	0
40	-0.025	0.125	0.005	0.124	0.013	0.176	0	-0.011	0.100	0.001	0.098	0.008	0.126	0	-0.006	0.077	0.002	0.075	0.005	0.089	0
50	-0.014	0.127	-0.004	0.125	0.012	0.178	0	-0.017	0.101	0.009	0.099	0.003	0.122	0	-0.003	0.070	-0.002	0.069	0.004	0.088	0
100	-0.037	0.115	0.024	0.115	-0.008	0.170	0	-0.011	0.099	0.004	0.097	-0.006	0.124	0	-0.009	0.075	0.004	0.073	0.000	0.089	0

Note: The estimators compared are the simple and winsorized estimators defined in Section 3.5.4. For $J = 1$ all the estimators reduce to the simple ARMA-SV estimator. RMSE is the estimated root mean square error based on the simulation. NIV stands for the number of inadmissible parameter values produced by the estimators (over 1000).

Table 3.2. Comparison of different estimation methods for an SV(2) model: bias and RMSE
Model: $M_1 = (0.30, 0.60, 0.025, 2.5)$

	$T = 500$				$T = 2000$			
	ϕ_1	ϕ_2	σ_y	σ_v	ϕ_1	ϕ_2	σ_y	σ_v
True value	0.30	0.60	0.025	2.5	0.30	0.60	0.025	2.5
				Bias				
GMM-6M-E	-0.248	-0.761	0.049	2.625	-0.281	-0.799	0.060	2.234
GMM-6M-E-R	-0.252	-0.773	0.049	2.625	-0.279	-0.807	0.060	2.234
GMM-6M-NW	-0.526	-0.482	1.586	0.060	-0.551	-0.604	2.201	-0.407
GMM-6M-NW-R	-0.456	-0.538	1.586	0.060	-0.541	-0.612	2.201	-0.407
GMM-24M-E	0.149	-0.413	0.049	3.108	0.202	-0.491	0.055	3.051
GMM-24M-E-R	0.123	-0.432	0.049	3.108	0.144	-0.538	0.055	3.051
GMM-24M-NW	-0.528	-0.612	1.961	0.608	-0.574	-0.747	2.851	-0.030
GMM-24M-NW-R	-0.486	-0.648	1.961	0.608	-0.568	-0.752	2.851	-0.030
Bayesian-MCMC	0.772	-0.825	0.314	-2.269	0.731	-0.737	0.349	-2.298
EDV	-0.081	-0.250	0.650	-1.096	-0.048	-0.146	0.680	-0.981
EDV-R	-0.024	-0.244	0.678	-1.340	-0.017	-0.135	0.724	-1.181
ARMA-SV	0.016	-0.032	0.004	0.001	0.004	-0.008	0.001	0.002
R-ARMA-SV	0.009	-0.025	0.004	0.001	0.004	-0.008	0.001	0.002
W-ARMA-SV($J = 10$)	-0.006	-0.012	0.004	0.004	0.000	-0.005	0.001	0.002
				RMSE				
GMM-6M-E	1.042	0.901	0.089	4.198	1.011	0.938	0.095	3.979
GMM-6M-E-R	1.027	0.895	0.089	4.198	0.997	0.935	0.095	3.979
GMM-6M-NW	0.883	0.737	2.401	2.139	0.770	0.697	3.016	1.284
GMM-6M-NW-R	0.798	0.702	2.401	2.139	0.754	0.690	3.016	1.284
GMM-24M-E	0.544	0.587	0.071	4.237	0.690	0.731	0.063	4.424
GMM-24M-E-R	0.518	0.570	0.071	4.237	0.638	0.697	0.063	4.424
GMM-24M-NW	0.993	0.866	3.039	2.982	0.962	0.870	3.805	2.214
GMM-24M-NW-R	0.963	0.844	3.039	2.982	0.957	0.866	3.805	2.214
Bayesian-MCMC	0.883	0.847	0.515	2.272	0.804	0.740	0.470	2.299
EDV	0.356	0.399	1.377	0.446	0.301	0.343	0.840	0.508
EDV-R	0.542	0.546	1.310	0.664	0.431	0.457	0.980	0.688
ARMA-SV	0.186	0.179	0.017	0.179	0.084	0.080	0.008	0.090
R-ARMA-SV	0.191	0.184	0.017	0.178	0.084	0.080	0.008	0.090
W-ARMA-SV($J = 10$)	0.143	0.139	0.017	0.177	0.075	0.072	0.008	0.089

Notes:

1. GMM-6M and GMM-24M are the generalized method of moment estimators with six moments and 24 moments, respectively.
2. E stands for the efficient GMM estimation where we used the inverse of the covariance matrix as the weighting matrix.
3. R stands for the restricted estimation proposed in Section 3.5.3 where the estimates are restrained on the space of acceptable parameter solutions.
4. NW stands for the GMM estimation where we used the inverse of Newey West covariance matrix as the weighting matrix.
5. Bayesian-MCMC is the Bayesian estimator based on Markov Chain Monte Carlo methods.
6. EDV is the extension of Dufour and Valéry (2006) method proposed in Section 3.5.1.
7. ARMA-SV is the simple ARMA-based estimator proposed in Section 3.5.2.
8. W-ARMA-SV is the winsorized ARMA-SV estimator based on OLS proposed in Section 3.5.4.
9. Number of inadmissible values for each estimator are reported in Table 3.6.

Table 3.3. Comparison of different estimation methods for an SV(2) model: bias and RMSE
Model: $M_2 = (0.95, -0.85, 0.025, 2.5)$

	$T = 500$				$T = 2000$			
	ϕ_1	ϕ_2	σ_y	σ_v	ϕ_1	ϕ_2	σ_y	σ_v
True value	0.95	-0.85	0.025	2.5	0.95	-0.85	0.025	2.5
				Bias				
GMM-6M-E	-0.452	0.291	0.105	-1.058	-0.455	0.315	0.139	-1.080
GMM-6M-E-R	-0.442	0.275	0.105	-1.058	-0.445	0.302	0.139	-1.080
GMM-6M-NW	-1.102	0.802	1.569	-0.444	-0.971	0.719	1.571	-0.387
GMM-6M-NW-R	-1.101	0.802	1.569	-0.444	-0.971	0.719	1.571	-0.387
GMM-24M-E	-0.027	1.130	0.254	6.401	-0.053	1.496	0.420	5.953
GMM-24M-E-R	-0.225	0.988	0.254	6.401	-0.386	1.274	0.420	5.953
GMM-24M-NW	-1.119	0.737	2.610	-0.355	-0.953	0.685	2.708	-0.324
GMM-24M-NW-R	-1.117	0.734	2.610	-0.355	-0.953	0.685	2.708	-0.324
Bayesian-MCMC	0.312	0.489	1.323	-2.202	0.163	0.697	26.80	-2.282
EDV	-0.717	0.175	0.693	-1.175	-0.600	0.035	1.035	-1.454
EDV-R	-0.725	-0.094	2.437	-2.389	-0.513	-0.129	3.309	-2.451
ARMA-SV	-0.001	0.004	0.000	-0.017	0.000	0.001	0.000	0.001
R-ARMA-SV	-0.001	0.004	0.000	-0.017	0.000	0.001	0.000	0.001
W-ARMA-SV($J = 10$)	-0.001	0.002	0.000	-0.023	-0.001	0.001	0.000	0.001
				RMSE				
GMM-6M-E	0.892	0.617	0.122	2.196	0.839	0.589	0.151	1.866
GMM-6M-E-R	0.866	0.565	0.122	2.196	0.813	0.543	0.151	1.866
GMM-6M-NW	1.152	0.830	2.240	0.883	1.041	0.736	2.243	0.673
GMM-6M-NW-R	1.152	0.829	2.240	0.883	1.041	0.736	2.243	0.673
GMM-24M-E	0.672	1.373	0.409	6.795	0.417	1.637	0.615	6.536
GMM-24M-E-R	0.789	1.160	0.409	6.795	0.642	1.368	0.615	6.536
GMM-24M-NW	1.262	0.805	3.703	1.555	1.101	0.727	3.809	1.216
GMM-24M-NW-R	1.258	0.797	3.703	1.555	1.101	0.727	3.809	1.216
Bayesian-MCMC	0.437	0.545	2.138	2.208	0.256	0.699	32.51	2.282
EDV	0.426	0.155	0.320	0.323	0.274	0.221	0.338	0.587
EDV-R	1.005	0.313	3.654	0.602	0.877	0.194	6.545	0.427
ARMA-SV	0.035	0.041	0.002	0.187	0.017	0.019	0.001	0.089
R-ARMA-SV	0.035	0.041	0.002	0.187	0.017	0.019	0.001	0.089
W-ARMA-SV($J = 10$)	0.035	0.032	0.002	0.201	0.017	0.015	0.001	0.094

Notes:

1. GMM-6M and GMM-24M are the generalized method of moment estimators with six moments and 24 moments, respectively.
2. E stands for the efficient GMM estimation where we used the inverse of the covariance matrix as the weighting matrix.
3. R stands for the restricted estimation proposed in Section 3.5.3 where the estimates are restrained on the space of acceptable parameter solutions.
4. NW stands for the GMM estimation where we used the inverse of Newey West covariance matrix as the weighting matrix.
5. Bayesian-MCMC is the Bayesian estimator based on Markov Chain Monte Carlo methods.
6. EDV is the extension of Dufour and Valéry (2006) method proposed in Section 3.5.1.
7. ARMA-SV is the simple ARMA-based estimator proposed in Section 3.5.2.
8. W-ARMA-SV is the winsorized ARMA-SV estimator based on OLS proposed in Section 3.5.4.
9. Number of inadmissible values for each estimator are reported in Table 3.6.

Table 3.4. Comparison of different estimation methods for an SV(2) model: bias and RMSE
Model: $M_3 = (0.45, 0.45, 0.25, 2.5)$

	$T = 500$				$T = 2000$			
	ϕ_1	ϕ_2	σ_y	σ_v	ϕ_1	ϕ_2	σ_y	σ_v
True value	0.45	0.45	0.25	2.5	0.45	0.45	0.25	2.5
				Bias				
GMM-6M-E	-0.526	-0.627	0.556	2.600	-0.755	-0.613	0.687	1.670
GMM-6M-E-R	-0.495	-0.656	0.556	2.600	-0.728	-0.635	0.687	1.670
GMM-6M-NW	-0.361	-0.610	3.259	-0.229	-0.317	-0.685	3.082	-0.354
GMM-6M-NW-R	-0.359	-0.611	3.259	-0.229	-0.317	-0.685	3.082	-0.354
GMM-24M-E	-0.221	-0.346	0.610	2.817	-0.519	-0.409	0.811	3.574
GMM-24M-E-R	-0.184	-0.375	0.610	2.817	-0.488	-0.434	0.811	3.574
GMM-24M-NW	-0.342	-0.739	4.926	0.578	-0.275	-0.764	4.894	0.189
GMM-24M-NW-R	-0.342	-0.739	4.926	0.578	-0.275	-0.764	4.894	0.189
Bayesian-MCMC	0.615	-0.670	3.552	-2.274	0.560	-0.588	3.414	-2.303
EDV	-0.098	-0.208	7.498	-1.130	-0.077	-0.110	8.394	-1.006
EDV-R	-0.034	-0.198	8.274	-1.379	-0.029	-0.108	8.973	-1.218
ARMA-SV	0.083	-0.096	0.035	-0.015	0.026	-0.029	0.013	-0.003
R-ARMA-SV	0.072	-0.084	0.040	-0.022	0.025	-0.029	0.014	-0.003
W-ARMA-SV($J = 10$)	-0.077	0.053	0.040	0.031	-0.018	0.011	0.014	0.009
				RMSE				
GMM-6M-E	1.143	0.839	0.843	4.379	1.088	0.787	0.811	3.534
GMM-6M-E-R	1.114	0.814	0.843	4.379	1.062	0.769	0.811	3.534
GMM-6M-NW	0.784	0.677	4.481	1.267	0.715	0.729	4.278	1.029
GMM-6M-NW-R	0.780	0.676	4.481	1.267	0.715	0.729	4.278	1.029
GMM-24M-E	0.795	0.608	0.919	4.200	0.995	0.659	1.017	4.897
GMM-24M-E-R	0.741	0.572	0.919	4.200	0.960	0.630	1.017	4.897
GMM-24M-NW	1.013	0.846	5.986	2.766	0.887	0.850	6.040	2.166
GMM-24M-NW-R	1.012	0.846	5.986	2.766	0.887	0.850	6.040	2.166
Bayesian-MCMC	0.750	0.695	7.950	2.277	0.672	0.592	4.395	2.304
EDV	0.399	0.456	13.432	0.461	0.343	0.392	10.642	0.513
EDV-R	0.580	0.604	13.893	0.662	0.507	0.530	12.179	0.687
ARMA-SV	0.383	0.363	0.171	0.204	0.211	0.200	0.077	0.100
R-ARMA-SV	0.488	0.460	0.173	0.247	0.218	0.207	0.077	0.102
W-ARMA-SV($J = 10$)	0.203	0.195	0.173	0.187	0.155	0.148	0.077	0.094

Notes:

1. GMM-6M and GMM-24M are the generalized method of moment estimators with six moments and 24 moments, respectively.
2. E stands for the efficient GMM estimation where we used the inverse of the covariance matrix as the weighting matrix.
3. R stands for the restricted estimation proposed in Section 3.5.3 where the estimates are restrained on the space of acceptable parameter solutions.
4. NW stands for the GMM estimation where we used the inverse of Newey West covariance matrix as the weighting matrix.
5. Bayesian-MCMC is the Bayesian estimator based on Markov Chain Monte Carlo methods.
6. EDV is the extension of Dufour and Valéry (2006) method proposed in Section 3.5.1.
7. ARMA-SV is the simple ARMA-based estimator proposed in Section 3.5.2.
8. W-ARMA-SV is the winsorized ARMA-SV estimator based on OLS proposed in Section 3.5.4.
9. Number of inadmissible values for each estimator are reported in Table 3.6.

Table 3.5. Comparison of different estimation methods for an SV(2) model: bias and RMSE
Model: $M_4 = (0.00, 0.90, 0.025, 2.5)$

	$T = 500$				$T = 2000$			
	ϕ_1	ϕ_2	σ_y	σ_v	ϕ_1	ϕ_2	σ_y	σ_v
True value	0.00	0.90	0.025	2.5	0.00	0.90	0.025	2.5
				Bias				
GMM-6M-E	-0.426	-1.078	0.134	2.337	-0.280	-1.010	0.176	1.997
GMM-6M-E-R	-0.419	-1.087	0.134	2.337	-0.272	-1.021	0.176	1.997
GMM-6M-NW	-0.284	-0.884	2.927	-0.028	-0.145	-0.997	3.359	-0.055
GMM-6M-NW-R	-0.266	-0.898	2.927	-0.028	-0.145	-0.997	3.359	-0.055
GMM-24M-E	0.317	-0.817	0.153	4.405	-0.039	-0.816	0.181	4.296
GMM-24M-E-R	0.296	-0.831	0.153	4.405	-0.046	-0.826	0.181	4.296
GMM-24M-NW	-0.355	-1.013	3.753	0.599	-0.169	-1.094	4.831	0.673
GMM-24M-NW-R	-0.344	-1.023	3.753	0.599	-0.169	-1.094	4.831	0.673
Bayesian-MCMC	1.001	-1.099	1.246	-2.289	0.951	-1.020	1.241	-2.313
EDV	-0.248	-0.467	1.212	-1.281	-0.358	-0.455	1.719	-1.293
EDV-R	-0.103	-0.149	3.423	-1.471	-0.205	-0.209	4.119	-1.510
ARMA-SV	-0.005	-0.013	0.004	-0.002	-0.001	-0.004	0.001	0.001
R-ARMA-SV	-0.005	-0.013	0.004	-0.002	-0.001	-0.004	0.001	0.001
W-ARMA-SV($J = 10$)	-0.006	-0.016	0.004	0.012	-0.002	-0.004	0.001	0.003
				RMSE				
GMM-6M-E	0.937	1.180	0.218	3.958	0.786	1.101	0.270	3.517
GMM-6M-E-R	0.932	1.178	0.218	3.958	0.774	1.100	0.270	3.517
GMM-6M-NW	0.655	0.956	4.103	1.558	0.603	1.031	4.552	1.261
GMM-6M-NW-R	0.637	0.954	4.103	1.558	0.603	1.031	4.552	1.261
GMM-24M-E	0.863	0.942	0.494	5.491	0.724	0.925	0.244	5.510
GMM-24M-E-R	0.846	0.938	0.494	5.491	0.715	0.922	0.244	5.510
GMM-24M-NW	0.847	1.116	4.948	2.750	0.845	1.151	5.944	2.490
GMM-24M-NW-R	0.843	1.114	4.948	2.750	0.845	1.151	5.944	2.490
Bayesian-MCMC	1.108	1.114	4.054	2.292	1.038	1.023	1.849	2.315
EDV	0.401	0.443	1.421	0.412	0.447	0.479	1.413	0.478
EDV-R	0.425	0.464	9.023	0.750	0.495	0.522	8.593	0.705
ARMA-SV	0.031	0.030	0.017	0.183	0.014	0.014	0.008	0.091
R-ARMA-SV	0.031	0.030	0.017	0.183	0.014	0.014	0.008	0.091
W-ARMA-SV($J = 10$)	0.030	0.028	0.017	0.176	0.013	0.013	0.008	0.088

Notes:

1. GMM-6M and GMM-24M are the generalized method of moment estimators with six moments and 24 moments, respectively.
2. E stands for the efficient GMM estimation where we used the inverse of the covariance matrix as the weighting matrix.
3. R stands for the restricted estimation proposed in Section 3.5.3 where the estimates are restrained on the space of acceptable parameter solutions.
4. NW stands for the GMM estimation where we used the inverse of Newey West covariance matrix as the weighting matrix.
5. Bayesian-MCMC is the Bayesian estimator based on Markov Chain Monte Carlo methods.
6. EDV is the extension of Dufour and Valéry (2006) method proposed in Section 3.5.1.
7. ARMA-SV is the simple ARMA-based estimator proposed in Section 3.5.2.
8. W-ARMA-SV is the winsorized ARMA-SV estimator based on OLS proposed in Section 3.5.4.
9. Number of inadmissible values for each estimator are reported in Table 3.6.

Table 3.6. Comparison of different estimation methods for an SV(2) model: Number of inadmissible values

	$T = 500$				$T = 2000$			
	M_1	M_2	M_3	M_4	M_1	M_2	M_3	M_4
GMM-6M-E	40	26	59	17	44	24	43	25
GMM-6M-NW	142	6	5	35	26	0	0	0
GMM-24M-E	29	947	84	193	64	994	120	94
GMM-24M-NW	141	14	6	47	36	0	0	2
Bayesian-MCMC	0	0	0	0	0	0	0	0
EDV	213	988	245	766	156	992	189	660
ARMA-SV	17	0	140	0	0	0	4	0
W-ARMA-SV-OLS($J = 10$)	0	0	0	0	0	0	0	0

Notes:

1. GMM-6M and GMM-24M are the generalized method of moment estimators with six moments and 24 moments, respectively.
2. E stands for the efficient GMM estimation where we used the inverse of the covariance matrix as the weighting matrix.
3. NW stands for the GMM estimation where we used the inverse of Newey West covariance matrix as the weighting matrix.
4. Bayesian-MCMC is the Bayesian estimator based on Markov Chain Monte Carlo methods.
5. EDV is the extension of Dufour and Valéry (2006) method proposed in Section 3.5.1.
6. ARMA-SV is the simple ARMA-based estimator proposed in Section 3.5.2.
7. W-ARMA-SV-OLS is the winsorized ARMA-SV estimator based on OLS proposed in Section 3.5.4.

Table 3.7. Comparison of different estimation methods with respect to relative time for an SV(2) model using simulated data

Relative computing time with respect to ARMA-SV estimator		
	$T = 500$	$T = 2000$
GMM-6M-E	734.81	717.67
GMM-6M-NW	1019.49	1467.50
GMM-24M-E	1752.19	1785.62
GMM-24M-NW	3091.74	4059.37
Bayesian-MCMC	55750.12	127080.14
EDV	0.99	0.99
ARMA-SV	1.00	1.00
W-ARMA-SV-OLS($J = 10$)	1.38	1.36

Notes:

1. GMM-6M and GMM-24M are the generalized method of moment estimators with six moments and 24 moments, respectively.
2. E stands for the efficient GMM estimation where we used the inverse of the covariance matrix as the weighting matrix.
3. NW stands for the GMM estimation where we used the inverse of Newey West covariance matrix as the weighting matrix.
4. Bayesian-MCMC is the Bayesian estimator based on Markov Chain Monte Carlo methods.
5. EDV is the extension of Dufour and Valéry (2006) method proposed in Section 3.5.1.
6. ARMA-SV is the simple ARMA-based estimator proposed in Section 3.5.2.
7. W-ARMA-SV-OLS is the winsorized ARMA-SV estimator based on OLS proposed in Section 3.5.4.

Table 3.8. Summary statistics

S&P 500 index, 1928 - 2016, number of observations: 23372								
Series	Mean	Std. Dev.	Kurtosis	Skewness	Range	Max	Min	LB(10)
y_t	0.00	0.50	21.98	-0.43	16.62	6.66	-9.95	104.4
y_t^2	0.25	1.14	2647.09	37.17	99.08	99.08	0.00	7338.5
$\log(\lvert y_t \rvert)$	-1.73	1.25	5.07	-0.96	13.60	2.30	-11.30	5180.9
$y_t^* = \log(y_t^2) - \mu$	0.00	2.49	5.07	-0.96	27.20	8.05	-19.15	5180.9

Notes:

1. $y_t = r_t - \hat{\mu}_r$ is the residual return, y_t^2 is the squared of residual return and y_t^* is the residual of log square of residual return.
2. LB(10) is the heteroskedasticity-corrected Ljung - Box statistics with 10 lags. The critical values for LB(10) are: 15.99 (10%), 18.31 (5%), and 23.21 (1%).

Table 3.9. Asymptotic and finite sample inference for SV(p) models based on ARMA-type estimators

S&P 500 index, 1928 - 2016, number of observations: 23372							
$p=1$							
	Coefficient	Std. error	t-stat	Asymptotic tests	Local Monte Carlo tests		
					$N=19$	$N=99$	$N=999$
ϕ_1	0.9938	(0.0357)	27.84	0.00	0.05	0.01	0.001
σ_y	0.3356	(0.0167)	20.06	0.00	0.05	0.01	0.001
σ_v	0.6533	(0.0623)	10.48	0.00	0.05	0.01	0.001
Time (in seconds)				0.69	1.5	4.5	38.2
$p=2$							
	Coefficient	Std. error	t-stat	Asymptotic tests	Local Monte Carlo tests		
					$N=19$	$N=99$	$N=999$
ϕ_1	0.6887	(0.0719)	9.58	0.00	0.05	0.01	0.001
ϕ_2	0.2863	(0.0734)	3.90	0.00	0.05	0.01	0.001
σ_y	0.3356	(0.0167)	20.06	0.00	0.05	0.01	0.001
σ_v	0.6166	(0.3204)	1.92	0.03	0.10	0.06	0.075
Time (in seconds)				0.70	2.6	10.3	95.1
$p=3$							
	Coefficient	Std. error	t-stat	Asymptotic tests	Local Monte Carlo tests		
					$N=19$	$N=99$	$N=999$
ϕ_1	0.5477	(0.1204)	4.55	0.00	0.05	0.01	0.001
ϕ_2	-0.4264	(0.0936)	-4.55	0.00	0.05	0.01	0.001
ϕ_3	0.8489	(0.0122)	69.67	0.00	0.05	0.01	0.001
σ_y	0.3356	(0.0167)	20.06	0.00	0.05	0.01	0.001
σ_v	0.6211	(0.3993)	1.56	0.06	0.10	0.09	0.082
Time (in seconds)				0.79	12.2	60.2	622.1
$p=4$							
	Coefficient	Std. error	t-stat	Asymptotic tests	Local Monte Carlo tests		
					$N=19$	$N=99$	$N=999$
ϕ_1	0.3633	(0.2153)	1.69	0.05	0.05	0.01	0.001
ϕ_2	-0.0251	(0.2117)	-0.12	0.45	0.85	0.88	0.865
ϕ_3	0.6305	(0.0167)	37.68	0.00	0.05	0.01	0.001
ϕ_4	0.0005	(0.0162)	0.03	0.49	0.70	0.65	0.623
σ_y	0.3356	(0.0167)	20.06	0.00	0.05	0.01	0.001
σ_v	0.6133	(0.9210)	0.67	0.25	0.20	0.15	0.185
Time (in seconds)				0.97	20.7	105.0	1237.2

Notes:

1. Except for ϕ_1 and ϕ_2 parameters of SV(3) model and σ_y in all models, we test each coefficient is zero against a right-sided alternative.
2. We cannot test $\phi_1 = 0$ and $\phi_2 = 0$ in SV(3) model since putting each of these restrictions leads to some of the eigenvalues of the latent AR(3) model are outside the unit circle, hence non-stationarity. In these cases, the ARMA-based estimation is infeasible. So we test $\phi_1 = 0.2$ and $\phi_2 = -0.4$ against a right-sided and a left-sided alternative, respectively.
3. We test $\sigma_y = 0.01$ against a right-sided alternative since when $\sigma_y = 0$, SV models are unidentified.

Table 3.10. W-ARMA-SV-OLS estimates with different level of winsorization

S&P 500 index: 1996 - 2016, number of observations: 5222											
	J	10	20	30	40	50	60	70	80	90	100
SV(1)	$\hat{\phi}_1$	1.0119	0.9897	0.9930	0.9885	0.9835	0.9853	0.9835	0.9802	0.9777	0.9765
	$\hat{\sigma}_y$	0.3894	0.3894	0.3894	0.3894	0.3894	0.3894	0.3894	0.3894	0.3894	0.3894
	$\hat{\sigma}_v$	0.6109	0.8051	0.7455	0.8329	0.7702	0.8803	0.8361	0.8371	0.9014	0.9257
SV(2)	$\hat{\phi}_1$	0.7586	0.5736	0.5322	0.4797	0.3333	0.3447	0.3165	0.3342	0.3433	0.3315
	$\hat{\phi}_2$	0.1646	0.3541	0.4239	0.4808	0.6483	0.6341	0.6655	0.6490	0.6393	0.6514
	$\hat{\sigma}_y$	0.3894	0.3894	0.3894	0.3894	0.3894	0.3894	0.3894	0.3894	0.3894	0.3894
	$\hat{\sigma}_v$	0.7422	0.6961	0.6659	0.6492	0.5933	0.5988	0.5883	0.5923	0.5954	0.5917
SV(3)	$\hat{\phi}_1$	0.0440	0.0571	0.0983	0.0688	0.1407	0.1241	0.1004	0.1112	0.1281	0.1234
	$\hat{\phi}_2$	0.6099	0.5507	0.5341	0.5150	0.4115	0.2656	0.2886	0.2664	0.2451	0.2510
	$\hat{\phi}_3$	0.3212	0.3821	0.3558	0.4010	0.4176	0.5772	0.5839	0.5930	0.5972	0.5967
	$\hat{\sigma}_y$	0.3894	0.3894	0.3894	0.3894	0.3894	0.3894	0.3894	0.3894	0.3894	0.3894
	$\hat{\sigma}_v$	0.5181	0.5098	0.5242	0.5188	0.5550	0.5557	0.5433	0.5488	0.5543	0.5522
SV(4)	$\hat{\phi}_1$	0.5241	0.2733	0.2659	0.1763	0.0435	0.0235	0.0224	0.0244	0.0254	0.0243
	$\hat{\phi}_2$	-0.4048	0.5450	0.5395	0.5184	-0.2331	-0.1759	-0.1734	-0.1778	-0.1750	-0.1731
	$\hat{\phi}_3$	0.2491	0.1954	0.1867	0.2025	0.2863	0.3270	0.3297	0.3323	0.3301	0.3332
	$\hat{\phi}_4$	0.5563	0.0131	0.0329	0.1147	0.8431	0.7699	0.7661	0.7649	0.7638	0.7608
	$\hat{\sigma}_y$	0.3894	0.3894	0.3894	0.3894	0.3894	0.3894	0.3894	0.3894	0.3894	0.3894
	$\hat{\sigma}_v$	0.7420	0.5426	0.5441	0.5386	0.6434	0.6275	0.6268	0.6279	0.6277	0.6264
S&P 500 index: 2006 - 2016, number of observations: 2703											
	J	10	20	30	40	50	60	70	80	90	100
SV(1)	$\hat{\phi}_1$	1.0039	0.9888	0.9877	0.9863	0.9788	0.9758	0.9743	0.9697	0.9672	0.9666
	$\hat{\sigma}_y$	0.3548	0.3548	0.3548	0.3548	0.3548	0.3548	0.3548	0.3548	0.3548	0.3548
	$\hat{\sigma}_v$	1.0975	1.2290	1.1250	1.2759	1.2356	1.3744	1.3282	1.2798	1.3584	1.3879
SV(2)	$\hat{\phi}_1$	0.3869	0.2879	0.3418	0.2698	0.3030	0.3266	0.2916	0.2948	0.2980	0.2994
	$\hat{\phi}_2$	0.5298	0.6301	0.6144	0.6710	0.6550	0.6399	0.6708	0.6609	0.6577	0.6567
	$\hat{\sigma}_y$	0.3548	0.3548	0.3548	0.3548	0.3548	0.3548	0.3548	0.3548	0.3548	0.3548
	$\hat{\sigma}_v$	1.1228	1.1081	1.0958	1.0936	1.0892	1.0882	1.0853	1.0893	1.0898	1.0897
SV(3)	$\hat{\phi}_1$	0.1565	0.0551	0.1940	0.1381	0.1423	0.0972	0.0586	0.0649	0.0666	0.0672
	$\hat{\phi}_2$	0.6473	0.4297	0.3155	0.2576	0.2048	0.2277	0.2456	0.2421	0.2398	0.2378
	$\hat{\phi}_3$	0.1448	0.4745	0.4404	0.5396	0.6006	0.6568	0.6801	0.6705	0.6714	0.6725
	$\hat{\sigma}_y$	0.3548	0.3548	0.3548	0.3548	0.3548	0.3548	0.3548	0.3548	0.3548	0.3548
	$\hat{\sigma}_v$	1.0657	1.0279	1.0558	1.0505	1.0412	1.0123	1.0034	1.0089	1.0089	1.0091
SV(4)	$\hat{\phi}_1$	-0.2498	0.0402	0.0791	0.0767	-0.0319	0.0180	0.0492	0.0281	0.0283	0.0279
	$\hat{\phi}_2$	0.3377	0.2618	0.3093	0.2324	-0.0692	0.0571	0.0724	0.0887	0.0886	0.0875
	$\hat{\phi}_3$	0.2763	0.4629	0.3971	0.5099	0.4766	0.5435	0.5419	0.4830	0.4827	0.4832
	$\hat{\phi}_4$	0.5648	0.1741	0.1604	0.1381	0.5549	0.3215	0.2782	0.3387	0.3388	0.3399
	$\hat{\sigma}_y$	0.3548	0.3548	0.3548	0.3548	0.3548	0.3548	0.3548	0.3548	0.3548	0.3548
	$\hat{\sigma}_v$	1.0316	1.0447	1.0495	1.0363	1.0532	1.0425	1.0448	1.0488	1.0488	1.0488
S&P 500 index: 2006 - 2010, number of observations: 1259											
	J	10	20	30	40	50	60	70	80	90	100
SV(1)	$\hat{\phi}_1$	1.0107	0.9828	0.9912	0.9872	0.9852	0.9818	0.9795	0.9787	0.9773	0.9733
	$\hat{\sigma}_y$	0.4194	0.4194	0.4194	0.4194	0.4194	0.4194	0.4194	0.4194	0.4194	0.4194
	$\hat{\sigma}_v$	0.8886	1.1144	0.9922	1.1888	1.0528	1.2163	1.1949	1.1547	1.2106	1.3133
SV(2)	$\hat{\phi}_1$	0.2461	0.3429	0.3335	0.3157	0.3672	0.3654	0.3596	0.3937	0.3980	0.3739
	$\hat{\phi}_2$	0.7337	0.6123	0.6223	0.6543	0.6007	0.6218	0.6240	0.5877	0.5857	0.5810
	$\hat{\sigma}_y$	0.4194	0.4194	0.4194	0.4194	0.4194	0.4194	0.4194	0.4194	0.4194	0.4194
	$\hat{\sigma}_v$	0.9250	0.9550	0.9533	0.9409	0.9488	0.9346	0.9365	0.9424	0.9413	0.9589
SV(3)	$\hat{\phi}_1$	0.0058	0.4596	0.4604	0.4468	0.4845	0.4056	0.4055	0.4471	0.4352	0.2878
	$\hat{\phi}_2$	0.6765	-0.0689	-0.0420	-0.0255	-0.1603	-0.0541	-0.0559	-0.1146	-0.1798	-0.0495
	$\hat{\phi}_3$	0.3245	0.5134	0.4910	0.4958	0.5957	0.5986	0.6012	0.6210	0.7196	0.6936
	$\hat{\sigma}_y$	0.4194	0.4194	0.4194	0.4194	0.4194	0.4194	0.4194	0.4194	0.4194	0.4194
	$\hat{\sigma}_v$	0.8345	0.9509	0.9497	0.9419	0.9339	0.9009	0.9001	0.9013	0.8718	0.8885
SV(4)	$\hat{\phi}_1$	1.4611	0.5749	0.5764	0.5647	0.4380	0.4352	0.4402	0.4280	0.2058	0.2008
	$\hat{\phi}_2$	0.1345	-0.4170	-0.3963	-0.3766	-0.2509	-0.1999	-0.2068	-0.1842	-0.0734	-0.0629
	$\hat{\phi}_3$	-0.5136	0.3481	0.3195	0.3226	0.2848	0.3012	0.3017	0.2819	0.3499	0.3494
	$\hat{\phi}_4$	-0.1333	0.3796	0.3904	0.3809	0.4413	0.3799	0.3822	0.3936	0.4400	0.4332
	$\hat{\sigma}_y$	0.4194	0.4194	0.4194	0.4194	0.4194	0.4194	0.4194	0.4194	0.4194	0.4194
	$\hat{\sigma}_v$	1.1298	1.0109	1.0115	1.0083	0.9849	0.9780	0.9781	0.9778	0.9434	0.9437

Table 3.11. Summary statistics of full sample of experiment - 1

S&P 500 index, September 01, 2005 to August 31, 2010, $T = 1258$								
Series	Mean	Std. Dev.	Kurtosis	Skewness	Range	Max	Min	LB(10)
y	0.00	0.67	11.39	-0.17	8.83	4.63	-4.20	46.9
y^2	0.45	1.44	99.59	8.58	21.41	21.41	0.00	1117.5
y^*	0.00	2.67	4.80	-0.93	19.56	6.11	-13.45	473.3
RV5	0.00	0.00	108.51	8.05	0.01	0.01	0.00	4275.8
RV5-SS	0.00	0.00	108.51	8.05	0.01	0.01	0.00	4275.8
BV5	0.00	0.00	91.24	7.73	0.01	0.01	0.00	4362.1
BV5-SS	0.00	0.00	91.24	7.73	0.01	0.01	0.00	4362.1
MedRV	0.00	0.00	76.43	7.43	0.00	0.00	0.00	4457.5
TSRK	0.00	0.00	152.67	9.47	0.01	0.01	0.00	3729.8
RK	0.00	0.00	54.33	6.35	0.01	0.01	0.00	4273.5
RSV5	0.00	0.00	79.42	7.12	0.00	0.00	0.00	3706.0
RSV5-SS	0.00	0.00	79.42	7.12	0.00	0.00	0.00	3706.0
Log-RV5	-4.12	0.52	3.18	0.60	3.17	-2.11	-5.28	7569.5
Log-RV5-SS	-4.12	0.52	3.18	0.60	3.17	-2.11	-5.28	7569.5
Log-BV5	-4.21	0.52	3.23	0.62	3.35	-2.22	-5.57	7831.3
Log-BV5-SS	-4.21	0.52	3.23	0.62	3.35	-2.22	-5.57	7831.3
Log-MedRV	-4.51	0.56	3.22	0.54	3.45	-2.54	-5.99	7470.3
Log-TSRK	-4.14	0.51	3.31	0.67	3.19	-2.08	-5.27	8126.4
Log-RK	-4.20	0.57	3.09	0.40	3.44	-2.29	-5.73	5562.8
Log-RSV5	-4.47	0.57	3.02	0.48	3.48	-2.45	-5.93	6017.0
Log-RSV5-SS	-4.47	0.57	3.02	0.48	3.48	-2.45	-5.93	6017.0

Notes:

1. $y_t = r_t - \hat{\mu}_r$ is the residual return, y_t^2 is the squared of residual return and y_t^* is the residual of log squared of residual return.
2. RV5 is the 5-minute realized variance, BV5 is the 5-minute bi-power variation, RSV5 is the 5-minute realized semi-variance, RK is the realized kernel, TSRK is the two-scale realized kernels, and MedRV is the median realized volatility.
3. SS denotes the use of 1-minute subsamples in the calculation of realized volatility estimators.
4. LB(10) is the heteroskedasticity-corrected Ljung-Box statistics with 10 lags. The critical values for LB(10) are: 15.99 (10%), 18.31 (5%), and 23.21 (1%).

Table 3.12. Summary statistics of full sample of experiment - 2

S&P 500 index, January 01, 2005 to December 31, 2009, $T = 1259$

Series	Mean	Std. Dev.	Kurtosis	Skewness	Range	Max	Min	LB(10)
y	0.00	0.65	12.78	-0.18	8.83	4.62	-4.20	53.7
y^2	0.42	1.43	102.64	8.78	21.39	21.39	0.00	1181.7
y^*	0.00	2.69	6.45	-1.14	24.16	6.22	-17.94	466.2
RV5	0.00	0.00	111.55	8.21	0.01	0.01	0.00	4483.6
RV5-SS	0.00	0.00	111.55	8.21	0.01	0.01	0.00	4483.6
BV5	0.00	0.00	93.70	7.88	0.01	0.01	0.00	4547.5
BV5-SS	0.00	0.00	93.70	7.88	0.01	0.01	0.00	4547.5
MedRV	0.00	0.00	77.39	7.50	0.00	0.00	0.00	4571.3
TSRK	0.00	0.00	160.47	9.74	0.01	0.01	0.00	4028.3
RK	0.00	0.00	55.33	6.44	0.01	0.01	0.00	4425.0
RSV5	0.00	0.00	82.90	7.32	0.00	0.00	0.00	3957.9
RSV5-SS	0.00	0.00	82.90	7.32	0.00	0.00	0.00	3957.9
Log-RV5	-4.18	0.53	3.34	0.77	3.17	-2.11	-5.28	8036.3
Log-RV5-SS	-4.18	0.53	3.34	0.77	3.17	-2.11	-5.28	8036.3
Log-BV5	-4.27	0.53	3.40	0.80	3.35	-2.22	-5.57	8215.8
Log-BV5-SS	-4.27	0.53	3.40	0.80	3.35	-2.22	-5.57	8215.8
Log-MedRV	-4.57	0.56	3.37	0.71	3.45	-2.54	-5.99	7726.2
Log-TSRK	-4.20	0.51	3.46	0.85	3.19	-2.08	-5.27	8598.9
Log-RK	-4.25	0.57	3.28	0.53	3.53	-2.29	-5.82	5826.7
Log-RSV5	-4.53	0.57	3.17	0.65	3.48	-2.45	-5.93	6520.8
Log-RSV5-SS	-4.53	0.57	3.17	0.65	3.48	-2.45	-5.93	6520.8

Notes:

1. $y_t = r_t - \hat{\mu}_r$ is the residual return, y_t^2 is the squared of residual return and y_t^* is the residual of log squared of residual return.
2. RV5 is the 5-minute realized variance, BV5 is the 5-minute bi-power variation, RSV5 is the 5-minute realized semi-variance, RK is the realized kernel, TSRK is the two-scale realized kernels, and MedRV is the median realized volatility.
3. SS denotes the use of 1-minute subsamples in the calculation of realized volatility estimators.
4. LB(10) is the heteroskedasticity-corrected Ljung-Box statistics with 10 lags. The critical values for LB(10) are: 15.99 (10%), 18.31 (5%), and 23.21 (1%).

Table 3.13. Forecasting squared return: relative MSE and associated MCS p-value during moderate volatility regimes

	$1-day$			$2-day$			$1-week$			$2-week$			$3-week$			$1-month$		
S&P 500	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}
SV(1)	0.196	0.12	0.06	0.749	0.16	0.07	1.114	0.35	0.33	1.162	0.39	0.27	1.116	0.31	0.28	1.064	0.07	0.22
SV(2)	0.194	0.12	0.06	0.742	0.16	0.07	1.124	0.33	0.32	1.160	0.26	0.24	1.117	0.07	0.18	1.062	0.07	0.24
SV(3)	**0.180****	1.00	1.00	**0.676****	1.00	1.00	0.982*	0.65	0.83	1.026*	0.64	0.83	**0.985****	1.00	1.00	**0.967****	1.00	1.00
SV(3)*	0.185	0.12	0.07	0.696	0.16	0.08	1.014*	0.65	0.66	1.040*	0.64	0.67	0.998*	0.95	0.94	0.973**	0.97	0.95
GARCH(1,1)	0.197	0.12	0.06	0.750	0.16	0.07	1.104	0.35	0.34	1.102	0.49	0.40	1.056	0.39	0.52	1.005*	0.82	0.82
GARCH(1,2)	0.205	0.12	0.06	0.777	0.16	0.07	1.144	0.33	0.27	1.132	0.49	0.31	1.076	0.36	0.38	1.017*	0.63	0.72
GARCH(2,1)	0.197	0.12	0.06	0.750	0.16	0.07	1.104	0.35	0.37	1.102	0.49	0.37	1.056	0.36	0.38	1.005*	0.82	0.82
GARCH(2,2)	0.202	0.12	0.06	0.767	0.16	0.07	1.130	0.33	0.30	1.122	0.49	0.35	1.068	0.36	0.38	1.012*	0.82	0.82
GARCH(3,3)	0.204	0.12	0.06	0.775	0.16	0.07	1.143	0.33	0.28	1.131	0.49	0.33	1.075	0.36	0.38	1.015*	0.82	0.80
GJR(1,1)	0.214	0.12	0.05	0.811	0.13	0.07	1.193	0.23	0.21	1.173	0.26	0.21	1.111	0.31	0.33	1.043	0.38	0.51
GJR(2,2)	0.214	0.12	0.04	0.811	0.13	0.07	1.193	0.23	0.18	1.173	0.26	0.19	1.111	0.31	0.29	1.044	0.38	0.46
GJR(3,3)	0.214	0.06	0.03	0.813	0.07	0.07	1.196	0.07	0.13	1.175	0.07	0.16	1.113	0.07	0.24	1.046	0.07	0.39
EGARCH(1,1)	0.215	0.06	0.03	0.816	0.07	0.07	1.201	0.07	0.11	1.180	0.07	0.12	1.117	0.07	0.19	1.049	0.07	0.32
EGARCH(2,2)	0.215	0.06	0.03	0.815	0.07	0.07	1.198	0.07	0.12	1.179	0.07	0.13	1.117	0.07	0.20	1.049	0.07	0.29
EGARCH(3,3)	0.214	0.12	0.04	0.813	0.13	0.07	1.196	0.23	0.16	1.176	0.07	0.15	1.114	0.07	0.22	1.048	0.07	0.34
HAR-RV5	1.000	0.06	0.02	1.000	0.07	0.07	1.000*	0.65	0.77	1.000*	0.64	0.83	1.000*	0.95	0.94	1.000**	0.97	0.95
HAR-RV5-SS	1.000	0.06	0.02	1.000	0.07	0.07	1.000*	0.65	0.66	1.000*	0.64	0.83	1.000*	0.95	0.94	1.000**	0.97	0.95
HAR-BV5	1.114	0.06	0.02	1.097	0.07	0.07	1.018*	0.65	0.52	1.024*	0.64	0.67	1.038	0.39	0.43	1.047	0.38	0.50
HAR-BV5-SS	1.114	0.06	0.03	1.097	0.07	0.07	1.018*	0.65	0.46	1.024*	0.64	0.59	1.038	0.36	0.38	1.047*	0.54	0.58
HAR-MedRV	1.016	0.06	0.03	0.977	0.07	0.07	0.973*	0.65	0.83	0.999*	0.64	0.83	1.015*	0.54	0.76	1.035*	0.54	0.60
HAR-TSRK	1.135	0.06	0.02	1.111	0.07	0.07	1.043	0.44	0.42	1.028*	0.64	0.58	1.034*	0.54	0.60	1.046	0.07	0.39
HAR-RK	1.076	0.06	0.02	1.067	0.07	0.07	1.022*	0.65	0.46	1.058	0.49	0.43	1.038	0.36	0.38	1.026*	0.63	0.69
HAR-RSV5	0.941	0.06	0.02	0.960	0.07	0.07	0.973*	0.65	0.83	0.996*	0.64	0.83	0.990**	0.95	0.95	1.007*	0.82	0.82
HAR-RSV5-SS	0.941	0.06	0.02	0.960	0.16	0.07	**0.973****	1.00	1.00	**0.996****	1.00	1.00	0.990**	0.95	0.95	1.007*	0.82	0.82

Notes: The sample period is from September 01, 2005 to August 31, 2010 and the number of observations is $T = 1258$. The in-sample is from September 01, 2005 to August 31, 2008 ($T = 753$) and the out-of-sample is from September 01, 2008 to August 31, 2010 ($T = 505$). The in-sample include most volatile part of late-2000s financial crisis. HAR stands for Heterogenous Autoregressive model, RV5 is the 5-minute realized variance, BV5 is the 5-minute bi-power variation, RSV5 is the 5-minute realized semi-variance, RK is the realized kernel, TSRK is the two-scale realized kernels, and MedRV is the median realized volatility. SS denotes the use of 1-minute subsamples in the calculation of realized volatility estimators. The forecasts of SV(3)* model is based on the restricted W-ARMA-SV-OLS estimation. These are relative to the reference model HAR-RV5 and values smaller than unity indicate better forecast performance than the HAR-RV5 model. p^M_{MCS} and p^R_{MCS} are associated with $MCS\_T_{\max,\mathcal{M}} = \max_{i\in\mathcal{M}} t_{i,\cdot}$ and $MCS\_T_{R,\mathcal{M}} = \max_{i,j\in\mathcal{M}} \mid t_{i,j} \mid$, respectively. The forecasts in superior model sets $\hat{\mathcal{M}}^*_{5\%}$ and $\hat{\mathcal{M}}^*_{50\%}$ are defined by the average of $p_{MCS} \geq 0.95$ and the average of $p_{MCS} \geq 0.50$, respectively. The forecasts in $\hat{\mathcal{M}}^*_{5\%}$ and $\hat{\mathcal{M}}^*_{50\%}$ are identified by two and one asterisks, respectively. Boldface color font highlights the best model.

Table 3.14. Forecasting squared return: relative MAE and associated MCS p-value during moderate volatility regimes

	$1-day$			$2-day$			$1-week$			$2-week$			$3-week$			$1-month$		
S&P 500	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}
SV(1)	0.307	0.00	0.00	0.670	0.00	0.00	0.796	0.28	0.18	0.816	0.00	0.06	0.814	0.34	0.45	0.770	0.10	0.32
SV(2)	0.293	0.47	0.45	0.639*	0.81	0.75	0.787	0.41	0.33	0.799*	0.66	0.62	0.807*	0.65	0.64	0.762*	0.92	0.92
SV(3)	0.291*	0.60	0.60	**0.634****	1.00	1.00	**0.764****	1.00	1.00	0.784*	0.66	0.62	0.796*	0.65	0.64	0.779	0.10	0.32
SV(3)*	**0.290****	1.00	1.00	0.636*	0.81	0.75	0.767	0.47	0.47	**0.779****	1.00	1.00	**0.787****	1.00	1.00	**0.759****	1.00	1.00
GARCH(1,1)	0.372	0.00	0.00	0.816	0.00	0.00	0.989	0.00	0.00	1.000	0.00	0.00	1.013	0.00	0.00	0.957	0.00	0.00
GARCH(1,2)	0.354	0.00	0.00	0.777	0.00	0.00	0.945	0.00	0.00	0.943	0.00	0.00	0.949	0.00	0.00	0.893	0.00	0.00
GARCH(2,1)	0.372	0.00	0.00	0.816	0.00	0.00	0.989	0.00	0.00	1.000	0.00	0.00	1.013	0.00	0.00	0.957	0.00	0.00
GARCH(2,2)	0.351	0.00	0.00	0.772	0.00	0.00	0.940	0.00	0.00	0.941	0.00	0.00	0.951	0.00	0.00	0.893	0.00	0.00
GARCH(3,3)	0.354	0.00	0.00	0.777	0.00	0.00	0.946	0.00	0.00	0.945	0.00	0.00	0.950	0.00	0.00	0.893	0.00	0.00
GJR(1,1)	0.324	0.00	0.00	0.712	0.00	0.00	0.866	0.00	0.00	0.859	0.00	0.01	0.859	0.00	0.03	0.802	0.03	0.09
GJR(2,2)	0.324	0.00	0.00	0.713	0.00	0.00	0.867	0.00	0.00	0.860	0.00	0.00	0.859	0.00	0.01	0.803	0.03	0.05
GJR(3,3)	0.327	0.00	0.00	0.717	0.00	0.00	0.872	0.00	0.00	0.865	0.00	0.00	0.864	0.00	0.00	0.808	0.00	0.00
EGARCH(1,1)	0.326	0.00	0.00	0.717	0.00	0.00	0.872	0.00	0.00	0.866	0.00	0.00	0.866	0.00	0.00	0.810	0.00	0.00
EGARCH(2,2)	0.329	0.00	0.00	0.724	0.00	0.00	0.877	0.00	0.00	0.872	0.00	0.00	0.873	0.00	0.00	0.817	0.00	0.00
EGARCH(3,3)	0.327	0.00	0.00	0.720	0.00	0.00	0.876	0.00	0.00	0.869	0.00	0.00	0.871	0.00	0.00	0.816	0.00	0.00
HAR-RV5	1.000	0.00	0.00	1.000	0.00	0.00	1.000	0.00	0.00	1.000	0.00	0.00	1.000	0.00	0.00	1.000	0.00	0.00
HAR-RV5-SS	1.000	0.00	0.00	1.000	0.00	0.00	1.000	0.00	0.00	1.000	0.00	0.00	1.000	0.00	0.00	1.000	0.00	0.00
HAR-BV5	1.017	0.00	0.00	1.023	0.00	0.00	1.003	0.00	0.00	1.013	0.00	0.00	1.018	0.00	0.00	1.024	0.00	0.00
HAR-BV5-SS	1.017	0.00	0.00	1.023	0.00	0.00	1.003	0.00	0.00	1.013	0.00	0.00	1.018	0.00	0.00	1.024	0.00	0.00
HAR-MedRV	0.966	0.00	0.00	0.983	0.00	0.00	0.968	0.00	0.00	0.991	0.00	0.00	0.998	0.00	0.00	1.012	0.00	0.00
HAR-TSRK	1.032	0.00	0.00	1.027	0.00	0.00	1.018	0.00	0.00	1.018	0.00	0.00	1.024	0.00	0.00	1.027	0.00	0.00
HAR-RK	1.032	0.00	0.00	1.041	0.00	0.00	1.022	0.00	0.00	1.026	0.00	0.00	1.024	0.00	0.00	1.016	0.00	0.00
HAR-RSV5	0.970	0.00	0.00	0.979	0.00	0.00	0.981	0.00	0.00	0.990	0.00	0.00	0.994	0.00	0.00	1.009	0.00	0.00
HAR-RSV5-SS	0.970	0.00	0.00	0.979	0.00	0.00	0.981	0.00	0.00	0.990	0.00	0.00	0.994	0.00	0.00	1.009	0.00	0.00

Notes: The sample period is from September 01, 2005 to August 31, 2010 and the number of observations is $T = 1258$. The in-sample is from September 01, 2005 to August 31, 2008 ($T = 753$) and the out-of-sample is from September 01, 2008 to August 31, 2010 ($T = 505$). The in-sample include most volatile part of late-2000s financial crisis. HAR stands for Heterogenous Autoregressive model, RV5 is the 5-minute realized variance, BV5 is the 5-minute bi-power variation, RSV5 is the 5-minute realized semi-variance, RK is the realized kernel, TSRK is the two-scale realized kernels, and MedRV is the median realized volatility. SS denotes the use of 1-minute subsamples in the calculation of realized volatility estimators. The forecasts of SV(3)* model is based on the restricted W-ARMA-SV-OLS estimation. These are relative to the reference model HAR-RV5 and values smaller than unity indicate better forecast performance than the HAR-RV5 model. p^M_{MCS} and p^R_{MCS} are associated with $MCS\_T_{\max,\mathcal{M}} = \max_{i\in\mathcal{M}} t_{i,\cdot}$ and $MCS\_T_{R,\mathcal{M}} = \max_{i,j\in\mathcal{M}} \mid t_{i,j} \mid$, respectively. The forecasts in superior model sets $\hat{\mathcal{M}}^*_{5\%}$ and $\hat{\mathcal{M}}^*_{50\%}$ are defined by the average of $p_{MCS} \geq 0.95$ and the average of $p_{MCS} \geq 0.50$, respectively. The forecasts in $\hat{\mathcal{M}}^*_{5\%}$ and $\hat{\mathcal{M}}^*_{50\%}$ are identified by two and one asterisks, respectively. Boldface color font highlights the best model.

Table 3.15. Forecasting squared return: relative R2LOG and associated MCS p-value during moderate volatility regimes

	$1-day$			$2-day$			$1-week$			$2-week$			$3-week$			$1-month$		
S&P 500	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}
SV(1)	0.558	0.00	0.00	0.755	0.00	0.00	0.823	0.00	0.01	0.807	0.00	0.01	0.782	0.01	0.06	0.783	0.02	0.10
SV(2)	0.489	0.12	0.05	0.692	0.54	0.42	0.756	0.19	0.20	0.752	0.08	0.13	0.744	0.03	0.12	0.751*	0.74	0.74
SV(3)	0.477	0.12	0.09	0.683	0.54	0.42	0.742	0.19	0.20	0.734	0.08	0.13	0.737	0.03	0.12	0.757	0.02	0.10
SV(3)*	**0.476****	1.00	1.00	**0.682****	1.00	1.00	**0.739****	1.00	1.00	**0.726****	1.00	1.00	**0.724****	1.00	1.00	**0.739****	1.00	1.00
GARCH(1, 1)	0.796	0.00	0.00	1.151	0.00	0.00	1.231	0.00	0.00	1.212	0.00	0.00	1.174	0.00	0.00	1.133	0.00	0.00
GARCH(1, 2)	0.755	0.00	0.00	1.091	0.00	0.00	1.166	0.00	0.00	1.146	0.00	0.00	1.107	0.00	0.00	1.063	0.00	0.00
GARCH(2, 1)	0.796	0.00	0.00	1.151	0.00	0.00	1.231	0.00	0.00	1.212	0.00	0.00	1.174	0.00	0.00	1.133	0.00	0.00
GARCH(2, 2)	0.753	0.00	0.00	1.088	0.00	0.00	1.164	0.00	0.00	1.145	0.00	0.00	1.109	0.00	0.00	1.066	0.00	0.00
GARCH(3, 3)	0.757	0.00	0.00	1.093	0.00	0.00	1.169	0.00	0.00	1.149	0.00	0.00	1.110	0.00	0.00	1.064	0.00	0.00
GJR(1, 1)	0.662	0.00	0.00	0.957	0.00	0.00	1.022	0.00	0.00	1.000	0.00	0.00	0.962	0.00	0.00	0.914	0.01	0.01
GJR(2, 2)	0.663	0.00	0.00	0.959	0.00	0.00	1.024	0.00	0.00	1.002	0.00	0.00	0.964	0.00	0.00	0.916	0.01	0.00
GJR(3, 3)	0.672	0.00	0.00	0.971	0.00	0.00	1.036	0.00	0.00	1.014	0.00	0.00	0.974	0.00	0.00	0.926	0.00	0.00
EGARCH(1, 1)	0.673	0.00	0.00	0.974	0.00	0.00	1.041	0.00	0.00	1.020	0.00	0.00	0.981	0.00	0.00	0.934	0.00	0.00
EGARCH(2, 2)	0.679	0.00	0.00	0.988	0.00	0.00	1.051	0.00	0.00	1.032	0.00	0.00	0.994	0.00	0.00	0.950	0.00	0.00
EGARCH(3, 3)	9.811	0.00	0.00	14.473	0.00	0.00	15.225	0.00	0.00	15.150	0.00	0.00	14.703	0.00	0.00	14.289	0.00	0.00
HAR-RV5	1.000	0.00	0.00	1.000	0.00	0.00	1.000	0.00	0.00	1.000	0.00	0.00	1.000	0.00	0.00	1.000	0.00	0.00
HAR-RV5-SS	1.000	0.00	0.00	1.000	0.00	0.00	1.000	0.00	0.00	1.000	0.00	0.00	1.000	0.00	0.00	1.000	0.00	0.00
HAR-BV5	0.979	0.00	0.00	0.990	0.00	0.00	0.991	0.00	0.00	0.991	0.00	0.00	1.000	0.00	0.00	1.002	0.00	0.00
HAR-BV5-SS	0.979	0.00	0.00	0.990	0.00	0.00	0.991	0.00	0.00	0.991	0.00	0.00	1.000	0.00	0.00	1.002	0.00	0.00
HAR-MedRV	0.947	0.00	0.00	0.967	0.00	0.00	0.972	0.00	0.00	0.981	0.00	0.00	0.982	0.00	0.00	0.989	0.00	0.00
HAR-TSRK	0.983	0.00	0.00	0.994	0.00	0.00	1.001	0.00	0.00	1.003	0.00	0.00	1.004	0.00	0.00	1.006	0.00	0.00
HAR-RK	1.023	0.00	0.00	1.019	0.00	0.00	1.011	0.00	0.00	1.008	0.00	0.00	1.004	0.00	0.00	1.007	0.00	0.00
HAR-RSV5	0.958	0.00	0.00	0.976	0.00	0.00	0.994	0.00	0.00	1.001	0.00	0.00	1.000	0.00	0.00	1.005	0.00	0.00
HAR-RSV5-SS	0.958	0.00	0.00	0.976	0.00	0.00	0.994	0.00	0.00	1.001	0.00	0.00	1.000	0.00	0.00	1.005	0.00	0.00

Notes: The sample period is from September 01, 2005 to August 31, 2010 and the number of observations is $T = 1258$. The in-sample is from September 01, 2005 to August 31, 2008 ($T = 753$) and the out-of-sample is from September 01, 2008 to August 31, 2010 ($T = 505$). The in-sample include most volatile part of late-2000s financial crisis. HAR stands for Heterogenous Autoregressive model, RV5 is the 5-minute realized variance, BV5 is the 5-minute bi-power variation, RSV5 is the 5-minute realized semi-variance, RK is the realized kernel, TSRK is the two-scale realized kernels, and MedRV is the median realized volatility. SS denotes the use of 1-minute subsamples in the calculation of realized volatility estimators. The forecasts of SV(3)* model is based on the restricted W-ARMA-SV-OLS estimation. These are relative to the reference model HAR-RV5 and values smaller than unity indicate better forecast performance than the HAR-RV5 model. p^M_{MCS} and p^R_{MCS} are associated with $MCS\_T_{\max,\mathcal{M}} = \max_{i\in\mathcal{M}} t_{i,\cdot}$ and $MCS\_T_{R,\mathcal{M}} = \max_{i,j\in\mathcal{M}} \mid t_{i,j} \mid$, respectively. The forecasts in superior model sets $\hat{\mathcal{M}}^*_{5\%}$ and $\hat{\mathcal{M}}^*_{50\%}$ are defined by the average of $p_{MCS} \geq 0.95$ and the average of $p_{MCS} \geq 0.50$, respectively. The forecasts in $\hat{\mathcal{M}}^*_{5\%}$ and $\hat{\mathcal{M}}^*_{50\%}$ are identified by two and one asterisks, respectively. Boldface color font highlights the best model.

Table 3.16. Forecasting squared return: relative MSE and associated MCS p-value during high volatility regimes

	$1-day$			$2-day$			$1-week$			$2-week$			$3-week$			$1-month$		
S&P 500	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}
SV(1)	0.199	0.10	0.07	0.753	0.15	0.08	1.122	0.37	0.34	1.169	0.12	0.22	1.123	0.11	0.20	1.071	0.06	0.15
SV(2)	0.197	0.10	0.07	0.744	0.15	0.08	1.132	0.30	0.28	1.165	0.12	0.26	1.123	0.05	0.13	1.068	0.06	0.17
SV(3)	**0.181****	1.00	1.00	**0.677****	1.00	1.00	0.982*	0.67	0.74	1.027*	0.73	0.89	**0.974****	1.00	1.00	**0.956****	1.00	1.00
SV(3)*	0.186	0.13	0.07	0.695	0.15	0.08	1.017*	0.67	0.69	1.043*	0.73	0.63	1.002*	0.67	0.78	0.982*	0.68	0.72
GARCH(1,1)	0.197	0.13	0.07	0.743	0.15	0.08	1.097	0.43	0.38	1.096*	0.67	0.54	1.051	0.49	0.39	1.003*	0.47	0.72
GARCH(1,2)	0.206	0.10	0.07	0.773	0.15	0.07	1.142	0.30	0.25	1.129	0.42	0.34	1.075	0.49	0.38	1.018*	0.42	0.63
GARCH(2,1)	0.197	0.13	0.07	0.743	0.15	0.08	1.097	0.43	0.37	1.096	0.48	0.44	1.051*	0.54	0.53	1.003*	0.47	0.72
GARCH(2,2)	0.203	0.10	0.07	0.763	0.15	0.08	1.128	0.37	0.32	1.121	0.48	0.40	1.067	0.49	0.39	1.013*	0.47	0.72
GARCH(3,3)	0.205	0.10	0.07	0.771	0.15	0.08	1.140	0.30	0.27	1.125	0.48	0.38	1.070	0.49	0.39	1.014*	0.47	0.72
GJR(1,1)	0.215	0.10	0.05	0.811	0.10	0.06	1.196	0.12	0.19	1.176	0.12	0.18	1.115	0.13	0.30	1.049	0.15	0.43
GJR(2,2)	0.216	0.10	0.04	0.811	0.10	0.06	1.197	0.12	0.15	1.177	0.12	0.16	1.115	0.13	0.25	1.049	0.15	0.37
GJR(3,3)	0.216	0.10	0.03	0.813	0.10	0.06	1.199	0.12	0.12	1.179	0.12	0.14	1.117	0.11	0.21	1.051	0.13	0.31
EGARCH(1,1)	0.217	0.10	0.03	0.817	0.10	0.06	1.205	0.12	0.07	1.184	0.09	0.09	1.123	0.05	0.14	1.055	0.06	0.21
EGARCH(2,2)	0.217	0.10	0.03	0.815	0.10	0.06	1.202	0.12	0.09	1.182	0.09	0.10	1.121	0.05	0.15	1.054	0.06	0.23
EGARCH(3,3)	0.216	0.10	0.03	0.813	0.10	0.06	1.201	0.12	0.10	1.180	0.12	0.12	1.119	0.11	0.18	1.053	0.13	0.27
HAR-RV5	1.000	0.10	0.02	1.000	0.10	0.06	1.000*	0.67	0.69	1.000*	0.73	0.89	1.000*	0.67	0.78	1.000*	0.68	0.72
HAR-RV5-SS	1.000	0.10	0.02	1.000	0.10	0.06	1.000*	0.67	0.74	1.000*	0.80	0.92	1.000*	0.54	0.71	1.000*	0.47	0.72
HAR-BV5	1.113	0.10	0.02	1.097	0.10	0.06	1.017*	0.67	0.55	1.024*	0.73	0.69	1.037	0.49	0.39	1.043	0.42	0.55
HAR-BV5-SS	1.113	0.10	0.02	1.097	0.10	0.06	1.017*	0.67	0.46	1.024*	0.73	0.63	1.037	0.49	0.39	1.043	0.42	0.54
HAR-MedRV	1.013	0.10	0.02	0.975	0.10	0.06	**0.972****	1.00	1.00	0.999*	0.80	0.92	1.015*	0.54	0.59	1.031	0.42	0.56
HAR-TSRK	1.140	0.10	0.02	1.114	0.10	0.06	1.044	0.44	0.41	1.029*	0.73	0.60	1.034*	0.54	0.46	1.043	0.15	0.43
HAR-RK	1.073	0.10	0.02	1.065	0.10	0.06	1.022*	0.67	0.46	1.057	0.48	0.44	1.036	0.49	0.39	1.024*	0.42	0.61
HAR-RSV5	0.947	0.10	0.02	0.963	0.10	0.06	0.974*	0.67	0.74	0.996*	0.80	0.92	0.990*	0.67	0.78	1.006*	0.47	0.72
HAR-RSV5-SS	0.947	0.10	0.02	0.963	0.15	0.07	0.974**	0.95	0.95	**0.996****	1.00	1.00	0.990*	0.82	0.82	1.006*	0.47	0.72

Notes: The sample period is from January 01, 2005 to December 31, 2009 and the number of observations is $T = 1259$. The in-sample is from January 01, 2005 to December 31, 2007 ($T = 754$) and the out-of-sample is from January 01, 2008 to December 31, 2009 ($T = 505$). In this setting, we forecast a highly volatile period. HAR stands for Heterogenous Autoregressive model, RV5 is the 5-minute realized variance, BV5 is the 5-minute bi-power variation, RSV5 is the 5-minute realized semi-variance, RK is the realized kernel, TSRK is the two-scale realized kernels, and MedRV is the median realized volatility. SS denotes the use of 1-minute subsamples in the calculation of realized volatility estimators. The forecasts of SV(3)* model is based on the restricted W-ARMA-SV-OLS estimation. These are relative to the reference model HAR-RV5 and values smaller than unity indicate better forecast performance than the HAR-RV5 model. p^M_{MCS} and p^R_{MCS} are associated with $MCS\_T_{\max,\mathcal{M}} = \max_{i\in\mathcal{M}} t_{i,\cdot}$ and $MCS\_T_{R,\mathcal{M}} = \max_{i,j\in\mathcal{M}} \mid t_{i,j} \mid$, respectively. The forecasts in superior model sets $\hat{\mathcal{M}}^*_{5\%}$ and $\hat{\mathcal{M}}^*_{50\%}$ are defined by the average of $p_{MCS} \geq 0.95$ and the average of $p_{MCS} \geq 0.50$, respectively. The forecasts in $\hat{\mathcal{M}}^*_{5\%}$ and $\hat{\mathcal{M}}^*_{50\%}$ are identified by two and one asterisks, respectively. Boldface color font highlights the best model.

Table 3.17. Forecasting squared return: relative MAE and associated MCS p-value during high volatility regimes

	$1-day$			$2-day$			$1-week$			$2-week$			$3-week$			$1-month$		
S&P 500	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}
SV(1)	0.313	0.00	0.00	0.661	0.00	0.00	0.797	0.23	0.15	0.825	0.00	0.05	0.831	0.24	0.37	0.801	0.02	0.42
SV(2)	0.300	0.26	0.25	0.632*	0.72	0.67	0.789	0.31	0.24	0.807*	0.67	0.62	0.824*	0.65	0.66	0.792*	0.80	0.80
SV(3)	**0.297****	1.00	1.00	**0.626****	1.00	1.00	**0.763****	1.00	1.00	0.792*	0.67	0.62	0.803*	0.81	0.81	0.799*	0.62	0.72
SV(3)*	0.297*	0.70	0.70	0.627*	0.77	0.77	0.765	0.41	0.41	**0.787****	1.00	1.00	**0.800****	1.00	1.00	**0.784****	1.00	1.00
GARCH(1,1)	0.339	0.00	0.00	0.719	0.00	0.00	0.881	0.00	0.01	0.903	0.00	0.01	0.924	0.00	0.01	0.882	0.00	0.01
GARCH(1,2)	0.330	0.00	0.00	0.701	0.00	0.00	0.862	0.00	0.01	0.871	0.00	0.01	0.884	0.00	0.01	0.842	0.00	0.01
GARCH(2,1)	0.339	0.00	0.00	0.719	0.00	0.00	0.881	0.00	0.01	0.903	0.00	0.01	0.924	0.00	0.01	0.882	0.00	0.01
GARCH(2,2)	0.328	0.00	0.00	0.697	0.00	0.00	0.858	0.00	0.01	0.870	0.00	0.01	0.886	0.00	0.01	0.843	0.00	0.01
GARCH(3,3)	0.329	0.00	0.00	0.698	0.00	0.00	0.858	0.00	0.01	0.868	0.00	0.01	0.881	0.00	0.01	0.838	0.00	0.01
GJR(1,1)	0.319	0.00	0.00	0.677	0.00	0.00	0.832	0.04	0.04	0.835	0.00	0.05	0.843	0.21	0.24	0.801	0.02	0.42
GJR(2,2)	0.319	0.03	0.02	0.678	0.01	0.02	0.832	0.04	0.03	0.835	0.00	0.05	0.843	0.21	0.20	0.801	0.02	0.42
GJR(3,3)	0.320	0.00	0.00	0.679	0.00	0.00	0.834	0.00	0.01	0.837	0.00	0.02	0.845	0.00	0.06	0.802	0.00	0.16
EGARCH(1,1)	0.321	0.00	0.00	0.682	0.00	0.00	0.838	0.00	0.01	0.841	0.00	0.01	0.850	0.00	0.01	0.807	0.00	0.03
EGARCH(2,2)	0.323	0.00	0.00	0.685	0.00	0.00	0.841	0.00	0.01	0.844	0.00	0.01	0.852	0.00	0.01	0.810	0.00	0.01
EGARCH(3,3)	0.322	0.00	0.00	0.683	0.00	0.00	0.839	0.00	0.01	0.842	0.00	0.01	0.850	0.00	0.01	0.808	0.00	0.01
HAR-RV5	1.000	0.00	0.00	1.000	0.00	0.00	1.000	0.00	0.01	1.000	0.00	0.01	1.000	0.00	0.01	1.000	0.00	0.01
HAR-RV5-SS	1.000	0.00	0.00	1.000	0.00	0.00	1.000	0.00	0.01	1.000	0.00	0.01	1.000	0.00	0.01	1.000	0.00	0.01
HAR-BV5	1.022	0.00	0.00	1.025	0.00	0.00	1.004	0.00	0.01	1.015	0.00	0.01	1.019	0.00	0.01	1.022	0.00	0.01
HAR-BV5-SS	1.022	0.00	0.00	1.025	0.00	0.00	1.004	0.00	0.01	1.015	0.00	0.01	1.019	0.00	0.01	1.022	0.00	0.01
HAR-MedRV	0.971	0.00	0.00	0.985	0.00	0.00	0.969	0.00	0.01	0.991	0.00	0.01	0.999	0.00	0.01	1.009	0.00	0.01
HAR-TSRK	1.042	0.00	0.00	1.033	0.00	0.00	1.021	0.00	0.01	1.022	0.00	0.01	1.028	0.00	0.01	1.028	0.00	0.01
HAR-RK	1.021	0.00	0.00	1.033	0.00	0.00	1.018	0.00	0.01	1.023	0.00	0.01	1.019	0.00	0.01	1.011	0.00	0.01
HAR-RSV5	0.979	0.00	0.00	0.989	0.00	0.00	0.987	0.00	0.01	0.994	0.00	0.01	0.997	0.00	0.01	1.010	0.00	0.01
HAR-RSV5-SS	0.979	0.00	0.00	0.989	0.00	0.00	0.987	0.00	0.01	0.994	0.00	0.01	0.997	0.00	0.01	1.010	0.00	0.01

Notes: The sample period is from January 01, 2005 to December 31, 2009 and the number of observations is $T = 1259$. The in-sample is from January 01, 2005 to December 31, 2007 ($T = 754$) and the out-of-sample is from January 01, 2008 to December 31, 2009 ($T = 505$). In this setting, we forecast a highly volatile period. HAR stands for Heterogenous Autoregressive model, RV5 is the 5-minute realized variance, BV5 is the 5-minute bi-power variation, RSV5 is the 5-minute realized semi-variance, RK is the realized kernel, TSRK is the two-scale realized kernels, and MedRV is the median realized volatility. SS denotes the use of 1-minute subsamples in the calculation of realized volatility estimators. The forecasts of SV(3)* model is based on the restricted W-ARMA-SV-OLS estimation. These are relative to the reference model HAR-RV5 and values smaller than unity indicate better forecast performance than the HAR-RV5 model. p^M_{MCS} and p^R_{MCS} are associated with $MCS\_T_{\max,\mathcal{M}} = \max_{i\in\mathcal{M}} t_{i,\cdot}$ and $MCS\_T_{R,\mathcal{M}} = \max_{i,j\in\mathcal{M}} \mid t_{i,j} \mid$, respectively. The forecasts in superior model sets $\hat{\mathcal{M}}^*_{5\%}$ and $\hat{\mathcal{M}}^*_{50\%}$ are defined by the average of $p_{MCS} \geq 0.95$ and the average of $p_{MCS} \geq 0.50$, respectively. The forecasts in $\hat{\mathcal{M}}^*_{5\%}$ and $\hat{\mathcal{M}}^*_{50\%}$ are identified by two and one asterisks, respectively. Boldface color font highlights the best model.

Table 3.18. Forecasting squared return: relative R2LOG and associated MCS p-value during high volatility regimes

	$1-day$			$2-day$			$1-week$			$2-week$			$3-week$			$1-month$		
S&P 500	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}
SV(1)	0.526	0.00	0.00	0.710	0.00	0.00	0.797	0.00	0.02	0.828	0.00	0.00	0.809	0.00	0.02	0.844	0.00	0.01
SV(2)	0.469	0.30	0.23	0.657*	0.47	0.60	0.738	0.30	0.38	0.760	0.23	0.16	0.754	0.24	0.23	0.782	0.32	0.20
SV(3)	0.462	0.38	0.38	0.655*	0.47	0.60	0.728	0.30	0.38	0.720	0.26	0.26	0.703*	0.60	0.60	0.730	0.32	0.20
SV(3)*	**0.461****	1.00	1.00	**0.654****	1.00	1.00	**0.725****	1.00	1.00	**0.717****	1.00	1.00	**0.700****	1.00	1.00	**0.725****	1.00	1.00
GARCH(1,1)	0.641	0.00	0.00	0.918	0.00	0.00	1.016	0.00	0.00	1.025	0.00	0.00	1.002	0.00	0.00	0.983	0.00	0.00
GARCH(1,2)	0.615	0.00	0.00	0.882	0.00	0.00	0.975	0.00	0.00	0.979	0.00	0.00	0.952	0.00	0.00	0.928	0.00	0.00
GARCH(2,1)	0.641	0.00	0.00	0.918	0.00	0.00	1.016	0.00	0.00	1.025	0.00	0.00	1.002	0.00	0.00	0.983	0.00	0.00
GARCH(2,2)	0.614	0.00	0.00	0.879	0.00	0.00	0.973	0.00	0.00	0.979	0.00	0.00	0.955	0.00	0.00	0.932	0.00	0.00
GARCH(3,3)	0.612	0.00	0.00	0.878	0.00	0.00	0.971	0.00	0.00	0.975	0.00	0.00	0.948	0.00	0.00	0.923	0.00	0.00
GJR(1,1)	0.564	0.00	0.00	0.809	0.00	0.00	0.893	0.00	0.00	0.892	0.00	0.00	0.863	0.00	0.00	0.834	0.05	0.05
GJR(2,2)	0.565	0.00	0.00	0.811	0.00	0.00	0.895	0.00	0.00	0.893	0.00	0.00	0.865	0.00	0.00	0.835	0.00	0.01
GJR(3,3)	0.569	0.00	0.00	0.816	0.00	0.00	0.900	0.00	0.00	0.898	0.00	0.00	0.869	0.00	0.00	0.839	0.00	0.01
EGARCH(1,1)	0.576	0.00	0.00	0.826	0.00	0.00	0.914	0.00	0.00	0.912	0.00	0.00	0.885	0.00	0.00	0.855	0.00	0.00
EGARCH(2,2)	0.582	0.00	0.00	0.833	0.00	0.00	0.921	0.00	0.00	0.918	0.00	0.00	0.893	0.00	0.00	0.861	0.00	0.00
EGARCH(3,3)	0.577	0.00	0.00	0.828	0.00	0.00	0.914	0.00	0.00	0.913	0.00	0.00	0.886	0.00	0.00	0.856	0.00	0.00
HAR-RV5	1.000	0.00	0.00	1.000	0.00	0.00	1.000	0.00	0.00	1.000	0.00	0.00	1.000	0.00	0.00	1.000	0.00	0.00
HAR-RV5-SS	1.000	0.00	0.00	1.000	0.00	0.00	1.000	0.00	0.00	1.000	0.00	0.00	1.000	0.00	0.00	1.000	0.00	0.00
HAR-BV5	0.988	0.00	0.00	0.995	0.00	0.00	0.997	0.00	0.00	0.999	0.00	0.00	1.004	0.00	0.00	1.002	0.00	0.00
HAR-BV5-SS	0.988	0.00	0.00	0.995	0.00	0.00	0.997	0.00	0.00	0.999	0.00	0.00	1.004	0.00	0.00	1.002	0.00	0.00
HAR-MedRV	0.951	0.00	0.00	0.966	0.00	0.00	0.972	0.00	0.00	0.984	0.00	0.00	0.987	0.00	0.00	0.986	0.00	0.00
HAR-TSRK	0.995	0.00	0.00	1.001	0.00	0.00	1.010	0.00	0.00	1.011	0.00	0.00	1.012	0.00	0.00	1.012	0.00	0.00
HAR-RK	1.018	0.00	0.00	1.012	0.00	0.00	1.004	0.00	0.00	1.005	0.00	0.00	0.999	0.00	0.00	1.000	0.00	0.00
HAR-RSV5	0.973	0.00	0.00	0.988	0.00	0.00	1.003	0.00	0.00	1.006	0.00	0.00	1.005	0.00	0.00	1.008	0.00	0.00
HAR-RSV5-SS	0.973	0.00	0.00	0.988	0.00	0.00	1.003	0.00	0.00	1.006	0.00	0.00	1.005	0.00	0.00	1.008	0.00	0.00

Notes: The sample period is from January 01, 2005 to December 31, 2009 and the number of observations is $T = 1259$. The in-sample is from January 01, 2005 to December 31, 2007 ($T = 754$) and the out-of-sample is from January 01, 2008 to December 31, 2009 ($T = 505$). In this setting, we forecast a highly volatile period. HAR stands for Heterogenous Autoregressive model, RV5 is the 5-minute realized variance, BV5 is the 5-minute bi-power variation, RSV5 is the 5-minute realized semi-variance, RK is the realized kernel, TSRK is the two-scale realized kernels, and MedRV is the median realized volatility. SS denotes the use of 1-minute subsamples in the calculation of realized volatility estimators. The forecasts of SV(3)* model is based on the restricted W-ARMA-SV-OLS estimation. These are relative to the reference model HAR-RV5 and values smaller than unity indicate better forecast performance than the HAR-RV5 model. p^M_{MCS} and p^R_{MCS} are associated with $MCS\_T_{\max,\mathcal{M}} = \max_{i\in\mathcal{M}} t_{i,\cdot}$ and $MCS\_T_{R,\mathcal{M}} = \max_{i,j\in\mathcal{M}} \mid t_{i,j} \mid$, respectively. The forecasts in superior model sets $\hat{\mathcal{M}}^*_{5\%}$ and $\hat{\mathcal{M}}^*_{50\%}$ are defined by the average of $p_{MCS} \geq 0.95$ and the average of $p_{MCS} \geq 0.50$, respectively. The forecasts in $\hat{\mathcal{M}}^*_{5\%}$ and $\hat{\mathcal{M}}^*_{50\%}$ are identified by two and one asterisks, respectively. Boldface color font highlights the best model.

Table 3.19. Forecasting realized volatility: relative MSE and associated MCS p-value during moderate volatility regimes

	$1-day$			$2-day$			$1-week$			$2-week$			$3-week$			$1-month$		
S&P 500	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}
HAR-RV	1.000**	0.95	0.95	1.000*	0.69	0.69	1.000*	0.81	0.84	1.000	0.24	0.31	1.000	0.21	0.23	1.000	0.15	0.13
SV(1)	**0.991****	1.00	1.00	**0.972****	1.00	1.00	0.986*	0.81	0.84	0.969	0.24	0.31	0.949	0.21	0.23	0.893	0.15	0.13
SV(2)	0.996**	0.95	0.95	0.986*	0.69	0.69	0.981*	0.81	0.84	0.963	0.24	0.31	0.944	0.21	0.23	0.877	0.15	0.13
SV(3)	1.087	0.16	0.17	1.059	0.24	0.32	**0.972****	1.00	1.00	**0.931****	1.00	1.00	**0.881****	1.00	1.00	**0.833****	1.00	1.00
FTSE100	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}
HAR-RV	1.000*	0.54	0.54	1.000	0.46	0.44	1.000*	0.91	0.92	1.000*	0.89	0.83	1.000	0.37	0.41	1.000	0.27	0.13
SV(1)	1.004*	0.54	0.54	0.943*	0.89	0.89	1.009	0.41	0.66	**0.975****	1.00	1.00	**0.953****	1.00	1.00	0.950	0.27	0.14
SV(2)	**0.975****	1.00	1.00	**0.941****	1.00	1.00	**0.996****	1.00	1.00	0.981*	0.89	0.83	0.955*	0.85	0.78	0.947	0.27	0.26
SV(3)	1.046	0.27	0.21	1.019	0.46	0.47	1.001*	0.91	0.92	0.982*	0.89	0.83	0.958*	0.85	0.78	**0.943****	1.00	1.00
NASDAQ100	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}
HAR-RV	1.000*	0.82	0.82	1.000*	0.93	0.95	1.000*	0.69	0.68	1.000	0.19	0.36	1.000	0.38	0.53	1.000	0.18	0.26
SV(1)	1.018*	0.80	0.79	1.001*	0.93	0.95	0.989*	0.69	0.68	1.020	0.12	0.16	1.021	0.14	0.22	0.984	0.09	0.14
SV(2)	**0.987****	1.00	1.00	**0.996****	1.00	1.00	**0.976****	1.00	1.00	0.985	0.19	0.36	0.976	0.38	0.53	0.943	0.18	0.26
SV(3)	1.087	0.29	0.25	1.070*	0.60	0.64	1.011*	0.69	0.68	**0.970****	1.00	1.00	**0.965****	1.00	1.00	**0.920****	1.00	1.00
N225	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}
HAR-RV	1.000	0.38	0.32	1.000	0.16	0.15	1.000**	0.99	0.99	1.000	0.30	0.15	1.000*	0.76	0.73	1.000*	0.78	0.70
SV(1)	0.943*	0.67	0.67	0.956	0.16	0.15	1.060	0.26	0.53	1.061	0.30	0.13	1.051	0.23	0.26	1.024	0.21	0.29
SV(2)	**0.937****	1.00	1.00	**0.922****	1.00	1.00	0.998**	0.99	0.99	0.940	0.30	0.15	0.973*	0.94	0.94	0.980*	0.78	0.70
SV(3)	1.203	0.32	0.19	1.084	0.16	0.15	**0.996****	1.00	1.00	**0.907****	1.00	1.00	**0.973****	1.00	1.00	**0.975****	1.00	1.00
SSMI20	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}
HAR-RV	1.000*	0.65	0.60	1.000*	0.80	0.77	**1.000****	1.00	1.00	1.000*	0.66	0.55	1.000	0.37	0.27	1.000	0.14	0.12
SV(1)	0.968*	0.65	0.60	**0.978****	1.00	1.00	1.003**	0.98	0.97	0.984*	0.66	0.55	0.958	0.37	0.27	0.943	0.14	0.12
SV(2)	**0.963****	1.00	1.00	0.990*	0.80	0.77	1.006**	0.98	0.97	0.974*	0.66	0.55	0.948	0.37	0.27	0.924	0.26	0.26
SV(3)	1.128	0.29	0.24	1.127	0.43	0.46	1.011**	0.98	0.97	**0.961****	1.00	1.00	**0.932****	1.00	1.00	**0.912****	1.00	1.00

Notes: The sample period is from September 01, 2005 to August 31, 2010 where the in-sample is from September 01, 2005 to August 31, 2008 and the out-of-sample is from September 01, 2008 to August 31, 2010. The in-sample include the most volatile part of the late-2000s financial crisis. HAR stands for the Heterogenous Autoregressive model, and we used 5-minute RV following the results of Liu et al. (2015). These are relative to the reference model HAR-RV and values smaller than unity indicate better forecast performance than HAR-RV model. p^M_{MCS} and p^R_{MCS} are associated with $MCS\_T_{\max,\mathcal{M}} = \max_{i\in\mathcal{M}} t_{i,\cdot}$ and $MCS\_T_{R,\mathcal{M}} = \max_{i,j\in\mathcal{M}} \mid t_{i,j} \mid$, respectively. The forecasts in superior model sets $\hat{\mathcal{M}}^*_{5\%}$ and $\hat{\mathcal{M}}^*_{50\%}$ are defined by the average of $p_{MCS} \geq 0.95$ and the average of $p_{MCS} \geq 0.50$, respectively. The forecasts in $\hat{\mathcal{M}}^*_{5\%}$ and $\hat{\mathcal{M}}^*_{50\%}$ are identified by two and one asterisks, respectively. Boldface color font highlights the best model.

Table 3.20. Forecasting realized volatility: relative MAE and associated MCS p-value during moderate volatility regimes

	$1-day$			$2-day$			$1-week$			$2-week$			$3-week$			$1-month$		
S&P 500	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}
HAR-RV	1.000	0.01	0.01	1.000	0.10	0.06	1.000	0.18	0.13	1.000	0.15	0.17	1.000	0.05	0.05	1.000	0.05	0.06
SV(1)	0.936*	0.89	0.89	0.950	0.48	0.48	0.916	0.29	0.26	0.849*	0.85	0.79	0.808	0.17	0.18	0.775	0.21	0.23
SV(2)	**0.934****	1.00	1.00	**0.944****	1.00	1.00	0.905	0.36	0.36	0.847*	0.85	0.79	0.807	0.15	0.18	0.771	0.18	0.23
SV(3)	0.968	0.04	0.08	0.983	0.10	0.07	**0.895****	1.00	1.00	**0.843****	1.00	1.00	**0.785****	1.00	1.00	**0.745****	1.00	1.00
FTSE100	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}
HAR-RV	1.000	0.22	0.29	1.000	0.12	0.12	1.000	0.05	0.02	1.000	0.05	0.03	1.000	0.01	0.00	1.000	0.01	0.00
SV(1)	0.986	0.44	0.44	0.987	0.12	0.16	0.960	0.05	0.04	0.862	0.29	0.20	0.826	0.04	0.02	0.789	0.04	0.04
SV(2)	**0.971****	1.00	1.00	**0.950****	1.00	1.00	0.928	0.25	0.25	0.854	0.31	0.31	0.820	0.04	0.03	0.783	0.04	0.04
SV(3)	1.015	0.17	0.23	0.952*	0.92	0.92	**0.902****	1.00	1.00	**0.848****	1.00	1.00	**0.807****	1.00	1.00	**0.771****	1.00	1.00
NASDAQ100	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}
HAR-RV	1.000	0.32	0.39	1.000	0.10	0.11	1.000	0.22	0.12	1.000	0.10	0.04	1.000	0.13	0.08	1.000	0.11	0.05
SV(1)	**0.971****	1.00	1.00	0.966	0.12	0.17	0.944	0.27	0.19	0.883	0.11	0.07	0.866	0.17	0.18	0.859	0.11	0.05
SV(2)	0.976*	0.80	0.80	**0.943****	1.00	1.00	0.932	0.30	0.30	0.868	0.18	0.18	0.848*	0.72	0.72	0.832	0.11	0.10
SV(3)	1.003	0.18	0.24	0.965	0.23	0.23	**0.920****	1.00	1.00	**0.860****	1.00	1.00	**0.847****	1.00	1.00	**0.822****	1.00	1.00
N225	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}
HAR-RV	1.000	0.06	0.03	1.000	0.01	0.01	1.000	0.05	0.09	1.000	0.09	0.06	1.000	0.13	0.05	1.000	0.09	0.05
SV(1)	0.968	0.07	0.07	0.952	0.02	0.02	0.937	0.05	0.09	0.935	0.09	0.08	0.915	0.13	0.07	0.898	0.11	0.18
SV(2)	**0.950****	1.00	1.00	**0.927****	1.00	1.00	**0.915****	1.00	1.00	0.880	0.32	0.32	0.868*	0.50	0.50	0.865**	0.97	0.97
SV(3)	1.060	0.06	0.03	1.003	0.02	0.02	0.944	0.05	0.09	**0.866****	1.00	1.00	**0.860****	1.00	1.00	**0.865****	1.00	1.00
SSMI20	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}
HAR-RV	1.000*	0.49	0.54	1.000*	0.56	0.66	1.000	0.17	0.13	1.000	0.01	0.01	1.000	0.04	0.01	1.000	0.04	0.02
SV(1)	0.995*	0.49	0.54	0.990*	0.56	0.66	0.975	0.17	0.13	0.890	0.02	0.02	0.861	0.05	0.05	0.831	0.13	0.10
SV(2)	**0.987****	1.00	1.00	0.984**	0.98	0.98	0.974	0.17	0.13	0.891	0.01	0.01	0.868	0.04	0.03	0.824	0.13	0.10
SV(3)	1.025	0.40	0.37	**0.984****	1.00	1.00	**0.934****	1.00	1.00	**0.862****	1.00	1.00	**0.839****	1.00	1.00	**0.813****	1.00	1.00

Notes: The sample period is from September 01, 2005 to August 31, 2010 where the in-sample is from September 01, 2005 to August 31, 2008 and the out-of-sample is from September 01, 2008 to August 31, 2010. The in-sample include the most volatile part of the late-2000s financial crisis. HAR stands for the Heterogenous Autoregressive model, and we used 5-minute RV following the results of Liu et al. (2015). These are relative to the reference model HAR-RV and values smaller than unity indicate better forecast performance than HAR-RV model. p^M_{MCS} and p^R_{MCS} are associated with $MCS\_T_{\max,\mathcal{M}} = \max_{i\in\mathcal{M}} t_{i,\cdot}$ and $MCS\_T_{R,\mathcal{M}} = \max_{i,j\in\mathcal{M}} \mid t_{i,j} \mid$, respectively. The forecasts in superior model sets $\hat{\mathcal{M}}^*_{5\%}$ and $\hat{\mathcal{M}}^*_{50\%}$ are defined by the average of $p_{MCS} \geq 0.95$ and the average of $p_{MCS} \geq 0.50$, respectively. The forecasts in $\hat{\mathcal{M}}^*_{5\%}$ and $\hat{\mathcal{M}}^*_{50\%}$ are identified by two and one asterisks, respectively. Boldface color font highlights the best model.

Table 3.21. Forecasting realized volatility: relative R2LOG and associated MCS p-value during moderate volatility regimes

	$1-day$			$2-day$			$1-week$			$2-week$			$3-week$			$1-month$		
S&P 500	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}
HAR-RV	1.000	0.01	0.02	1.000	0.07	0.05	1.000	0.07	0.08	1.000	0.27	0.22	1.000	0.16	0.11	1.000	0.07	0.07
SV(1)	0.921	0.46	0.46	0.933	0.09	0.09	0.939	0.07	0.09	0.887	0.51	0.37	0.853	0.16	0.19	0.782	0.24	0.27
SV(2)	**0.906****	1.00	1.00	**0.915****	1.00	1.00	0.909	0.39	0.39	0.878	0.51	0.39	0.852	0.16	0.19	0.783	0.23	0.27
SV(3)	0.990	0.04	0.09	0.973	0.07	0.07	**0.896****	1.00	1.00	**0.861****	1.00	1.00	**0.809****	1.00	1.00	**0.743****	1.00	1.00
FTSE100	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}
HAR-RV	1.000	0.02	0.02	1.000	0.00	0.01	1.000	0.00	0.01	1.000	0.04	0.02	1.000	0.02	0.00	1.000	0.04	0.01
SV(1)	0.959	0.03	0.03	0.977	0.00	0.01	0.940	0.00	0.01	0.873	0.04	0.03	0.834	0.02	0.00	0.786	0.04	0.02
SV(2)	**0.923****	1.00	1.00	**0.927****	1.00	1.00	0.893*	0.68	0.68	0.849	0.11	0.11	0.812	0.02	0.01	0.770	0.04	0.02
SV(3)	1.009	0.02	0.02	0.964	0.14	0.14	**0.885****	1.00	1.00	**0.829****	1.00	1.00	**0.784****	1.00	1.00	**0.744****	1.00	1.00
NASDAQ100	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}
HAR-RV	1.000	0.13	0.13	1.000	0.06	0.10	1.000	0.05	0.09	1.000	0.19	0.16	1.000	0.18	0.10	1.000	0.14	0.07
SV(1)	**0.944****	1.00	1.00	0.959	0.06	0.10	0.968	0.05	0.09	0.925	0.19	0.16	0.912	0.18	0.10	0.857	0.14	0.12
SV(2)	0.952*	0.66	0.66	**0.936****	1.00	1.00	0.934	0.40	0.40	0.896	0.36	0.36	0.874	0.20	0.20	0.816	0.29	0.29
SV(3)	0.999	0.08	0.08	0.964	0.06	0.10	**0.920****	1.00	1.00	**0.884****	1.00	1.00	**0.858****	1.00	1.00	**0.801****	1.00	1.00
N225	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}
HAR-RV	1.000	0.00	0.00	1.000	0.00	0.00	1.000	0.01	0.03	1.000	0.07	0.03	1.000	0.09	0.05	1.000	0.10	0.07
SV(1)	0.942	0.03	0.03	0.951	0.00	0.00	0.970	0.01	0.03	0.945	0.07	0.05	0.937	0.09	0.06	0.918	0.10	0.08
SV(2)	**0.897****	1.00	1.00	**0.914****	1.00	1.00	**0.902****	1.00	1.00	0.859*	0.89	0.89	**0.843****	1.00	1.00	**0.834****	1.00	1.00
SV(3)	1.128	0.00	0.00	1.102	0.00	0.00	0.967	0.03	0.03	**0.857****	1.00	1.00	0.861	0.35	0.35	0.854	0.31	0.31
SSMI20	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}
HAR-RV	1.000	0.17	0.10	1.000	0.01	0.10	1.000	0.08	0.06	1.000	0.11	0.08	1.000	0.00	0.01	1.000	0.02	0.00
SV(1)	0.969	0.17	0.10	0.984	0.01	0.10	0.979	0.08	0.06	0.931	0.11	0.08	0.887	0.02	0.02	0.832	0.02	0.01
SV(2)	**0.948****	1.00	1.00	**0.957****	1.00	1.00	0.955	0.09	0.09	0.919	0.11	0.08	0.877	0.00	0.01	0.813	0.02	0.01
SV(3)	1.015	0.17	0.10	1.007	0.01	0.10	**0.915****	1.00	1.00	**0.881****	1.00	1.00	**0.830****	1.00	1.00	**0.772****	1.00	1.00

Notes: The sample period is from September 01, 2005 to August 31, 2010 where the in-sample is from September 01, 2005 to August 31, 2008 and the out-of-sample is from September 01, 2008 to August 31, 2010. The in-sample include the most volatile part of the late-2000s financial crisis. HAR stands for the Heterogenous Autoregressive model, and we used 5-minute RV following the results of Liu et al. (2015). These are relative to the reference model HAR-RV and values smaller than unity indicate better forecast performance than HAR-RV model. p^M_{MCS} and p^R_{MCS} are associated with $MCS\_T_{\max,\mathcal{M}} = \max_{i\in\mathcal{M}} t_{i,\cdot}$ and $MCS\_T_{R,\mathcal{M}} = \max_{i,j\in\mathcal{M}} \mid t_{i,j} \mid$, respectively. The forecasts in superior model sets $\hat{\mathcal{M}}^*_{5\%}$ and $\hat{\mathcal{M}}^*_{50\%}$ are defined by the average of $p_{MCS} \geq 0.95$ and the average of $p_{MCS} \geq 0.50$, respectively. The forecasts in $\hat{\mathcal{M}}^*_{5\%}$ and $\hat{\mathcal{M}}^*_{50\%}$ are identified by two and one asterisks, respectively. Boldface color font highlights the best model.

Table 3.22. Forecasting realized volatility: relative MSE and associated MCS p-value during high volatility regimes

	$1-day$			$2-day$			$1-week$			$2-week$			$3-week$			$1-month$		
S&P 500	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}
HAR-RV	1.000*	0.94	0.95	1.000*	0.69	0.72	1.000*	0.84	0.89	1.000	0.23	0.31	1.000	0.18	0.23	1.000	0.13	0.13
SV(1)	**0.994****	1.00	1.00	**0.974****	1.00	1.00	0.992*	0.84	0.89	0.981	0.23	0.31	0.967	0.18	0.23	0.914	0.13	0.13
SV(2)	1.002*	0.94	0.95	0.989*	0.69	0.72	0.989*	0.84	0.89	0.974	0.23	0.31	0.960	0.18	0.23	0.898	0.13	0.13
SV(3)	1.082	0.18	0.21	1.061	0.26	0.35	**0.980****	1.00	1.00	**0.940****	1.00	1.00	**0.895****	1.00	1.00	**0.852****	1.00	1.00
FTSE100	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}
HAR-RV	1.000*	0.67	0.63	1.000	0.46	0.39	**1.000****	1.00	1.00	1.000*	0.90	0.93	1.000*	0.70	0.69	1.000	0.18	0.17
SV(1)	0.997*	0.67	0.63	0.949*	0.93	0.93	1.015	0.46	0.67	**0.988****	1.00	1.00	**0.970****	1.00	1.00	0.968	0.18	0.17
SV(2)	**0.977****	1.00	1.00	**0.948****	1.00	1.00	1.004*	0.93	0.93	0.993*	0.90	0.93	0.972*	0.88	0.85	0.963	0.20	0.20
SV(3)	1.058	0.18	0.15	1.024	0.46	0.42	1.008*	0.93	0.93	0.991*	0.90	0.93	0.973*	0.88	0.85	**0.960****	1.00	1.00
NASDAQ100	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}
HAR-RV	1.000*	0.78	0.79	1.000**	0.99	0.99	1.000*	0.71	0.77	1.000	0.22	0.40	1.000	0.37	0.56	1.000	0.19	0.31
SV(1)	1.014*	0.78	0.79	**0.994****	1.00	1.00	0.992*	0.71	0.77	1.033	0.12	0.16	1.038	0.13	0.22	1.004	0.07	0.13
SV(2)	**0.976****	1.00	1.00	0.996**	0.99	0.99	**0.981****	1.00	1.00	0.992	0.22	0.40	0.988	0.37	0.56	0.957	0.19	0.31
SV(3)	1.089	0.33	0.25	1.083*	0.54	0.59	1.020*	0.65	0.69	**0.977****	1.00	1.00	**0.976****	1.00	1.00	**0.934****	1.00	1.00
N225	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}
HAR-RV	1.000	0.41	0.33	1.000	0.17	0.16	**1.000****	1.00	1.00	1.000	0.32	0.23	1.000**	0.97	0.97	1.000*	0.88	0.88
SV(1)	0.948*	0.62	0.62	0.966	0.17	0.16	1.072	0.25	0.50	1.080	0.32	0.18	1.071	0.23	0.29	1.044	0.19	0.28
SV(2)	**0.942****	1.00	1.00	**0.929****	1.00	1.00	1.008**	0.99	0.98	0.960	0.32	0.23	0.991**	0.97	0.97	0.999*	0.77	0.85
SV(3)	1.210	0.30	0.17	1.092	0.17	0.16	1.006**	0.99	0.98	**0.927****	1.00	1.00	**0.990****	1.00	1.00	**0.994****	1.00	1.00
SSMI20	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}
HAR-RV	1.000*	0.62	0.59	1.000*	0.74	0.74	**1.000****	1.00	1.00	1.000*	0.72	0.70	1.000	0.28	0.31	1.000	0.10	0.11
SV(1)	**0.958****	1.00	1.00	**0.985****	1.00	1.00	1.015*	0.85	0.86	0.997*	0.62	0.61	0.977	0.28	0.31	0.970	0.10	0.11
SV(2)	0.961*	0.83	0.83	0.998*	0.63	0.73	1.022*	0.83	0.82	0.986*	0.72	0.70	0.961	0.28	0.31	0.945	0.27	0.27
SV(3)	1.128	0.23	0.18	1.115	0.39	0.42	1.016*	0.85	0.86	**0.973****	1.00	1.00	**0.948****	1.00	1.00	**0.934****	1.00	1.00

Notes: The sample period is from January 01, 2005 to December 31, 2009 where the in-sample is from January 01, 2005 to December 31, 2007 and the out-of-sample is from January 01, 2008 to December 31, 2009. In this setting, we forecast a highly volatile period. HAR stands for the Heterogenous Autoregressive model, and we used 5-minute RV following the results of Liu et al. (2015). These are relative to the reference model HAR-RV and values smaller than unity indicate better forecast performance than HAR-RV model. p^M_{MCS} and p^R_{MCS} are associated with $MCS\_T_{\max,\mathcal{M}} = \max_{i\in\mathcal{M}} t_{i,\cdot}$ and $MCS\_T_{R,\mathcal{M}} = \max_{i,j\in\mathcal{M}} \mid t_{i,j} \mid$, respectively. The forecasts in superior model sets $\hat{\mathcal{M}}^*_{5\%}$ and $\hat{\mathcal{M}}^*_{50\%}$ are defined by the average of $p_{MCS} \geq 0.95$ and the average of $p_{MCS} \geq 0.50$, respectively. The forecasts in $\hat{\mathcal{M}}^*_{5\%}$ and $\hat{\mathcal{M}}^*_{50\%}$ are identified by two and one asterisks, respectively. Boldface color font highlights the best model.

Table 3.23. Forecasting realized volatility: relative MAE and associated MCS p-value during high volatility regimes

	1 − *day*			2 − *day*			1 − *week*			2 − *week*			3 − *week*			1 − *month*		
S&P 500	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}
HAR-RV	1.000	0.05	0.02	1.000	0.07	0.06	1.000	0.29	0.27	1.000	0.25	0.25	1.000	0.09	0.10	1.000	0.11	0.11
SV(1)	0.945*	0.67	0.67	0.952*	0.53	0.53	0.919*	0.62	0.67	0.862*	0.85	0.90	0.831	0.44	0.40	0.812	0.31	0.26
SV(2)	**0.939****	1.00	1.00	**0.947****	1.00	1.00	0.914*	0.86	0.86	**0.859****	1.00	1.00	0.828	0.44	0.40	0.805	0.31	0.26
SV(3)	0.968	0.06	0.11	0.989	0.07	0.06	**0.912****	1.00	1.00	0.860**	0.95	0.95	**0.812****	1.00	1.00	**0.782****	1.00	1.00
FTSE100	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}
HAR-RV	1.000	0.19	0.23	1.000	0.13	0.19	1.000	0.12	0.06	1.000	0.11	0.04	1.000	0.02	0.01	1.000	0.06	0.02
SV(1)	0.991	0.26	0.26	0.992	0.13	0.19	0.964	0.12	0.09	0.892	0.20	0.13	0.859	0.06	0.03	0.834	0.16	0.11
SV(2)	**0.972****	1.00	1.00	**0.958****	1.00	1.00	0.942	0.36	0.36	0.885	0.22	0.22	0.855	0.06	0.03	0.829	0.16	0.11
SV(3)	1.026	0.12	0.15	0.967*	0.64	0.64	**0.923****	1.00	1.00	**0.876****	1.00	1.00	**0.842****	1.00	1.00	**0.819****	1.00	1.00
NASDAQ100	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}
HAR-RV	1.000	0.32	0.33	1.000	0.20	0.18	1.000	0.33	0.26	1.000	0.17	0.06	1.000	0.30	0.20	1.000	0.13	0.09
SV(1)	**0.971****	1.00	1.00	0.963	0.32	0.27	0.940*	0.70	0.63	0.904	0.17	0.07	0.886	0.46	0.39	0.902	0.13	0.09
SV(2)	0.971**	0.98	0.98	**0.950****	1.00	1.00	0.937*	0.70	0.63	0.889	0.17	0.12	0.872*	0.68	0.68	0.871	0.15	0.15
SV(3)	0.998	0.28	0.25	0.976	0.32	0.27	**0.929****	1.00	1.00	**0.879****	1.00	1.00	**0.870****	1.00	1.00	**0.862****	1.00	1.00
N225	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}
HAR-RV	1.000	0.11	0.08	1.000	0.05	0.03	1.000	0.06	0.12	1.000	0.15	0.09	1.000	0.13	0.06	1.000	0.16	0.13
SV(1)	0.970	0.35	0.35	0.967	0.05	0.03	0.953	0.06	0.12	0.954	0.15	0.10	0.941	0.13	0.06	0.935	0.16	0.15
SV(2)	**0.962****	1.00	1.00	**0.944****	1.00	1.00	**0.931****	1.00	1.00	0.904	0.31	0.31	0.895	0.27	0.27	0.902*	0.88	0.88
SV(3)	1.075	0.04	0.03	1.009	0.05	0.03	0.948	0.45	0.45	**0.889****	1.00	1.00	**0.881****	1.00	1.00	**0.900****	1.00	1.00
SSMI20	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}
HAR-RV	1.000*	0.84	0.83	1.000**	0.99	0.99	1.000	0.11	0.10	1.000	0.02	0.01	1.000	0.09	0.03	1.000	0.11	0.05
SV(1)	0.991**	0.95	0.95	0.998**	0.99	0.99	0.986	0.11	0.10	0.911	0.02	0.01	0.879	0.11	0.11	0.863	0.15	0.15
SV(2)	**0.991****	1.00	1.00	1.000**	0.98	0.99	0.991	0.08	0.10	0.912	0.02	0.01	0.886	0.09	0.07	0.849	0.40	0.40
SV(3)	1.036	0.24	0.25	**0.996****	1.00	1.00	**0.945****	1.00	1.00	**0.883****	1.00	1.00	**0.861****	1.00	1.00	**0.844****	1.00	1.00

Notes: The sample period is from January 01, 2005 to December 31, 2009 where the in-sample is from January 01, 2005 to December 31, 2007 and the out-of-sample is from January 01, 2008 to December 31, 2009. In this setting, we forecast a highly volatile period. HAR stands for the Heterogenous Autoregressive model, and we used 5-minute RV following the results of Liu et al. (2015). These are relative to the reference model HAR-RV and values smaller than unity indicate better forecast performance than HAR-RV model. p^M_{MCS} and p^R_{MCS} are associated with $MCS\_T_{\max,\mathcal{M}} = \max_{i\in\mathcal{M}} t_{i\cdot}$ and $MCS\_T_{R,\mathcal{M}} = \max_{i,j\in\mathcal{M}} \mid t_{i,j} \mid$, respectively. The forecasts in superior model sets $\hat{\mathcal{M}}^*_{5\%}$ and $\hat{\mathcal{M}}^*_{50\%}$ are defined by the average of $p_{MCS} \geq 0.95$ and the average of $p_{MCS} \geq 0.50$, respectively. The forecasts in $\hat{\mathcal{M}}^*_{5\%}$ and $\hat{\mathcal{M}}^*_{50\%}$ are identified by two and one asterisks, respectively. Boldface color font highlights the best model.

Table 3.24. Forecasting realized volatility: relative R2LOG and associated MCS p-value during high volatility regimes

	$1-day$			$2-day$			$1-week$			$2-week$			$3-week$			$1-month$		
S&P 500	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}
HAR-RV	1.000	0.00	0.00	1.000	0.01	0.02	1.000	0.43	0.34	1.000	0.39	0.35	1.000	0.18	0.11	1.000	0.18	0.10
SV(1)	0.925	0.25	0.25	0.922	0.40	0.40	0.933	0.43	0.48	0.914	0.54	0.44	0.915	0.18	0.11	0.870	0.18	0.10
SV(2)	**0.899****	1.00	1.00	**0.913****	1.00	1.00	**0.920****	1.00	1.00	0.908	0.54	0.44	0.897	0.18	0.11	0.850	0.18	0.10
SV(3)	0.968	0.00	0.02	0.981	0.01	0.02	0.921**	0.98	0.98	**0.885****	1.00	1.00	**0.844****	1.00	1.00	**0.788****	1.00	1.00
FTSE100	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}
HAR-RV	1.000	0.01	0.01	1.000	0.00	0.02	1.000	0.01	0.14	1.000	0.07	0.10	1.000	0.03	0.04	1.000	0.05	0.06
SV(1)	0.970	0.01	0.01	0.964	0.00	0.02	0.972	0.01	0.14	0.958	0.07	0.10	0.940	0.03	0.04	0.912	0.05	0.06
SV(2)	**0.929****	1.00	1.00	**0.931****	1.00	1.00	**0.942****	1.00	1.00	0.938	0.15	0.15	0.918	0.03	0.04	0.889	0.05	0.06
SV(3)	1.052	0.01	0.01	1.008	0.00	0.02	0.954*	0.62	0.62	**0.915****	1.00	1.00	**0.886****	1.00	1.00	**0.860****	1.00	1.00
NASDAQ100	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}
HAR-RV	1.000	0.14	0.10	1.000	0.21	0.14	1.000	0.43	0.42	1.000	0.14	0.14	1.000	0.07	0.06	1.000	0.06	0.06
SV(1)	**0.937****	1.00	1.00	0.952	0.21	0.14	0.958	0.43	0.42	0.960	0.14	0.14	0.952	0.07	0.06	0.931	0.06	0.06
SV(2)	0.944*	0.64	0.64	**0.934****	1.00	1.00	0.941**	0.97	0.97	0.916	0.27	0.27	0.893	0.07	0.06	0.851	0.10	0.10
SV(3)	1.015	0.08	0.05	0.980	0.21	0.14	**0.940****	1.00	1.00	**0.897****	1.00	1.00	**0.865****	1.00	1.00	**0.825****	1.00	1.00
N225	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}
HAR-RV	1.000	0.05	0.03	1.000	0.02	0.05	1.000	0.36	0.36	1.000	0.41	0.38	1.000*	0.67	0.62	1.000*	0.72	0.73
SV(1)	0.970	0.05	0.04	0.984	0.02	0.05	1.012	0.02	0.14	1.009	0.12	0.15	1.042	0.12	0.19	1.060	0.10	0.23
SV(2)	**0.936****	1.00	1.00	**0.953****	1.00	1.00	**0.951****	1.00	1.00	0.916*	0.84	0.84	**0.924****	1.00	1.00	**0.947****	1.00	1.00
SV(3)	1.182	0.00	0.00	1.136	0.00	0.00	1.005	0.26	0.34	**0.913****	1.00	1.00	0.936*	0.67	0.64	0.965*	0.72	0.73
SSMI20	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}
HAR-RV	1.000	0.18	0.18	1.000	0.26	0.30	1.000	0.35	0.35	1.000	0.29	0.28	1.000	0.02	0.04	1.000	0.02	0.03
SV(1)	**0.957****	1.00	1.00	0.984	0.26	0.30	0.996	0.17	0.18	0.990	0.11	0.13	0.959	0.02	0.04	0.933	0.02	0.03
SV(2)	0.958*	0.92	0.92	**0.973****	1.00	1.00	0.982	0.35	0.35	0.963	0.29	0.28	0.918	0.02	0.04	0.867	0.16	0.16
SV(3)	1.061	0.07	0.05	1.048	0.26	0.21	**0.950****	1.00	1.00	**0.929****	1.00	1.00	**0.879****	1.00	1.00	**0.843****	1.00	1.00

Notes: The sample period is from January 01, 2005 to December 31, 2009 where the in-sample is from January 01, 2005 to December 31, 2007 and the out-of-sample is from January 01, 2008 to December 31, 2009. In this setting, we forecast a highly volatile period. HAR stands for the Heterogenous Autoregressive model, and we used 5-minute RV following the results of Liu et al. (2015). These are relative to the reference model HAR-RV and values smaller than unity indicate better forecast performance than HAR-RV model. p^M_{MCS} and p^R_{MCS} are associated with $MCS\_T_{\max,\mathcal{M}} = \max_{i\in\mathcal{M}} t_{i,\cdot}$ and $MCS\_T_{R,\mathcal{M}} = \max_{i,j\in\mathcal{M}} \mid t_{i,j} \mid$, respectively. The forecasts in superior model sets $\hat{\mathcal{M}}^*_{5\%}$ and $\hat{\mathcal{M}}^*_{50\%}$ are defined by the average of $p_{MCS} \geq 0.95$ and the average of $p_{MCS} \geq 0.50$, respectively. The forecasts in $\hat{\mathcal{M}}^*_{5\%}$ and $\hat{\mathcal{M}}^*_{50\%}$ are identified by two and one asterisks, respectively. Boldface color font highlights the best model.

3.12.3 Figures

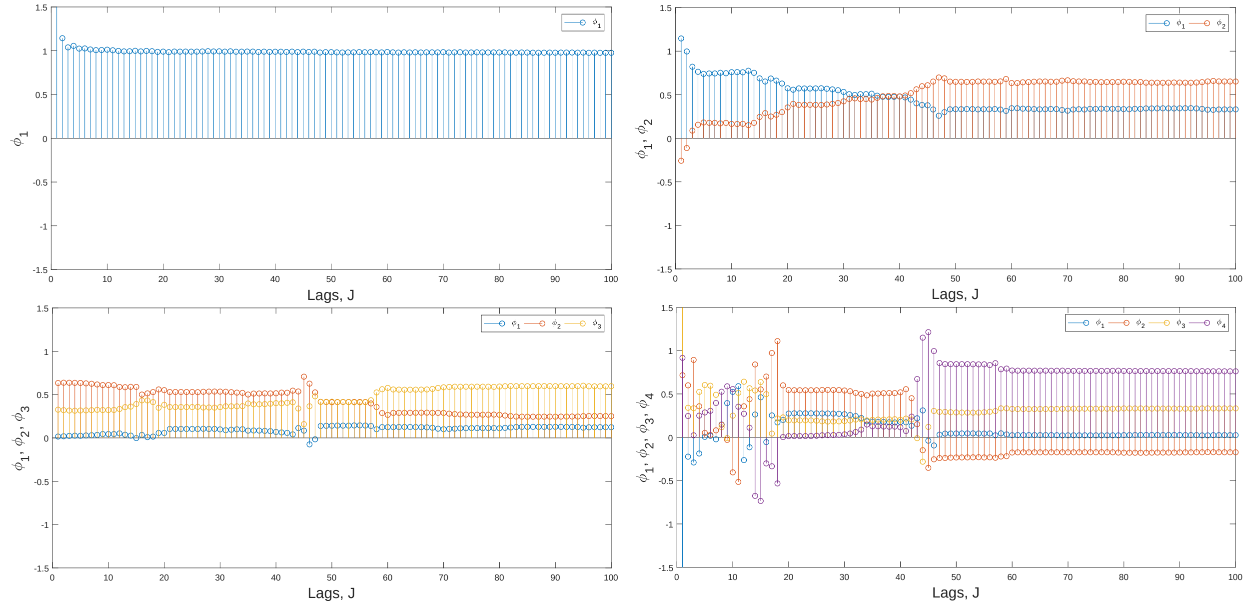

Figure 3.1. S&P 500: 1996-2016. W-ARMA-SV-OLS estimators of volatility persistence parameters (ϕ's) as a function of the number of lags (J). Four SV models are considered.

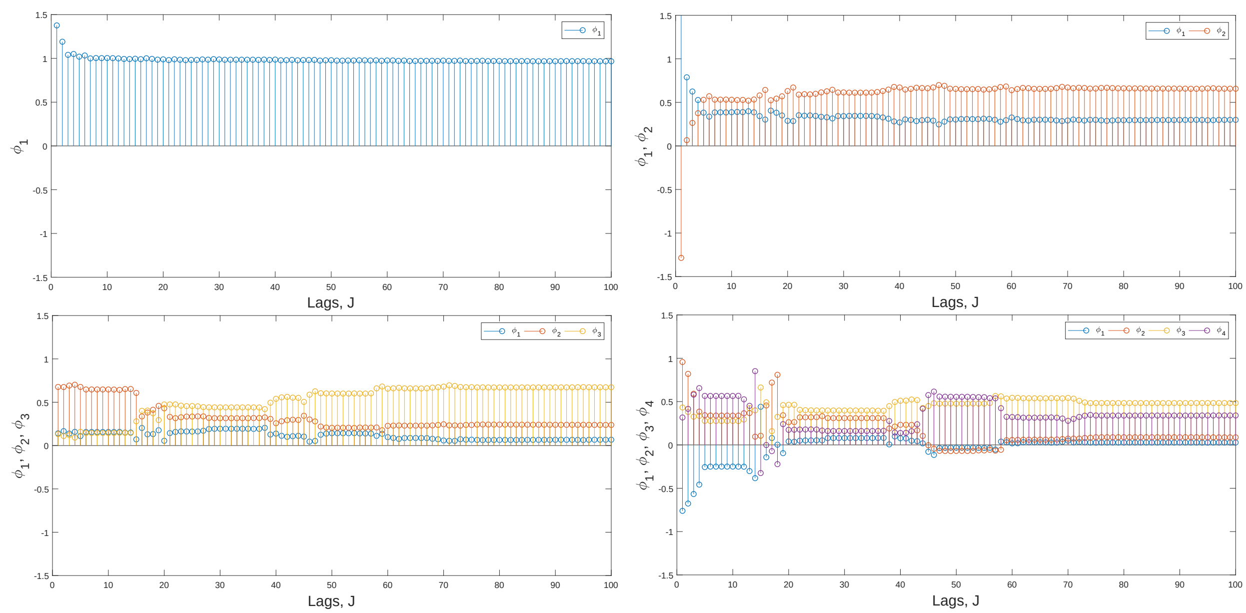

Figure 3.2. S&P 500: 2006-2016. W-ARMA-SV-OLS estimators of volatility persistence parameters (ϕ's) as a function of the number of lags (J). Four SV models are considered.

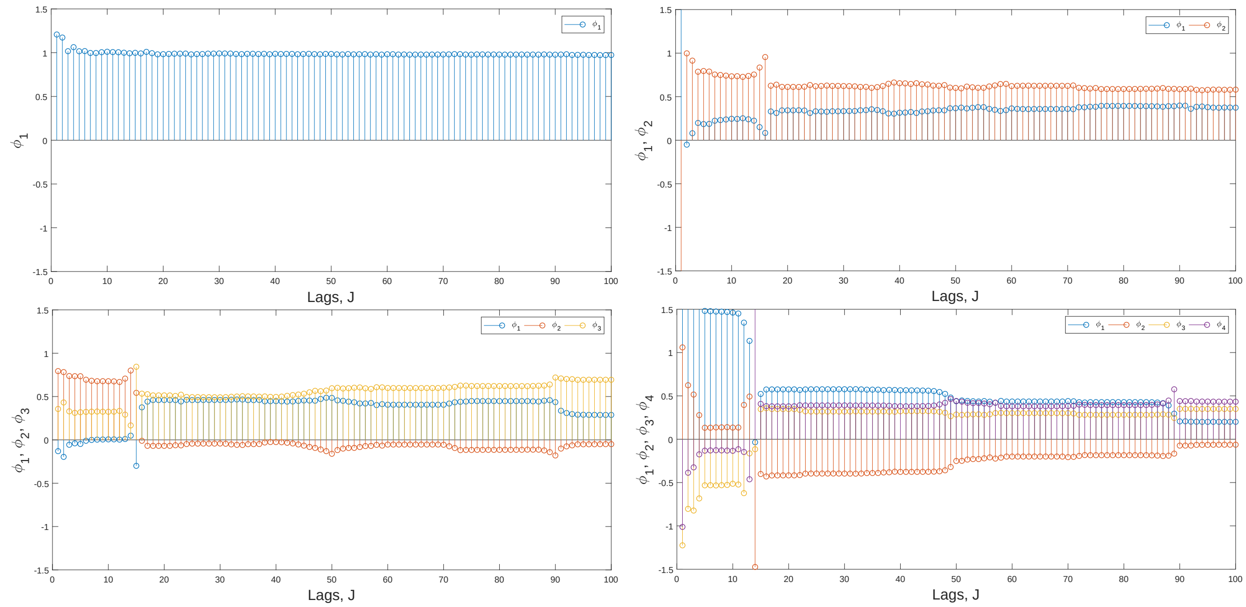

Figure 3.3. S&P 500: 2006-2010. W-ARMA-SV-OLS estimators of volatility persistence parameters (ϕ's) as a function of the number of lags (J). Four SV models are considered.

3.12.4 Forecasting with SV(p) models

As discussed earlier, SV(p) models can be written as a linear state-space model without losing any information. The state-space representation of SV(p) models is given by

$$\begin{aligned} y_t^* &= w_t + \epsilon_t, \\ w_t &= \sum_{j=1}^{p} \phi_j w_{t-j} + v_t, \end{aligned} \tag{3.12.93}$$

where the distribution ϵ_t is approximated by a normal distribution with mean 0 and variance $\pi^2/2$. Using the standard notations of Hamilton (1994), the model defined in (3.12.93) can be rewritten as following:

$$\begin{aligned} y_t &= A' x_t + H' \xi_t + w_t, \\ \xi_{t+1} &= F \xi_t + \nu_{t+1}, \end{aligned} \tag{3.12.94}$$

with $y_t = y_t^*$, $A' = 0$, $x_t = 1$, $H' = (1, 0, \ldots, 0)$ is a $1 \times p$ vector, $w_t = \epsilon_t$, $R = \mathbb{E}(w_t w_t') = \pi^2/2$,

$$\xi_t = \begin{bmatrix} w_t \\ w_{t-1} \\ w_{t-2} \\ \vdots \\ w_{t-p-1} \end{bmatrix}, \quad F = \begin{bmatrix} \phi_1 & \phi_2 & \cdots & \phi_{p-1} & \phi_p \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}, \quad \nu_{t+1} = \begin{bmatrix} v_{t+1} \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

$$Q = \mathbb{E}(\nu_t \nu_t') = \begin{bmatrix} \sigma_v^2 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix},$$

where F and Q are $p \times p$ matrices, and ξ_t are ν_{t+1} are $p \times 1$ vectors. Now using (3.12.94), the Kalman filter can be applied as follows:

- Initialization:

$$\begin{aligned} \hat{\xi}_{1|0} &= \mathbb{E}(\xi_1) = \mathbf{0}_{(p\times 1)}, \\ \mathbf{P}_{1|0} &= \mathbb{E}([(\xi_1 - \mathbb{E}(\xi_1)][(\xi_1 - \mathbb{E}(\xi_1)]') = diag[\sigma_v^2, \ldots, \sigma_v^2]_{(p\times p)}, \end{aligned} \tag{3.12.95}$$

  where $\mathbf{P}_{1|0}$ is the MSE associated with $\hat{\xi}_{1|0}$.

- Sequential updating:

$$\begin{aligned} \hat{\xi}_{t|t} &= \hat{\xi}_{t|t-1} + \mathbf{P}_{t|t-1} H (H' \mathbf{P}_{t|t-1} H + R)^{-1} \times (y_t - H' \hat{\xi}_{t|t-1}), \\ \mathbf{P}_{t|t} &= \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1} H (H' \mathbf{P}_{t|t-1} H + R)^{-1} \mathbf{P}_{t|t-1} H'. \end{aligned} \tag{3.12.96}$$

- In-sample prediction:

$$\begin{aligned}
&\hat{\xi}_{t+1|t} = F\hat{\xi}_{t|t-1} + F\mathbf{P}_{t|t-1}H(H'\mathbf{P}_{t|t-1}H + R)^{-1} \times (y_t - H'\hat{\xi}_{t|t-1}), \\
&\mathbf{P}_{t+1|t} = F\mathbf{P}_{t|t}F' + Q.
\end{aligned} \tag{3.12.97}$$

Given (3.12.97), the forecast of y_{t+1} and the MSE of forecast error are given by

$$\begin{aligned}
&\hat{y}_{t+1|t} = H'\hat{\xi}_{t+1|t}, \\
&\mathbb{E}([y_{t+1} - \hat{y}_{t+1|t}][y_{t+1} - \hat{y}_{t+1|t}]') = H'\mathbf{P}_{t+1|t}H + R.
\end{aligned} \tag{3.12.98}$$

- Out-of-sample h-step-ahead forecasting:

$$\begin{aligned}
&\hat{\xi}_{T+h|T} = F^h\hat{\xi}_{T|T}, \\
&\hat{y}_{T+h|T} = H'\hat{\xi}_{T+h|T} = H'F^h\hat{\xi}_{T|T}.
\end{aligned} \tag{3.12.99}$$

The h-step-ahead forecast is computed by (3.12.99) with the simple estimates plugged in.

3.12.5 Forecasting with GARCH models

GARCH Model: The generalized autoregressive conditional heteroskedastic (GARCH) model is an extension of the ARCH model by Engle (1982). If a series exhibits volatility clustering, this suggests that past variances might be predictive of the current variance. The GARCH(p,q) model is an autoregressive moving average model for conditional variances, with p GARCH coefficients associated with lagged variances, and q ARCH coefficients associated with lagged squared innovations or lagged squared residual returns. The GARCH(p,q) model of residual return is

$$\begin{aligned}
&y_t = \sigma_t z_t, \quad z_t \sim i.i.d \;\; N(0,1), \\
&\sigma_t^2 = \omega + \beta_1\sigma_{t-1}^2 + \ldots + \beta_p\sigma_{t-p}^2 + \alpha_1 y_{t-1}^2 + \ldots + \alpha_q y_{t-q}^2,
\end{aligned}$$

where y_t is the residual return observed at time t and σ_t is the corresponding volatility. For stationarity and positivity, the GARCH model has the following constraints:

- $\omega > 0$,

- $\beta_i \geq 0$, $\alpha_j \geq 0$
- $\sum_{i=1}^{p} \beta_i + \sum_{j=1}^{q} \alpha_j < 1$.

The h-step-ahead forecast of the GARCH(1, 1) model is computed according to:

$$\begin{aligned}
&\hat{\sigma}^2_{t+h|t} = \hat{\omega} + \hat{\beta}_1 \hat{\sigma}^2_{t+h-1|t} + \hat{\alpha}_1 \hat{y}^2_{t+h-1|t},\\
&\hat{y}^2_{t+h|t} = \hat{\sigma}^2_{t+h|t} \quad if \quad h > 0,\\
&\hat{y}^2_{t+h|t} = y^2_{t+h} \quad \hat{\sigma}^2_{t+h|t} = \sigma^2_{t+h} \quad if \quad h \leq 0.
\end{aligned}$$

EGARCH Model: The exponential GARCH (EGARCH) model was developed by Nelson (1991). It is a GARCH variant that models the logarithm of the conditional variance process. In addition to modeling the logarithm, the EGARCH model has additional leverage terms to capture asymmetry in volatility clustering. The EGARCH(p, q) model has p GARCH coefficients associated with lagged log variance terms, q ARCH coefficients associated with the magnitude of lagged standardized innovations, and q leverage coefficients associated with signed, lagged standardized innovations. The form of the EGARCH(p, q) model is

$$\begin{aligned}
y_t &= \sigma_t z_t, \quad z_t \sim i.i.d \;\; N(0,1),\\
\log \sigma_t^2 &= \omega + \sum_{i=1}^{p} \beta_i \log \sigma^2_{t-i} + \sum_{j=1}^{q} \alpha_j \left(|z_{t-j}| - \mathbb{E}\left(|z_{t-j}|\right)\right) + \sum_{j=1}^{q} \gamma_j z_{t-j},
\end{aligned}$$

where $z_t := y_t \sigma_t^{-1}$ and to ensure stationarity, all roots of the GARCH coefficient polynomial, $(1 - \beta_1 L - \ldots - \beta_p L^p)$, must lie outside the unit circle. The h-step-ahead forecast of the EGARCH(1, 1) model is computed according to:

$$\log \hat{\sigma}^2_{t+h|t} = \hat{\omega} + \hat{\beta}_1 \log \hat{\sigma}^2_{t+h-1|t} + \hat{\alpha}_1 \left(|\hat{z}_{t+h-1|t}| - \mathbb{E}\left(|\hat{z}_{t+h-1|t}|\right)\right) + \gamma_1 \hat{z}_{t+h-1|t}.$$

GJR Model: The GJR-GARCH, or just GJR, model of Glosten et al. (1993) allows the conditional variance to respond differently to the past negative and positive innovations. The GJR(p, q) model has p GARCH coefficients associated with lagged variances, q ARCH coefficients associated with lagged squared innovations, and q leverage coefficients associated with the square of negative lagged innovations. The GJR(p, q) model may be expressed as:

$$y_t = \sigma_t z_t, \quad z_t \sim i.i.d \ \ N(0,1),$$
$$\log \sigma_t^2 = \omega + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2 + \sum_{j=1}^{q} (\alpha_j + \gamma_j I_{[y_{t-j}<0]}) y_{t-j}^2,$$

where the indicator function $I[y_{t-j} < 0]$ equals 1 if $y_{t-j} < 0$, and 0 otherwise. Thus, the leverage coefficients are applied to negative innovations, giving negative changes additional weight. For stationarity and positivity, the GJR model has the following constraints:

- $\omega > 0$
- $\beta_i \geq 0$, $\alpha_j \geq 0$
- $\alpha_j + \gamma_j \geq 0$
- $\sum_{i=1}^{p} \beta_i + \sum_{j=1}^{q} (\alpha_j + \frac{1}{2}\gamma_j) < 1$

The GARCH model is nested in the GJR model. If all leverage coefficients are zero, then the GJR model reduces to the GARCH model. The recursive formula for the h-step-ahead forecast of the GJR-GARCH(1,1) model is calculated as:

$$\hat{\sigma}_{t+h|t}^2 = \hat{\omega} + \left(\hat{\alpha}_1 + \frac{\hat{\gamma}_1}{2} + \hat{\beta}_1\right) \hat{\sigma}_{t+h-1|t}^2.$$

3.12.6 Heterogenous Autoregressive model of Realized Volatility

Heterogenous Autoregressive model of Realized Volatility (HAR-RV) model proposed by Corsi (2009). In financial markets, either traders are perceived to be heterogeneous in the sense of a different horizon of investments [Müller et al. (1997)] or information arrival is heterogeneous [Andersen and Bollerslev (1998)]. HAR-RV model takes into account the long memory feature, and among the models proposed to forecast volatility, it stands out in terms of performance and simplicity.

A generalized version of HAR-RV model that we used here is as follows:

$$\log RV_{t+1}^{(d)} = c + \beta^{(d)} \log RV_t^{(d)} + \beta^{(w)} \log RV_t^{(w)} + \beta^{(m)} \log RV_t^{(m)} + u_{t+1}^{d} \tag{3.12.100}$$

where

$$\log RV_t^{(w)} = \frac{1}{5}\sum_{j=0}^{4} \log RV_{t-j}^{(d)},$$

$$\log RV_t^{(m)} = \frac{1}{22}\sum_{j=0}^{21} \log RV_{t-j}^{(d)}.$$

This class of models can be estimated with ordinary least squares. For the details of forecasting in HAR-RV model, see Corsi (2009).

Chapter 4

High-frequency instruments and identification-robust inference for stochastic volatility models

Abstract

We introduce a novel class of generalized stochastic volatility (GSV) models, which utilize high-frequency (HF) information (realized volatility (RV) measures). GSV models can accommodate nonstationary volatility, various distributional assumptions, and exogenous regressors in the latent volatility equation. Instrumental variable methods are employed to provide a unified framework for GSV models' analysis (estimation and inference). We consider the parameter inference problem in GSV models with nonstationary volatility and develop identification-robust methods for joint hypotheses involving the volatility persistence parameter and the autocorrelation parameter of the composite error (or the noise ratio). For inference about the volatility persistence parameter, projection techniques are applied. The proposed tests include Anderson-Rubin-type (AR) tests, a dynamic version of the split-sample (SS) procedure, and point-optimal versions of these tests (AR^* and SS^*). For distributional theory, three different sets of assumptions are considered: (1) for Gaussian errors, we provide exact tests and confidence sets; (2) for a wide class of parametric non-Gaussian errors (possibly heavy-tailed), we establish that exact Monte Carlo procedures can be applied using the statistics considered; (3) under weaker distributional assumptions, we show these tests are asymptotically valid. A comprehensive Monte Carlo study indicates that the proposed tests outperform the usual asymptotic test regarding size and exhibit excellent power in empirically realistic settings. We apply our inference methods to IBM's price and option data (2009-2013). We consider 175 different instruments (IV's) spanning 22 classes and analyze their ability to describe the low-frequency volatility. The IV's are compared based on the average length of confidence intervals, which are produced by the proposed tests. The superior instrument set mostly consists of 5-minute HF realized measures, and these IV's produce confidence sets where the volatility persistence parameter lies roughly between 0.85 and 1.0. This outcome suggests that the volatility process is highly persistent and close to unit-root. We find RVs with higher frequency produce wider confidence intervals compared to RVs with slightly lower frequency, showing that these confidence intervals adjust to absorb market microstructure noise or discretization error. Further, when we consider irrelevant or weak IV's (jumps and signed jumps), the proposed tests produce unbounded confidence intervals. Although jumps contain little information content regarding the low-frequency volatility, we find evidence that there may be a nonlinear relationship between jumps and the low-frequency volatility.

Key words: Stochastic volatility, Realized variance, High frequency data, Identification robust test.
Journal of Economic Literature classification: C15, C22, C53, C58, C32.

4.1 Introduction

In stochastic volatility (SV) models [proposed by Taylor (1982, 1986)], the return variation dynamics is modelled as a latent autocorrelated stochastic process. Estimation and inference are challenging in SV models due to the inherent problem of evaluating the likelihood function.[1] As a result, a variety of alternative methods have been proposed to estimate SV models.[2] For a review of the SV literature; see Ghysels et al. (1996), Broto and Ruiz (2004), Shephard (2005), Ahsan and Dufour (2018).

This paper proposes generalized stochastic volatility models (GSV), where volatility is modelled as a latent stochastic process. Instrumental-variable methods can be used to estimate GSV models, where a standard solution is to replace the unobserved volatility by a proxy. Hence, we need valid instruments (IV's) for the latent volatility. The choice of instruments plays a crucial role. As a result, we consider broad classes of IV's for the latent log volatility; we use high-frequency (HF) realized measures as instruments. To the best of our knowledge, this paper is the first to propose instrumental-variable (IV) regression in the context of SV models. GSV models can accommodate nonstationary volatility, various distributional assumptions, and exogenous regressors in the latent volatility equation.

This study considers the problem of testing hypotheses and building confidence sets for the volatility persistence parameter, which captures the volatility clustering. This parameter plays

[1] The marginal likelihood of SV models is given by a high dimensional integral, which makes the estimation by conventional maximum likelihood (ML) infeasible. This is a general feature of most nonlinear latent variable models because the latent variables must be integrated out of the joint density for the observed and latent processes, leading to an integral of high dimensionality.

[2] Major references include: the Quasi-Maximum Likelihood (QML) [Harvey et al. (1994); Ruiz (1994)], the Generalized Method of Moments (GMM) [Melino and Turnbull (1990); Andersen and Sørensen (1996)], the Efficient Method of Moments (EMM) [Gallant and Tauchen (1996); Andersen et al. (1999)], the Maximum Likelihood Monte Carlo (MLMC) [Sandmann and Koopman (1998)], the Simulated Maximum Likelihood (SML) [Danielsson and Richard (1993); Danielsson (1994); Durham (2006); Liesenfeld and Jung (2000); Richard and Zhang (2007)], method base on linear-representation (LiR) [Francq and Zakoïan (2006)], the closed-form moment-based estimator (DV) [Dufour and Valéry (2006)], the ARMA-based winsorized estimator (W-ARMA-SV) [Ahsan and Dufour (2019)] and Bayesian methods based on Markov Chain Monte Carlo (MCMC) methods [Jacquier et al. (1994), Kim et al. (1998), Chib et al. (2002), Flury and Shephard (2011)].

a crucial role in many areas of financial economics. *First*, asset allocation theories have shown that this parameter can reflect the persistence in the risk premium, *e.g.*, when there is a high persistence in volatility, a rational investor should frequently and permanently change the weighting of assets whenever a volatility shock arrives. *Second*, the volatility persistence parameter's confidence set determines the conditional volatility forecast interval given the current volatility, which is important for risk management, option pricing, and asset pricing:

- Accurate estimation of the tails of the return distribution are of particular importance for risk management tools (Value at Risk and Expected Shortfall); see Taylor (1999).
- The volatility forecast interval is important for option pricing; see Hansen (1994).
- The accurate confidence interval estimation of volatility has consequences for the forecasts of the conditional mean (the prediction interval of returns) through projection techniques; see Baillie and Bollerslev (1992).

We are interested in testing some general restrictions on the volatility persistence parameter (including non-stationarity of the volatility process by testing for a unit root) in log-squared low-frequency returns in a model setup, which utilizes high-frequency information (realized volatility (RV) measures). Indeed, we let the latent volatility process's autoregressive root be close or equal to one. Nonstationarity in the volatility process has been well documented for macroeconomic and financial time series data; see Pagan and Schwert (1990*a*, 1990*b*), Loretan and Phillips (1994), McConnell and Perez-Quiros (2000), Blanchard and Simon (2001), Busetti and Taylor (2003), Sensier and Dijk (2004), Cavaliere and Taylor (2007). For instance, nonstationary volatility arises when the variance is trending (upward or downward) or undergoes structural breaks. Several studies note that the empirical estimate of the dominant root of the SV-type process is close to unit circle; see Harvey et al. (1994), Hansen (1995), Broto and Ruiz (2004). Hansen (1995) is the only study that proposed robust regressions in the mean equation under nonstationary stochastic volatility, whereas Harvey et al. (1994) estimated SV models imposing a unit root in the variance equation. Besides, we want to build a valid confidence set of the persistence parameter that can be used to determine the volatility forecast interval and/or the distribution of the volatility forecasts in our proposed model setup.

Previous attempts on hypothesis testing for the volatility persistence parameter are limited, these include: Harvey et al. (1994), Wright (1999), Dufour and Valéry (2009) and Ahsan and Dufour (2019). Harvey et al. (1994) considered a classical unit root test in a QML setup, which suffers from large size distortions. Wright (1999) proposed to use the unit root test of Perron and Ng (1996) that is based on large-sample approximations and is not reliable in finite samples (requires extremely large samples) and different parameter settings. These inference procedures are based on large-sample approximations (asymptotic standard errors), and it is known that when a time series is nearly nonstationary, the asymptotic standard error can be markedly different, and asymptotic approximations are very unreliable in finite samples. In the context of a standard SV model, simulation results (in this paper, see Table 4.2) show that tests based on asymptotic standard errors fail to control the type I errors when the volatility persistence parameter approaches to the unit circle. This assertion is also supported by Harvey et al. (1994)) and Wright (1999).

Dufour and Valéry (2009) and Ahsan and Dufour (2019) developed both exact and asymptotic tests for no persistence (or no clustering) hypothesis, which are primarily based on stationarity (requires time invariance of unconditional variances and autocovariances) and normality assumptions. Applying these procedures in real data may be problematic since the latent log volatility process can be highly persistent. The formal hypothesis testing problem for the persistence parameter (concerning size and power) in the latent nonstationary stochastic volatility equation with additional measurements for volatility has not been studied in the literature, *i.e.*, all these previous studies did not exploit high-frequency information.

To be more specific, the other contributions of the paper can be summarized as follows.

First, we consider a variety of IV's for the latent log volatility, including: realized volatility (RV) measures at a different sampling frequency (*e.g.*, 1-second or 5-minute sampling), sampling scheme (calendar time or tick time), and functional form (*e.g.*, jumps or kernel). We also consider subsampled versions of some of these HF IV's; these include realized semivariance, realized range RV, nearest neighbor truncated RV, and HF principal component factors. Realized volatility measures (non-parametric volatility estimates from HF data) have received much attention among practitioners as an accurate measure of the true latent volatility under ideal market assumptions [see Andersen and Bollerslev (1998), Barndorff-Nielsen and Shephard (2001)]. Hence, we use RV measures as IV's for the daily latent volatility, in contrast with recent studies, where RV has been incorporated in traditional volatility models (GARCH or SV) by adding a measurement equation that connects the daily volatility measure and the realized volatility. It is worthwhile to note that several studies in the SV literature, such as those by Takahashi et al. (2009) and Koopman and Scharth (2012), model realized volatility and daily returns simultaneously, assuming that the realized volatility includes the market microstructure noise but still contains a great deal of information regarding the latent volatility. On the other hand, daily returns contain less noise but may not have sufficient information about the latent volatility. In GARCH-type framework, examples of such models are the Multiplicative Error Model (MEM) model [Engle and Gallo (2006)], the HEAVY (High-frEquency-bAsed VolatilitY) model [Shephard and Sheppard (2010), Noureldin et al. (2012)] and the Realized GARCH model [Hansen et al. (2012)].

Second, we propose inference methods which are robust to *weak instruments* since potential HF IV's may be weak due to *discretization errors* or *market microstructure noise*.[3] The discretization error is present in the estimates of the volatility since we only observe prices at intermittent and discrete points in time. The market microstructure noise is due to bid/ask bounces, the different price impact of different types of trades, limited liquidity, or other types of market frictions. These noises may lead to a divergence between the observed price process and the true or latent "frictionless equilibrium" price process. The literature on constructing consistent volatility proxy using HF data is considerable. These include maximum likelihood estimation [Aït-Sahalia et al. (2005)], quasi-maximum likelihood estimation [Xiu (2010)], Two Scales Realized Volatility [Zhang et al. (2005)], Multi-Scale Realized Volatility [Zhang (2006)], Realized Kernels [Hansen and Lunde (2006), Barndorff-Nielsen et al. (2008, 2011)], and Pre-Averaging volatility estimation [Jacod et al. (2009)]. Other relevant references include Bandi and Russell (2006), Fan and Wang (2007), Gatheral and Oomen (2010), Kalnina and Linton (2008), Li and Mykland (2007), and Aït-Sahalia et al. (2011). Thus incorporating noisy RV es-

[3] In IV regressions, when IV's are not valid (the identification conditions are not satisfied), the standard asymptotic theory for estimators and test statistics typically collapses. Further, when IV's are weak, the limiting distributions of standard test statistics - like Student, Wald, likelihood ratio and Lagrange multiplier criteria - have non-standard distributions and often depend heavily on nuisance parameters; see Phillips (1989), Bekker (1994), Dufour (1997), Staiger and Stock (1997), and Wang and Zivot (1998). In particular, standard Wald-type procedures based on the use of asymptotic standard errors are very unreliable in the presence of weak identification.

timates may lead to weak identification. As a result, standard inference procedures may produce invalid confidence tests and sets.

As pointed out by Dufour (1997), the statistical inference should be based on proper pivots, especially when a model involves locally almost unidentified parameters, *i.e.*, in the presence of weak IV's. The proposed inference methods include Anderson-Rubin-type (AR) tests, dynamic versions of split-sample (SS) procedure [Dufour and Jasiak (2001)], and point-optimal versions of these tests (AR^* and SS^*). The AR test is considered robust to weak IV's because the test has the correct size in cases where IV's are weak and/or strong. The SS procedure is an alternative to the AR test, where one can estimate the optimal IV's as well as any nuisance parameter. Further, appropriately splitting the sample into two parts, one for estimation of optimal IV's and nuisance parameters and the other for testing, also ensures exogeneity of the constructed IV's and the validity of the tests. Point-optimal versions of these tests gain power by exploiting the differences in the error covariance matrices under the null and the alternative; see King (1980), Dufour and King (1991), and Andrews et al. (2006).

Third, we consider a joint testing problem where we make an inference jointly on both the volatility persistence parameter and the autocorrelation parameter of the composite error (or the noise ratio). Hence, for inference on general (possibly nonlinear) transformations of model parameters [single parameter or a subvector], projection techniques can be applied [see Dufour (1989), Dufour (1990), Dufour and Jasiak (2001), Dufour and Taamouti (2005, 2007)].

Fourth, the proposed inference procedures are also robust to *dynamics*, *i.e.*, nonstationarity. Under the null hypothesis (even with nonstationary stochastic volatility) and appropriate assumptions on IV's, these tests can become pivotal functions with the possibility of exact inference.

Fifth, we employ three different sets of assumptions for the error distribution:

1. Assuming Gaussian errors, we provide confidence sets and tests based on standard Fisher critical values for the AR and SS test statistics. For point-optimal versions, we propose to use the Monte Carlo test (MCT) method [see Dwass (1957), Barnard (1963) and Dufour (2006)].

2. We assume that the conditional distribution of scale transformed error is completely specified up to an unknown scale factor, under which the Monte Carlo tests (MCT) method can apply for exact statistical inference. This assumption enables us to deal with non-standard error distributions. For example, even when errors have a heavy-tailed distribution, such as Cauchy distribution or more generally the family of stable distributions, which may not have moments and thus makes statistical inference complicated, our procedures provide exact solutions.

3. We show that the asymptotic validity of these procedures under quite general distributional assumptions.

Sixth, we study the statistical properties of the proposed inference procedures by simulation experiments. We find that the usual asymptotic t-tests fail to control the level, whereas the proposed tests control the level and show excellent power. These findings hold for several empirically realistic simulation setups, where the simulated DGPs are incorrectly specified due to the violation of independence assumption and/or misspecification of error distributions together with either weak, low- or high-frequency instruments.

Finally, we apply the proposed procedures to IBM's price and option data (2009-2013). We consider 175 different instruments spanning 22 different classes and look at their ability to describe the low-frequency volatility. The average length of confidence intervals produced by the proposed tests is used to examine the strength of the IV's. The superior instrument set constitutes of 1-, 5- and 10-minute high-frequency realized measures and call option implied volatilities. These IV's produce confidence sets where the persistence parameter lies roughly between 0.85 and 1.0. This result shows that the latent volatility process of IBM is highly persistent and close to unit-root.

Further, we find RVs with higher frequency produce wider confidence intervals than RVs with slightly lower frequency, pointing out that these confidence intervals adjust to incorporate the microstructure noise or discretization error. We also find jumps and signed jumps have no or little information content regarding the low-frequency volatility, whereas their log squared versions have a strong identification strength. When we consider irrelevant or weak instruments, such as jumps and signed jumps, the proposed procedures produce unbounded (valid) confidence sets with a non-zero probability.

This paper proceeds as follows. Section 4.2 specifies models and assumptions. Section 4.3 proposes finite-sample identification-robust inference procedures, whereas Section 4.4 extends finite-sample procedures with non-standard error distributions. Section 4.5 develops the asymptotic validity of the proposed tests. Section 4.6 presents the simulation study, and Section 4.7 presents the empirical applications. Section 4.8 offers conclusions. Proofs, Figures, and Tables are reported in the Appendix.

4.2 Framework

This paper presents extensions of the standard log-normal autoregressive SV model, which is described below following Taylor (1986), Shephard (1996), and Ghysels et al. (1996). $\mathbb{N}_0$ refers to the non-negative integers.

Assumption 4.2.1. LOG-NORMAL STOCHASTIC VOLATILITY MODEL. *The process* $\{s_t : t \in \mathbb{N}_0\}$ *follows an SV model of the type:*

$$s_t = \sigma_t z_t, \tag{4.2.1}$$

$$\log(\sigma_t^2) = \mu + \phi \log(\sigma_{t-1}^2) + v_t, \tag{4.2.2}$$

where s_t *is the return observed at time* t*, and* σ_t *is the corresponding volatility. The* z_t*'s and* v_t*'s, are i.i.d.* $\mathrm{N}(0,1)$ *and* $\mathrm{N}(0,\sigma_v^2)$ *random variables, respectively and* ϕ*,* μ*,* σ_v *are the fixed parameters of the model.*

Assumption 4.2.2. STATIONARITY. *The process* $l_t = \left(s_t, \log(\sigma_t^2)\right)'$ *is strictly stationary.*

The above assumption implies that the log-volatility follows a stationary AR(1) process with $|\phi| < 1$ and the process is initialized with $\log(\sigma_0^2) \sim N[\mu/(1-\phi), \sigma_v^2/(1-\phi^2)]$. The SV model consists of two stochastic processes, where s_t [$s_t := r_t - \mu_r$ is residual return of an asset with μ_r is the mean of return (r_t)] describes the dynamics of returns and $\log(\sigma_t^2)$ captures the dynamics of latent log volatilities. The latent process $\log(\sigma_t^2)$ in (4.2.2) can be interpreted as the random and uneven flow of new information in financial markets, while ϕ is the persistence in the volatility.

Now transforming s_t by taking logarithms of the squares, we can write the measurement equation of the model as following:

$$\begin{aligned}\log(s_t^2) &= \log(\sigma_t^2)+\log(z_t^2)=\mathbb{E}\left[\log(z_t^2)\right]+\log(\sigma_t^2)+\left\{\log(z_t^2)-\mathbb{E}\left[\log(z_t^2)\right]\right\}\\ &= \mathbb{E}\left[\log(z_t^2)\right]+\log(\sigma_t^2)+\epsilon_t \end{aligned} \tag{4.2.3}$$

where

$$\epsilon_t := \log(z_t^2)-\mathbb{E}\left[\log(z_t^2)\right]. \tag{4.2.4}$$

This transformation entails no information loss since the distribution of z_t is symmetric [see Remark 1 of Francq and Zakoïan (2006)]. Under the standard normality assumption for z_t, the transformed errors ϵ_t are i.i.d. according to the distribution of a centered $\log(\chi^2_{(1)})$ random variable with $\mathbb{E}\left[\log(z_t^2)\right] \simeq -1.2704$, $\sigma_\epsilon^2 := \mathbb{E}[\epsilon_t^2] = \mathrm{Var}\left(\log(z_t^2)\right) = \pi^2/2$ and $\mathbb{E}[\epsilon_t^4] = \pi^4 + 3\sigma_\epsilon^2$ [see Abramowitz and Stegun (1970)]. Notice that the model expressed by (4.2.3) can be written as

$$y_t = w_t + \epsilon_t \tag{4.2.5}$$

where

$$y_t := \log(s_t^2)-\mathbb{E}\left[\log(z_t^2)\right], \quad w_t := \log(\sigma_t^2). \tag{4.2.6}$$

Combining (4.2.2) and (4.2.5), we have a linear state space representation for the SV model. Given initial condition of the variables, the SV model [in Assumption 4.2.1] can be written as following

$$\text{State Transition Equation:} \quad w_t = \mu + \phi w_{t-1} + v_t \tag{4.2.7}$$

$$\text{Measurement Equation:} \quad y_t = w_t + \epsilon_t \tag{4.2.8}$$

where w_t is the logarithm of latent daily volatility, y_t is a logarithm of daily squared returns, the matrix X_t is a set of exogenous variables which may predict the latent volatility as well as capture the leverage effect [X_t also includes the constant term in the model], while v and ϵ are the disturbances.

It is evident from (4.2.7)-(4.2.8) that using any proxy for latent volatility (*e.g.*, replacing w_t by

y_t) will induce a measurement error problem. Further, the latent volatility process introduces a moving average of measurement errors. We could alleviate this type of problem by using an IV regression where we replace the unobserved variables by their proxies.

In the following assumption, we introduce generalized stochastic volatility models, where we incorporate valid IV's $\bar{Z}_{t-2}$ which are related to w_{t-1} and uncorrelated to ϵ_{t-1}.

Assumption 4.2.3. GENERALIZED STOCHASTIC VOLATILITY MODEL. *The process* $\{y_t : t \in \mathbb{N}_0\}$ *satisfies the following equations:*

$$\begin{aligned} \textit{State Transition Equation:} \quad w &= \phi w_{-1} + X\beta + v && (4.2.9) \\ \textit{Measurement Equation:} \quad y &= w + \epsilon && (4.2.10) \\ \textit{Instrument Equation:} \quad w_{-1} &= \bar{Z}_{-2}\bar{\pi} + u_{-1} && (4.2.11) \end{aligned}$$

where $w = (w_1, \dots, w_T)'$, $w_{-1} = (w_0, \dots, w_{T-1})'$, $y = (y_1, \dots, y_T)'$ *are* $T \times 1$ *vector,* $X = [X_1', \dots, X_T']'$ *is a* $T \times k$ *matrix of exogenous explanatory variables which may predict the latent volatility as well as capture the leverage effect,* $\bar{Z}_{-2} = [\bar{Z}_{-1}', \dots, \bar{Z}_{T-2}']'$ *is a* $T \times m$ *matrix of of variables related to* w_{-1}, *while* $\epsilon = (\epsilon_1, \dots, \epsilon_T)'$, $v = (v_1, \dots, v_T)'$, $u_{-1} = (u_0, \dots, u_{T-1})'$ *are* $T \times 1$ *vector of disturbances. The matrices of unknown coefficients* ϕ, β , *and* $\bar{\pi}$ *have dimensions respectively* 1×1, $k \times 1$, *and* $m \times 1$.

We do not impose any stationary restriction on the latent volatility process. The assumption that the latent autoregressive volatility process is first-order is not essential to the analysis. Indeed, higher-level dynamics could be allowed, but in this paper we focus on the first-order case.

To derive finite distributional theory for test statistics (proposed in Section 4.3), we employ the following assumptions.

Assumption 4.2.4. INDEPENDENCE. *The* $T \times k$ *matrix* X *and* $T \times m$ *matrix* $\bar{Z}_{-2}$ *are independent of* $T \times 1$ *vectors* v *and* ϵ.

Assumption 4.2.5. FULL RANK. $\operatorname{rank}(X) = k$, $1 \le \operatorname{rank}(\bar{Z}_{-2}) = m < T$, $1 \le \operatorname{rank}[Z_{-2}, X_1, X_2] = l + k < T$, *where* Z_{-2}, X_1, *and* X_2 *are* $T \times l$, $T \times k_1$, *and* $T \times k_2$ *matrices, respectively with* $k = k_1 + k_2$ *and* $m = l + k_2$.

Assumption 4.2.6. GAUSSIAN NOISE. *The ϵ_t's and v_t's are i.i.d.* $\mathrm{N}(0,\sigma_\epsilon^2)$ *and* $\mathrm{N}(0,\sigma_v^2)$ *random variables, respectively.*

In order to handle common variables (*e.g.*, the constant term) in equations (4.2.9) and (4.2.11), Assumption 4.2.5 allows for the presence of common columns in the matrices $\bar{Z}_{-2}$ and X. If $\bar{Z}_{-2}$ and X have k_2 columns in common ($0 \le k_2 < m$) then the other k_1 columns of X are linearly independent of $\bar{Z}_{-2}$. It is important to note that no restriction is imposed on the distribution of u and it may follow any distribution (heteroskedastic or autocorrelated) since no statistical property of u has effects on the validity of the tests proposed in this paper.

Note that we change the distributional assumption of ϵ_t by an i.i.d. $\mathrm{N}(0,\sigma_\epsilon^2)$ distribution. This is consistent with several previous studies where the distribution of ϵ_t is approximated by a normal distribution characterized by a mean of zero and a variance of $\pi^2/2$ [see Harvey et al. (1994), Ruiz (1994), Breidt and Carriquiry (1996), Harvey and Shephard (1996), Kim et al. (1998), Chib et al. (2002), Knight et al. (2002), Broto and Ruiz (2004), Omori et al. (2007)]. We relax the above assumptions in Sections 4.4 and 4.5.

The IV regression requires valid IV's for the observable volatility proxy y_t, which is typically the low-frequency (LF) daily squared return. As a result, IV's are also connected to the logarithm of latent daily volatility [see equation (4.2.11)]. To find valid IV's, we first look at the properties of the observed volatility proxy y_t. If y_t is autocorrelated with a sufficiently long lag and the ϵ_t's are uncorrelated, then the lag values of observed proxy $(y_{t-2}, y_{t-3}, y_{t-4}, \ldots)$ are potential clean IV's for y_{t-1}. It is important to not introduce y's with too high lags as IV's, because this requires truncating the sample in order to observe IV's for each date used in the estimation, and the good statistical properties of the IV method begins to break down. We can also use realized volatility as IV's ($\bar{Z}_{t-2}$ contains past realized volatilities) since HF price data contain valuable information regarding the latent volatility. In the Section 4.7 below, we consider not only daily and HF IV's but also consider option implied volatility as IV's.

4.3 Finite-sample procedures

In this section, we consider the problem of testing the volatility persistence parameter in a GSV model given in Assumption 4.2.3, *i.e.*, testing restriction about volatility clustering. We propose four finite-sample procedures, which are valid under Assumptions 4.2.4-4.2.6. Let us now consider the null hypothesis:

$$H_0 : \phi = \phi_0 \,. \tag{4.3.1}$$

We consider an instrument substitution method, which is based on replacing unobserved variables with a set of IV's. First, we substitute (4.2.10) into (4.2.9):

$$y = \phi y_{-1} + X\beta + v + \epsilon - \phi \epsilon_{-1} \,. \tag{4.3.2}$$

Subtracting $\phi_0 y_{-1}$ on both sides of (4.3.2), we then get:

$$y - \phi_0 y_{-1} = (\phi - \phi_0) y_{-1} + X\beta + v + \epsilon - \phi \epsilon_{-1} \,. \tag{4.3.3}$$

Since $\mathbb{E}[y_{t-1}\epsilon_{t-1}] \neq 0$, we need to find IV's for w_{-1} to solve this endogeneity problem. Substituting (4.2.10) into (4.2.11), we have

$$y_{-1} = \bar{Z}_{-2}\bar{\pi} + \eta_{-1} \,, \tag{4.3.4}$$

where $\eta_{-1} := \epsilon_{-1} + u_{-1}$. From Assumption 4.2.4, we can see that $\bar{Z}_{-2}$ is independent of ϵ_{-1}. We substitute (4.3.4) into (4.3.3):

$$\begin{aligned} y - \phi_0 y_{-1} &= \bar{Z}_{-2}\bar{\pi}(\phi - \phi_0) + X\beta + (\phi - \phi_0)\eta_{-1} + v + \epsilon - \phi\epsilon_{-1} \\ &= \bar{Z}_{-2}\bar{\pi}(\phi - \phi_0) + X\beta + (\phi - \phi_0)[\epsilon_{-1} + u_{-1}] + v + \epsilon - \phi\epsilon_{-1} \\ &= \bar{Z}_{-2}\bar{\pi}(\phi - \phi_0) + X\beta + (\phi - \phi_0) u_{-1} + v + \epsilon - \phi_0\epsilon_{-1} \end{aligned}$$

or equivalently,

$$y - \phi_0 y_{-1} = \bar{Z}_{-2}\bar{\pi}(\phi - \phi_0) + X\beta + \xi \tag{4.3.5}$$

where

$$\xi := (\phi - \phi_0) u_{-1} + v + \epsilon - \phi_0 \epsilon_{-1} \,. \tag{4.3.6}$$

Using Assumption 4.2.5, we can write (4.3.5) as

$$y-\phi_0 y_{-1} = Z_{-2}\delta + X\beta_* + \xi \tag{4.3.7}$$

where

$$\delta := \bar{\pi}_1(\phi-\phi_0)\,, \quad \beta_* := (\beta_1', \beta_{2*}')'\,, \quad \beta_{2*} := \beta_2 + \bar{\pi}_2(\phi-\phi_0)\,, \quad \bar{\pi} := (\bar{\pi}_1', \bar{\pi}_2')'\,, \tag{4.3.8}$$

and $\bar{\pi}_i$ is a $k_i \times 1$ vector.

4.3.1 Anderson-Rubin-type procedure

Since $\epsilon_t - \phi_0\epsilon_{t-1}$ is an MA(1) process, thus ξ's are serially correlated. However, when $\phi = \phi_0 = 0$, ξ is distributed $\mathrm{N}(0, \sigma_\xi^2 I_T)$ where $\sigma_\xi^2 = \sigma_v^2 + \sigma_\epsilon^2$. As a result, the model in equation (4.3.7) satisfies all the assumptions of the classical linear model when $\phi_0 = 0$. Furthermore, since $\delta = 0$ when $\phi = \phi_0$, we can test H_0 by a standard F-test of the following null hypothesis:

$$H_0^* : \delta = 0\,. \tag{4.3.9}$$

This F-statistic has the form

$$AR(\phi_0) = \frac{(y-\phi_0 y_{-1})'(M[X]-M[X,Z_{-2}])(y-\phi_0 y_{-1})/l}{(y-\phi_0 y_{-1})'M[X,Z_{-2}](y-\phi_0 y_{-1})/(T-l-k)} \tag{4.3.10}$$

where $M(A) = I - A(A'A)^{-1}A'$.

$AR(\phi_0)$ can be interpret as an Anderson-Rubin-type statistic. When the normality assumption holds $\left[\xi \sim \mathrm{N}(0, \sigma_\xi^2 I_T)\right]$, and X and Z_{-2} are exogenous, we have $AR(\phi_0) \sim F(l, T-l-k)$, and $H_0(\phi_0)$ can be tested by using a critical region of the form $\left\{AR(\phi_0) > f(\alpha)\right\}$ where $f(\alpha) = F_\alpha(l, T-l-k)$ is the $(1-\alpha)$-quantile of the $F(l, T-l-k)$ distribution. A confidence set with level $1-\alpha$ for ϕ is then given by

$$C_\phi(\alpha) = \left\{\phi_0 : AR(\phi_0) \le F_\alpha(l, T-l-k)\right\} = \left\{\phi : Q(\phi) \le 0\right\} \tag{4.3.11}$$

where $Q(\phi) = \phi' A\phi + b'\phi + c$, $A = y_{-1}' H y_{-1}$, $b = -2y_{-1}' H y$, $c = y' H y$, $H = M[X] - [1 + f(\alpha)(l/T -$

$l-k)]M[X,Z_{-2}]$, and $f(\alpha)=F_{\alpha}(l,T-l-k)$; see Dufour and Taamouti (2005).[4]

Unfortunately, this property does not extend to a more general $AR(\phi_0)$ statistic where $\phi_0 \neq 0$ because in this case under the H_0, the composite error ξ_t is not i.i.d.. When $\phi_0 \neq 0$, it is easy to see that the model (4.3.7) under H_0 does not satisfy all the assumptions of the classical linear model. In this case, under the null hypothesis, $\xi = \upsilon + \epsilon - \phi_0\epsilon_{-1}$ is an MA(1) process which makes the standard t-tests and F-tests are invalid because the standard errors are wrong. We could correct the standard errors by a Generalized Least Squares (GLS) type transformation. The model defined by equation (4.3.7) can be transformed under the H_0 to a model such that the AR-type tests will be valid, and the distribution of the test statistic will follow the F-distribution. Now, under H_0,

$$\xi = \upsilon + \epsilon - \phi_0\epsilon_{-1}$$

is an MA(1) process. Under Assumption 4.2.6, $\xi \sim \mathrm{N}\big[0, \sigma_{\xi}^2 \Sigma(\rho)\big]$ where

$$\Sigma(\rho) := \begin{pmatrix} 1 & -\rho & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ -\rho & 1 & -\rho & 0 & & & & \vdots \\ 0 & -\rho & 1 & -\rho & \ddots & & & \vdots \\ \vdots & 0 & \ddots & \ddots & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & \ddots & 0 & \vdots \\ \vdots & & & \ddots & -\rho & 1 & -\rho & 0 \\ \vdots & & & & 0 & -\rho & 1 & -\rho \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 & -\rho & 1 \end{pmatrix}, \tag{4.3.12}$$

$$\sigma_{\xi}^2 := (1+\phi_0^2)\sigma_{\epsilon}^2 + \sigma_{\upsilon}^2, \tag{4.3.13}$$

$$\rho := \frac{-\mathrm{Cov}(\xi_t \xi_{t-1})}{\mathrm{Var}(\xi_t)} = \frac{\phi_0 \sigma_{\epsilon}^2}{(1+\phi_0^2)\sigma_{\epsilon}^2 + \sigma_{\upsilon}^2}. \tag{4.3.14}$$

Clearly, ρ is a function of ϕ_0, σ_{υ}^2, and σ_{ϵ}^2. $\Sigma(\rho)$ is a *Toeplitz* matrix (or diagonal-constant matrix) with dimension $T \times T$. Because $\Sigma(\rho)$ is a symmetric positive-definite matrix, there exists a

[4]When the disturbances are i.i.d with finite fourth-order moments, the AR-statistic converges under H_0 to a χ^2 distributed random variable when the sample size gets large. This large sample distribution of the AR-statistic does not depend on the value of $\bar{\pi}$ which makes it a more reliable statistic for practical purposes than the Wald statistic.

$T \times T$ matrix C, such that $C\Sigma(\rho)C' = I_T$. If the $\Sigma(\rho)$ matrix is known, then we can propose the following transformation. Multiply equation (4.3.7) by C to make the error covariance matrix to an identity matrix. However, ρ is not known. On setting the noise ratio

$$\lambda := \sigma_\epsilon^2/\sigma_v^2 \ \in [0,\ \infty), \tag{4.3.15}$$

we can write ρ as

$$\rho(\phi_0, \lambda) = \frac{\phi_0 \lambda}{(1+\phi_0^2)\lambda + 1}. \tag{4.3.16}$$

We can do a joint test such that under the null ρ is known. In any economic model, the disturbances contain important information, and particularly in the context of serially correlated models, researchers may be interested in joint inference. Consider the following null hypothesis:

$$H_0(\phi_0, \lambda_0) : \phi = \phi_0, \quad \lambda = \lambda_0. \tag{4.3.17}$$

Under $H_0(\phi_0, \lambda_0)$, we can write

$$\rho_0 := \frac{\phi_0 \lambda_0}{(1+\phi_0^2)\lambda_0 + 1} \in [-1/2, 1/2], \tag{4.3.18}$$

and the joint null hypothesis [in (4.3.17)] becomes

$$\bar{H}_0(\phi_0, \rho_0) : \phi = \phi_0, \quad \rho = \rho_0. \tag{4.3.19}$$

Under $\bar{H}_0(\phi_0, \rho_0)$, we have:

$$\lambda_0 = \frac{\rho_0}{\phi_0 - \rho_0(1+\phi_0)^2} \ \in [0,\ \infty). \tag{4.3.20}$$

See Table 4.1 for the corresponding values of λ_0.

Since ρ_0 is known under $H_0(\phi_0, \lambda_0)$ or $\bar{H}_0(\phi_0, \rho_0)$, we can consider the following transformed model:

$$C_0(y - \phi_0 y_{-1}) = C_0 Z_{-2}\delta + C_0 X \beta_* + C_0 \xi \tag{4.3.21}$$

where $C_0 = C(\rho_0)$ is a $T \times T$ matrix such that $C_0 \Sigma(\rho_0) C_0' = I_T$. The variance-covariance matrix of $\xi^* := C_0(\xi)$ is now an i.i.d. $N(0, \sigma_\xi^2 I_T)$ distribution. The F-statistic for testing $\delta = 0$ (or $\phi = \phi_0$) in (4.3.21) is:

$$AR(\phi_0, \rho_0) = \frac{y(\phi_0, \rho_0)'(M_{C_0}[X] - M_{C_0}[X, Z_{-2}])y(\phi_0, \rho_0)/l}{y(\phi_0, \rho_0)' M_{C_0}[X, Z_{-2}]y(\phi_0, \rho_0)/(T - l - k)} \tag{4.3.22}$$

where $y(\phi_0, \rho_0) = C_0(y - \phi_0 y_{-1})$, $M_{C_0}[A] = I - A[A'\Sigma(\rho_0)^{-1}A]^{-1}A'\Sigma(\rho_0)^{-1}$. A central feature of most situations where IV methods are required come from the fact that IV's may be used to solve an endogeneity or an errors-in-variables problem. It is very rare that one can or should use all the possible valid IV's. A drawback of the AR method is that it loses power when too many IV's are used. However, the AR procedure is *robust to missing IV's* (or *instrument exclusion*) [see Dufour and Taamouti (2007)]. Alternative methods of inference aimed at being robust to weak identification [Wang and Zivot (1998), Kleibergen (2002), Moreira (2003)] do not enjoy this type of robustness. In the case of feasible GLS-type transformations, where ρ is replace by an estimate $\hat{\rho}$, the test statistic is no longer F-distributed, but it converges under $\bar{H}_0(\phi_0, \rho_0)$ to a χ^2 distribution in large samples. The tests and confidence sets obtained by the instrument substitution method can be interpreted as likelihood ratio (LR) procedures (based on appropriately chosen reduced form alternatives), or equivalently as profile likelihood techniques [for further discussion of such techniques, see Bates and Watts (1988), Meeker and Escobar (1995) and Chen and Jennrich (1996)].

4.3.2 Anderson-Rubin-type point-optimal procedure (AR^*)

In this section, we propose a point-optimal (PO) version of AR-type tests. PO tests provide simple and effective methods for creating exact small sample tests with excellent power properties in a wide variety of problems in linear regression. The empirical evidence in the literature indicates that in general, PO tests often outperform other testing methods in terms of power. Besides, exact small-sample critical values for PO tests can be computed in most cases. Thus, one does not have to rely on the asymptotic properties of the test statistic to make inferences. For a general review of PO tests, the reader may consult King (1980), King (1987) and Dufour and King (1991).

Following Dufour and King (1991), a PO test of $\rho = \rho_0$ against $\rho = \rho_1$ under Gaussian assumptions given as

$$S(\rho_0,\rho_1) = \frac{\hat{\xi}'\Sigma(\rho_0)^{-1}\hat{\xi}}{\tilde{\xi}'\Sigma(\rho_1)^{-1}\tilde{\xi}}, \tag{4.3.23}$$

where $|\rho_0| \leq 1/2$, $|\rho_1| \leq 1/2$, and $\hat{\xi}$ and $\tilde{\xi}$ are the GLS residual vectors corresponding to covariance matrices $\Sigma(\rho_0)$ and $\Sigma(\rho_1)$, respectively. The test rejects the null for large values of $S(\rho_0,\rho_1)$. However, the choice of ρ_1 is important. To obtain a test of $\rho = \rho_0$ against $\rho > \rho_0$, we select a value of ρ_1, such that $\rho_0 < \rho_1 \leq 1/2$ and apply the test based on $S(\rho_0,\rho_1)$. Similarly, testing $\rho = \rho_0$ against $\rho < \rho_0$, we select ρ_1, such that $-1/2 \leq \rho_1 < \rho_0$. For example, we may choose ρ_1 such that $\rho_1 = \rho_0 - \bar{\Delta}$ where $0 < \bar{\Delta} < 1$. The test based on (4.3.23) is point-optimal, and it gains power by exploiting the differences in the error covariance matrices under the null and the alternative.

As pointed out by King (1987), a PO test can be viewed as a partition of the sample space into two regions, a rejection region and a non-rejection region. If the observed sample falls in the rejection region, the null is rejected. Otherwise, the null is not rejected. Consider an AR-type PO test statistic $\overline{AR}(\phi_0,\rho_0,\rho_1)$ similar to (4.3.23) for $\rho = \rho_0$ against $\rho = \rho_1$ (under $\phi = \phi_0$):

$$\overline{AR}(\phi_0,\rho_0,\rho_1) = \frac{y(\phi_0,\rho_0)' M_{C_0}[X]\, y(\phi_0,\rho_0)}{y(\phi_0,\rho_1)' M_{C_1}[X, Z_{-2}]\, y(\phi_0,\rho_1)} \tag{4.3.24}$$

where

$$y(\phi_0,\rho_0) = C_0(y - \phi_0 y_{-1}), \quad y(\phi_0,\rho_1) = C_1(y - \phi_0 y_{-1}), \quad M_{C_i}[A] = I - A\big[A'\Sigma(\rho_i)^{-1}A\big]^{-1}A'\Sigma(\rho_i)^{-1}, i = 0,1.$$

Note that it is difficult to derive the analytical null distribution of (4.3.24) even under the Gaussian assumption, while the MCT method described in Section 4.4 can be implemented and confidence set for ϕ and ρ with level $(1-\alpha)$ is obtained by inverting the tests.

It is worth noting that $\overline{AR}(\phi_0,\rho_0,\rho_1)$ can become degenerate in the limit. Thus we consider a monotonic transformation of $\overline{AR}(\phi_0,\rho_0,\rho_1)$, which is given as:

$$AR^*(\phi_0,\rho_0,\rho_1) = T\big[\overline{AR}(\phi_0,\rho_0,\rho_1) - 1\big]. \tag{4.3.25}$$

For finite-sample inference, both $\overline{AR}(\phi_0,\rho_0,\rho_1)$ and $AR^*(\phi_0,\rho_0,\rho_1)$ lead to identical results since a monotonic transformation does not change the rank of the statistic in the MCT method. On the other hand, $AR^*(\phi_0, \rho_0, \rho_1)$ is more appropriate for proving the asymptotic validity.

4.3.3 Split-sample-type procedure

Finite-sample inferences similar to the previous section may alternatively be obtained by applying a split-sample technique. If ρ (or λ) can be estimated from data, then the estimated ρ tends to be closer to the true one than those that are arbitrarily selected, and thus the power of the test can be improved. However, if we re-use the data (which is used to estimate ρ) then the test statistic is no longer F-distributed. Thus, we employ the split-sample technique, which splits the sample into two parts. The first subsample is used to construct IV's and an estimate of ρ, and the second subsample is used to implement the test. Note that we can estimate ρ as well as the optimal IV's using the first subsample. As a result, the number of IV's can be reduced to the number of endogenous variables.

It is a natural thing to replace $\bar{Z}_{-2}\bar{\pi}$ by $\bar{Z}_{-2}\widehat{\pi}$, where $\widehat{\pi}$ is an estimator of $\bar{\pi}$. One could use $\widehat{\pi} = (\bar{Z}_{-2}{}'\bar{Z}_{-2})^{-1}\bar{Z}_{-2}{}'y_{-1}$, the least squares estimate of $\bar{\pi}$ based on (4.2.11). Then we have:

$$y - \phi_0 y_{-1} = \bar{Z}_{-2}\widehat{\pi}(\phi - \phi_0) + X\beta + [\xi + \bar{Z}_{-2}(\bar{\pi} - \widehat{\pi})(\phi - \phi_0)] = \hat{y}_{-1}\delta_* + X\beta + \bar{\xi} \tag{4.3.26}$$

where

$$\delta_* := (\phi - \phi_0), \quad \hat{y}_{-1} := \bar{Z}_{-2}\widehat{\pi}, \quad \bar{\xi} := v + \epsilon - \phi_0\epsilon_{-1} + [u_{-1} + \bar{Z}_{-2}(\bar{\pi} - \widehat{\pi})](\phi - \phi_0).$$

Again, the null hypothesis ($\phi = \phi_0$) may be assessed by testing $H_0^{**} : \delta_* = 0$ in (4.3.26). Here the standard AR-statistic for H_0^{**} is obtained by replacing Z_{-2} by $\hat{y}_{-1}$ in (4.3.10):

$$AR(\phi_0; \hat{y}_{-1}) = \frac{(y - \phi_0 y_{-1})'(M[X] - M[X, \hat{y}_{-1}])(y - \phi_0 y_{-1})/l}{(y - \phi_0 y_{-1})'M[X, \hat{y}_{-1}](y - \phi_0 y_{-1})/(T - l - k)}. \tag{4.3.27}$$

This test statistic is valid only when $\phi = \phi_0 = 0$, since in this case, $\hat{y}_{-1}$ and $\bar{\xi}$ are independent, and conditional on $\hat{y}_{-1}$, the equation (4.3.26) satisfies all the assumptions of the classical linear model. Thus, the null distribution of the statistic $AR(0; \hat{y}_{-1})$ for testing $\phi_0 = 0$

is $F(l, T-l-k)$. Unfortunately, this property does not extend to a more general statistic $AR(\phi_0; \hat{y}_{-1})$ where $\phi_0 \neq 0$ because $\hat{y}_{-1}$ and $\bar{\xi}$ are not independent in this case. In order to deal with more general hypotheses, we need to take care of two things:

1. get an estimate $\widetilde{\pi}$ of $\bar{\pi}$ such that $\hat{y}_{-1}$ $(= \bar{Z}_{-2}\widetilde{\pi})$ and $\bar{\xi}$ are independent;

2. since under the null $\bar{\xi}$ is an MA(1) process, we need an estimate of $\rho(\phi_0) = \text{Cov}(\bar{\xi}_t \bar{\xi}_{t-1})/\text{Var}(\bar{\xi}_t)$ for the GLS-type transformation.

In particular, the split-sample procedure is as follows. Split the sample into subsample (1) with sample size T_1: $y^{(1)}, X^{(1)}, \bar{Z}^{(1)}$ and subsample (2) with sample size T_2: $y^{(2)}, X^{(2)}, \bar{Z}^{(2)}$ where

$$y = \begin{pmatrix} y^{(1)} \\ y^{(2)} \end{pmatrix}, \quad X = \begin{pmatrix} X^{(1)} \\ X^{(2)} \end{pmatrix}, \quad Z = \begin{pmatrix} \bar{Z}^{(1)} \\ \bar{Z}^{(2)} \end{pmatrix}. \tag{4.3.28}$$

we use the first subsample to estimate $\bar{\pi}$ using

$$\widetilde{\pi}^{(1)} = \left(\bar{Z}_{-2}^{(1)\prime} \bar{Z}_{-2}^{(1)}\right)^{-1} \bar{Z}_{-2}^{(1)\prime} y_{-1}^{(1)}$$

and second subsample to construct the following regression:

$$y^{(2)} - \phi_0 y_{-1}^{(2)} = \bar{Z}_{-2}^{(2)} \widetilde{\pi}^{(1)} (\phi - \phi_0) + X^{(2)}\beta + \bar{\xi}^{(2)} = \hat{y}_{-1}^{(2)} \delta_* + X^{(2)}\beta + \bar{\xi}^{(2)} \tag{4.3.29}$$

where

$$\delta_* := (\phi - \phi_0), \quad \hat{y}_{-1}^{(2)} := \bar{Z}_{-2}^{(2)} \widetilde{\pi}^{(1)}, \quad \bar{\xi}^{(2)} := v^{(2)} + \epsilon^{(2)} - \phi_0 \epsilon_{-1}^{(2)} + \left[u_{-1}^{(2)} + \bar{Z}_{-2}^{(2)}(\bar{\pi} - \widehat{\pi})\right](\phi - \phi_0).$$

In equation (4.3.29), $\hat{y}_{-1}^{(2)}$ and $\bar{\xi}^{(2)}$ are independent. Now under the null hypothesis $(\phi = \phi_0)$, $\bar{\xi}^{(2)} = v^{(2)} + \epsilon^{(2)} - \phi_0 \epsilon_{-1}^{(2)}$ is an MA(1) process and the variance covariance matrix of $\bar{\xi}^{(2)}$ is $\sigma^2_{\bar{\xi}^{(2)}} \Sigma_{\bar{\xi}^{(2)}}$ where $\sigma^2_{\bar{\xi}^{(2)}} = (1 + \phi_0^2)\sigma^2_{\epsilon^{(2)}} + \sigma^2_{v^{(2)}}$ and $\Sigma_{\bar{\xi}^{(2)}}$ is a *Toeplitz* matrix [similar to equation (4.3.12)] with $\rho^{(2)} = \text{Cov}(\bar{\xi}_t^{(2)} \bar{\xi}_{t-1}^{(2)})/\text{Var}(\bar{\xi}_t^{(2)})$. Since $\Sigma_{\bar{\xi}^{(2)}}$ is a $T_2 \times T_2$ symmetric positive-definite matrix, there exists a $T_2 \times T_2$ matrix $C\big(\rho^{(2)}\big)$, such that $C\big(\rho^{(2)}\big) \Sigma_{\bar{\xi}^{(2)}} C\big(\rho^{(2)}\big)' = I_{T_2}$. If the $\rho^{(2)}$ is known then we can multiply equation (4.3.29) by $C(\rho^{(2)})$ to make the error covariance matrix as an identity matrix. If we use an estimates of $\rho^{(2)}$ from second subsample then the test statistic is no longer

F-distributed, it is converges under H_0 to a χ^2 distributed random variable in large sample. In order to solve this problem, we use an estimate of $\rho^{(2)}$ from the first subsample that is $\hat{\rho}^{(1)}$. This transformation gives us the following test statistic:

$$SS\big(\phi_0;\hat{y}_{-1}^{(2)},\hat{\rho}^{(1)}\big)=\frac{y^{(2)}\big(\phi_0,\hat{\rho}^{(1)}\big)'\big(M_{C(\hat{\rho}^{(1)})}[X^{(2)}]-M_{C(\hat{\rho}^{(1)})}[X^{(2)},\hat{y}_{-1}^{(2)}]\big)y^{(2)}\big(\phi_0,\hat{\rho}^{(1)}\big)/l}{y^{(2)}\big(\phi_0,\hat{\rho}^{(1)}\big)'M_{C(\hat{\rho}^{(1)})}[X^{(2)},\hat{y}_{-1}^{(2)}]y^{(2)}\big(\phi_0,\hat{\rho}^{(1)}\big)/(T_2-l-k)} \tag{4.3.30}$$

where

$$y^{(2)}\big(\phi_0,\hat{\rho}^{(1)}\big)=C\big(\hat{\rho}^{(1)}\big)\big(y^{(2)}-\phi_0 y_{-1}^{(2)}\big),\qquad M_{C(\hat{\rho}^{(1)})}[A]=I-A\big[A'\Sigma(\hat{\rho}^{(1)})^{-1}A\big]^{-1}A'\Sigma(\hat{\rho}^{(1)})^{-1}.$$

This test statistic follows a $F(l,T_2-l-k)$ distribution when $\phi=\phi_0$. Consequently, the critical region

$$SS\big(\phi_0;\hat{y}_{-1}^{(2)},\hat{\rho}^{(1)}\big)>F_\alpha(l,T_2-l-k)$$

has size α. Furthermore,

$$\bar{C}_\phi(\alpha)=\Big\{\phi_0 : SS\big(\phi_0;\hat{y}_{-1}^{(2)},\hat{\rho}^{(1)}\big)\le F_\alpha(l,T_2-l-k)\Big\}$$

is a confidence set for ϕ with size $1-\alpha$ and this confidence set takes a form similar to (4.3.11). A test statistic for H_0: $\phi=\phi_0$ and $\rho=\rho_0$ is:

$$SS\big(\phi_0,\rho_0;\hat{y}_{-1}^{(2)}\big)=\frac{y^{(2)}\big(\phi_0,\rho_0\big)'\big(M_{C_0}[X^{(2)}]-M_{C_0}[X^{(2)},\hat{y}_{-1}^{(2)}]\big)y^{(2)}\big(\phi_0,\rho_0\big)/l}{y^{(2)}\big(\phi_0,\rho_0\big)'M_{C_0}[X^{(2)},\hat{y}_{-1}^{(2)}]y^{(2)}\big(\phi_0,\rho_0\big)/(T_2-l-k)} \tag{4.3.31}$$

where

$$y^{(2)}\big(\phi_0,\rho_0\big)=C_0\big(y^{(2)}-\phi_0 y_{-1}^{(2)}\big),\qquad M_{C_0}[A]=I-A\big[A'\Sigma(\rho_0)^{-1}A\big]^{-1}A'\Sigma(\rho_0)^{-1}.$$

It is noteworthy that we should be careful about the order of the subsamples (1) and (2). The order does not matter in a static model but it does in a dynamic model. If we use the second subsample to estimate parameters, then the estimators include $y^{(2)}$ and $y_{-1}^{(2)}$, which have past errors inside. As a result, $\hat{y}_{-1}^{(1)}=\bar{Z}_{-2}^{(1)}\tilde{\pi}^{(2)}$ and $\bar{\xi}^{(1)}$ are not independent and the inference procedure does not control its level correctly. Therefore we should use the first part of the sample to get the estimates. A crucial issue in the split-sample tests is how to determine a splitting ratio, $\tau = T_1/T$. SS-type tests are depend on the choice of split ratio, τ and power of these tests are inversely related with τ; see Dufour and Jasiak (2001). However this ratio does not affect the validity of the test.

4.3.4 Split-sample-type point-optimal procedure (SS^*)

We now propose a split-sample version of the point-optimal procedure. The split-sample methods gives us additional flexibility for inference since we can estimate the nuisance parameter from the first sample and use it with the second sample to do the inference. Now consider a split-sample version of test statistic similar to (4.3.23) for $\rho = \rho_0$ against $\rho = \rho_1$ (under $\phi = \phi_0$):

$$\overline{SS}(\phi_0, \rho_0, \rho_1; \hat{y}_{-1}^{(2)}) = \frac{y^{(2)}(\phi_0, \rho_0)' M_{C_0}[X] y^{(2)}(\phi_0, \rho_0)}{y^{(2)}(\phi_0, \rho_1)' M_{C_1}[X, \hat{y}_{-1}^{(2)}] y^{(2)}(\phi_0, \rho_1)} \tag{4.3.32}$$

where

$$y^{(2)}(\phi_0, \rho_i) = C_i\big(y^{(2)} - \phi_0 y_{-1}^{(2)}\big), \quad M_{C_i}[A] = I - A\big[A'\Sigma(\rho_i)^{-1}A\big]^{-1} A'\Sigma(\rho_i)^{-1}, \quad i = 0, 1.$$

We can also replace ρ_1 by $\hat{\rho}^{(1)}$ to construct test that controls the level:

$$SS^*\big(\phi_0, \rho_0; \hat{y}_{-1}^{(2)}, \hat{\rho}^{(1)}\big) = \frac{y^{(2)}(\phi_0, \rho_0)' M_{C_0}[X] y^{(2)}(\phi_0, \rho_0)}{y^{(2)}(\phi_0, \hat{\rho}^{(1)})' M_{C(\hat{\rho}^{(1)})}[X, \hat{y}_{-1}^{(2)}] y^{(2)}(\phi_0, \hat{\rho}^{(1)})} \tag{4.3.33}$$

or, for testing $\phi = \phi_0$,

$$SS^*\big(\phi_0; \hat{y}_{-1}^{(2)}, \hat{\rho}^{(1)}\big) = \frac{y^{(2)}(\phi_0, \hat{\rho}^{(1)})' M_{C(\hat{\rho}^{(1)})}[X] y^{(2)}(\phi_0, \hat{\rho}^{(1)})}{y^{(2)}(\phi_0, \hat{\rho}^{(1)})' M_{C(\hat{\rho}^{(1)})}[X, \hat{y}_{-1}^{(2)}] y^{(2)}(\phi_0, \hat{\rho}^{(1)})} \tag{4.3.34}$$

where

$$y^{(2)}(\phi_0, \rho_0) = C(\rho_0)\big(y^{(2)} - \phi_0 y_{-1}^{(2)}\big), \quad y^{(2)}(\phi_0, \hat{\rho}^{(1)}) = C\big(\hat{\rho}^{(1)}\big)\big(y^{(2)} - \phi_0 y_{-1}^{(2)}\big),$$

$$M_{C(\hat{\rho}^{(1)})}[A] = I - A\big[A'\Sigma(\hat{\rho}^{(1)})^{-1}A\big]^{-1} A'\Sigma(\hat{\rho}^{(1)})^{-1}.$$

Again it is difficult to derive the analytical null distribution of (4.3.32) or (4.3.33) under the Gaussian assumption, while the MCT method described in Section 4.4 can be implemented and a confidence set for ϕ and ρ with level $(1-\alpha)$ is obtained by inverting the tests.

Further, note that $\overline{SS}(\phi_0, \rho_0, \rho_1; \hat{y}_{-1}^{(2)})$ can become degenerate in the limit whereas a monotonic transformation of $\overline{SS}(\phi_0, \rho_0, \rho_1; \hat{y}_{-1}^{(2)})$, given by

$$SS^*\big(\phi_0, \rho_0, \rho_1; \hat{y}_{-1}^{(2)}\big) = T_2\big[SS^*(\phi_0, \rho_0, \rho_1; \hat{y}_{-1}^{(2)}) - 1\big]. \tag{4.3.35}$$

is more for asymptotic theory.

Again, for finite-sample inference, both $\overline{SS}(\phi_0, \rho_0, \rho_1; \hat{y}_{-1}^{(2)})$ and $SS^*\big(\phi_0, \rho_0, \rho_1; \hat{y}_{-1}^{(2)}\big)$ lead to identical results since a monotonic transformation does not change the rank of the statistic in the MCT method.

4.3.5 Inference on general transformations

In Sections 4.3.1-4.3.4, we make joint inference on $(\phi, \rho)'$. These tests are based on extensions of Anderson-Rubin statistics and designed to test hypotheses fixing the entire vector of the endogenous (or unobserved) regressor coefficients. When one is interested in its subsets, or more generally in any functions of the parameters, projection technique can be applied; see Dufour (1989), Dufour and Jasiak (2001), Dufour and Taamouti (2005, 2007).

Let $\theta := (\phi, \rho)'$ for notational convenience. A confidence set associated with one of the tests for $H_0(\theta_0) : \theta = \theta_0$ in the previous subsections can be written as

$$C_\alpha(\theta) = \big\{\theta_0 \mid H_0(\theta_0) \text{ is not rejected}\big\}. \tag{4.3.36}$$

If the test has level α, the confidence set $C_\alpha(\theta)$ has level $1-\alpha$. Note that all the four tests are based on pivotal functions and have size α. Thus, the confidence sets in (4.3.36) from these tests have size $1-\alpha$.

Now consider an arbitrary (possibly nonlinear) transformation $\delta = g(\theta)$ of θ, then a confidence set of δ, with the level at least $1-\alpha$, can be constructed as

$$C_\alpha(\delta) = \left\{\delta_0 \mid \delta_0 = g(\theta) \text{ for some } \theta \in C_\alpha(\theta)\right\}. \tag{4.3.37}$$

Since $\theta \in C_\alpha(\theta)$ implies $\delta = g(\theta) \in C_\alpha(\delta)$, and further,

$$Pr\left[\delta \in C_\alpha(\delta)\right] \geq Pr\left[\theta \in C_\alpha(\theta)\right] \geq 1-\alpha,$$

so that $C_\alpha(\delta)$ has level $1-\alpha$. We reject $H_0(\delta_0): \delta = \delta_0$ when $\delta_0 \notin C_\alpha(\delta)$ and get a test of level α.

One can use numerical optimization technique or grid search over economically or statistically plausible parameter space to implement the projection method. However, if the parameter transformation of interest is a linear scalar function, an analytical expression for $C_\alpha(\delta)$ is available in Dufour and Taamouti (2005).

If $\delta = \phi$ where $\theta = (\phi, \rho)'$, the projection method can be implemented more efficiently. Let $F(\theta_0)$ and c_α denote a test statistic used in confidence set in (4.3.36) and a corresponding critical value, respectively. Then, the confidence set in (4.3.37) is rewritten as

$$C_\alpha(\phi) = \left\{\phi_0 \mid \inf_{\rho \in \bar{\rho}} F(\phi_0, \rho) \leq c_\alpha\right\} \tag{4.3.38}$$

where $\bar{\rho}$ is the parameter space for ρ. An alternative projection technique improves efficiency by restricting ρ. The procedure can be described as follows.

1. Construct $C_{\alpha_1}(\rho \mid \phi_0)$, a confidence set for ρ under $H_0: \phi = \phi_0$ with level $(1-\alpha_1)$.

2. Reject $H_0: \phi = \phi_0$ if $C_{\alpha_1}(\rho \mid \phi_0) = \emptyset$, or

$$\inf_{\rho \in C_{\alpha_1}(\rho \mid \phi_0)} F(\phi_0, \rho) > c_{\alpha_2},$$

   where $\alpha = \alpha_1 + \alpha_2$ and c_{α_2} is a critical value chosen in the same manner as c_α but with α_2 instead of α. By Bonferroni inequality, the test has level α, and it can be inverted to get confidence set for ϕ with level $1-\alpha$.

Since the infimum is computed over $C_{\alpha_1}(\rho \mid \phi_0)$, this procedure is expected to be more efficient. Furthermore, it is worthwhile noting that, even though the simultaneous confidence set $C_\alpha(\theta)$ for θ may be interpreted as a confidence set based on inverting LR-type tests for $\theta = \theta_0$ [see Meeker and Escobar (1995) or Chen and Jennrich (1996)], projection-based confidence sets, such as $C_\alpha(\delta)$, are not (strictly speaking) LR confidence sets. For more details and further discussion about the projection technique; see Dufour (1989, 1990), Chaudhuri et al. (2010), Chaudhuri and Zivot (2011).

4.4 Finite-sample procedures with possibly non-Gaussian errors

In this section, we extend the exact tests proposed in the previous section, by allowing non-Gaussian distributions. The use of Gaussian assumptions, when the volatility distributions are not normal, can be hazardous; such a practice could lead us to invalid inferences, a wrong choice of portfolio, the underestimation of extreme losses, and hugely mispriced derivative products. An apparent reason is that Gaussian errors are not flexible enough to capture the fat tail commonly observed in financial return distributions. In the past, many researchers used non-Gaussian distributions to get better model fits; see Liesenfeld and Jung (2000) and Chib et al. (2002) in the context of SV models, and Bollerslev (1987) in the context of GARCH-type models.

Under the non-Gaussian assumptions, we can build an exact test based on the MCT technique. We can take the observed test statistic (derived under Gaussian assumptions) and perform simulations to obtain an exact test. In order to do that, we need the null distribution of the test statistic under non-Gaussian errors. Under the Assumption 4.2.6, the GLS transformed composite error $\xi^* \sim \mathrm{N}(0, \sigma_\xi^2 I_T)$, where $\sigma_\xi^2 = (1+\phi_0^2)\sigma_\epsilon^2 + \sigma_v^2$. We need the following assumption about the transformed composite error to get the exact inference under non-Gaussian errors.

Assumption 4.4.1. Conditional Scale model of transformed composite error.

$$\xi^* = \sigma_\xi \vartheta, \tag{4.4.1}$$

where σ_ξ is a (possibly random) scalar such that $P[\sigma_\xi \neq 0] = 1$, and the conditional distribution *of ϑ is completely or incompletely specified such that*

$$\vartheta \mid \overline{X} := (\vartheta_1, \ldots, \vartheta_T) \sim \mathcal{F}(\upsilon), \tag{4.4.2}$$

where $\mathcal{F}(\cdot)$ represents a known distribution function and $\overline{X} = [X, Z_{-2}]$.

We consider both the case where the error distribution does not involve nuisance parameters,

$$\vartheta \mid \overline{X} \sim \mathcal{F}(\upsilon_0), \quad \text{where } \upsilon_0 \text{ is specified} \tag{4.4.3}$$

and the one where it does,

$$\vartheta \mid \overline{X} \sim \mathcal{F}(\upsilon), \quad \text{where } \upsilon \text{ is unknown.} \tag{4.4.4}$$

The above assumption includes the Gaussian distribution, all elliptically symmetric distributions, such as the multivariate t, and cases where $\vartheta_1, \ldots, \vartheta_T$ are i.i.d. according to any given distribution.

In the following proposition, we characterize the null distribution of $AR(\phi_0, \rho_0)$ given in (4.3.22) under the above assumption.

Proposition 4.4.1. NULL DISTRIBUTION OF AR-TEST STATISTIC UNDER NON-GAUSSIAN ERRORS. *Suppose equation (4.3.21) and Assumption 4.4.1 hold. If $\phi = \phi_0$ and $\rho = \rho_0$, we have*

$$AR(\phi_0, \rho_0) = \kappa \frac{\vartheta'(M_{C_0}[X] - M_{C_0}[X, Z_{-2}])\vartheta}{\vartheta' M_{C_0}[X, Z_{-2}]\vartheta}, \tag{4.4.5}$$

where $\kappa = (T - l - k)/l$, $\vartheta = y(\phi_0, \rho_0) = C_0(y - \phi_0 y_{-1})$, $M_{C_0}[A] = I - A[A'\Sigma(\rho_0)^{-1}A]^{-1}A'\Sigma(\rho_0)^{-1}$, and the conditional distribution of ϑ is given in Assumption 4.4.1.

Proposition 4.4.1 covers the null distribution of $AR(\phi_0, \rho_0)$. It is easy to see that the null distribution of the other proposed test statistic under non-Gaussian errors can be derived in the same way upon employing Assumption 4.4.1. Proposition 4.4.1 means that the conditional null distribution of $AR(\phi_0, \rho_0)$ given $\overline{X}$, only depends on the distribution of ϑ. If the distribution of $\vartheta \mid \overline{X}$ can be simulated, one can get exact tests based on $AR(\phi_0, \rho_0, \vartheta \mid \overline{X})$ through the MCT method [see Dufour (2006)], even if this distribution is non-Gaussian. Furthermore, the exact test obtained in this way is robust to weak IV's as well as if the distribution does not have moments (*e.g.*, the Cauchy distribution).

The MCT technique was originally proposed by Dwass (1957) for implementing permutation tests and did not involve nuisance parameters. This technique was also independently proposed by Barnard (1963); for a review, see Dufour and Khalaf (2001), and for a general discussion and proofs, see Dufour (2006). It has the great attraction of providing exact (randomized) tests based on any statistic whose finite-sample distribution may be intractable but can be simulated. Here we have briefly summarized the procedure.

Let $S(Y,\overline{X})$ be a test statistic which can be rewritten in the form

$$S(Y,\overline{X}) = \bar{S}(\vartheta,\overline{X}) \tag{4.4.6}$$

under the null hypothesis, where ϑ is defined by (4.4.2) and the distribution of ϑ is known. For example, $S(Y,\overline{X})$ could be the AR-type statistic considered in Proposition 4.4.1. Then the conditional distribution of $S(Y,\overline{X})$, given $\overline{X}$, is completely determined by the matrix $\overline{X}$ and the conditional distribution of ϑ given $\overline{X}$, *i.e.*, $S(Y,\overline{X})$ is pivotal. We can then proceed as follows to obtain an exact critical region.

1. Compute the statistic $S^{(0)}$ (based on data), where $S^{(0)} = AR^{(0)}(\phi_0,\rho_0)$.
2. By Monte Carlo methods, draw N i.i.d. replications of $\vartheta : \vartheta_{(j)} = [\vartheta_1^{(j)},\ldots,\vartheta_T^{(j)}]$, $j = 1,\ldots,N$.
3. From each simulated error matrix $\vartheta_{(j)}$, compute the statistics, $S^{(j)} = \bar{S}(\vartheta_{(j)},X)$, $j = 1,\ldots,N$, according to the fully specified distribution of $\vartheta \mid \overline{X}$. For instance, in the case of the AR statistic underlying Proposition 4.4.1, calculate
$$AR^{(j)} := AR(\vartheta_{(j)}) = \frac{\vartheta'_{(j)}\left(M_{C_0}[X] - M_{C_0}[X,Z_{-2}]\right)\vartheta_{(j)}}{\vartheta'_{(j)}M_{C_0}[X,Z_{-2}]\vartheta_{(j)}}, \quad 1,\ldots,N. \tag{4.4.7}$$
4. Compute the MC p-value $\hat{p}_N[S] := p_N(S^{(0)};S)$, where
$$p_N(x,S) := \frac{NG_N(x;S)+1}{N+1}, \tag{4.4.8}$$

$$G_N(x;S) := \frac{1}{N}\sum_{j=1}^{N} I_{[0,\infty)}(S^{(j)} - x), \quad I_{[0,\infty)}(x) = \begin{cases} 1 & \text{if } x \in [0,\infty) \\ 0 & \text{if } x \notin [0,\infty) \end{cases}. \tag{4.4.9}$$

In other words, $p_N(S^{(0)};S) = [NG_N(S^{(0)};S)+1]/(N+1)$ where $NG_N(S^{(0)};S)$ is the number of simulated values which are greater than or equal to $S^{(0)}$. When $S^{(0)}, S^{(1)}, \ldots, S^{(N)}$ are all distinct [an event with probability one when the vector $(S^{(0)}, S^{(1)}, \ldots, S^{(N)})'$ has an absolutely continuous distribution], $\hat{R}_N(S^{(0)}) = N+1-NG_N(S^{(0)};S)$ is the rank of $S^{(0)}$ in the series $S^{(0)}, S^{(1)}, \ldots, S^{(N)}$.

5. The MC critical region is: $\hat{p}_N[S] \leq \alpha,\ 0 < \alpha < 1$. If α^* and N such that $\alpha(N+1)$ is an integer and the distribution of S is continuous under the null hypothesis, then under null,

$$P[\hat{p}_N[S] \leq \alpha] = \alpha \tag{4.4.10}$$

The above algorithm is valid for any fully specified distribution of ϑ and we reject the null hypothesis $H_0(\phi_0, \rho_0)$ at level α when $\hat{p}_N[AR^{(0)}(\phi_0, \rho_0)] \leq \alpha$.

Under the null hypothesis $H_\phi(\phi_0, \rho_0)$, $P\big[\hat{p}_N[AR^{(0)}(\phi_0, \rho_0)] \leq \alpha\big] = \alpha$, so that we have a test with level α. If the distribution of the test statistic is not continuous, the MC test procedure can easily be adapted by using "tie-breaking" method described in Dufour (2006).[5] Correspondingly, a confidence set with level $1-\alpha$ for (ϕ, ρ) is given by the set of all values (ϕ_0, ρ_0) which are not rejected by the above MC test. More precisely, the set

$$C_{(\phi,\rho)}(\alpha) = \big\{(\phi_0, \rho_0) : \hat{p}_N[AR^{(0)}(\phi_0, \rho_0)] > \alpha\big\} \tag{4.4.11}$$

is a confidence set with level $1-\alpha$ for (ϕ_0, ρ_0).

Consider now the case where the distribution of ϑ involves a nuisance parameter υ and $\upsilon \in \Phi_0$.

1. Let $S^{(0)}$ be the observed test statistic (based on data).

[5]Without the correction for continuity, the algorithm proposed for statistics with continuous distributions yields a conservative test, *i.e.*, the probability of rejection under the null hypothesis is not larger than the nominal level.

2. For each $\upsilon \in \Phi_0$, by Monte Carlo methods, draw N i.i.d. replications of $\vartheta : \vartheta_{(j)} = [\vartheta_1^{(j)}, \dots, \vartheta_T^{(j)}]$, $j = 1, \dots, N$ and compute the statistics, $S^{(j)}(\upsilon) = \bar{S}(\vartheta_{(j)}(\upsilon), X)$, $j = 1, \dots, N$.

3. Using these simulations we compute the MC p-value $\hat{p}_N[S] := p_N(S^{(0)}; S)$, where

$$\hat{p}_N\big[x; S \mid \upsilon\big] := \frac{N\hat{G}_N\big[x; S \mid \upsilon\big] + 1}{N+1}. \tag{4.4.12}$$

4. The p-value function $\hat{p}_N[S \mid \upsilon]$ as a function of υ is maximized over the parameter values compatible with the Φ_0, and H_0 is rejected if

$$\sup_{\upsilon \in \Phi_0} \hat{p}_N[S \mid \upsilon] \leq \alpha. \tag{4.4.13}$$

   If the number of simulated statistics N is chosen such that $\alpha(N+1)$ is an integer, then we have under H_0:

$$P\Big[\sup_{\upsilon \in \Phi_0} \big\{\hat{p}_N[S \mid \upsilon]\big\} \leq \alpha\Big] \leq \alpha, \tag{4.4.14}$$

   The test defined by $\hat{p}_N[S \mid \upsilon] \leq \alpha$ has size α for known υ . Treating υ as a nuisance parameter and Φ_0 is a nuisance parameter set consistent with null, the test is *exact at level* α; for a proof, see Dufour (2006).

Because of the maximization in the critical region (4.4.13) the test is called a *maximized Monte Carlo* (MMC) test. MMC tests provide valid inference under general regularity conditions such as almost-unidentified models or time series processes involving unit-roots. In particular, even though the moment conditions defining the estimator are derived under the stationarity assumption, this does not question in any way the validity of maximized MC tests, unlike the parametric bootstrap whose distributional theory is based on strong regularity conditions. Only the power of MMC tests may be affected. However, the simulated p-value function is not continuous, thus standard gradient-based methods cannot be used to maximize it. But search methods applicable to non-differentiable functions are applicable, *e.g.*, simulated annealing [see Goffe et al. (1994)]. A simplified approximate version of the MMC procedure can alleviate its computational load whenever a consistent point or set estimate of υ is available; for further discussion, see Dufour (2006).

4.5 Asymptotic distributional theory

In this section, we relax the Assumptions 4.2.4-4.2.6 and 4.4.1, and show that under weaker distributional assumptions on X, Z_{-2} and ξ, the proposed procedures remain "asymptotically valid". More precisely, we wish to show that if Assumption 4.2.4-4.2.6 hold jointly with a specific distributional assumption on ξ^*/σ_ξ [*e.g.*, $\xi^*/\sigma_\xi \sim \mathrm{N}(0, I_T)$] yields tests whose probability of type I error converges to the nominal level of the test as $T \to \infty$ under any parameter configuration compatible with the null hypothesis (pointwise asymptotic validity).

All our results up to now have been established for a given sample size of T. To formulate asymptotic properties, we need to consider a sequence of tests indexed by T. Consider the following sequence

$$\big\{S(T) := [y(T), y_{-1}(T), X(T), Z_{-2}(T), \xi(T)], T \geq T_0\big\}, \tag{4.5.1}$$

and rewrite the test statistic (4.3.10) in the following form:

$$AR_T(\phi_0) = \kappa(T)\frac{y_T'\big(M[Q_{1T}] - M[Q_T]\big)y_T}{y_T' M[Q_T] y_T / T}, \tag{4.5.2}$$

where $y_T = \big(y(T) - \phi_0 y_{-1}(T)\big)$, $Q_T = [Q_{1T}, Q_{2T}]$, $Q_{1T} = X(T)$, $Q_{2T} = Z_{-2}(T)$, $\kappa(T) = (T - l - k)/lT$, and k and l are the number of columns in Q_{1T} and Q_{2T}, respectively.

We examine the asymptotic distribution of $AR_T(\phi_0)$ under the following assumptions (where $\Longrightarrow$ refers to weak convergence as the sample size tends to infinity).

Assumption 4.5.1. *The sequence* $\big(S(T), T \geq T_0\big)$ *given in (4.5.1) belongs to a class* $\mathcal{Z}$ *of stochastic processes such that for each process in* $\mathcal{Z}$ *the following limits hold:*

1. $\frac{\xi'(T)\xi(T)}{T} \xrightarrow[T\longrightarrow\infty]{p} \sigma_\xi^2 > 0$, *where* σ_ξ^2 *is the same for all processes in* $\mathcal{Z}$*;*

2. *There exists a sequence of* $m \times m$*, nonsingular matrices* D_T *such that:*

   (A) $D_T' Q_T' Q_T D_T \xrightarrow[T\longrightarrow\infty]{p} \Sigma_{QQ} = \begin{pmatrix} \Sigma_{Q_1Q_1} & \Sigma_{Q_1Q_2} \\ \Sigma_{Q_2Q_1} & \Sigma_{Q_2Q_2} \end{pmatrix}$,
   where Σ_{QQ} *and* $\Sigma_{Q_1Q_1}$ *are* $m \times m$ *and* $k \times k$ *nonsingular matrices, respectively;*

(B) $D_T' Q_T' \xi(T) \Longrightarrow q \sim \mathrm{N}\big(0, \sigma_\xi^2 \Sigma_{QQ}\big)$,
where $q = \big(q_1', q_2'\big)'$, q_1 *and* q_2 *are* $k \times 1$ *and* $l \times 1$ *random vectors, respectively.*

It should be emphasized that Assumption 4.5.1 satisfies the condition

$$q_2 \mid q_1 \sim \mathrm{N}\big(\Sigma_{Q_2Q_1}\Sigma_{Q_1Q_1}^{-1} q_1, \sigma_\xi^2 \Sigma_{q_2|q_1}\big)$$

where $\Sigma_{q_2|q_1} = \Sigma_{Q_2Q_2} - \Sigma_{Q_2Q_1}\Sigma_{Q_1Q_1}^{-1}\Sigma_{Q_2Q_2}$. Thus the asymptotic distribution of $\big(q'\Sigma_{QQ}^{-1}q - q_1'\Sigma_{Q_1Q_1}^{-1}q_1\big)/\sigma_\xi^2$ is a $\chi_{(l)}^2$ distributed random variable. Note that the normality of the sub-vector of q_1 is not required, the conditional normality of q_2 given q_1 is sufficient.

Further, in the above Assumption 4.5.1(2), we allow both stationary and nonstationary regressors by adjusting the scaling matrix D_T, which is typical of the form, $D_T = diag\big[T^{-d_1}, \ldots, T^{-d_m}\big]$, where $d_i > 0$ for $i = 1, \ldots, m$ relying on the degree of nonstationarity of the regressors. For example, if $X(T)$ and $Z_{-2}(T)$ are stationary then $d_i = 0.5$ for $i = 1, \ldots, m$. However, if $X(T)$ and $Z_{-2}(T)$ are nonstationary and are integrated of order one, then the corresponding d_i should be one. The following proposition establishes the asymptotic validity of the AR procedure.

Proposition 4.5.1. Asymptotic validity of AR-type test. *Under the Assumption 4.5.1 and the null hypothesis in (4.3.1), the statistic $AR_T(\phi_0)$ in (4.5.2) has the same limiting distribution for all processes in $\mathcal{Z}$, i.e., $AR_T(\phi_0) \Longrightarrow \chi_{(l)}^2/l$.*

Similarly, one can show that the joint test defined in (4.3.22) has the null distribution of $AR_T(\phi_0, \rho_0) \Longrightarrow \chi_{(l)}^2/l$. Now we consider the test statistic of the AR-type PO procedure, which is rewritten in the following form:

$$AR_T^*(\phi_0, \rho_0, \rho_1) = T\left[\frac{y_T(\phi_0, \rho_0)' M[\hat{Q}_{1T}] y_T(\phi_0, \rho_0)}{y_T(\phi_0, \rho_1)' M[\tilde{Q}_T] y_T(\phi_0, \rho_1)} - 1\right], \tag{4.5.3}$$

where $y_T(\phi_0, \rho_0) = C(\rho_0)\big(y(T) - \phi_0 y_{-1}(T)\big)$, $y_T(\phi_0, \rho_1) = C(\rho_1)\big(y(T) - \phi_0 y_{-1}(T)\big)$, $\hat{Q}_{1T} = C(\rho_0)X(T)$, $\tilde{Q}_T = [\tilde{Q}_{1T}, \tilde{Q}_{2T}]$, $\tilde{Q}_{1T} = C(\rho_1)X(T)$, $\tilde{Q}_{2T} = C(\rho_1)Z_{-2}(T)$, k is the number of columns in $\hat{Q}_{1T}$ or $\tilde{Q}_{2T}$, l is the number of columns in $\tilde{Q}_{2T}$ and $m = l + k$. In order to prove the asymptotic validity of the $AR_T^*(\phi_0, \rho_0, \rho_1)$ that defined in (4.3.25), we need following assumption:

Assumption 4.5.2. *The sequence* $(S(T), T \geq T_0)$ *given in (4.5.1) belongs to a class $\mathcal{Z}$ of stochastic processes such that for each process in $\mathcal{Z}$ the following limits hold:*

1. $\frac{\hat{\xi}'(T)\hat{\xi}(T)}{T} \xrightarrow[T\longrightarrow\infty]{p} \sigma_{\xi}^2 > 0$, *where* σ_{ξ}^2 *is the same for all processes in* $\mathcal{Z}$*;*
2. $\frac{\tilde{\xi}'(T)\tilde{\xi}(T)}{T} \xrightarrow[T\longrightarrow\infty]{p} \sigma_{\xi}^2 > 0$, *where* σ_{ξ}^2 *is the same for all processes in* $\mathcal{Z}$*;*
3. *There exists a sequence of* $m \times m$*, nonsingular matrices* D_T *such that:*

   (A) $D_T'\tilde{Q}_T'\tilde{Q}_T D_T \xrightarrow[T\longrightarrow\infty]{p} \Sigma_{\tilde{Q}\tilde{Q}} = \begin{pmatrix} \Sigma_{\tilde{Q}_1\tilde{Q}_1} & \Sigma_{\tilde{Q}_1\tilde{Q}_2} \\ \Sigma_{\tilde{Q}_2\tilde{Q}_1} & \Sigma_{\tilde{Q}_2\tilde{Q}_2} \end{pmatrix}$,
   where $\Sigma_{\tilde{Q}\tilde{Q}}$ *and* $\Sigma_{\tilde{Q}_1\tilde{Q}_1}$ *are* $m \times m$ *and* $k \times k$ *nonsingular matrices, respectively;*

   (B) $D_{T1}'\hat{Q}_{T1}'\hat{Q}_{T1} D_{T1} \xrightarrow[T\longrightarrow\infty]{p} \Sigma_{\hat{Q}_1\hat{Q}_1}$,
   where $\Sigma_{\hat{Q}_1\hat{Q}_1}$ *is a* $k \times k$ *nonsingular matrix;*

   (C) $D_T'\tilde{Q}_T'\tilde{\xi}(T) \Longrightarrow \tilde{q} \sim \mathrm{N}\big(0, \sigma_{\xi}^2\Sigma_{\tilde{Q}\tilde{Q}}\big)$,
   where $\tilde{q} = (\tilde{q}_1', \tilde{q}_2')'$, $\tilde{q}_1$ *and* $\tilde{q}_2$ *are* $k \times 1$ *and* $l \times 1$ *random vectors, respectively.*

   (D) $D_T'\hat{Q}_{1T}'\hat{\xi}(T) \Longrightarrow \hat{q}_1 \sim \mathrm{N}\big(0, \sigma_{\xi}^2\Sigma_{\hat{Q}_1\hat{Q}_1}\big)$,
   where $\hat{q}_1$ *is a* $k \times 1$ *random vector.*

The following proposition establishes the asymptotic validity of the AR^* optimal procedure.

Proposition 4.5.2. Asymptotic validity of AR-type point-optimal test. *Under the Assumption 4.5.2 and the null hypothesis in (4.3.19) against a fixed alternative* $\rho = \rho_1$*, the statistic* $AR_T^*(\phi_0, \rho_0, \rho_1)$ *in (4.5.3) has the same limiting distribution for all processes in* $\mathcal{Z}$*, i.e.,* $AR_T^*(\phi_0, \rho_0, \rho_1) \Longrightarrow \chi^2_{(l)}$.

We consider the following sequences for the split-sample methods, where each element of the sequence (4.5.1) is split into the first and second subsamples with size T_1 and T_2 ($T = T_1 + T_2$), respectively:

$$\begin{aligned} &\big\{S^{(1)}(T) := [y^{(1)}(T), y_{-1}^{(1)}(T), X^{(1)}(T), Z_{-2}^{(1)}(T), \bar{\xi}^{(1)}(T)], \\ &S^{(2)}(T) := [y^{(2)}(T), y_{-1}^{(2)}(T), X^{(2)}(T), Z_{-2}^{(2)}(T), \bar{\xi}^{(2)}(T)], T > T_0\big\}. \end{aligned} \tag{4.5.4}$$

The split-sample test statistic in (4.3.31) can be constructed from (4.5.4) as follows:

$$SS_T(\phi_0,\rho_0) = \kappa(T_2)\frac{y_T^{*\prime}\big(M[Q_{1T}^*] - M[\hat{Q}_T^*]\big)y_T^*}{y_T^{*\prime}M[\hat{Q}_T^*]y_T^*/T_2}, \tag{4.5.5}$$

where $y_T^* = C(\rho_0)\big(y^{(2)}(T)-\phi_0 y_{-1}^{(2)}(T)\big)$, $\hat{Q}_T^* = [Q_{1T}^*, \hat{Q}_{2T}^*]$, $Q_{1T}^* = C(\rho_0)X^{(2)}(T)$, $\hat{Q}_{2T}^* = C(\rho_0)\hat{y}_{-1}^{(2)}(T)$, $\hat{y}_{-1}^{(2)} = \bar{Z}_{-2}^{(2)}\tilde{\pi}^{(1)}$, $\tilde{\pi}^{(1)} = \big(\bar{Z}_{-2}^{(1)\prime}\bar{Z}_{-2}^{(1)}\big)^{-1}\bar{Z}_{-2}^{(1)\prime}y_{-1}^{(1)}$, $\kappa(T_2) = (T_2 - l - k)/lT_2$, and k and l are the number of columns in Q_{1T}^* and $\hat{Q}_{2T}^*$, respectively. We will examine the asymptotic distribution of $SS_T(\phi_0,\rho_0)$ under the following assumptions.

Assumption 4.5.3. *$T_1/T \longrightarrow \tau \in (0,1)$ as $T \longrightarrow \infty$.*

Assumption 4.5.4. *The sequence $(S^{(1)}(T), S^{(2)}(T), T > T_0)$ given in (4.5.4) belongs to a class $\mathcal{Z}$ of stochastic processes such that for each process in $\mathcal{Z}$ the following limits hold:*

1. *$\frac{\bar{\xi}_T^{*(2)\prime}\bar{\xi}_T^{*(2)}}{T_2} \xrightarrow[T_2\longrightarrow\infty]{p} \sigma_\xi^2 > 0$, where $\bar{\xi}_T^{*(2)} := C(\rho_0)\bar{\xi}^{(2)}(T)$, and σ_ξ^2 is the same for all processes in $\mathcal{Z}$;*

2. *Conditional on the first subsample, there exists a sequence of $m\times m$, nonsingular matrices D_T such that:*

   (A) $D_T'\hat{Q}_T^{*\prime}\hat{Q}_T^* D_T \xrightarrow[T_2\longrightarrow\infty]{p} \Sigma_{\hat{Q}^*\hat{Q}^*} = \begin{pmatrix} \Sigma_{Q_1^*Q_1^*} & \Sigma_{Q_1^*\hat{Q}_2^*} \\ \Sigma_{\hat{Q}_2^*Q_1^*} & \Sigma_{\hat{Q}_2^*\hat{Q}_2^*}\end{pmatrix}$,
   where $\Sigma_{\hat{Q}^*\hat{Q}^*}$ and $\Sigma_{Q_1^*Q_1^*}$ are $m\times m$ and $k\times k$ nonsingular matrices, respectively;

   (B) $D_T'\hat{Q}_T^{*\prime}\bar{\xi}_T^{*(2)} \Longrightarrow \hat{q}^* \sim \mathrm{N}\big(0,\sigma_\xi^2\Sigma_{\hat{Q}^*\hat{Q}^*}\big)$,
   where $\hat{q}^* = (q_1^{*\prime}, \hat{q}_2^{*\prime})'$, q_1^* and $\hat{q}_2^*$ are $k\times 1$ and $l\times 1$ random vectors, respectively, such that $\tilde{q}^* := \hat{q}^{*\prime}\Sigma_{\hat{Q}^*\hat{Q}^*}^{-1}\hat{q}^* - q_1^{*\prime}\Sigma_{Q_1^*Q_1^*}^{-1}q_1^*$ has an absolutely continuous (non-degenerate) distribution on $\mathbb{R}$, which is the same for all processes in $\mathcal{Z}$.

It should be noted that $\hat{Q}_{2T}^*$ depends on $\tilde{\pi}^{(1)}$, which is estimated from the first subsample. However, conditioning on the first subsample, we can get rid of this unnecessary randomness. Assumption 4.5.4(2B) implies that $\Sigma_{\hat{Q}^*\hat{Q}^*}$ and $\hat{q}^*$ are depend on $S^{(1)}(T)$, while $\tilde{q}^*$ is not depend on $S^{(1)}(T)$. To see this, we consider an example where $D_T = T_2^{-1/2}I_{T_2}$ and $\hat{q}^* = q\big(\tilde{\pi}^{(1)}\big)$ follows a normal distribution. Thus, given $\widetilde{\pi}^{(1)}$,

$$\operatorname*{plim}_{T_2\to\infty}\frac{Q_T(\widetilde{\pi}^{(1)})'Q_T(\widetilde{\pi}^{(1)})}{T_2}=\Sigma_{QQ}(\widetilde{\pi}^{(1)})=\Sigma_{\hat{Q}^*\hat{Q}^*},$$

$$\frac{Q_T(\widetilde{\pi}^{(1)})'Q_T(\widetilde{\pi}^{(1)})}{\sqrt{T_2}}\Longrightarrow q(\widetilde{\pi}^{(1)})=\hat{q}^*\sim N(0,\ \sigma_\xi^2\Sigma_{\hat{Q}^*\hat{Q}^*}),$$

then, $\big(\hat{q}^{*\prime}\Sigma^{-1}_{\hat{Q}^*\hat{Q}^*}\hat{q}^*\big)/\sigma_\xi^2\sim\chi^2_m$, where χ^2_m is the χ^2 distribution with m degrees of freedom. As a result, even though $\Sigma_{\hat{Q}^*\hat{Q}^*}$ and $\hat{q}^*$ rely on $\widetilde{\pi}^{(1)}$, $\hat{q}^{*\prime}\Sigma^{-1}_{\hat{Q}^*\hat{Q}^*}\hat{q}^*$ does not depend on $\widetilde{\pi}^{(1)}$.

In the finite-sample distributional theory of split-sample procedures, independence of $\bar{\xi}^*_t=C(\rho_0)\bar{\xi}_t$ over $t=1,\dots,T$ is assumed. In asymptotic theory, however, a similar restriction on dependence of $\bar{\xi}^*_t$ (*e.g.*, α-mixing assumption) is implicitly imposed by Assumption 4.5.4(2B); if dependence between $\bar{\xi}^{*(1)}_t$ in the estimates and $\bar{\xi}^{*(2)}_t$ is too strong, then the limiting distribution $\hat{q}^*$ would rely on nuisance parameters governing the dependence, and as a result the assumption cannot be satisfied. The following proposition proves the asymptotic validity of SS procedure.

Proposition 4.5.3. Asymptotic validity of SS-type test. *Under the Assumptions 4.5.3-4.5.4 and the null hypothesis in (4.3.19), the statistic $SS_T(\phi_0,\rho_0)$ in (4.5.5) has the same limiting distribution for all processes in $\mathcal{Z}$, i.e., $SS_T(\phi_0,\rho_0)\Longrightarrow\chi^2_{(l)}/l$.*

Similarly, one can also prove the asymptotic validity of SS^* procedure.

4.6 Simulation study

In this section, we compare the performance of our proposed tests to the asymptotic t-type test. The standard SV model [given in (4.2.1)-(4.2.2)] has the following state-space representation:

$$w_t=\mu+\phi w_{t-1}+v_t,\qquad y_t=w_t+\epsilon_t,\qquad y_t:=\log(s_t^2)-\mathbb{E}\left[\log(z_t^2)\right],\tag{4.6.1}$$

where $w_t:=\log(\sigma_t^2)$, and the v_t's and ϵ_t's are i.i.d. $\mathrm{N}(0,\sigma_v^2)$ and $\log(\chi^2_{(1)})$ random variables, respectively and they are orthogonal to each other. With an instrument equation, the DGP in

(4.6.1) is:

$$y_t = \mu + \phi y_{t-1} + \xi_t, \quad \xi_t := v_t + \epsilon_t - \phi\epsilon_{t-1}, \quad v_t \sim \text{i.i.d. } N(0,\sigma_v^2), \quad \epsilon_t \sim \text{i.i.d. } \log(\chi^2_{(1)}), \tag{4.6.2}$$

$$y_{t-1} = Z'_{t-2}\bar{\pi}_1 + \eta_{t-1}, \quad \eta_{t-1} := \epsilon_{t-1} + u_{t-1}, \quad u_t \sim \text{i.i.d. } N(0,\sigma_u^2), \tag{4.6.3}$$

where $\bar{\pi}_1$ is an l-vector of first-stage coefficients, Z_{t-2} is an l-vector of independent N(0,1) variables, and the vector (ξ_t, η_{t-1}) has zero mean,

$$\text{Var}(\xi_t) = (1+\phi^2)\sigma_\epsilon^2 + \sigma_v^2, \quad \text{Var}(\eta_{t-1}) = \sigma_\epsilon^2 + \sigma_u^2, \quad \text{and } \text{Cov}(\xi_t, \eta_{t-1}) = -\phi\sigma_\epsilon^2.$$

We construct $\bar{\pi}_1$ as:

$$\bar{\pi}_1 = \frac{||\bar{\lambda}||\sqrt{(\sigma_\epsilon^2 + \sigma_u^2)}}{\sqrt{Tl}} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}, \tag{4.6.4}$$

so that $||\bar{\lambda}^2|| = \frac{T\bar{\pi}_1'\bar{\pi}_1}{\sigma_\epsilon^2 + \sigma_u^2}$. Since $\text{Var}(Z_{t-2}) = I_l$ and $\text{Var}(\eta) = \sigma_\epsilon^2 + \sigma_u^2$, $||\bar{\lambda}^2||$ is the concentration parameter in this model.

Note that the DGP given by (4.6.2)-(4.6.3) is a GSV model with no exogenous explanatory variable. This DGP is designed to broadly mimic the features of financial returns used in our empirical application. From (4.6.3), it is evident that this DGP violates the independence assumption. However, the instrument set Z_{t-2} is uncorrelated with η_{t-1}.

Except for the Section 4.6.2.4, in all experiments, we set: $\sigma_\epsilon^2 = \pi^2/2$, so that μ, ϕ and σ_v, are the only parameters that will vary. We use 10,000 replication to compute the empirical level and powers and employ 99 replications for PO tests based on the MCT procedure. For all tests, the nominal level is fixed at 5%. Thus, under the null hypothesis, the rejection rates should be less than (or close to) 5% for tests to be valid. Except for the analysis of asymptotic tests (Section 4.6.1), the sample sizes are $T = 100, \ 200$. For the split-sample tests, we employ the split ratio $\tau = 0.2$ and use OLS to construct the instrument set. Note that SS-type tests depend on the choice of τ, and the power of these tests is inversely related to τ [see Dufour and Jasiak (2001)]. Therefore, we set $\tau = 0.2$ to gain relatively more power.

4.6.1 Test size of asymptotic t-test

In this section, we evaluate the performance the asymptotic t-type test of $H_0 : \phi = \phi_0$. The simulated DGP is (4.6.2) with $|\phi| < 1$ and $\epsilon_t \sim$ i.i.d. $\log(\chi^2_{(1)})$ [it is the log-normal SV model]. We set $\mu = 0$, $\sigma_v = 2$ and $\phi \in [0, 1]$. For sample sizes, $T = \{100, 200, 300, 400, 500, 1000, 2000\}$ are used.

Table 4.2 reports the size of asymptotic t-type tests for $H_0(\phi) : \phi = \phi_0$. The test statistic is calculated using the simple winsorized estimator of Ahsan and Dufour (2019) [equations (3.8)-(3.9) with $J = 10$]. This estimator is more efficient compared to conventional methods (QMLE, GMM) and as efficient as the Bayesian procedure. In addition to that, it is extremely time-efficient and it produces empirical estimates which are similar to the Bayesian estimates. For the details of this asymptotic t-test, see Section 6.1 of Ahsan and Dufour (2019).

We can see from the results that the t-test (which is based on the asymptotic standard error) fails to control the level when $\phi \longrightarrow 1$. Size distortions are severe and equal upto 37.2% when $\phi = 1$. These size distortions do not go away even in larger samples ($T = 1000, 2000$), especially when $\phi > 0.999$, *i.e.*, ϕ is close to the unit circle.

4.6.2 Performance of the proposed tests

We will now examine the performance of the tests proposed in Sections 4.3.1-4.3.4. To simplify the exposition, we focus on four misspecified model setups (these are empirically motivated).

$M1$. The DGP is given in (4.6.2)-(4.6.3) with $\epsilon_t \sim$ i.i.d. $N(0, \pi^2/2)$. The instrument set Z_{t-2} includes weak IV's, which are related to past lags of the LF volatility proxy.

$M2$. The DGP is given in (4.6.2)-(4.6.3) and the instrument set Z_{t-2} includes weak IV's, which are related to past lags of the LF volatility proxy.

$M3$. The DGP is given in (4.6.2) and the instrument set Z_{t-2} includes past lags of the LF volatility proxy. It is the standard log-normal SV model, where we use past lags of observed volatility proxy (y_{t-1}) as IV's.

$M4$. The DGP is given in (4.6.2) with $|\phi| < 1$ and $\epsilon_t \sim$ i.i.d. $N(0, \sigma^2_\epsilon)$. The instrument set Z_{t-2} includes HF IV's, *i.e.*, realized volatility.

From the above setups, it is easy to see that the model $M1$ violates the Assumption 4.2.4, whereas $M2$-$M4$ are misspecified under the Assumptions 4.2.4 and 4.2.6.

For models $M1$-$M3$, we consider the joint tests $[H_0 : (\phi,\ \rho) = (\phi_0,\ \rho_0)]$. The test statistics $(AR,\ AR^*,\ SS,\ SS^*)$ are given in equations (4.3.22), (4.3.25), (4.3.31) and (4.3.35). For the model $M4$, we use a plug-in estimator for ρ and consider the single restriction tests $[H_0 : \phi = \phi_0]$. The considered test statistics $(AR,\ AR^*,\ SS,\ SS^*)$ are given in equations (4.3.22) and (4.3.25) with $\rho_0 = \rho_1 = \hat{\rho}$, and equations (4.3.30) and (4.3.34).

For the weak IV's robustness check (in Sections 4.6.2.1 and 4.6.2.2), we simulate model $M1$ and $M2$ with $\mu = 2$. We consider the concentration parameter $\bar{\lambda}^2 \in \{0, 0.1, 10\}$ with $\sigma_\epsilon^2 = \pi^2/2$ and $\sigma_u^2 = 0.1$. Thus, given $\bar{\lambda}^2 \in \{0, 0.1, 10\}$, the corresponding values of the first stage coefficients $\bar{\pi}_1[1, i] = \{0, 0.05, 0.50\}, i = \{1, \cdots, l\}$ for $T = 100$ and $\bar{\pi}_1[1, l] = \{0, 0.04, 0.35\}, i = \{1, \cdots, l\}$ for $T = 200$. The simulated models use different values of ϕ and ρ. These values are $\phi = \{0.50, 0.75, 0.90, 1.00\}$ and $\rho = \{0.1, 0.2, 0.3\}$. Thus, given $\rho = 0.1$ and $\phi = \{0.50, 0.75, 0.90, 1.00\}$, the corresponding values of λ $[= \rho/(\phi - \rho(1+\phi)^2)]$ are $\{0.27, 0.17, 0.14, 0.13\}$. Since we fix $\sigma_\epsilon^2 = \pi^2/2$, given $\lambda = \{0.27, 0.17, 0.14, 0.13\}$ the corresponding values of σ_v $[= \sigma_\epsilon/\sqrt{\lambda}]$ are $\{4.30, 5.41, 5.96, 6.28\}$. Similarly, for $\rho = \{0.2, 0.3\}$, we have different set of values for λ and σ_v. As a result, a restriction on ρ implies a restriction on λ or σ_v. For example, a joint null $(\phi_0, \rho_0) = (0.5, 0.1)$ is same as $(\phi_0, \lambda_0) = (0.5, 0.27)$ or $(\phi_0, \sigma_{v0}) = (0.5, 4.30)$. For PO tests, we set the alternative to $\rho_1 = 0.30$ in Sections 4.6.2.1 and 4.6.2.2 (power comparison experiments).

4.6.2.1 Test performance under $M1$ with weak instruments

We simulate the model $M1$. The generated instrument set Z_{-2} is related to past lags of the LF volatility proxy y_{-1}, so it is not independent of the error distributions of v and ϵ. The simulated DGP is incorrectly specified under the Assumption 4.2.4. The results are presented in Tables 4.3-4.4 and confirm the theoretical contributions of Sections 4.3.1-4.3.4 even with model misspecification. Our findings can be summarized as follows.

First, from Table 4.3, the levels of the proposed tests $(AR,\ AR^*,\ SS,\ SS^*)$ are well controlled: rejection frequencies are less than (or close to) 5%. This result holds whether the identification is completely failed $[\bar{\lambda}^2 = 0]$, weak $[\bar{\lambda}^2 \in \{0, 0.1\}]$, partial $[\bar{\lambda}^2 \in \{0.1, 10\}]$, or strong $[\bar{\lambda}^2 = 10]$. This represents a substantial improvement over the asymptotic test. However, the AR test exhibits minor size distortion, possibly due to model misspecification. Further, the SS controls the level correctly, but in most cases this test is undersized and it increases with the number of IV's. The optimal tests perfectly control the level.

Second, from Table 4.4, all tests exhibit excellent power as long as identification is not very weak. Note that, in our joint tests, we have an additional restriction under the null hypothesis on the parameter of the error distribution. This restriction works as an additional source of power. We also see that in all cases [weak or strong IV's], the AR^* and SS^* tests have more power compared to the AR and SS tests. As expected, these tests' power increases with the sample size and the concentration parameter and decreases as the number of IV's increases.

4.6.2.2 Test performance under $M2$ with weak instruments

The model $M2$ is considered. The generated instrument set Z_{-2} is related to past lags of the LF volatility proxy y_{-1}, so it is not independent of the error distributions of v and ϵ. This violates the Assumption 4.2.4. Further, since $\epsilon_t \sim$ i.i.d. $\log(\chi^2_{(1)})$, the simulated DGP is also misspecified under the Assumption 4.2.6. This DGP represents an SV model with an instrument equation. The results are presented in Tables 4.5-4.6. The results confirm that the tests proposed in Sections 4.3.1-4.3.4 are valid and robust to these misspecifications. The main findings are the following.

First, from Table 4.5, the empirical levels of the proposed tests are almost identical to those obtained when the model is only misspecified under Assumption 4.2.4 [compare Table 4.3 with 4.5]: rejection frequencies are less than (or close to) 5%, whether identification is completely failed [$\bar{\lambda}^2 = 0$], weak [$\bar{\lambda}^2 \in \{0, 0.1\}$], partial [$\bar{\lambda}^2 \in \{0.1, 10\}$], or strong [$\bar{\lambda}^2 = 10$], for all sample sizes considered. The optimal tests based on MCT method have better level control.

Second, from Table 4.6, the misspecification of the error distribution does not affect the power of these tests [compare Table 4.6 with Table 4.4].

Third, as the sample size increases, the rejection frequencies of these tests increase and in many cases, reach 100%. Overall, these tests appear to be reasonably robust to a misspecification of the error distribution, even with small samples.

4.6.2.3 Test performance under $M3$ with low-frequency instruments

We simulate model $M3$ with $\mu = 2$, $\phi \in (0.5,\ 1]$ and $\sigma_v \in (0.94,\ 3.14)$. This DGP corresponds to the standard log-normal SV model. We use past lags of y_{t-1} as IV's, so the instrument set Z_{t-2} is not independent of the error distributions of v and ϵ. As a result, the simulated DGP is incorrectly specified under the Assumptions 4.2.4 and 4.2.6. In this setting, for PO tests, we set the alternative to $\rho_1 = 0.35$. The results are appear in Table 4.7 and the main findings are the following.

First, in both samples ($T = 100, 200$), the levels of the proposed tests (AR, AR^*, SS , SS^*) are well controlled, even when $\phi = 1$. As the number of IV's increases, the SS test under rejects (when $l = 10$).

Second, all these tests exhibit excellent power (see from the second part of Table 4.7). Since, we set the alternative hypothesis to $(\phi,\ \rho) = (0.5,\ 0.35)$, as a result PO tests can gain power from the differences in covariance structure, *i.e.*, when $\rho = 0.25, 0.30$. From the results, in all cases, AR and SS tests have more power compare to their counterpart AR^* and SS^* when $l = 1$. Again, as expected, these tests' power increases with the sample size and decreases as the number of IV's increases.

Third, we also simulate the same DGP with $\epsilon_t \sim$ i.i.d. $N(0, \pi^2/2)$ and results are almost identical [compare Table 4.8 with Table 4.7]: rejection frequencies are similar.

4.6.2.4 Test performance under $M4$ with high-frequency instruments

In this experiment, we simulate the model $M4$ [DGP:(4.6.2) with $|\phi| < 1$ and $\epsilon_t \sim$ i.i.d. $N(0, \sigma_\epsilon^2)$] at a higher frequency and use these HF observations to construct RV estimates, which are employed as IV's. We consider the parameter inference for the LF model. Note that the HF model parameters are different from the LF model parameters. Therefore, making an inference about ϕ or ρ (or both) in the LF model requires functional relationships between HF and LF parameters under temporal aggregation, *e.g.*, $\phi_{lf} = f(\phi_{hf})$, where ϕ_{hf} and ϕ_{lf} are the HF and LF parameter, receptively. It should be noted that the log-normal SV model [DGP: (4.6.2) with $|\phi| < 1$ and $\epsilon_t \sim$ i.i.d. $\log(\chi^2_{(1)})$] is not closed under temporal aggregation [see Meddahi and Renault (2004)], whereas the $M4$ model is closed under temporal aggregation.

Since we assume stationarity of the latent HF volatility process ($|\phi_{hf}| < 1$), the HF process y_t given in model $M4$ admits an ARMA(1, 1) representation [see Proposition 3.1 of Ahsan and Dufour (2019)], which is given by

$$y_t = \mu_{hf} + \phi_{hf} y_{t-1} + \bar{\eta}_t - \theta_{hf} \bar{\eta}_{t-1}, \tag{4.6.5}$$

with $\bar{\eta}_t - \theta_{hf} \bar{\eta}_{t-1} = v_t + \epsilon_t - \phi_{hf} \epsilon_{t-1}$. The moving average parameter θ_{hf} and the white noise variance $\sigma^2_{hf,\bar{\eta}}$ are related to ϕ_{hf}, $\sigma^2_{hf,v}$ and $\sigma^2_{hf,\epsilon}$ through non-linear equations:

$$(1 + \theta^2_{hf})\sigma^2_{hf,\bar{\eta}} = \sigma^2_{hf,v} + (1 + \phi^2_{hf})\sigma^2_{hf,\epsilon}, \qquad -\theta_{hf}\sigma^2_{hf,\bar{\eta}} = -\phi_{hf}\sigma^2_{hf,\epsilon}. \tag{4.6.6}$$

Note that there are multiple solutions for θ_{hf} and $\sigma^2_{hf,\bar{\eta}}$ and some of which result in an invertible process.[6] Thus for temporal aggregation of model $M4$, we can exploit the well-known results for ARMA process.

An m-period nonoverlapping aggregates of y_t [given in (4.6.5)] is defined by

$$Y_T = \sum_{t=m(T-1)+1}^{mT} y_t = (1 + B + \cdots + B^{m-1}) y_{mT} = \sum_{j=0}^{m-1} B^j y_{mT}, \tag{4.6.7}$$

where m is the fixed order of aggregation and T is the aggregate time unit. The time series y_t and Y_T will be called the basic HF and the aggregate LF time series, respectively [$m = 1$ implies no aggregation]. If the HF time series y_t follows an ARMA(1, 1) model, then the LF series Y_T in (4.6.7) follows an ARMA(1, 1) model but the relationship between the parameters of both models is complicated; see Ahsanullah and Wei (1984). If y_t follows the ARMA(1, 1) model

$$(1 - \phi_{hf} B) y_t = \mu_{hf} + (1 - \theta_{hf} B) \bar{\eta}_t, \tag{4.6.8}$$

[6] Equating coefficients and making substitutions leads to $\sigma^2_{hf,\bar{\eta}} = \sigma^2_{hf,\epsilon}\phi_{hf}/\theta_{hf}$ and θ_{hf} is a solution to the quadratic equation

$$\theta^2_{hf} - \theta_{hf}\tilde{k} + 1 = 0, \quad \text{where} \quad \tilde{k} = (\sigma^2_{hf,v} + \sigma^2_{hf,\epsilon}(1 + \phi^2_{hf}))/(\sigma^2_{hf,\epsilon}\phi_{hf}).$$

It can be shown that $\tilde{k}^2 - 4 = (\tilde{k} - 2)(\tilde{k} + 2)$ is positive since $\tilde{k} > 2$ is equivalent to $\sigma^2_{hf,v} + \sigma^2_{hf,\epsilon}(1 - \phi_{hf})^2 > 0$. The induced model is invertible if $|\theta_{hf}| < 1$ which after some algebra is shown to be true for the root $(\tilde{k} + (\tilde{k}^2 - 4)^{1/2})/2$ when $0 < \phi_{hf} < 1$ and for the root $(\tilde{k} - (\tilde{k}^2 - 4)^{1/2})/2$ when $-1 < \phi_{hf} < 0$.

then Y_T follows the ARMA(1, 1) model

$$(1-\phi_{lf}B)Y_T = \mu_{lf} + (1-\theta_{lf}B)\zeta_T \tag{4.6.9}$$

with

$$\phi_{lf} = \phi_{hf}^m, \quad \mu_{lf} = m\Big(\big(1-\phi_{hf}^m\big)\big(1-\phi_{hf}\big)\Big)\mu_{hf}, \tag{4.6.10}$$

and θ_{lf} is the root of the quadratic equation:

$$\theta_{lf}^2 + \psi_1\theta_{lf} + 1 = 0 \tag{4.6.11}$$

where

$$\begin{aligned}
&\psi_1 = \psi_2/\psi_3,\\
&\psi_2 = \sum_{i=0}^{m-1}\Big(1+(\phi_{hf}-\theta_{hf})\sum_{j=0}^{i-1}\phi_{hf}^j\Big)^2 + \sum_{i=m}^{2(m-1)}\Big((\phi_{hf}-\theta_{hf})\sum_{j=i-m}^{m-2}\phi_{hf}^j - \theta_{hf}\phi_{hf}^{m-1}\Big)^2 + \Big(\theta_{hf}\phi_{hf}^{m-1}\Big)^2,\\
&\psi_3 = \sum_{i=0}^{m-2}\Big(1+(\phi_{hf}-\theta_{hf})\sum_{j=0}^{i-1}\phi_{hf}^j\Big)\Big((\phi_{hf}-\theta_{hf})\sum_{j=i}^{m-2}\phi_{hf}^j - \theta_{hf}\phi_{hf}^{m-1}\Big) - \Big(1+(\phi_{hf}-\theta_{hf})\sum_{j=0}^{m-2}\phi_{hf}^j\Big)\theta_{hf}\phi_{hf}^{m-1},
\end{aligned}$$

and $\theta_{lf} = (-\psi_1 \pm \sqrt{\psi_1^2-4})/2$ such that $|\theta_{lf}| < 1$ to ensure invertibility of the LF model. Further, $\sigma_{lf,\zeta}^2 = \psi_2\sigma_{hf,\bar{\eta}}^2/(1+\theta_{lf}^2)$. LF ARMA parameters $(\theta_{lf}, \sigma_{lf,\zeta}^2)$ are related to LF state-space parameters $(\phi_{lf}, \sigma_{lf,v}^2, \sigma_{lf,\epsilon}^2)$ through non-linear equations [similar to (4.6.6)].

It is easy to see that the LF parameter $\rho_{lf} = \frac{-\mathrm{Cov}(\zeta_T\zeta_{T-1})}{\mathrm{Var}(\zeta_T)}$ has multiple solutions in terms of HF parameters. Thus, we use a plug-in estimator for ρ_{lf} and consider the test $\phi_{lf} = \phi_0$. Using LF observations, we can estimate ρ_{lf} by using equations (3.8)-(3.9) with $J = 10$ of Ahsan and Dufour (2019).[7] Note that using a plug-in estimator for ρ_{lf} may lead to some inconsequential size distortion.

We consider several LF values of $\phi_{lf} \in (0, 0.999)$. Equal-spaced HF intraday data are considered with frequency = {30s, 1m, 5m, 10m}, where s and m stand for second and minute. Therefore, within a day (trading hours = 6.5) the number of HF observations are $N\Delta t = \{780, 390, 78, 39\}$. For each frequency, we generate data from model $M4$, which is the HF model with parameters $\mu_{hf} = 10^{-6}$, $\sigma_{hf,\epsilon} = 3.5$, $\phi_{hf} = \phi_{lf}^{N\Delta t}$ and $\sigma_{hf,v} = \sigma_v/\sqrt{N\Delta t}$ with $\sigma_v = 0.15$. The

[7]The simple estimation method of Ahsan and Dufour (2019) is not only applicable for the $M4$ model but also applicable for a variety of state-space models.

HF sample size T_{hf} is equal to $T \times N\Delta t$, where T is LF sample size. For example, for 1m frequency, HF values of ϕ_{hf} are related to LF parameter by $\phi_{lf} = \phi_{hf}^{390}$. Thus, in order to generate nearly nonstationary LF volatility process, we use large values of ϕ_{hf}, *e.g.*, in case of 30s frequency, $\phi_{hf} = 0.999999$ is corresponds to $\phi_{lf} = 0.999$.

The simulation results for tests $\phi_{lf} = \phi_0$ are displayed in Tables 4.9-4.10. The following conclusions emerge from these tables.

First, we see from Table 4.9 that in all cases of HF IV's (these are the logarithms of RVs), the proposed tests (AR, AR^*, SS, SS^*) controls the levels very well: rejection frequencies are less than (or close to) 5%. This results holds whether sample sizes are different ($T = 100, 200$), or the instrument set contains different number of IV's ($l = 1, 5, 10$). However, as the number of IV's increases, the SS test under rejects (when $l = 10$).

Second, from Table 4.10, in all cases of HF IV's (30s, 1m, 5m, 10m), the proposed tests (AR, AR^*, SS, SS^*) have excellent power against alternative (where null is $\phi_{lf} = 0$): up to 100%, 100%, 99.7%, and 99.7%, respectively and the power of these tests increases with the sample size, and decreases as the number of IV's increases.

Third, all these tests have excellent power across different sampling frequency, *e.g.*, 1-minute or 5-minute. However, SS-type tests have less power compared to AR-type tests. In particular, as the number of IV's increases, SS tests have less power than other tests.

4.7 Application to stock prices

In this section, we consider various types of financial data, discuss a large number of IV's, and examine the strength of these IV's. The proposed tests are implemented with various IV's and confidence intervals for the volatility persistence parameter ϕ are constructed by inverting the tests.

4.7.1 Data description

The LF daily prices are obtained from the CRSP database. The raw series p_t is converted to returns by the transformation $r_t := 100[\log(p_t) - \log(p_{t-1})]$ and the returns are converted to residual returns by $s_t := r_t - \hat{\mu}_r$, where $\hat{\mu}_r$ is the sample average of returns. The sample period is from January 1, 2009, to December 31, 2013 (1258 trading days). The daily volatility proxy is constructed by the transformation $y_t = \log(s_t^2) + 1.2704$. Initially, we consider daily IV's of nine stocks: General Electric Company (GE), IBM Common Stock (IBM), JPMorgan Chase & Co. (JPM), The Coca-Cola Co (KO), Pfizer Inc. (PFE), Exxon Mobil Corporation (XOM) and (2) The Procter and Gamble Company (PG), AT&T Inc. (T) and Walmart Inc. (WMT). After examining the strength of daily IV's (in Section 4.7.4.1), we proceed with IBM stock and consider realized measures and option implied volatilities as IV's.

IBM's tick price data are taken from the TAQ (Trade and Quote) database and option (American) data are sourced from the OptionMetrics database. The access to these databases (CRSP, TAQ, OptionMetrics) is done through the Wharton Research Data Services. Using the tick data, we construct a large number of HF IV's. Details of these HF IV's are given in the following section and computations are carried out using the MATLAB Oxford MFE Toolbox developed by Sheppard (2013).[8] From IBM American options, three classes of implied volatility (ImV) are considered: (1) call options; (2) put options; (3) both call and put options. For each class, we use all implied volatilities available at a given date to construct six ImV subclasses, which are mean, minimum, maximum, and three quantiles (q1, q2, q3).

4.7.2 High-frequency instruments of asset price variability

We consider the HF volatility measures as the choice of IV's for daily volatility. Depending on the sampling frequency and estimation techniques, we can build different HF realized measures of volatility. Realized volatility was introduced by Andersen and Bollerslev (1998), who documented that the sum of squared intraday returns, known as the realized variance, provides an accurate measure of daily volatility. For the theoretical foundation of RV; see Andersen, Bollerslev, Diebold and Labys (2001), Barndorff-Nielsen et al. (2002).

Let $p_t = \log(S_t)$ denote the logarithmic price, where S_t is the observed price (at time t) and $r_t = p_t - p_{t-1}$ denote the continuously compounded return from time $t-1$ to t. Assume that the logarithmic price process, p_t, may exhibit both stochastic volatility and jumps, so that, it

[8]The Oxford MFE Toolbox can be downloaded from the GitHub: https://github.com/bashtage/mfe-toolbox.

could belong to the class of continuous-time jump diffusion processes:

$$dp_t = \mu_t d_t + \sigma_t dW_t + \kappa_t dq_t, \quad 0 \le t \le T, \tag{4.7.1}$$

where μ_t is a continuous and locally bounded variation process, σ_t is a stochastic volatility process, W_t is the standard Brownian motion, dq_t is a counting process such that $dq_t = 1$ represents a jump at time t (and $dq_t = 0$ if no jump) with jump intensity $\tilde{\lambda}_t$. If p_{t-} denotes the price immediately prior to the jump at time t, then the magnitude of the random jump $\kappa_t = \Delta p_t = p_t - p_{t-}$. The process p_t consists of a continuous component and a pure jump component. The quadratic variation (QV) of this process is defined by

$$[r,r]_t = \int_0^t \sigma_s^2 dW_s + \sum_{0<s\le t} \kappa_s^2 \tag{4.7.2}$$

where the first component, called integrated volatility (IVol), comes from the continuous component of (4.7.1), and the second term is the contribution from discrete jumps. In the absence of jumps, the second term on the right-hand side disappears, and the QV is simply equal to the IVol. Here we consider several classes of HF IV's, which can be categorized as follows.

4.7.2.1 Classes of realized measures not robust to jumps

These classes of realized measures have been proposed to provide robustness to various types of market microstructure effects (bid-ask bounce, stale quotes, mis-reported prices) and improve the efficiency of volatility estimates. We consider five broad classes of realized measures, which are consistent estimators of the QV in the absence of jumps.

1. **Realized volatility (RV)**: RV is defined as the sum of squared intraday returns. By dividing an interval of time, *e.g.*, $[T_0, T_1]$, into n subintervals, $T_0 = t_{0,n} < t_{1,n} < \cdots < t_{n,n} = T_1$, we can define intraday returns, $r_{i,n} = p_{t_{i,n}} - p_{t_{i-1,n}}$, and then $RV_t = \sum_{i=1}^{n} r_{i,n}^2$. Andersen, Bollerslev, Diebold and Labys (2001) showed that the RV is a consistent estimator for the QV:
$$RV_t \xrightarrow{p} IVol_t = \int_0^t \sigma_s^2 dW_s.$$

2. **RV with optimal sampling (RVbr)**: A standard RV estimator with optimal sampling is

proposed by Bandi and Russell (2008), where the optimal sampling frequency is calculated using estimates of integrated quarticity and variance of the microstructure noise. This bias-corrected estimator removes the estimated impact of market microstructure noise. In the empirical applications below, we compute RVbr with an estimated optimal sampling frequency, which is the key feature of this estimator.

3. **Multi-scales RV (MSRV)**: The multi-scales RV by Zhang (2006) uses a combination of several high and lower frequencies to remove the noise and estimate the volatility. It is a generalization of two-scales RV [Zhang et al. (2005)] and can be defined as:

$$MSRV_t = [r,r]_t^{(K)} - \frac{\bar{n}_K}{\bar{n}_J}[r,r]_t^{(J)} \xrightarrow{p} IVol_t, \quad 1 \leq J < K \leq n,$$

   where J and K are the time scales and $\bar{n}_i = (n-i+1)/i$ with $i = J, K$.

4. **Realized kernels (RK)**: The realized kernel by Barndorff-Nielsen et al. (2008) is a robust measure of volatility, which ensures consistency and positive semi-definiteness. Several generalizations to handle more lags and various shapes of autocorrelation function are derived in Barndorff-Nielsen et al. (2011). In this paper, we use the latter variant, which is given by

$$RK = \sum_{h=-H}^{H} k\left(\frac{h}{H+1}\right)\gamma_h$$

   where $k(x)$ is the kernel function and $\gamma_h = \sum_{i=|h|+1}^{n} r_{i,n} r_{i-h,n}$. We consider four types of kernel functions: (1) Bartlett kernel [RKbart: $k(x) = 1-x$, flat-top, $n^{1/6}$ rate]; (2) Cubic kernel [RKcub: $k(x) = 1-3x^2+2x^3$, flat-top, $n^{1/4}$ rate]; (3) Parzen kernel [RKnfp: $k(x) = \{1-6x^2+6x^3 \text{ if } 0 \leq x \leq 1/2\ , 2(1-x)^3 \text{ if } 1/2 \leq x \leq 1\}$, non-flat-top, $n^{1/5}$ rate]; (4) Tukey-Hanning kernel with power 2 [RKth2: $k(x) = \sin^2\{\pi/2(1-x)^2\}$, flat-top, $n^{1/4}$ rate].

5. **Realized range RV (RRV)**: The realized range RV [Christensen and Podolskij (2007)] uses sums of normalized squared high-low ranges for intra-daily periods rather than sums of squared returns. As a result, it is based on extremes from the entire price path and provides more information than returns sampled at fixed time intervals. Decomposing the daily time interval into K *non-overlapping* intervals of size m_K, the estimator is given

by:

$$RRV^{(m_K,K)} = \frac{1}{\lambda_{2,m_K}} \sum_{i=1}^{K} s_i^{(m_K)^2} \xrightarrow{p} IVol$$

where the range of the price process over the ith interval is given by $s_i^{(m_K)} = \max_{0\le h,l\le m_K}\left(p_{\frac{i-1+h/mK}{K}} - p_{\frac{i-1+l/mK}{K}}\right)$, $i = 1,\ldots,K$, and $\lambda_{2,m_K} = E[\max_{0\le h,l\le m_K}(W_{h/m_K} - W_{l/m_K})^2]$ is the second moment of the range of a standard Brownian motion over the unit interval with m_K observed increments.

4.7.2.2 Classes of realized measures robust to jumps

In the presence of jumps, the RV is a consistent estimator of the QV [see Andersen and Bollerslev (1998), Andersen, Bollerslev, Diebold and Labys (2001), Barndorff-Nielsen et al. (2002)], which is a combination of IVol and jump variation (JV):

$$RV_t \xrightarrow{p} \underbrace{\int_0^t \sigma_s^2 dW_s}_{IVol_t} + \underbrace{\sum_{0<s\le t} \kappa_s^2}_{JV_t}. \tag{4.7.3}$$

We consider two classes of jump-robust realized measures:

1. **Bipower variation (BV)**: The most widely used estimator of IVol in the presence of jumps is the Bipower variation of Barndorff-Nielsen and Shephard (2004). It is the sum of adjacent absolute returns:

$$BV_t := \frac{\pi}{2}\sum_{i=2}^{n} |r_{i-1,n}||r_{i,n}| \xrightarrow{p} IVol_t = \int_0^t \sigma_s^2 dW_s. \tag{4.7.4}$$

2. **Nearest neighbor truncated RV**: Andersen et al. (2012) used nearest neighbor truncation approach to estimate the integrated volatility, where the median RV (MedRV) and minimum RV (MinRV) estimators use min or median of blocks of returns (MinRV with blocks of two returns and MedRV with blocks of three returns). The proposed estimators are:

$$MinRV_n = \frac{\pi}{\pi-2}\left(\frac{n}{n-1}\right)\sum_{i=1}^{n-1} [\min(|r_{i,n}|,\ |r_{i+1,n}|)]^2,$$

$$MedRV_n = \frac{\pi}{6-4\sqrt{3}+\pi}\left(\frac{n}{n-2}\right)\sum_{i=2}^{n-2}[\mathrm{med}(|r_{i-1,n}|,\ |r_{i,n}|,\ |r_{i+1,n}|)]^2.$$

4.7.2.3 Additional HF measures and jump variations

We also consider realized semivariance (RSV), JV, and signed JV (SJV) and squared logarithms of the latter (JV, SJV):

1. **Jump variation** Combining the results in equations (4.7.3) and (4.7.4), the contribution of the JV in the QV can be consistently estimated by

$$JV_t := RV_t - BV_t \xrightarrow{p} \sum_{0<s\leq t} \kappa_s^2; \tag{4.7.5}$$

see Barndorff-Nielsen and Shephard (2006).

2. **Realized semivariance** Barndorff-Nielsen et al. (2010) proposed RSV estimators that can capture the variation only due to negative or positive returns. These estimators are defined as:

$$RSV_t^+ := \sum_{j=1}^{n} r_{t_j}^2 \mathbf{1}_{\{r_{t_j}>0\}} \xrightarrow{p} \frac{1}{2}\int_0^t \sigma_s^2 dW_s + \sum_{0\leq s\leq t} \kappa_s^2 \mathbf{1}_{\{\kappa_s>0\}}, \tag{4.7.6}$$

$$RSV_t^- := \sum_{j=1}^{n} r_{t_j}^2 \mathbf{1}_{\{r_{t_j}<0\}} \xrightarrow{p} \frac{1}{2}\int_0^t \sigma_s^2 dW_s + \sum_{0\leq s\leq t} \kappa_s^2 \mathbf{1}_{\{\kappa_s<0\}}, \tag{4.7.7}$$

where the first term in the limit of both RSV^+ and RSV^- is one-half of the integrated variance. These estimators provide a complete decomposition of RV, in the sense that $RV = RSV^+ + RSV^-$.

3. **Signed jump variation** The variation due to the continuous component can be removed by subtracting one RSV from the other without estimating it separately. The remaining part is defined as the signed jump variation:

$$SJV_t := \lim_{n\to\infty}(RSV_t^+ - RSV_t^-) = \sum_{0\leq s\leq t} \kappa_s^2 \mathbf{1}_{\{\kappa_s>0\}} - \sum_{0\leq s\leq t} \kappa_s^2 \mathbf{1}_{\{\kappa_s<0\}}. \tag{4.7.8}$$

4.7.3 Final instrument set, econometric model and test statistics

The final instrument set also includes principal component factors (PCF) and daily log volatility of y_t. The three largest principal component factors are extracted from HF IV's; see Table 4.11 for details. Formally, PCF-based identification-robust inference in the context of IV regressions was considered by Kapetanios et al. (2016) to deal with the problem of many IV's. Note that we use logarithms of RV-RSVP and PCF classes of IV's; see Table 4.11.

We consider one hundred and seventy-five IVs, which can be divided into 22 classes. The description of these IV's are given in Table 4.11. The HF subclass includes different sampling frequencies [tick, second and minute], sampling scheme [tick or business], and sub-sampling. We use 1-minute sub-sampling [ss] in the calculation of several HF measures.

For empirical analysis, we consider the following GSV model:

$$w_t = \mu + \phi w_{t-1} + v_t\,, \qquad y_t = w_t + \epsilon_t\,, \qquad v_t \sim \text{i.i.d. } N(0, \sigma_v^2)\,, \quad \epsilon_t \sim \text{i.i.d. } \log(\chi^2_{(1)})\,, \tag{4.7.9}$$

$$y_{t-1} = \bar{\pi}_0 + Z'_{t-2}\bar{\pi}_1 + \eta_{t-1}\,, \quad \eta_{t-1} := \epsilon_{t-1} + u_{t-1}\,, \quad u_t \sim \text{i.i.d. } N(0, \sigma_u^2)\,, \tag{4.7.10}$$

where $w_t = \log(\sigma_t^2)$, $y_t = \log(s_t^2) + 1.2704$ with $s_t := r_t - \mu_r$ is residual return of an asset with μ_r is the mean of return $r_t = 100[\log(p_t) - \log(p_{t-1})]$ and Z_{t-2} is the set of IV's.

For inference, we consider joint tests $(\phi,\ \rho) = (\phi_0,\ \rho_0)$. The inference procedures AR, AR^*, SS, and SS^* are proposed in Sections 4.3.1-4.3.4 and corresponding test statistics are given in equations (4.3.22), (4.3.25), (4.3.31) and (4.3.35). We use $\tau = 0.2$ for SS-type tests and employ 99 Monte Carlo replications for PO type procedures.

4.7.4 Results

In this section, we examine the strength of IV's; after that, we build projection-based confidence sets for the volatility persistence parameter of IBM stock by employing numerous types of instruments. Using the lengths of these identification-robust confidence sets, we identify several crucial empirical stylized facts.

4.7.4.1 Strength of IV's

We investigate the strength of daily IV's since a pressing concern with an IV approach is the possible use of weak IV's, which can produce biased estimators [bias towards OLS estimates] and hypothesis tests with large size distortion. The existing econometric literature defines weak IV's based on the strength of the first-stage equation [Bekker (1994), Dufour (2003), Staiger and Stock (1997), and Stock and Yogo (2005)]. Following Stock and Yogo (2005), we employ the first-stage F-statistic to detect whether IV's are weak or not.

Tests based on F-statistics that whether daily IV's [past lags of the endogenous variable] all have zero coefficients are reported in Table 4.12 with corresponding critical values associated with the desired maximum level of size distortion. From the table, we can see that many F-statistics are less than the corresponding critical value associated with the maximum asymptotic size of a Wald test [these critical values are obtained using weak-IV asymptotic distributions]. These results suggest that IV estimates are biased towards OLS estimates, and we need to use weak instrument robust inference methods: see Dufour (1997) for more details about the Wald test.

Now, we wish to check if the HF and other IV's are weak or not. We consider IBM stock and different classes of IV's. Results with other stocks are qualitatively similar and omitted to conserve space. Table 4.13 reports the first-stage F-statistics of all IV's. From the results, we can draw several conclusions: (1) most of the HF IV's are strong for IBM, but exceptions are JV and SJV HF classes, ImV-mean subclass, and daily IV's; (2) if we consider multiple IV's, then Wald-type tests fail to control the level in many cases; (3) in most cases, the value of F-statistic (that measures the strength of IV's) is maximum, when we consider only one instrument irrespective of it is weak or strong.

4.7.4.2 Projection-based confidence sets

To construct a projection-based confidence interval for the volatility persistence parameter ϕ, we first construct a confidence interval for λ with level $(1-\alpha_1)$, denoted as $C_{\alpha_1}(\lambda)$. We parametrize the noise ratio λ rather than ρ since this is the more natural choice. We set $\alpha_1 = 0.05$, and compute λ using the simple winsorized method proposed by Ahsan and Dufour (2019). We use equations (3.8)-(3.9) with $J = 10$ of Ahsan and Dufour (2019) to estimate σ_v^2 and the corresponding standard error (SE). By setting $\sigma_\epsilon^2 = \pi^2/2$, the SE of $\hat{\lambda} = \sigma_\epsilon^2/\hat{\sigma_v}^2$ is computed using the delta method. The estimated 95% confidence interval for the nuisance parameter λ is $C_{0.05}(\lambda) = [33.943,\ 61.154]$ with $\hat{\lambda} = 47.548$ and $\text{SE}(\hat{\lambda}) = 6.935$. For each value of λ in the confidence interval $C_{\alpha_1}(\lambda)$, we then construct $(1-\alpha_2)$ confidence intervals for ϕ given λ [denoted by $C_{\alpha_2}(\phi|\lambda)$] by inverting a test robust to weak IV's proposed in Sections 4.3.1-4.3.4. By Bonferroni's inequality, this confidence interval has coverage of at least $100(1-\alpha)$%, where $\alpha = \alpha_1 + \alpha_2$. If we use $\alpha_2 = 0.05$, then a 90% confidence interval for ϕ that does not depend on λ can be obtained by

$$C_{0.10}(\phi) = \bigcup_{\lambda \in C_{0.05}(\lambda)} C_{0.05}(\phi|\lambda).$$

The projection method is thoroughly discussed in Section 4.3.5. Note that we employ grid testing during the test inversion, in which a series of tests $[H_0 : \phi = \phi_0,\ \lambda = \lambda_0$, where $\phi_0 \in [0,\ 1]$, $\lambda_0 \in C_{\alpha_1}(\lambda)]$ performed. Note that we restrict ϕ_0 in the most relevant part of the parameter space, *i.e.*, $\phi_0 \in [0,\ 1]$.

We use $\alpha_1 = \alpha_2$, which is the rule typically employed in the literature on simultaneous inference (*e.g.*, in Bonferroni-type procedures) and test combination; see Miller (1981), Savin (1984). Cavanagh et al. (1995) suggest a refinement of the Bonferroni method that makes it less conservative than the basic approach. The idea is to shrink the confidence interval for λ so that the refined interval is a subset of the original (unrefined) interval. This consequently shrinks the Bonferroni confidence interval for ϕ, achieving an exact test of the desired significance level. However, it is important to note that α should be selected a priori, not on the basis of the results yielded by different choices of α_1 for a given sample.

As pointed out by Dufour (1997), when IV's are arbitrary weak, then confidence sets with correct coverage probability must have an infinite length with positive probability.[9] As a result, the length of a weak instrument robust confidence interval can summarize the corresponding instrument's identification strength. Since we restrict $\phi_0 \in [0,\ 1]$, then an irrelevant (no identi-

[9] Dufour (1997) showed that if the IV's are not correlated with the regressor [irrelevant IV's], then the corresponding parameter is not identified, and any value of the parameter is consistent with data. A valid confidence set in such a case must be infinite, at least with probability equal to the coverage. Most empirical applications use the conventional Wald confidence interval, which is always finite. As a result, the Wald confidence interval has a low coverage probability and should not be used when IV's are weak.

fication) instrument for the regressor should produce a confidence interval with length equal to 1.

From an identification-robust confidence interval, we define the precision (or informational efficiency) of an instrument set i as follows:

$$d_i := 1 - (ub_i - lb_i) \tag{4.7.11}$$

where ub and lb are the upper and lower bound of the confidence set, and $ub - lb$ is the length of the confidence set. The definition d_i implies that if i is a weak instrument then it will produce d_i close to 0 and if i is a strong instrument then it will produce d_i close to 1. For example, a large value of d_i implies that the corresponding instrument set is highly informative about the parameter ϕ.

Figure 4.1 shows the precision measure d_i of different classes of IV's, where the instrument set consists of a constant and a lag of the corresponding instrument. For each class, we consider average, median, minimum, and maximum precision measures across the proposed inference methods. The following inferences emerge from Figure 4.1. *First*, except for JV and SJV classes, all HF classes are considered as strong instruments, *i.e.*, these classes produce very high d_i values. These results hold in all precision measures and across four inference methods. *Second*, JV and SJV classes have many weak and irrelevant (no identification) IV's because average and median precision measures of these classes are low and zero, respectively. These results suggest that JV and SJV classes have no or little predictive power regarding the latent daily volatility. However, log squared JV and SJV IV's are informative about the volatility clustering. This finding suggests that the second moment of jumps or signed jumps is correlated with the latent daily volatility proxy. *Third*, both PCF and ImV classes have some relevant IV's. However, all ImV classes include some weak IV's. *Fourth*, according to SS-type tests, the daily instrument is uninformative regarding the latent daily volatility proxy. That is, SS and SS^* produce d_i equal to 0.047 and 0, indicating weak and no identification, respectively.

Figure 4.2 shows the precision measure of different subclasses of HF IV's. On average, all HF subclasses produce confidence intervals with similar lengths, *e.g.*, on average, both 1s and 5m produce almost similar identification-robust confidence intervals. Note that the instrument equation (4.7.10) connects the daily volatility proxy to the instrument set. With an instrument set containing a constant and a lag of HF instrument, (4.7.10) yields:

$$y_{t-1} = \bar{\pi}_0 + \bar{\pi}_1 Z_{t-2} + \eta_{t-1}$$

where y_{t-1} is the daily volatility proxy and Z_{t-2} is the selected RV instrument. The constant term $\bar{\pi}_0$ captures the bias in the RV estimate due to the non-trading hours and microstructure noise. If the bias-correction term $\bar{\pi}_0$ is negative, RV has an upward bias that may be due to the market microstructure noise, and if $\bar{\pi}_0$ is positive, it has a downward bias due to the non-trading hours; see Takahashi et al. (2009). Due to this bias-correction term, the proposed inference methods produce confidence intervals robust to the non-trading hours and microstructure noise even with a very high sampling frequency.

To formalize, we define the notion of the average precision of an instrument set i over the proposed inference methods by

$$\bar{d}_{i,s} := \frac{1}{S}\sum_{i=1}^{S} d_i \tag{4.7.12}$$

where $s \in S$ and S is the set of identification-robust inference methods. We use this measure to rank the information content of instruments. Table 4.14 reports the projection-based 90% confidence intervals for ϕ using strong IV's, *i.e.*, based on $\bar{d}_{i,s}$. Panel A includes superior IV's while panel B and C include IV's which produce slightly larger confidence sets compared to the IV's in panel A. Panel A mostly includes HF IV's, and 70% of these are 5m subclass. This finding proves that HF RV does provide an additional gain in predicting the LF volatility proxy. The best instrument is RSVN-5m-ss. The average implied volatility that extracts from IBM call options is also a strong instrument. This finding is in line with Christensen and Prabhala (1998), who find that implied volatility has large explanatory power regarding past volatility. We also find that confidence sets with 30s RVs [Panel C: RSVN-30s, RV-30s, BV-30s, MSRV-30s] are spacious than confidence sets with 5m RVs [Panel A and B] and conclude that the effect of market microstructure noise leads to slightly wider confidence sets. It is well-known that the market microstructure noise becomes progressively more dominant as the sampling frequency increases; see Zhang et al. (2005), Bandi and Russell (2008), and Hansen and Lunde (2006). Thus, our result suggests that the proposed inference methods produce valid confidence sets even with noisy RVs at a higher frequency. Further, 85% of the time, Panel A and B include IV's with frequency 1m, 5m, and 10m. These confidence sets are less sensitive to the market microstructure noise.

Table 4.14 also gives several other conclusions. *First*, we can infer from these confidence sets that the persistence parameter lies roughly between 0.85 and 1.0 for IBM. This outcome indicates that the volatility process is highly persistent, close to unit-root, consistent with the empirical literature. These confidence sets include $\phi = 1$, implying that these sets are also robust to nonstationarity. *Second*, in all cases, simulation-based point-optimal confidence sets [AR^* and SS^*] are conservative compared to the corresponding AR-type confidence sets [AR and SS].

Table 4.15 presents the projection-based 90% confidence intervals for ϕ using weak IV's, *i.e.*, based on $\bar{d}_{i,s}$. Panel A of Table 4.15 contains IV's with no identification. As a result, these IV's produce unbounded confidence intervals. These confidence intervals cover the entire set of $\phi \in [0,\ 1]$. Panel A comprises mostly by JV and SJV HF classes and ImV-max subclass. Note that under no identification, all values of ϕ are observationally equivalent, which implies that the proposed test statistics yield valid confidence sets that are unbounded with a non-zero probability. Consequently, the proposed tests are robust to weak identification. From Panel C, we find that the LF daily instrument produces a valid confidence set. However, the length of this set is larger compared to HF confidence sets given in Table 4.14. We also see that in some cases, SS-type tests produce confidence intervals that are entirely different from those that are provided by AR-type tests. This finding could be because SS-type tests are computed from the second part of the sample and may be affected by an unmodeled structural change.

In Table 4.16, we report the estimated confidence intervals, where the instrument set includes a constant and several lags of an instrument, $l = 1,\ 3,\ 5$. In this setup, we use the first set of strong IV's [Table 4.14 - Panel A], ImV-C-q3, and 1-day. In most cases, we find that all confidence intervals for ϕ (AR, AR^*, SS, SS^*) are getting wider as l increases. The average length of these confidence intervals when $l = 3,\ 5$ are larger than the confidence intervals were when $l = 1$. Therefore, we do not see any apparent gains by adding more lags in the instrument set. The only exception is the LF daily instrument, where the average length of confidence intervals is shorter than before. This result implies that we should use more daily lags as IV's to get a smaller confidence set. We also construct several confidence sets where the instrument set includes a constant and various combinations of strong IV's. We report these confidence sets in Table 4.17. The conclusion is similar to Table 4.16, *i.e.*, no apparent gains from combining strong IV's.

Finally, these confidence sets can be extended to allow for non-Gaussian error distributions [where the conditional distribution of scale transformed error has a non-Gaussian error distribution] using the MCT procedure described in Section 4.4. Furthermore, these confidence intervals formed from a range of accepted values due to grid testing; thus, it is easy to get a nonparametric estimate of ϕ by applying the Hodges-Lehmann principal.

4.8 Conclusion

This paper has introduced a novel class of GSV models, which can use high-frequency information content and accommodate nonstationary volatility. We employ IV methods to provide a unified framework for the analysis of GSV models.

In the framework of GSV class models, we have studied the problem of testing hypotheses and building confidence sets for the volatility persistence parameter. This parameter has an intrinsic interest because it measures the latent volatility process's persistence, *i.e.*, "volatility clustering of asset returns". We proposed more reliable identification-robust finite-sample procedures, which are robust to weak IV's and/or nonstationary latent volatility. We also showed that these finite-sample procedures (based on a Gaussian assumption on the errors) remain asymptotically valid under weaker distributional assumptions. We then study the statistical properties of the proposed tests in simulation experiments. These tests outperform the asymptotic t-type test in terms of size and exhibit excellent power.

We applied these methods to IBM's price and option data and observed several empirical facts. The superior instrument set constitutes of HF realized measures and call option implied volatilities. These IV's produce confidence sets, which show that the latent volatility process of IBM is close to unit-root. We find RVs at higher frequency produce spacious confidence intervals than RVs at slightly lower frequencies, pointing out that these confidence intervals adjust to incorporate the microstructure noise. We also find jumps and signed jumps have no or little information content regarding the low-frequency volatility, whereas their log squared versions have a strong identification strength. When we consider irrelevant or weak instruments, the proposed procedures give unbounded confidence intervals.

Finally, it is easy to see that the inference methods used in this paper can be adapted to other situations, *e.g.*, measurement error in ARMA-type models, or noisy realized measures in HAR volatility modeling. The extension to multivariate models and parameter estimation in the GSV framework are topics of ongoing research.

4.9 Appendix

4.9.1 Proofs

PROOF OF PROPOSITION 4.4.1 When $\phi = \phi_0$ and $\rho = \rho_0$, on multiplying the two sides of (4.3.21) by $M_{C_0}[X] - M_{C_0}[X, Z_{-2}]$ and $M_{C_0}[X]$, we see that:

$$\begin{aligned}(M_{C_0}[X] - M_{C_0}[X, Z_{-2}])C_0(y - \phi_0 y_{-1}) &= \sigma_\xi(M_{C_0}[X] - M_{C_0}[X, Z_{-2}])\vartheta\,,\\ M_{C_0}[X]C_0(y - \phi_0 y_{-1}) &= \sigma_\xi M_{C_0}[X]\vartheta\,.\end{aligned} \tag{4.9.1}$$

Thus, the AR-statistic in (4.3.22) can be rewritten as:

$$AR(\phi_0, \rho_0) = \frac{\sigma_\xi^2 \vartheta'(M_{C_0}[X] - M_{C_0}[X, Z_{-2}])\vartheta / l}{\sigma_\xi^2 \vartheta' M_{C_0}[X]\vartheta / (T - l - k)} = \frac{\vartheta'(M_{C_0}[X] - M_{C_0}[X, Z_{-2}])\vartheta / l}{\vartheta' M_{C_0}[X]\vartheta / (T - l - k)}\,.$$

Hence, the null conditional distribution of $AR(\phi_0, \rho_0)$, given $\overline{X}$, only depends on distribution of ϑ. If normality holds conditional on $\overline{X}$, *i.e.*, $\vartheta \mid X \sim \mathrm{N}(0, I_T)$, we have $\vartheta'(M_{C_0}[X] - M_{C_0}[X, Z_{-2}])\vartheta \sim \chi^2_{(l)}$ and $\vartheta' M_{C_0}[X]\vartheta \sim \chi^2_{(T-l-k)}$. Since $M_{C_0}[X, Z_{-2}](M_{C_0}[X] - M_{C_0}[X, Z_{-2}]) = 0$, hence $\vartheta'(M_{C_0}[X] - M_{C_0}[X, Z_{-2}])\vartheta$ and $\vartheta' M_{C_0}[X]\vartheta$ are independent conditional on $\overline{X}$. Consequently, $AR(\phi_0, \rho_0) \sim F(l, T - l - k)$. □

PROOF OF PROPOSITION 4.5.1 Under the null hypothesis $\phi = \phi_0$,

$$AR_T(\phi_0) = \kappa(T)\frac{\Lambda_{1T} - \Lambda_{2T}}{\Lambda_{2T}/T}\,, \tag{4.9.2}$$

where

$$\Lambda_{1T} := \xi(T)' M[Q_{1T}]\xi(T)\,, \quad \Lambda_{2T} := \xi(T)' M[Q_T]\xi(T)\,, \quad \kappa(T) := \frac{T - l - k}{lT}\,.$$

Under the Assumption (4.5.1), we have

$$\kappa(T) \underset{T \to \infty}{\longrightarrow} \frac{1}{l}, \tag{4.9.3}$$

$$q_2 \mid q_1 \sim \mathrm{N}(\Sigma_{Q_2Q_1}\Sigma^{-1}_{Q_1Q_1} q_1, \sigma_\xi^2 \Sigma_{q_2|q_1}), \tag{4.9.4}$$

where $\Sigma_{q_2|q_1} = \Sigma_{Q_2Q_2} - \Sigma_{Q_2Q_1}\Sigma_{Q_1Q_1}^{-1}\Sigma_{Q_1Q_2}$. Then

$$(q_2 - \Sigma_{Q_2Q_1}\Sigma_{Q_1Q_1}^{-1}q_1)'\Sigma_{q_2|q_1}^{-1}(q_2 - \Sigma_{Q_2Q_1}\Sigma_{Q_1Q_1}^{-1}q_1) \sim \sigma_\xi^2\chi_{(l)}^2. \tag{4.9.5}$$

$$\begin{aligned}
\Lambda_{1T} - \Lambda_{2T} &= \xi(T)'M[Q_{1T}]\xi(T) - \xi(T)'M[Q_T]\xi(T) \\
&= \xi(T)'(I - P[Q_{1T}])\xi(T) - \xi(T)'(I - P[Q_T])\xi(T) \\
&= \xi(T)'Q_T(Q_T'Q_T)^{-1}Q_T'\xi(T) - \xi(T)'Q_{1T}(Q_{1T}'Q_{1T})^{-1}Q_{1T}'\xi(T) \\
&= \xi(T)'Q_TD_T(D_T'Q_T'Q_TD_T)^{-1}D_T'Q_T'\xi(T) - \xi(T)'D_{1T}Q_{1T}(D_{1T}'Q_{1T}'Q_{1T}D_{1T})^{-1}D_{1T}'Q_{1T}'\xi(T) \\
&\Longrightarrow q'\Sigma_{QQ}^{-1}q - q_1'\Sigma_{Q_1Q_1}^{-1}q_1.
\end{aligned} \tag{4.9.6}$$

Now using standard formulas of a partitioned matrix inverse for Σ_{QQ} and setting $S = q'\Sigma_{QQ}^{-1}q - q_1'\Sigma_{Q_1Q_1}^{-1}q_1$ [see Gentle (2007), Section 3.4.1], we have

$$\begin{aligned}
S &= q'\Sigma_{QQ}^{-1}q - q_1'\Sigma_{Q_1Q_1}^{-1}q_1 \\
&= (q_1', q_2')' \begin{bmatrix} \Sigma_{Q_1Q_1}^{-1} + \Sigma_{Q_1Q_1}^{-1}\Sigma_{Q_1Q_2}\Sigma_{q_2|q_1}^{-1}\Sigma_{Q_2Q_1}\Sigma_{Q_1Q_1}^{-1} & -\Sigma_{Q_1Q_1}^{-1}\Sigma_{Q_1Q_2}\Sigma_{q_2|q_1}^{-1} \\ -\Sigma_{q_2|q_1}^{-1}\Sigma_{Q_2Q_1}\Sigma_{Q_1Q_1}^{-1} & \Sigma_{q_2|q_1}^{-1} \end{bmatrix} \begin{pmatrix} q_1 \\ q_2 \end{pmatrix} \\
&\quad - q_1'\Sigma_{Q_1Q_1}^{-1}q_1 \\
&= q_1'\Sigma_{Q_1Q_1}^{-1}q_1 + q_1'\Sigma_{Q_1Q_1}^{-1}\Sigma_{Q_1Q_2}\Sigma_{q_2|q_1}^{-1}\Sigma_{Q_2Q_1}\Sigma_{Q_1Q_1}^{-1}q_1 - 2q_2'\Sigma_{Q_1Q_1}^{-1}\Sigma_{Q_1Q_2}\Sigma_{q_2|q_1}^{-1}q_2 + q_2'\Sigma_{q_2|q_1}^{-1}q_2 \\
&\quad - q_1'\Sigma_{Q_1Q_1}^{-1}q_1 \\
&= q_1'\Sigma_{Q_1Q_1}^{-1}\Sigma_{Q_1Q_2}\Sigma_{q_2|q_1}^{-1}\Sigma_{Q_2Q_1}\Sigma_{Q_1Q_1}^{-1}q_1 - 2q_2'\Sigma_{Q_1Q_1}^{-1}\Sigma_{Q_1Q_2}\Sigma_{q_2|q_1}^{-1}q_2 + q_2'\Sigma_{q_2|q_1}^{-1}q_2 \\
&= (q_2 - \Sigma_{Q_2Q_1}\Sigma_{Q_1Q_1}^{-1}q_1)'\Sigma_{q_2|q_1}^{-1}(q_2 - \Sigma_{Q_2Q_1}\Sigma_{Q_1Q_1}^{-1}q_1).
\end{aligned} \tag{4.9.7}$$

Thus, from (4.9.5), (4.9.6), and (4.9.7), we have

$$\Lambda_{1T} - \Lambda_{2T} \Longrightarrow \sigma_\xi^2\chi_{(l)}^2, \quad \text{and} \quad \frac{\Lambda_{2T}}{T} \xrightarrow[T\to\infty]{p} \sigma_\xi^2, \quad \text{hence} \quad AR_T(\phi_0) \Longrightarrow \frac{\chi_{(l)}^2}{l}.$$

□

PROOF OF PROPOSITION 4.5.2 Under the null hypothesis ($\phi = \phi_0, \rho = \rho_0$),

$$AR_T^*(\phi_0, \rho_0, \rho_1) = \frac{\Lambda_{1T} - \Lambda_{2T}}{\Lambda_{2T}/T}, \tag{4.9.8}$$

where

$$\Lambda_{1T} := \hat{\xi}(T)' M[\hat{Q}_{1T}]\hat{\xi}(T), \quad \Lambda_{2T} = \tilde{\xi}(T)' M[\tilde{Q}_T]\tilde{\xi}(T).$$

Under the Assumption 4.5.2, we have

$$\begin{aligned}
\Lambda_{2T}/T &= \tilde{\xi}(T)'\tilde{\xi}(T)/T - \tilde{\xi}(T)' P[\tilde{Q}_T]\tilde{\xi}(T)/T \\
&= \tilde{\xi}(T)'\tilde{\xi}(T)/T - \tilde{\xi}(T)'\tilde{Q}_T D_T (D_T' \tilde{Q}_T' \tilde{Q}_T D_T)^{-1} D_T' \tilde{Q}_T' \tilde{\xi}(T)/T \\
&= \tilde{\xi}(T)'\tilde{\xi}(T)/T - \tilde{\xi}(T)'\tilde{Q}_T D_T (D_T' \tilde{Q}_T' \tilde{Q}_T D_T)^{-1} D_T' \tilde{Q}_T' \tilde{\xi}(T)/T \\
&\xrightarrow[T\longrightarrow\infty]{p} \sigma_\xi^2,
\end{aligned} \tag{4.9.9}$$

where the last equality follows from

$$\tilde{\xi}(T)'\tilde{\xi}(T)/T \xrightarrow[T\longrightarrow\infty]{p} \sigma_\xi^2, \quad \tilde{\xi}(T)'\tilde{Q}_T D_T (D_T' \tilde{Q}_T' \tilde{Q}_T D_T)^{-1} D_T' \tilde{Q}_T' \tilde{\xi}(T)/T \Longrightarrow \frac{\sigma_\xi^2 \chi^2_{(l+k)}}{T} \xrightarrow[T\longrightarrow\infty]{} 0.$$

Now using restrictions under the null and alternative that $\hat{\xi}(T) = \tilde{\xi}(T) := \xi_T^* \sim N(0, I_T)$, we have

$$\begin{aligned}
\Lambda_{1T} - \Lambda_{2T} &= \hat{\xi}(T)' M[\hat{Q}_{1T}]\hat{\xi}(T) - \tilde{\xi}(T)' M[\tilde{Q}_T]\tilde{\xi}(T) \\
&= \xi_T^{*\prime} M[\hat{Q}_{1T}]\xi_T^* - \xi_T^{*\prime} M[\tilde{Q}_T]\xi_T^* \\
&= [\xi_T^{*\prime}\xi_T^* - \xi_T^{*\prime}\xi_T^*] + [\xi_T^{*\prime} P[\tilde{Q}_T]\xi_T^* - \xi_T^{*\prime} P[\hat{Q}_{1T}]\xi_T^*] \\
&= \xi_T^{*\prime} P[\tilde{Q}_T]\xi_T^* - \xi_T^{*\prime} P[\hat{Q}_{1T}]\xi_T^* \\
&= \xi_T^{*\prime}\tilde{Q}_T(\tilde{Q}_T'\tilde{Q}_T)^{-1}\tilde{Q}_T'\xi_T^* - \xi_T^{*\prime}\hat{Q}_{1T}(\hat{Q}_{1T}'\hat{Q}_{1T})^{-1}\hat{Q}_{1T}'\xi_T^* \\
&= \xi_T^{*\prime} Q_T[Q_T'\Sigma(\rho_1)^{-1}Q_T]^{-1}Q_T'\Sigma(\rho_1)^{-1}\xi_T^* - \xi_T^{*\prime} Q_{1T}[Q_{1T}'\Sigma(\rho_0)^{-1}Q_{1T}]^{-1}Q_{1T}'\Sigma(\rho_0)^{-1}\xi_T^* \\
&= \xi_T^{*\prime} Q_T D_T[D_T' Q_T'\Sigma(\rho_1)^{-1}Q_T D_T]^{-1} D_T' Q_T'\Sigma(\rho_1)^{-1}\xi_T^* \\
&\quad - \xi_T^{*\prime} Q_{1T} D_{1T}[D_{1T}' Q_{1T}'\Sigma(\rho_0)^{-1}Q_{1T} D_{1T}]^{-1} D_{1T}' Q_{1T}'\Sigma(\rho_0)^{-1}\xi_T^* \\
&= \xi_T^{*\prime}\overline{\Lambda}_1\xi_T^* - \xi_T^{*\prime}\overline{\Lambda}_0\xi_T^* \\
&= \overline{\overline{\Lambda}}_1 - \overline{\overline{\Lambda}}_0,
\end{aligned} \tag{4.9.10}$$

where $Q_T = [Q_{1T} \vdots Q_{2T}]$, $Q_{1T} = X(T)$, $Q_{2T} = Z_{-2}(T)$, and

$$\begin{aligned}
\overline{\Lambda}_1 &:= Q_T D_T [D_T' Q_T' \Sigma(\rho_1)^{-1} Q_T D_T]^{-1} D_T' Q_T' \Sigma(\rho_1)^{-1}, \\
\overline{\Lambda}_0 &:= Q_{1T} D_{1T} [D_{1T}' Q_{1T}' \Sigma(\rho_0)^{-1} Q_{1T} D_{1T}]^{-1} D_{1T}' Q_{1T}' \Sigma(\rho_0)^{-1}, \\
\overline{\overline{\Lambda}}_1 &:= \xi_T^{*\prime} \overline{\Lambda}_1 \xi_T^*, \qquad \overline{\overline{\Lambda}}_0 := \xi_T^{*\prime} \overline{\Lambda}_0 \xi_T^*.
\end{aligned}$$

Under the Assumption 4.5.2, we have

$$\begin{aligned}
\overline{\overline{\Lambda}}_1 &= \xi_T^{*\prime} \overline{\Lambda}_1 \xi_T^* \Longrightarrow \sigma_\xi^2 \chi^2_{(l+k)}, \\
\overline{\overline{\Lambda}}_0 &= \xi_T^{*\prime} \overline{\Lambda}_0 \xi_T^* \Longrightarrow \sigma_\xi^2 \chi^2_{(k)}.
\end{aligned}$$

Further, from the properties of quadratic forms [see Hogg and Craig (1958)], if $\overline{\overline{\Lambda}}_1 - \overline{\overline{\Lambda}}_0 \geq 0$, then

$$\overline{\overline{\Lambda}}_1 - \overline{\overline{\Lambda}}_0 \Longrightarrow \sigma_\xi^2 \chi^2_{(l)}. \tag{4.9.11}$$

Since $\overline{\Lambda}_1$ is a projection onto $[D_{1T}X(T),\ D_{2T}Z_{-2}(T)]$ plane and $\overline{\Lambda}_0$ is a projection onto $D_{1T}X(T)$, $\overline{\Lambda}_1 - \overline{\Lambda}_0$ is a projection onto $D_{2T}Z_{-2}(T)$, *i.e.*, it is a projection onto the orthogonal complement of $D_{1T}X(T)$ within $[D_{1T}X(T),\ D_{2T}Z_{-2}(T)]$. As a result, $\overline{\Lambda}_1 - \overline{\Lambda}_0$ is an idempotent and positive-semidefinite matrix. This implies

$$\overline{\overline{\Lambda}}_1 - \overline{\overline{\Lambda}}_0 = \xi_T^{*\prime} (\overline{\Lambda}_1 - \overline{\Lambda}_0) \xi_T^* \geq 0, \tag{4.9.12}$$

and therefore

$$\overline{\overline{\Lambda}}_1 - \overline{\overline{\Lambda}}_0 \Longrightarrow \sigma_\xi^2 \chi^2_{(l)}. \tag{4.9.13}$$

Hence from (4.9.9) and (4.9.13), we have

$$AR_T^*(\phi_0, \rho_0, \rho_1) \Longrightarrow \chi^2_{(l)}.$$

□

PROOF OF PROPOSITION 4.5.3 Under the null $\phi = \phi_0,\ \rho = \rho_0$,

$$SS_T(\phi_0, \rho_0) = \kappa(T_2)\frac{\Lambda^*_{1T} - \Lambda^*_{2T}}{\Lambda^*_{2T}/T_2},$$

where $\Lambda^*_{1T} = \bar{\xi}^{*(2)'}_T M[Q^*_{1T}]\bar{\xi}^{*(2)}_T$ and $\Lambda^*_{2T} = \bar{\xi}^{*(2)'}_T M[\hat{Q}^*_T]\bar{\xi}^{*(2)}_T$. Under the Assumptions 4.5.3-4.5.4,

$$\kappa(T_2) \underset{T\longrightarrow\infty}{\longrightarrow} \frac{1}{l}, \tag{4.9.14}$$

$$\hat{q}^*_2 \mid q^*_1 \sim \mathrm{N}(\Sigma_{\hat{Q}^*_2 Q^*_1}\Sigma^{-1}_{Q^*_1 Q^*_1} q^*_1, \sigma^2_\xi \Sigma_{\hat{q}^*_2|q^*_1}), \tag{4.9.15}$$

where $\Sigma_{\hat{q}^*_2|q^*_1} = \Sigma_{\hat{Q}^*_2\hat{Q}^*_2} - \Sigma_{\hat{Q}^*_2 Q^*_1}\Sigma^{-1}_{Q^*_1 Q^*_1}\Sigma_{Q^*_1\hat{Q}^*_2}$. Then

$$(\hat{q}^*_2 - \Sigma_{\hat{Q}^*_2 Q^*_1}\Sigma^{-1}_{Q^*_1 Q^*_1} q^*_1)'\Sigma^{-1}_{\hat{q}^*_2|q^*_1}(\hat{q}^*_2 - \Sigma_{\hat{Q}^*_2 Q^*_1}\Sigma^{-1}_{Q^*_1 Q^*_1} q^*_1) \sim \sigma^2_\xi \chi^2_{(l)}. \tag{4.9.16}$$

$$\begin{aligned}
\Lambda^*_{1T} - \Lambda^*_{2T} &= \bar{\xi}^{*(2)'}_T M[Q^*_{1T}]\bar{\xi}^{*(2)}_T - \bar{\xi}^{*(2)'}_T M[\hat{Q}^*_T]\bar{\xi}^{*(2)}_T \\
&= \bar{\xi}^{*(2)'}_T (I - P[Q^*_{1T}])\bar{\xi}^{*(2)}_T - \bar{\xi}^{*(2)'}_T (I - P[\hat{Q}^*_T])\bar{\xi}^{*(2)}_T \\
&= \bar{\xi}^{*(2)'}_T \hat{Q}^*_T(\hat{Q}^{*\prime}_T\hat{Q}^*_T)^{-1}\hat{Q}^{*\prime}_T\bar{\xi}^{*(2)}_T - \bar{\xi}^{*(2)'}_T Q^*_{1T}(Q^{*\prime}_{1T}Q^*_{1T})^{-1}Q^{*\prime}_{1T}\bar{\xi}^{*(2)}_T \\
&= \bar{\xi}^{*(2)'}_T \hat{Q}^*_T D_T(D'_T\hat{Q}^{*\prime}_T\hat{Q}^*_T D_T)^{-1}D'_T\hat{Q}^{*\prime}_T\bar{\xi}^{*(2)}_T - \bar{\xi}^{*(2)'}_T D_{1T}Q^*_{1T}(D'_{1T}Q^{*\prime}_{1T}Q^*_{1T}D_{1T})^{-1}D'_{1T}Q^{*\prime}_{1T}\bar{\xi}^{*(2)}_T \\
&\Longrightarrow \hat{q}^{*\prime}\Sigma^{-1}_{\hat{Q}^*\hat{Q}^*}\hat{q}^* - q^{*\prime}_1\Sigma^{-1}_{Q^*_1Q^*_1}q^*_1. \qquad (4.9.17)
\end{aligned}$$

Now using standard formulas of a partitioned matrix inverse for $\Sigma_{\hat{Q}^*\hat{Q}^*}$ and setting $\tilde{q}^* := \hat{q}^{*\prime}\Sigma^{-1}_{\hat{Q}^*\hat{Q}^*}\hat{q}^* - q^{*\prime}_1\Sigma^{-1}_{Q^*_1Q^*_1}q^*_1$ [see Gentle (2007), Section 3.4.1], we have

$$\begin{aligned}
\tilde{q}^* &= \hat{q}^{*\prime}\Sigma^{-1}_{\hat{Q}^*\hat{Q}^*}\hat{q}^* - q^{*\prime}_1\Sigma^{-1}_{Q^*_1Q^*_1}q^*_1 \\
&= (q^{*\prime}_1, \hat{q}^{*\prime}_2)'\begin{bmatrix} \Sigma^{-1}_{Q^*_1Q^*_1} + \Sigma^{-1}_{Q^*_1Q^*_1}\Sigma_{Q^*_1\hat{Q}^*_2}\Sigma^{-1}_{\hat{q}^*_2|q^*_1}\Sigma_{\hat{Q}^*_2Q^*_1}\Sigma^{-1}_{Q^*_1Q^*_1} & -\Sigma^{-1}_{Q^*_1Q^*_1}\Sigma_{Q^*_1\hat{Q}^*_2}\Sigma^{-1}_{\hat{q}^*_2|q^*_1} \\ -\Sigma^{-1}_{\hat{q}^*_2|q^*_1}\Sigma_{\hat{Q}^*_2Q^*_1}\Sigma^{-1}_{Q^*_1Q^*_1} & \Sigma^{-1}_{\hat{q}^*_2|q^*_1} \end{bmatrix}\begin{pmatrix} q^*_1 \\ \hat{q}^*_2 \end{pmatrix} \\
&\quad - q^{*\prime}_1\Sigma^{-1}_{Q^*_1Q^*_1}q^*_1 \\
&= q^{*\prime}_1\Sigma^{-1}_{Q^*_1Q^*_1}q^*_1 + q^{*\prime}_1\Sigma^{-1}_{Q^*_1Q^*_1}\Sigma_{Q^*_1\hat{Q}^*_2}\Sigma^{-1}_{\hat{q}^*_2|q^*_1}\Sigma_{\hat{Q}^*_2Q^*_1}\Sigma^{-1}_{Q^*_1Q^*_1}q^*_1 - 2\hat{q}^{*\prime}_2\Sigma^{-1}_{Q^*_1Q^*_1}\Sigma_{Q^*_1\hat{Q}^*_2}\Sigma^{-1}_{\hat{q}^*_2|q^*_1}\hat{q}^*_2 + \hat{q}^{*\prime}_2\Sigma^{-1}_{\hat{q}^*_2|q^*_1}\hat{q}^*_2
\end{aligned}$$

$$
\begin{aligned}
&- q_1^{*\prime} \Sigma_{Q_1^* Q_1^*}^{-1} q_1^* \\
&= q_1^{*\prime} \Sigma_{Q_1^* Q_1^*}^{-1} \Sigma_{Q_1^* \hat{Q}_2^*} \Sigma_{\hat{q}_2^* | q_1^*}^{-1} \Sigma_{\hat{Q}_2^* Q_1^*} \Sigma_{Q_1^* Q_1^*}^{-1} q_1^* - 2 \hat{q}_2^{*\prime} \Sigma_{Q_1^* Q_1^*}^{-1} \Sigma_{Q_1^* \hat{Q}_2^*} \Sigma_{\hat{q}_2^* | q_1^*}^{-1} \hat{q}_2^* + \hat{q}_2^{*\prime} \Sigma_{\hat{q}_2^* | q_1^*}^{-1} \hat{q}_2^* \\
&= (\hat{q}_2^* - \Sigma_{\hat{Q}_2^* Q_1^*} \Sigma_{Q_1^* Q_1^*}^{-1} q_1^*)' \Sigma_{\hat{q}_2^* | q_1^*}^{-1} (\hat{q}_2^* - \Sigma_{\hat{Q}_2^* Q_1^*} \Sigma_{Q_1^* Q_1^*}^{-1} q_1^*).
\end{aligned} \tag{4.9.18}
$$

Thus, from (4.9.16), (4.9.17), and (4.9.18), we have

$$
\Lambda_{1T}^* - \Lambda_{2T}^* \Longrightarrow \sigma_\xi^2 \chi_{(l)}^2, \quad \text{and} \quad \frac{\Lambda_{2T}^*}{T_2} \xrightarrow[T_2 \longrightarrow \infty]{p} \sigma_\xi^2, \quad \text{hence} \quad SS_T(\phi_0, \rho_0) \Longrightarrow \frac{\chi_{(l)}^2}{l}.
$$

□

4.9.2 Figures

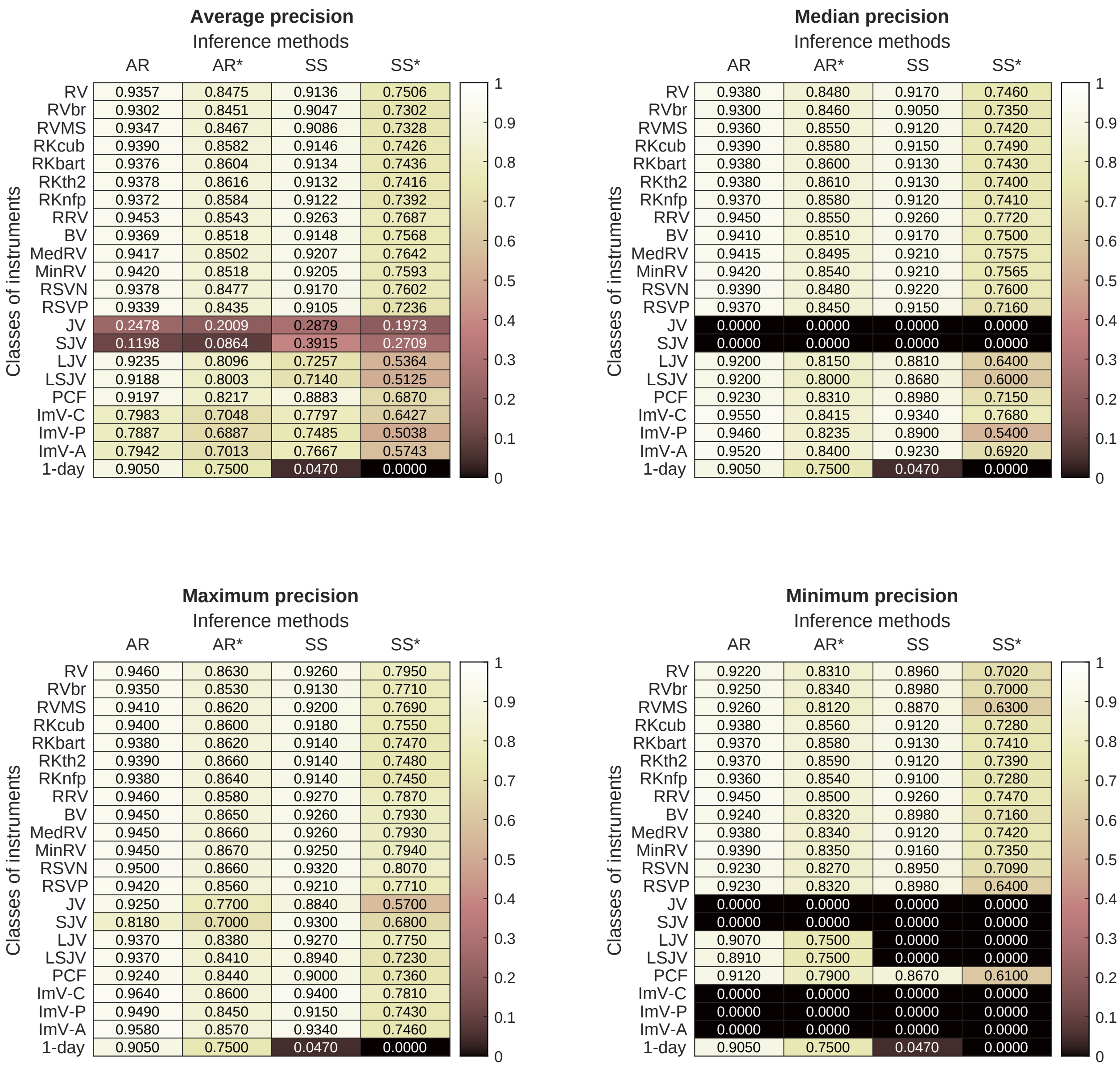

Figure 4.1. IBM: 2009-2013: Precision of different classes of instruments.

Note: The instrument set consists of a constant and a lag of instrument, $l = 1$. We use logarithms of RV-RSVP and PCF classes of instruments given in Table 4.11. The precision of an instrument set i is defined as $d_i = 1 - (ub_i - lb_i)$. For each class, we consider the average, median, minimum, and maximum precision measure across the proposed inference methods [AR, AR^*, SS and SS^*]. These inference procedures are proposed in Sections 4.3.1-4.3.4 and corresponding test statistics are given in equations (4.3.22), (4.3.25), (4.3.31) and (4.3.35). We use $\tau = 0.2$ for SS-type tests and employ 99 Monte Carlo replications for point-optimal type procedures.

Average precision

Inference methods

Subclasses of instruments	AR	AR*	SS	SS*
1s	0.9360	0.8415	0.9129	0.7449
5s	0.9356	0.8361	0.9169	0.7500
30s	0.9399	0.8485	0.9201	0.7685
1m	0.9313	0.8424	0.9053	0.7339
5m	0.9312	0.8404	0.9091	0.7336
10m	0.9336	0.8491	0.9094	0.7429
5m-ss	0.9360	0.8313	0.9062	0.6978
10m-ss	0.9237	0.8148	0.6803	0.5320
1t	0.9314	0.8483	0.9050	0.7346
5t	0.9314	0.8492	0.9068	0.7309
10t	0.9324	0.8537	0.9072	0.7345
20t	0.9336	0.8528	0.9075	0.7234
50t	0.9339	0.8502	0.9105	0.7292

Median precision

Inference methods

Subclasses of instruments	AR	AR*	SS	SS*
1s	0.9380	0.8410	0.9165	0.7455
5s	0.9380	0.8345	0.9210	0.7505
30s	0.9415	0.8470	0.9210	0.7665
1m	0.9310	0.8360	0.9030	0.7340
5m	0.9300	0.8380	0.9050	0.7380
10m	0.9300	0.8480	0.9070	0.7310
5m-ss	0.9370	0.8445	0.9080	0.7240
10m-ss	0.9315	0.8330	0.9010	0.6900
1t	0.9340	0.8540	0.9050	0.7375
5t	0.9335	0.8530	0.9085	0.7375
10t	0.9330	0.8555	0.9100	0.7355
20t	0.9330	0.8525	0.9065	0.7165
50t	0.9340	0.8510	0.9115	0.7315

Maximum precision

Inference methods

Subclasses of instruments	AR	AR*	SS	SS*
1s	0.9420	0.8580	0.9180	0.7510
5s	0.9390	0.8410	0.9220	0.7600
30s	0.9440	0.8580	0.9240	0.7850
1m	0.9420	0.8580	0.9180	0.7510
5m	0.9410	0.8560	0.9220	0.7600
10m	0.9440	0.8560	0.9240	0.7850
5m-ss	0.9450	0.8660	0.9250	0.7930
10m-ss	0.9420	0.8530	0.9190	0.7480
1t	0.9400	0.8660	0.9150	0.7550
5t	0.9410	0.8600	0.9140	0.7460
10t	0.9390	0.8610	0.9150	0.7510
20t	0.9400	0.8620	0.9180	0.7490
50t	0.9400	0.8590	0.9180	0.7530

Minimum precision

Inference methods

Subclasses of instruments	AR	AR*	SS	SS*
1s	0.9250	0.8190	0.9010	0.7350
5s	0.9260	0.8320	0.9030	0.7380
30s	0.9310	0.8450	0.9130	0.7570
1m	0.9220	0.8270	0.8950	0.7140
5m	0.9230	0.8330	0.8980	0.7000
10m	0.9260	0.8440	0.8980	0.7090
5m-ss	0.9250	0.7700	0.8840	0.5500
10m-ss	0.8900	0.7400	0.0000	0.0000
1t	0.9220	0.8270	0.8950	0.7140
5t	0.9230	0.8330	0.8980	0.7000
10t	0.9260	0.8440	0.8980	0.7090
20t	0.9290	0.8420	0.9010	0.7020
50t	0.9290	0.8340	0.9000	0.7020

Figure 4.2. IBM: 2009-2013: Precision of different subclasses of HF instruments.

Note: The instrument set consists of a constant and a lag of instrument, $l = 1$. We use logarithms of RV-RSVP and PCF classes of instruments given in Table 4.11. The precision of an instrument set i is defined as $d_i = 1 - (ub_i - lb_i)$. For each class, we consider the average, median, minimum, and maximum precision measure across the proposed inference methods [AR, AR^*, SS and SS^*]. These inference procedures are proposed in Sections 4.3.1-4.3.4 and corresponding test statistics are given in equations (4.3.22), (4.3.25), (4.3.31) and (4.3.35). We use $\tau = 0.2$ for SS-type tests and employ 99 Monte Carlo replications for point-optimal type procedures.

4.9.3 Tables

Table 4.1. Corresponding values of λ_0

ϕ_0 \ ρ_0	-0.5	-0.4999	-0.45	-0.4	-0.35	-0.3	-0.25	-0.2	-0.15	-0.1	-0.05	0	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.4999	0.5
-1	-	2499.50	4.50	2.00	1.17	0.75	0.50	0.33	0.21	0.13	0.06	0.00	-	-	-	-	-	-	-	-	-	-	-
-0.9999	-	2499.81	4.50	2.00	1.17	0.75	0.50	0.33	0.21	0.13	0.06	0.00	-	-	-	-	-	-	-	-	-	-	-
-0.95	-	-	4.79	2.12	1.23	0.79	0.53	0.35	0.23	0.13	0.06	0.00	-	-	-	-	-	-	-	-	-	-	-
-0.9	-	-	5.26	2.27	1.31	0.84	0.56	0.37	0.24	0.14	0.06	0.00	-	-	-	-	-	-	-	-	-	-	-
-0.8	-	-	7.26	2.78	1.55	0.97	0.64	0.42	0.27	0.16	0.07	0.00	-	-	-	-	-	-	-	-	-	-	-
-0.7	-	-	15.25	3.85	1.96	1.19	0.76	0.50	0.31	0.18	0.08	0.00	-	-	-	-	-	-	-	-	-	-	-
-0.6	-	-	-	7.14	2.82	1.56	0.96	0.61	0.38	0.22	0.09	0.00	-	-	-	-	-	-	-	-	-	-	-
-0.5	-	-	-	-	5.60	2.40	1.33	0.80	0.48	0.27	0.11	0.00	-	-	-	-	-	-	-	-	-	-	-
-0.4	-	-	-	-	-	5.77	2.27	1.19	0.66	0.35	0.15	0.00	-	-	-	-	-	-	-	-	-	-	-
-0.3	-	-	-	-	-	-	9.09	2.44	1.10	0.52	0.20	0.00	-	-	-	-	-	-	-	-	-	-	-
-0.2	-	-	-	-	-	-	-	-	3.41	1.04	0.34	0.00	-	-	-	-	-	-	-	-	-	-	-
-0.1	-	-	-	-	-	-	-	-	-	-	1.01	0.00	-	-	-	-	-	-	-	-	-	-	-
0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
0.1	-	-	-	-	-	-	-	-	-	-	-	0.00	1.01	-	-	-	-	-	-	-	-	-	-
0.2	-	-	-	-	-	-	-	-	-	-	-	0.00	0.34	1.04	3.41	-	-	-	-	-	-	-	-
0.3	-	-	-	-	-	-	-	-	-	-	-	0.00	0.20	0.52	1.10	2.44	9.09	-	-	-	-	-	-
0.4	-	-	-	-	-	-	-	-	-	-	-	0.00	0.15	0.35	0.66	1.19	2.27	5.77	-	-	-	-	-
0.5	-	-	-	-	-	-	-	-	-	-	-	0.00	0.11	0.27	0.48	0.80	1.33	2.40	5.60	∞	-	-	-
0.6	-	-	-	-	-	-	-	-	-	-	-	0.00	0.09	0.22	0.38	0.61	0.96	1.56	2.82	7.14	-	-	-
0.7	-	-	-	-	-	-	-	-	-	-	-	0.00	0.08	0.18	0.31	0.50	0.76	1.19	1.96	3.85	15.25	-	-
0.8	-	-	-	-	-	-	-	-	-	-	-	0.00	0.07	0.16	0.27	0.42	0.64	0.97	1.55	2.78	7.26	-	-
0.9	-	-	-	-	-	-	-	-	-	-	-	0.00	0.06	0.14	0.24	0.37	0.56	0.84	1.31	2.27	5.26	-	-
0.95	-	-	-	-	-	-	-	-	-	-	-	0.00	0.06	0.13	0.23	0.35	0.53	0.79	1.23	2.12	4.79	-	-
0.9999	-	-	-	-	-	-	-	-	-	-	-	0.00	0.06	0.13	0.21	0.33	0.50	0.75	1.17	2.00	4.50	2499.81	-
1	-	-	-	-	-	-	-	-	-	-	-	0.00	0.06	0.13	0.21	0.33	0.50	0.75	1.17	2.00	4.50	2499.50	∞

Table 4.2: Size of asymptotic t-type test for $H_0 : \phi = \phi_0$ (nominal level: 5%)

ϕ	$T = 100$	$T = 200$	$T = 300$	$T = 400$	$T = 500$	$T = 1000$	$T = 2000$
0.0000	0.2	0.1	0.0	0.0	0.0	0.0	0.0
0.1000	0.2	0.1	0.0	0.0	0.0	0.0	0.0
0.2000	0.2	0.1	0.0	0.0	0.1	0.2	0.4
0.3000	0.2	0.1	0.0	0.1	0.2	1.0	1.7
0.4000	0.2	0.3	0.4	0.7	1.0	2.2	2.4
0.5000	0.3	0.9	1.3	1.6	1.9	2.4	2.3
0.6000	0.9	1.7	2.0	2.2	2.4	2.2	2.0
0.7000	1.7	2.3	2.4	2.4	2.3	1.8	1.7
0.8000	2.6	2.4	2.6	2.3	2.3	2.0	1.9
0.9000	4.5	3.4	3.7	3.6	3.2	2.8	2.9
0.9500	7.6	5.5	5.1	4.7	4.4	3.4	3.3
0.9800	13.8	9.1	7.1	6.3	6.0	4.5	3.8
0.9850	15.2	10.3	8.1	7.3	6.9	5.1	4.2
0.9900	18.0	12.5	10.3	9.0	8.6	6.3	5.1
0.9950	20.9	16.4	14.6	12.9	12.4	9.0	7.0
0.9990	24.1	21.5	21.9	22.0	22.7	20.5	17.6
0.9995	24.7	22.2	23.4	24.0	25.5	24.9	23.6
0.9999	25.3	23.0	24.8	25.9	27.9	29.9	32.4
1.0000	30.5	29.2	29.9	30.1	31.4	34.3	37.2

Table 4.3. Size comparison of joint tests ($H_0: \phi = \phi_0, \rho = \rho_0$) under $M1$ with weak instruments, nominal level: 5%

	$T = 100$												
	ρ	0.1											
	ϕ	0.50			0.75			0.90			1.00		
	λ	0.27			0.17			0.14			0.13		
	σ_v	4.30			5.41			5.96			6.28		
	$\bar{\lambda}^2$	0	0.1	10	0	0.1	10	0	0.1	10	0	0.1	10
$l = 2$	$\bar{\pi}_1[1,1]$	0	0.05	0.50	0	0.05	0.50	0	0.05	0.50	0	0.05	0.50
	AR	5.1	5.0	5.4	5.0	5.0	4.6	5.1	5.1	4.5	5.2	5.1	8.3
	AR^*	4.8	4.7	4.7	4.7	4.7	3.9	4.6	4.7	3.8	4.6	4.7	7.0
	SS	1.4	1.4	1.3	1.3	1.4	1.2	1.3	1.4	1.2	1.3	1.2	2.5
	SS^*	4.6	4.7	4.7	4.6	4.6	3.5	4.7	4.4	3.7	4.7	4.7	6.3
$l = 5$	$\bar{\pi}_1[1,1]$	0	0.03	0.32	0	0.03	0.32	0	0.03	0.32	0	0.03	0.32
	AR	5.5	5.4	5.5	5.3	5.3	5.2	5.4	5.5	5.2	4.6	4.6	6.5
	AR^*	5.1	5.0	5.0	5.1	5.0	4.7	5.0	4.9	4.5	4.2	4.3	5.7
	SS	0.1	0.1	0.1	0.1	0.1	0.0	0.1	0.1	0.0	0.1	0.1	0.1
	SS^*	4.6	4.7	4.6	4.5	4.6	4.0	4.7	4.7	4.3	4.8	4.9	5.9
$l = 10$	$\bar{\pi}_1[1,1]$	0	0.02	0.22	0	0.02	0.22	0	0.02	0.22	0	0.02	0.22
	AR	4.8	4.8	4.8	5.0	5.0	4.8	5.1	5.1	4.7	5.0	5.1	6.4
	AR^*	4.4	4.4	4.3	4.5	4.5	4.2	4.5	4.5	4.3	4.4	4.4	5.6
	SS	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	SS^*	4.9	4.9	4.6	4.5	4.5	4.4	4.7	4.6	4.5	5.2	5.2	5.6

	$T = 100$												
	ρ	0.2											
	ϕ	0.50			0.75			0.90			1.00		
	λ	0.80			0.46			0.37			0.33		
	σ_v	2.48			3.29			3.64			3.85		
	$\bar{\lambda}^2$	0	0.1	10	0	0.1	10	0	0.1	10	0	0.1	10
$l = 2$	$\bar{\pi}_1[1,1]$	0	0.05	0.50	0	0.05	0.50	0	0.05	0.50	0	0.05	0.50
	AR	4.9	5.0	8.7	5.1	5.0	5.9	5.1	5.0	4.0	5.0	5.0	6.5
	AR^*	4.1	4.1	5.8	3.9	4.0	3.5	3.8	3.9	2.3	3.8	3.9	3.9
	SS	1.3	1.2	2.8	1.3	1.3	1.7	1.2	1.2	0.9	1.3	1.4	1.7
	SS^*	4.0	4.1	5.7	3.8	4.1	3.4	3.9	3.9	2.2	4.5	4.4	4.0
$l = 5$	$\bar{\pi}_1[1,1]$	0	0.03	0.32	0	0.03	0.32	0	0.03	0.32	0	0.03	0.32
	AR	5.6	5.6	7.1	5.3	5.3	5.8	5.4	5.4	5.2	4.7	4.7	5.6
	AR^*	4.0	4.1	4.8	3.8	3.9	3.7	3.9	3.8	3.1	3.2	3.2	3.2
	SS	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.0	0.1	0.1	0.1
	SS^*	4.1	4.2	5.2	4.1	4.0	3.7	4.2	4.2	3.0	4.0	4.1	4.0
$l = 10$	$\bar{\pi}_1[1,1]$	0	0.02	0.22	0	0.02	0.22	0	0.02	0.22	0	0.02	0.22
	AR	5.0	4.9	5.8	4.9	4.9	5.1	4.8	4.8	4.6	5.0	5.1	5.9
	AR^*	3.0	3.0	3.7	3.2	3.2	3.2	3.1	3.1	2.8	3.2	3.2	3.7
	SS	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	SS^*	4.4	4.3	4.8	4.1	4.0	4.0	4.0	4.0	3.5	4.8	4.8	4.5

	$T = 100$												
	ρ	0.3											
	ϕ	0.50			0.75			0.90			1.00		
	λ	2.40			1.07			0.84			0.75		
	σ_v	1.43			2.15			2.42			2.57		
	$\bar{\lambda}^2$	0	0.1	10	0	0.1	10	0	0.1	10	0	0.1	10
$l = 2$	$\bar{\pi}_1[1,1]$	0	0.05	0.50	0	0.05	0.50	0	0.05	0.50	0	0.05	0.50
	AR	4.8	5.1	12.3	5.0	5.2	9.4	5.1	5.2	4.8	5.3	5.2	5.5
	AR^*	2.6	2.6	5.7	2.7	2.7	3.4	2.5	2.5	1.3	2.6	2.7	1.7
	SS	1.2	1.3	3.8	1.1	1.2	2.8	1.3	1.2	0.9	1.5	1.5	1.3
	SS^*	2.9	3.0	5.4	2.8	3.0	3.3	2.8	2.8	1.2	3.5	3.4	1.8
$l = 5$	$\bar{\pi}_1[1,1]$	0	0.03	0.32	0	0.03	0.32	0	0.03	0.32	0	0.03	0.32
	AR	5.4	5.4	8.6	5.5	5.6	7.4	5.6	5.5	5.4	5.1	5.1	5.0
	AR^*	2.1	2.1	3.4	2.2	2.1	2.3	2.2	2.1	1.5	1.9	1.8	1.4
	SS	0.1	0.1	0.2	0.1	0.1	0.1	0.1	0.1	0.0	0.1	0.1	0.1
	SS^*	3.1	3.0	4.3	2.9	2.9	3.3	3.0	3.2	1.8	3.2	3.2	2.4
$l = 10$	$\bar{\pi}_1[1,1]$	0	0.02	0.22	0	0.02	0.22	0	0.02	0.22	0	0.02	0.22
	AR	4.5	4.5	6.2	4.9	4.9	5.8	4.8	4.9	5.0	5.3	5.3	5.7
	AR^*	1.4	1.4	1.8	1.4	1.4	1.5	1.4	1.4	1.1	1.4	1.4	1.4
	SS	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	SS^*	3.5	3.5	3.8	3.2	3.1	3.2	3.1	3.0	2.1	3.4	3.4	3.0

	$T = 200$												
	ρ	0.1											
	ϕ	0.50			0.75			0.90			1.00		
	λ	0.27			0.17			0.14			0.13		
	σ_v	4.30			5.41			5.96			6.28		
	$\bar{\lambda}^2$	0	0.1	10	0	0.1	10	0	0.1	10	0	0.1	10
$l = 2$	$\bar{\pi}_1[1,1]$	0	0.04	0.35	0	0.04	0.35	0	0.04	0.35	0	0.04	0.35
	AR	5.4	5.4	6.6	5.2	5.2	5.7	5.2	5.2	4.8	5.1	5.1	7.1
	AR^*	5.0	5.0	5.9	5.0	4.9	4.9	4.9	4.8	4.3	4.8	4.8	6.4
	SS	1.7	1.6	2.0	1.7	1.7	1.6	1.7	1.8	1.4	1.4	1.4	1.9
	SS^*	5.0	4.9	5.8	5.1	5.0	4.9	5.2	5.1	4.1	4.8	4.7	6.0
$l = 5$	$\bar{\pi}_1[1,1]$	0	0.02	0.22	0	0.02	0.22	0	0.02	0.22	0	0.02	0.22
	AR	4.9	4.9	5.8	5.0	4.9	5.1	5.0	5.0	4.8	4.9	5.0	6.2
	AR^*	4.6	4.5	5.2	4.6	4.6	4.6	4.6	4.6	4.3	4.6	4.7	5.7
	SS	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.0	0.1	0.1	0.1
	SS^*	5.1	4.9	5.3	4.8	4.9	4.4	5.0	4.7	4.0	4.7	4.6	5.7
$l = 10$	$\bar{\pi}_1[1,1]$	0	0.02	0.16	0	0.02	0.16	0	0.02	0.16	0	0.02	0.16
	AR	5.1	5.0	5.4	5.1	5.1	5.1	5.1	5.1	4.8	5.2	5.2	5.9
	AR^*	4.5	4.6	4.8	4.3	4.4	4.5	4.5	4.5	4.4	4.9	4.9	5.2
	SS	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	SS^*	4.6	4.7	4.8	4.7	4.8	4.4	4.9	4.9	4.1	4.4	4.3	4.9

	$T = 200$												
	ρ	0.2											
	ϕ	0.50			0.75			0.90			1.00		
	λ	0.80			0.46			0.37			0.33		
	σ_v	2.48			3.29			3.64			3.85		
	$\bar{\lambda}^2$	0	0.1	10	0	0.1	10	0	0.1	10	0	0.1	10
$l = 2$	$\bar{\pi}_1[1,1]$	0	0.04	0.35	0	0.04	0.35	0	0.04	0.35	0	0.04	0.35
	AR	5.5	5.4	11.3	5.4	5.4	8.7	5.4	5.3	5.6	5.3	5.3	6.3
	AR^*	4.0	4.1	8.3	4.0	4.0	6.1	3.9	3.9	3.6	4.1	4.1	4.1
	SS	1.8	1.7	4.0	1.6	1.6	3.0	1.5	1.7	1.4	1.5	1.5	1.7
	SS^*	4.5	4.2	7.8	4.5	4.5	5.8	4.4	4.3	3.6	4.2	4.3	4.1
$l = 5$	$\bar{\pi}_1[1,1]$	0	0.02	0.22	0	0.02	0.22	0	0.02	0.22	0	0.02	0.22
	AR	4.9	5.0	8.4	5.1	5.1	6.9	5.0	5.1	5.3	4.8	4.8	5.4
	AR^*	3.4	3.4	5.7	3.6	3.5	4.3	3.5	3.6	3.4	3.3	3.4	3.6
	SS	0.1	0.1	0.2	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
	SS^*	4.1	4.1	6.2	4.4	4.2	5.0	4.3	4.0	3.3	4.1	4.2	4.5
$l = 10$	$\bar{\pi}_1[1,1]$	0	0.02	0.16	0	0.02	0.16	0	0.02	0.16	0	0.02	0.16
	AR	5.1	5.1	6.8	5.0	5.0	6.0	5.0	5.0	5.1	5.3	5.2	5.7
	AR^*	3.0	3.0	4.0	3.2	3.2	3.7	3.3	3.3	3.1	3.5	3.5	3.5
	SS	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	SS^*	4.0	3.9	5.3	4.0	4.0	4.7	4.5	4.4	3.7	3.9	3.9	4.1

	$T = 200$												
	ρ	0.3											
	ϕ	0.50			0.75			0.90			1.00		
	λ	2.40			1.07			0.84			0.75		
	σ_v	1.43			2.15			2.42			2.57		
	$\bar{\lambda}^2$	0	0.1	10	0	0.1	10	0	0.1	10	0	0.1	10
$l = 2$	$\bar{\pi}_1[1,1]$	0	0.04	0.35	0	0.04	0.35	0	0.04	0.35	0	0.04	0.35
	AR	5.3	5.6	16.0	5.2	5.3	14.6	5.3	5.3	7.6	5.4	5.5	5.7
	AR^*	2.7	2.6	8.5	2.5	2.6	7.0	2.6	2.6	2.7	2.9	2.9	2.1
	SS	1.6	1.6	5.5	1.6	1.6	5.1	1.5	1.5	2.1	1.6	1.7	1.5
	SS^*	3.2	3.0	7.9	3.2	3.1	6.8	3.1	3.1	2.8	3.3	3.3	2.3
$l = 5$	$\bar{\pi}_1[1,1]$	0	0.02	0.22	0	0.02	0.22	0	0.02	0.22	0	0.02	0.22
	AR	5.0	5.1	10.5	5.1	5.1	9.7	5.1	5.2	6.4	5.0	5.1	5.2
	AR^*	1.8	1.9	4.3	1.8	1.9	3.7	2.1	2.0	2.2	1.9	1.8	1.7
	SS	0.1	0.1	0.3	0.1	0.1	0.3	0.1	0.1	0.1	0.2	0.2	0.1
	SS^*	3.2	3.2	5.5	3.2	3.3	4.9	3.3	3.1	2.7	2.9	3.1	2.9
$l = 10$	$\bar{\pi}_1[1,1]$	0	0.02	0.16	0	0.02	0.16	0	0.02	0.16	0	0.02	0.16
	AR	5.0	5.1	8.1	5.2	5.1	7.6	5.2	5.1	5.7	5.3	5.4	5.6
	AR^*	1.4	1.4	2.3	1.4	1.4	1.9	1.5	1.5	1.4	1.5	1.5	1.4
	SS	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	SS^*	2.8	2.8	4.3	3.1	3.0	4.3	3.1	3.0	3.0	2.8	3.0	2.7

Table 4.4. Power comparison of joint tests ($H_0: \phi = 0.50, \rho = 0.30$) under $M1$ with weak instruments, nominal level: 5%

		$T = 100$																																			
	ρ	0.1												0.2												0.3											
	ϕ	0.50			0.75			0.90			1.00			0.50			0.75			0.90			1.00			0.50			0.75			0.90			1.00		
	λ	0.27			0.17			0.14			0.13			0.80			0.46			0.37			0.33			2.40			1.07			0.84			0.75		
	σ_v	4.30			5.41			5.96			6.28			2.48			3.29			3.64			3.85			1.43			2.15			2.42			2.57		
	$\bar{\lambda}^2$	0	0.1	10	0	0.1	10	0	0.1	10	0	0.1	10	0	0.1	10	0	0.1	10	0	0.1	10	0	0.1	10	0	0.1	10	0	0.1	10	0	0.1	10	0	0.1	10
$l = 2$	$\bar{\pi}_1[1,1]$	0	0.05	0.50	0	0.05	0.50	0	0.05	0.50	0	0.05	0.50	0	0.05	0.50	0	0.05	0.50	0	0.05	0.50	0	0.05	0.50	0	0.05	0.50	0	0.05	0.50	0	0.05	0.50	0	0.05	0.50
	AR	7.6	8.0	29.9	12.6	14.7	91.4	16.4	20.8	99.7	20.2	26.8	100.0	6.4	6.6	21.8	10.9	12.8	86.9	15.5	19.4	99.4	19.7	26.5	100.0	4.8	5.1	12.3	8.9	10.1	76.5	13.3	17.1	98.6	19.1	25.1	100.0
	AR^*	0.8	0.8	1.5	63.7	64.0	75.3	98.4	98.5	99.3	100.0	100.0	100.0	0.3	0.3	1.1	25.7	26.3	64.9	89.6	89.5	98.4	99.5	99.5	100.0	2.6	2.6	5.7	5.1	5.8	62.1	8.1	10.4	96.4	12.5	17.2	99.8
	SS	2.1	2.3	13.6	3.9	4.6	69.8	5.5	6.8	90.7	7.2	9.3	89.7	1.6	1.6	8.6	3.1	3.9	61.2	5.2	6.0	88.0	6.7	8.7	88.6	1.2	1.3	3.8	2.4	2.8	46.1	4.4	5.3	81.9	6.5	8.5	86.2
	SS^*	0.8	0.8	1.7	57.4	57.5	67.2	96.5	96.4	97.7	99.7	99.7	99.8	0.4	0.4	1.4	24.5	25.1	54.8	84.9	85.1	94.4	98.5	98.5	99.3	2.9	3.0	5.4	4.6	5.3	48.0	7.4	8.4	83.0	9.9	11.8	87.3
$l = 5$	$\bar{\pi}_1[1,1]$	0	0.03	0.32	0	0.03	0.32	0	0.03	0.32	0	0.03	0.32	0	0.03	0.32	0	0.03	0.32	0	0.03	0.32	0	0.03	0.32	0	0.03	0.32	0	0.03	0.32	0	0.03	0.32	0	0.03	0.32
	AR	9.4	9.6	21.3	17.4	18.5	81.6	24.5	27.0	98.7	29.5	34.1	99.9	7.4	7.4	15.0	14.6	15.9	74.7	22.2	25.0	97.8	28.7	32.8	99.9	5.4	5.4	8.6	11.6	12.4	61.0	19.0	21.3	95.8	27.8	31.2	99.8
	AR^*	1.3	1.3	1.8	56.4	56.7	70.4	97.3	97.5	98.9	99.9	99.9	100.0	0.6	0.7	1.0	19.2	19.8	53.2	78.5	79.1	96.5	98.0	98.0	99.8	2.1	2.1	3.4	5.3	5.8	43.1	10.5	11.8	90.4	16.4	20.0	99.4
	SS	0.1	0.1	1.3	0.4	0.5	21.6	0.9	1.0	51.2	1.2	1.5	54.0	0.1	0.1	0.6	0.4	0.4	15.5	0.7	0.9	44.8	1.2	1.5	51.0	0.1	0.1	0.2	0.3	0.3	8.4	0.6	0.7	34.8	0.9	1.1	47.0
	SS^*	0.9	0.9	1.4	57.5	57.5	64.0	96.5	96.5	97.3	99.6	99.6	99.7	0.5	0.5	1.0	24.9	25.2	44.1	85.5	85.6	92.5	98.1	98.1	99.0	3.1	3.0	4.3	4.9	5.1	30.0	7.3	7.7	61.0	9.3	10.0	67.1
$l = 10$	$\bar{\pi}_1[1,1]$	0	0.02	0.22	0	0.02	0.22	0	0.02	0.22	0	0.02	0.22	0	0.02	0.22	0	0.02	0.22	0	0.02	0.22	0	0.02	0.22	0	0.02	0.22	0	0.02	0.22	0	0.02	0.22	0	0.02	0.22
	AR	10.1	10.3	17.8	23.1	24.0	74.5	35.3	37.1	97.5	43.2	46.6	99.8	7.1	7.2	12.0	18.9	19.6	65.1	31.6	33.5	96.1	41.9	45.4	99.8	4.5	4.5	6.2	13.9	14.2	48.6	26.3	28.0	91.6	39.6	42.9	99.6
	AR^*	1.4	1.3	1.6	46.3	46.6	62.1	94.6	94.9	97.9	99.7	99.7	99.9	0.7	0.7	0.9	15.6	15.9	41.3	64.5	65.4	93.4	94.1	94.4	99.6	1.4	1.4	1.8	5.2	5.4	28.1	12.7	14.0	81.5	22.6	24.7	98.3
	SS	0.0	0.0	0.0	0.0	0.0	1.7	0.0	0.0	10.3	0.1	0.1	14.0	0.0	0.0	0.0	0.0	0.0	1.0	0.1	0.1	7.6	0.1	0.1	12.8	0.0	0.0	0.0	0.0	0.0	0.4	0.0	0.0	4.9	0.0	0.1	10.6
	SS^*	0.8	0.8	1.3	57.3	57.4	61.0	96.3	96.3	96.7	99.6	99.6	99.7	0.4	0.4	0.7	25.2	25.2	36.1	84.9	84.9	89.1	98.1	98.1	98.7	3.5	3.5	3.8	5.2	5.0	18.4	7.6	8.0	40.0	9.3	10.0	47.6
		$T = 200$																																			
	ρ	0.1												0.2												0.3											
	ϕ	0.50			0.75			0.90			1.00			0.50			0.75			0.90			1.00			0.50			0.75			0.90			1.00		
	λ	0.27			0.17			0.14			0.13			0.80			0.46			0.37			0.33			2.40			1.07			0.84			0.75		
	σ_v	4.30			5.41			5.96			6.28			2.48			3.29			3.64			3.85			1.43			2.15			2.42			2.57		
	$\bar{\lambda}^2$	0	0.1	10	0	0.1	10	0	0.1	10	0	0.1	10	0	0.1	10	0	0.1	10	0	0.1	10	0	0.1	10	0	0.1	10	0	0.1	10	0	0.1	10	0	0.1	10
$l = 2$	$\bar{\pi}_1[1,1]$	0	0.04	0.35	0	0.04	0.35	0	0.04	0.35	0	0.04	0.35	0	0.04	0.35	0	0.04	0.35	0	0.04	0.35	0	0.04	0.35	0	0.04	0.35	0	0.04	0.35	0	0.04	0.35	0	0.04	0.35
	AR	7.9	8.3	38.2	12.8	15.2	97.0	17.3	22.1	100.0	20.7	27.5	100.0	6.6	7.0	28.3	11.2	13.3	94.5	16.3	20.6	100.0	20.8	27.0	100.0	5.3	5.6	16.0	9.5	10.7	88.1	14.4	18.3	99.9	20.2	26.5	100.0
	AR^*	0.2	0.2	0.6	90.8	90.7	94.2	100.0	100.0	100.0	100.0	100.0	100.0	0.0	0.0	0.2	52.3	52.8	82.6	99.7	99.7	100.0	100.0	100.0	100.0	2.7	2.6	8.5	5.1	6.3	78.6	8.7	11.9	99.6	13.2	18.8	100.0
	SS	2.6	2.5	19.0	4.3	4.9	82.7	6.2	7.9	95.9	7.3	9.7	89.9	2.2	2.1	12.3	3.7	4.1	75.6	5.6	7.4	94.6	7.3	9.5	89.7	1.6	1.6	5.5	3.0	3.1	62.5	5.0	6.4	92.1	7.4	9.3	88.4
	SS^*	0.3	0.3	0.7	85.1	85.2	89.0	99.9	99.9	100.0	100.0	100.0	100.0	0.0	0.0	0.3	47.9	48.4	73.6	99.3	99.3	99.8	100.0	100.0	100.0	3.2	3.0	7.9	5.3	5.5	66.5	7.9	9.5	92.8	10.8	13.0	89.6
$l = 5$	$\bar{\pi}_1[1,1]$	0	0.02	0.22	0	0.02	0.22	0	0.02	0.22	0	0.02	0.22	0	0.02	0.22	0	0.02	0.22	0	0.02	0.22	0	0.02	0.22	0	0.02	0.22	0	0.02	0.22	0	0.02	0.22	0	0.02	0.22
	AR	9.1	9.4	28.4	17.9	19.2	92.2	26.1	29.0	99.9	32.7	37.5	100.0	7.2	7.3	19.6	15.5	16.6	87.6	24.0	27.2	99.7	32.2	37.4	100.0	5.0	5.1	10.5	12.0	12.8	76.3	20.9	23.6	99.4	31.4	36.4	100.0
	AR^*	0.3	0.3	0.7	87.9	87.9	92.2	100.0	100.0	100.0	100.0	100.0	100.0	0.1	0.1	0.3	40.4	41.0	74.7	99.4	99.5	99.9	100.0	100.0	100.0	1.8	1.9	4.3	5.6	6.0	60.6	11.1	13.2	98.0	19.2	22.9	100.0
	SS	0.2	0.2	2.3	0.5	0.6	38.8	0.9	1.2	71.9	1.3	1.6	63.5	0.2	0.2	1.2	0.5	0.5	29.7	0.8	0.9	67.5	1.2	1.6	61.8	0.1	0.1	0.3	0.3	0.3	16.9	0.7	0.8	58.2	1.3	1.6	59.1
	SS^*	0.3	0.3	0.6	85.3	85.2	87.9	100.0	100.0	100.0	100.0	100.0	100.0	0.1	0.1	0.2	47.5	47.6	66.8	99.3	99.4	99.7	100.0	100.0	100.0	3.2	3.2	5.5	5.4	5.6	46.7	7.7	8.2	79.1	10.1	11.2	74.7
$l = 10$	$\bar{\pi}_1[1,1]$	0	0.02	0.16	0	0.02	0.16	0	0.02	0.16	0	0.02	0.16	0	0.02	0.16	0	0.02	0.16	0	0.02	0.16	0	0.02	0.16	0	0.02	0.16	0	0.02	0.16	0	0.02	0.16	0	0.02	0.16
	AR	11.1	11.1	24.0	24.8	26.0	88.0	38.2	40.9	99.8	48.3	51.8	100.0	8.0	7.8	15.8	20.6	21.8	81.7	34.7	37.3	99.5	47.7	51.3	100.0	5.0	5.1	8.1	15.1	15.9	67.7	29.3	31.4	98.7	46.4	50.1	100.0
	AR^*	0.4	0.4	0.6	84.4	84.5	89.9	100.0	100.0	100.0	100.0	100.0	100.0	0.1	0.1	0.3	33.5	33.9	66.4	98.1	98.3	99.8	100.0	100.0	100.0	1.4	1.4	2.3	5.5	5.8	45.4	14.0	15.5	95.5	26.8	30.2	99.9
	SS	0.0	0.0	0.1	0.0	0.0	6.3	0.1	0.1	28.1	0.2	0.2	26.5	0.0	0.0	0.0	0.0	0.0	3.8	0.1	0.1	23.2	0.2	0.2	25.0	0.0	0.0	0.0	0.0	0.0	1.4	0.1	0.1	15.7	0.1	0.2	23.1
	SS^*	0.3	0.3	0.4	85.5	85.6	87.3	100.0	100.0	100.0	100.0	100.0	100.0	0.1	0.1	0.1	47.6	47.6	61.1	99.4	99.4	99.6	100.0	100.0	100.0	2.8	2.8	4.3	5.4	5.3	31.2	8.1	8.5	62.5	10.5	11.2	59.5

Table 4.5. Size comparison of joint tests ($H_0 : \phi = \phi_0, \rho = \rho_0$) under $M2$ with weak instruments, nominal level: 5%

		$T = 100$																																			
	ρ	0.1												0.2												0.3											
	ϕ	0.50			0.75			0.90			1.00			0.50			0.75			0.90			1.00			0.50			0.75			0.90			1.00		
	λ	0.27			0.17			0.14			0.13			0.80			0.46			0.37			0.33			2.40			1.07			0.84			0.75		
	σ_v	4.30			5.41			5.96			6.28			2.48			3.29			3.64			3.85			1.43			2.15			2.42			2.57		
	$\bar{\lambda}^2$	0	0.1	10	0	0.1	10	0	0.1	10	0	0.1	10	0	0.1	10	0	0.1	10	0	0.1	10	0	0.1	10	0	0.1	10	0	0.1	10	0	0.1	10	0	0.1	10
	$\bar{\pi}_1[1,1]$	0	0.05	0.50	0	0.05	0.50	0	0.05	0.50	0	0.05	0.50	0	0.05	0.50	0	0.05	0.50	0	0.05	0.50	0	0.05	0.50	0	0.05	0.50	0	0.05	0.50	0	0.05	0.50	0	0.05	0.50
$l = 2$	AR	5.3	5.4	5.5	5.4	5.4	4.9	5.3	5.4	4.8	5.4	5.3	8.3	5.1	5.2	8.8	5.3	5.3	6.5	5.3	5.3	4.2	5.4	5.3	6.5	5.4	5.7	12.7	5.3	5.3	9.9	5.2	5.2	4.8	5.7	5.7	5.5
	AR^*	4.8	4.8	4.9	4.8	4.8	4.2	4.8	4.9	4.1	5.0	4.7	7.0	4.0	4.1	6.2	4.0	4.1	4.1	4.0	4.0	2.6	4.3	4.4	4.0	2.7	2.8	6.1	2.8	2.7	3.7	2.7	2.6	1.5	2.9	2.9	1.8
	SS	1.4	1.4	1.5	1.4	1.4	1.2	1.4	1.3	1.2	1.4	1.4	2.3	1.4	1.4	2.6	1.4	1.4	1.6	1.5	1.4	1.0	1.4	1.3	1.5	1.5	1.5	4.1	1.5	1.5	2.9	1.6	1.6	0.9	1.3	1.3	1.2
	SS^*	5.1	4.8	4.6	4.7	4.8	3.8	4.7	4.6	3.7	5.1	4.9	6.7	4.2	4.2	5.7	4.2	4.3	3.3	4.1	4.1	2.2	4.4	4.5	3.9	3.1	3.2	5.8	3.0	3.3	3.4	3.2	3.2	1.3	3.2	3.2	1.9
	$\bar{\pi}_1[1,1]$	0	0.03	0.32	0	0.03	0.32	0	0.03	0.32	0	0.03	0.32	0	0.03	0.32	0	0.03	0.32	0	0.03	0.32	0	0.03	0.32	0	0.03	0.32	0	0.03	0.32	0	0.03	0.32	0	0.03	0.32
$l = 5$	AR	5.1	5.1	5.2	5.1	5.1	4.6	5.2	5.1	4.5	4.9	5.0	6.5	5.1	5.2	6.6	5.1	5.1	5.4	5.0	5.0	4.4	5.0	5.0	5.6	5.0	5.1	8.8	5.1	5.2	7.1	5.2	5.3	5.0	5.1	5.1	5.2
	AR^*	4.7	4.7	4.6	4.6	4.6	4.2	4.5	4.7	4.2	4.4	4.4	5.4	3.8	3.7	4.5	3.5	3.6	3.4	3.4	3.5	2.7	3.4	3.4	3.2	2.2	2.3	3.3	2.0	2.0	2.4	2.0	2.1	1.4	2.0	2.1	1.6
	SS	0.0	0.0	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.0	0.1	0.1	0.0	0.1	0.2	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.2	0.1	0.1	0.2	0.1	0.1	0.1	0.1	0.1	0.1
	SS^*	5.2	5.0	4.9	4.9	4.7	4.3	5.0	5.0	4.2	4.8	4.8	5.8	4.6	4.6	5.1	4.4	4.2	3.7	4.6	4.5	3.1	4.2	4.2	4.0	3.1	3.2	4.3	3.3	3.3	3.3	3.3	3.3	1.9	3.2	3.1	2.5
	$\bar{\pi}_1[1,1]$	0	0.02	0.22	0	0.02	0.22	0	0.02	0.22	0	0.02	0.22	0	0.02	0.22	0	0.02	0.22	0	0.02	0.22	0	0.02	0.22	0	0.02	0.22	0	0.02	0.22	0	0.02	0.22	0	0.02	0.22
$l = 10$	AR	5.1	5.0	5.1	5.0	5.0	4.9	4.9	4.9	4.9	5.0	5.0	6.2	5.2	5.1	5.9	4.7	4.8	5.1	4.8	4.8	4.5	5.2	5.1	5.9	5.2	5.2	6.7	5.0	4.9	5.7	4.7	4.7	4.6	5.4	5.4	5.9
	AR^*	4.6	4.5	4.6	4.7	4.7	4.3	4.6	4.7	4.3	4.3	4.4	5.2	3.3	3.3	3.6	3.3	3.3	3.1	3.3	3.3	2.8	3.3	3.3	3.4	1.6	1.6	2.0	1.6	1.6	1.6	1.6	1.5	1.4	1.8	1.7	1.4
	SS	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	SS^*	5.0	5.0	5.1	4.9	4.9	4.7	4.8	4.7	4.5	4.8	4.9	5.8	4.6	4.5	5.1	4.2	4.3	4.0	4.3	4.5	3.4	4.3	4.2	4.5	3.2	3.3	3.9	3.3	3.4	3.4	3.2	3.3	2.5	3.3	3.3	3.0

		$T = 200$																																			
	ρ	0.1												0.2												0.3											
	ϕ	0.50			0.75			0.90			1.00			0.50			0.75			0.90			1.00			0.50			0.75			0.90			1.00		
	λ	0.27			0.17			0.14			0.13			0.80			0.46			0.37			0.33			2.40			1.07			0.84			0.75		
	σ_v	4.30			5.41			5.96			6.28			2.48			3.29			3.64			3.85			1.43			2.15			2.42			2.57		
	$\bar{\lambda}^2$	0	0.1	10	0	0.1	10	0	0.1	10	0	0.1	10	0	0.1	10	0	0.1	10	0	0.1	10	0	0.1	10	0	0.1	10	0	0.1	10	0	0.1	10	0	0.1	10
	$\bar{\pi}_1[1,1]$	0	0.04	0.35	0	0.04	0.35	0	0.04	0.35	0	0.04	0.35	0	0.04	0.35	0	0.04	0.35	0	0.04	0.35	0	0.04	0.35	0	0.04	0.35	0	0.04	0.35	0	0.04	0.35	0	0.04	0.35
$l = 2$	AR	4.9	4.8	6.2	4.7	4.8	5.3	4.8	4.8	4.6	4.7	4.7	6.8	5.1	4.9	11.2	5.0	5.1	8.5	5.0	5.0	5.4	4.7	4.8	6.0	5.1	5.2	16.1	5.0	5.0	14.4	5.1	5.1	7.5	4.8	4.8	5.4
	AR^*	4.7	4.7	5.5	4.6	4.5	4.6	4.6	4.6	4.0	4.7	4.7	6.0	3.8	4.0	8.2	3.9	4.1	5.6	3.9	4.0	3.6	3.7	3.8	3.9	2.5	2.5	8.8	2.4	2.6	6.8	2.5	2.5	2.8	2.4	2.3	1.9
	SS	1.4	1.5	2.0	1.5	1.5	1.6	1.6	1.6	1.2	1.3	1.4	2.1	1.3	1.4	4.1	1.7	1.5	2.9	1.7	1.6	1.7	1.3	1.3	1.8	1.3	1.5	5.9	1.6	1.6	5.4	1.8	1.7	2.2	1.4	1.4	1.4
	SS^*	4.8	4.8	5.8	4.7	4.9	4.7	4.7	4.6	4.1	4.7	4.6	5.9	3.9	4.3	8.5	4.1	4.2	5.8	3.8	4.0	3.4	4.0	4.0	3.9	2.9	3.1	8.4	3.2	3.1	7.1	3.3	3.4	2.8	2.8	2.7	2.2
	$\bar{\pi}_1[1,1]$	0	0.02	0.22	0	0.02	0.22	0	0.02	0.22	0	0.02	0.22	0	0.02	0.22	0	0.02	0.22	0	0.02	0.22	0	0.02	0.22	0	0.02	0.22	0	0.02	0.22	0	0.02	0.22	0	0.02	0.22
$l = 5$	AR	4.8	4.8	5.6	4.7	4.8	5.2	4.7	4.7	4.7	5.1	5.1	6.3	5.0	5.0	7.7	4.9	4.9	6.3	4.9	4.9	5.0	5.0	5.0	5.5	5.0	5.1	10.7	4.9	4.9	9.5	4.9	4.9	6.0	5.2	5.2	5.3
	AR^*	4.4	4.4	4.9	4.4	4.4	4.5	4.3	4.4	4.3	4.5	4.5	5.7	3.5	3.5	5.5	3.4	3.4	4.3	3.3	3.4	3.1	3.6	3.5	3.5	1.9	1.8	4.1	1.8	1.8	3.5	1.9	1.9	2.0	1.9	1.9	1.6
	SS	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.3	0.1	0.1	0.2	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.4	0.0	0.1	0.4	0.1	0.1	0.1	0.1	0.1	0.0
	SS^*	4.8	4.8	5.4	4.7	4.6	4.5	4.6	4.4	4.1	4.8	4.8	5.9	4.3	4.3	6.4	4.1	4.0	5.2	4.0	4.1	3.5	4.1	4.2	4.3	3.0	3.1	5.7	3.0	3.0	5.3	2.9	2.8	2.8	2.9	2.9	2.5
	$\bar{\pi}_1[1,1]$	0	0.02	0.16	0	0.02	0.16	0	0.02	0.16	0	0.02	0.16	0	0.02	0.16	0	0.02	0.16	0	0.02	0.16	0	0.02	0.16	0	0.02	0.16	0	0.02	0.16	0	0.02	0.16	0	0.02	0.16
$l = 10$	AR	5.2	5.2	5.6	5.3	5.2	5.2	5.2	5.2	5.1	5.0	5.0	5.9	5.2	5.3	7.2	5.5	5.5	6.5	5.3	5.3	5.4	5.1	5.2	5.6	4.9	4.9	8.2	5.4	5.4	8.2	5.5	5.6	6.1	5.6	5.7	5.8
	AR^*	4.9	4.9	5.1	4.9	4.9	4.8	4.7	4.7	4.5	4.6	4.6	5.3	3.4	3.4	4.4	3.4	3.4	3.7	3.4	3.3	3.1	3.1	3.3	3.2	1.5	1.5	2.3	1.6	1.5	2.1	1.5	1.5	1.5	1.4	1.4	1.3
	SS	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	SS^*	4.8	4.9	5.1	4.8	4.8	4.8	5.0	4.9	4.4	4.9	4.8	5.0	4.2	4.2	5.5	4.4	4.2	4.6	4.5	4.4	3.8	4.3	4.4	4.2	3.0	2.9	4.2	3.0	3.0	4.4	3.3	3.3	2.8	3.2	3.3	2.9

Table 4.6. Power comparison of joint tests ($H_0: \phi = 0.50, \rho = 0.30$) under $M2$ with weak instruments, nominal level: 5%

		$T=100$											
	ρ	0.1											
	ϕ	0.50			0.75			0.90			1.00		
	λ	0.27			0.17			0.14			0.13		
	σ_v	4.30			5.41			5.96			6.28		
	$\bar{\lambda}^2$	0	0.1	10	0	0.1	10	0	0.1	10	0	0.1	10
	$\bar{\pi}_1[1,1]$	0	0.05	0.50	0	0.05	0.50	0	0.05	0.50	0	0.05	0.50
$l=2$	AR	7.8	8.3	29.4	12.7	14.6	91.1	16.5	20.7	99.7	20.2	27.0	100.0
	AR^*	1.0	1.0	1.9	64.4	64.3	75.7	98.4	98.4	99.3	99.9	99.9	100.0
	SS	2.4	2.3	12.6	4.0	4.5	69.5	5.7	6.7	90.7	7.1	9.3	89.4
	SS^*	1.0	0.9	1.7	57.6	57.7	67.4	96.5	96.4	97.9	99.7	99.7	99.8
	$\bar{\pi}_1[1,1]$	0	0.03	0.32	0	0.03	0.32	0	0.03	0.32	0	0.03	0.32
$l=5$	AR	9.1	9.4	20.7	17.3	18.4	81.8	24.4	27.0	98.8	29.7	34.2	99.9
	AR^*	1.3	1.3	1.7	56.4	56.6	69.8	97.2	97.3	98.8	99.9	99.9	100.0
	SS	0.1	0.1	1.5	0.4	0.5	21.8	0.9	1.0	51.0	1.2	1.6	54.0
	SS^*	1.0	1.0	1.4	56.9	57.2	64.1	96.2	96.2	97.1	99.6	99.7	99.7
	$\bar{\pi}_1[1,1]$	0	0.02	0.22	0	0.02	0.22	0	0.02	0.22	0	0.02	0.22
$l=10$	AR	10.8	10.8	17.9	23.4	24.5	74.8	35.5	37.5	97.4	43.9	46.6	99.8
	AR^*	1.5	1.5	1.8	46.7	47.0	62.2	94.6	94.6	97.7	99.7	99.7	99.9
	SS	0.0	0.0	0.0	0.0	0.0	1.9	0.0	0.0	10.1	0.1	0.2	13.7
	SS^*	0.9	0.9	1.2	58.0	57.6	61.6	96.2	96.2	96.6	99.7	99.7	99.7

		$T=100$											
	ρ	0.2											
	ϕ	0.50			0.75			0.90			1.00		
	λ	0.80			0.46			0.37			0.33		
	σ_v	2.48			3.29			3.64			3.85		
	$\bar{\lambda}^2$	0	0.1	10	0	0.1	10	0	0.1	10	0	0.1	10
	$\bar{\pi}_1[1,1]$	0	0.05	0.50	0	0.05	0.50	0	0.05	0.50	0	0.05	0.50
$l=2$	AR	6.9	7.1	21.3	11.4	13.0	86.8	15.4	19.3	99.5	19.9	26.4	100.0
	AR^*	0.5	0.5	1.6	26.3	27.1	64.7	89.1	89.7	98.4	99.4	99.4	99.9
	SS	2.0	2.0	8.5	3.3	3.9	60.6	5.2	6.1	88.1	7.0	8.9	88.0
	SS^*	0.5	0.5	1.6	25.6	25.8	55.0	85.0	85.2	94.8	98.2	98.3	99.2
	$\bar{\pi}_1[1,1]$	0	0.03	0.32	0	0.03	0.32	0	0.03	0.32	0	0.03	0.32
$l=5$	AR	7.2	7.2	14.5	14.5	15.6	74.4	22.3	24.4	98.0	28.5	33.0	99.9
	AR^*	0.7	0.7	1.2	19.5	20.1	52.8	77.7	78.2	96.4	97.9	97.9	99.9
	SS	0.1	0.1	0.6	0.4	0.4	15.6	0.8	0.8	45.1	1.2	1.5	51.3
	SS^*	0.4	0.5	1.1	25.7	25.7	44.7	85.6	85.6	91.8	98.3	98.4	98.9
	$\bar{\pi}_1[1,1]$	0	0.02	0.22	0	0.02	0.22	0	0.02	0.22	0	0.02	0.22
$l=10$	AR	7.7	7.6	11.9	19.4	20.2	65.8	31.8	33.9	95.7	42.1	45.4	99.8
	AR^*	0.8	0.8	1.1	15.8	16.3	42.8	65.0	66.3	93.1	94.1	94.1	99.6
	SS	0.0	0.0	0.0	0.0	0.0	1.1	0.0	0.0	7.8	0.1	0.1	12.6
	SS^*	0.3	0.4	0.7	25.4	25.3	37.0	84.5	84.5	89.2	98.2	98.2	98.7

		$T=100$											
	ρ	0.3											
	ϕ	0.50			0.75			0.90			1.00		
	λ	2.40			1.07			0.84			0.75		
	σ_v	1.43			2.15			2.42			2.57		
	$\bar{\lambda}^2$	0	0.1	10	0	0.1	10	0	0.1	10	0	0.1	10
	$\bar{\pi}_1[1,1]$	0	0.05	0.50	0	0.05	0.50	0	0.05	0.50	0	0.05	0.50
$l=2$	AR	5.4	5.7	12.7	9.5	11.1	76.5	14.0	17.2	98.5	18.8	25.1	100.0
	AR^*	2.7	2.8	6.1	5.4	6.1	62.0	8.4	11.0	96.5	12.5	17.2	99.8
	SS	1.5	1.5	4.1	2.6	3.1	46.9	4.5	5.6	82.2	6.8	8.2	86.1
	SS^*	3.1	3.2	5.8	5.0	5.7	48.8	7.3	8.6	83.3	9.9	11.8	87.3
	$\bar{\pi}_1[1,1]$	0	0.03	0.32	0	0.03	0.32	0	0.03	0.32	0	0.03	0.32
$l=5$	AR	5.0	5.1	8.8	11.3	11.9	60.6	19.1	21.2	95.7	27.2	31.7	99.7
	AR^*	2.2	2.3	3.3	5.4	5.6	42.7	10.3	11.8	90.2	16.3	19.2	99.3
	SS	0.1	0.1	0.2	0.3	0.3	8.3	0.6	0.8	35.1	1.2	1.5	47.5
	SS^*	3.1	3.2	4.3	5.1	5.1	30.2	7.4	7.8	61.3	9.4	10.3	67.4
	$\bar{\pi}_1[1,1]$	0	0.02	0.22	0	0.02	0.22	0	0.02	0.22	0	0.02	0.22
$l=10$	AR	5.2	5.2	6.7	14.2	14.8	50.9	26.9	28.9	92.2	40.1	43.2	99.5
	AR^*	1.6	1.6	2.0	5.5	5.7	29.6	13.1	14.0	82.3	22.3	24.9	98.3
	SS	0.0	0.0	0.0	0.0	0.0	0.4	0.0	0.0	4.9	0.1	0.1	10.5
	SS^*	3.2	3.3	3.9	5.1	5.1	19.1	7.1	7.2	40.1	9.3	9.6	48.3

		$T=200$											
	ρ	0.1											
	ϕ	0.50			0.75			0.90			1.00		
	λ	0.27			0.17			0.14			0.13		
	σ_v	4.30			5.41			5.96			6.28		
	$\bar{\lambda}^2$	0	0.1	10	0	0.1	10	0	0.1	10	0	0.1	10
	$\bar{\pi}_1[1,1]$	0	0.04	0.35	0	0.04	0.35	0	0.04	0.35	0	0.04	0.35
$l=2$	AR	7.5	7.7	38.5	12.7	14.4	96.8	17.1	22.0	100.0	20.7	27.4	100.0
	AR^*	0.3	0.4	0.7	90.5	90.6	94.0	100.0	100.0	100.0	100.0	100.0	100.0
	SS	2.3	2.4	19.5	4.1	5.0	82.5	6.5	8.0	96.0	7.4	9.7	90.2
	SS^*	0.4	0.4	0.7	85.4	85.5	89.0	100.0	100.0	100.0	100.0	100.0	100.0
	$\bar{\pi}_1[1,1]$	0	0.02	0.22	0	0.02	0.22	0	0.02	0.22	0	0.02	0.22
$l=5$	AR	9.0	9.1	28.5	17.8	19.3	92.5	25.8	29.1	99.9	32.6	37.5	100.0
	AR^*	0.4	0.4	0.7	87.7	87.8	92.2	100.0	100.0	100.0	100.0	100.0	100.0
	SS	0.2	0.1	2.5	0.5	0.5	39.0	1.0	1.2	71.8	1.3	1.9	63.4
	SS^*	0.3	0.3	0.5	84.6	84.6	87.3	99.9	99.9	100.0	100.0	100.0	100.0
	$\bar{\pi}_1[1,1]$	0	0.02	0.16	0	0.02	0.16	0	0.02	0.16	0	0.02	0.16
$l=10$	AR	11.1	11.1	24.3	24.9	26.3	88.0	38.3	40.8	99.7	48.6	51.9	100.0
	AR^*	0.6	0.6	0.9	84.0	83.9	89.9	100.0	100.0	100.0	100.0	100.0	100.0
	SS	0.0	0.0	0.1	0.1	0.0	6.4	0.1	0.1	27.9	0.2	0.2	26.8
	SS^*	0.3	0.3	0.5	85.2	85.2	87.4	99.9	99.9	100.0	100.0	100.0	100.0

		$T=200$											
	ρ	0.2											
	ϕ	0.50			0.75			0.90			1.00		
	λ	0.80			0.46			0.37			0.33		
	σ_v	2.48			3.29			3.64			3.85		
	$\bar{\lambda}^2$	0	0.1	10	0	0.1	10	0	0.1	10	0	0.1	10
	$\bar{\pi}_1[1,1]$	0	0.04	0.35	0	0.04	0.35	0	0.04	0.35	0	0.04	0.35
$l=2$	AR	6.4	6.4	28.2	11.1	12.6	94.3	16.0	20.7	100.0	20.4	26.9	100.0
	AR^*	0.0	0.0	0.4	52.7	53.4	82.3	99.7	99.8	100.0	100.0	100.0	100.0
	SS	1.8	1.9	12.8	3.6	4.2	75.4	5.8	7.6	94.4	7.2	9.4	89.8
	SS^*	0.1	0.1	0.3	48.6	49.1	73.5	99.2	99.2	99.8	100.0	100.0	100.0
	$\bar{\pi}_1[1,1]$	0	0.02	0.22	0	0.02	0.22	0	0.02	0.22	0	0.02	0.22
$l=5$	AR	6.9	7.0	19.7	15.5	16.5	87.4	23.9	27.2	99.9	32.0	37.0	100.0
	AR^*	0.1	0.1	0.4	41.3	41.8	74.2	99.2	99.2	100.0	100.0	100.0	100.0
	SS	0.1	0.1	1.2	0.4	0.5	29.8	0.9	1.1	66.8	1.3	1.7	61.5
	SS^*	0.0	0.0	0.3	48.1	48.3	66.4	99.2	99.2	99.6	100.0	100.0	100.0
	$\bar{\pi}_1[1,1]$	0	0.02	0.16	0	0.02	0.16	0	0.02	0.16	0	0.02	0.16
$l=10$	AR	7.9	8.0	16.1	20.9	21.7	81.6	35.0	37.2	99.4	47.6	51.0	100.0
	AR^*	0.2	0.2	0.4	33.6	34.0	66.2	98.3	98.3	99.8	100.0	100.0	100.0
	SS	0.0	0.0	0.0	0.0	0.0	3.8	0.1	0.1	22.5	0.1	0.2	25.3
	SS^*	0.0	0.1	0.2	47.5	47.5	60.9	99.4	99.4	99.7	100.0	100.0	100.0

		$T=200$											
	ρ	0.3											
	ϕ	0.50			0.75			0.90			1.00		
	λ	2.40			1.07			0.84			0.75		
	σ_v	1.43			2.15			2.42			2.57		
	$\bar{\lambda}^2$	0	0.1	10	0	0.1	10	0	0.1	10	0	0.1	10
	$\bar{\pi}_1[1,1]$	0	0.04	0.35	0	0.04	0.35	0	0.04	0.35	0	0.04	0.35
$l=2$	AR	5.1	5.2	16.1	9.2	10.3	87.9	14.4	18.5	99.9	19.8	26.5	100.0
	AR^*	2.5	2.5	8.8	4.8	6.1	78.5	8.8	11.8	99.6	13.2	18.7	100.0
	SS	1.3	1.5	5.9	2.9	3.3	62.5	5.1	6.4	91.6	6.9	9.0	88.6
	SS^*	2.9	3.1	8.4	5.3	5.4	65.7	7.9	9.5	92.6	10.3	13.1	89.9
	$\bar{\pi}_1[1,1]$	0	0.02	0.22	0	0.02	0.22	0	0.02	0.22	0	0.02	0.22
$l=5$	AR	5.0	5.1	10.7	12.0	12.6	76.3	20.6	23.6	99.4	31.4	36.3	100.0
	AR^*	1.9	1.8	4.1	5.3	5.7	60.2	10.9	12.6	98.0	19.2	23.0	100.0
	SS	0.1	0.1	0.4	0.3	0.3	17.6	0.7	0.9	57.8	1.3	1.7	59.0
	SS^*	3.0	3.1	5.7	4.9	5.1	46.8	7.3	8.0	78.8	10.1	11.2	75.0
	$\bar{\pi}_1[1,1]$	0	0.02	0.16	0	0.02	0.16	0	0.02	0.16	0	0.02	0.16
$l=10$	AR	4.9	4.9	8.2	14.9	15.7	67.7	30.0	31.8	98.6	46.4	50.0	100.0
	AR^*	1.5	1.5	2.3	5.9	6.3	45.4	14.2	15.5	95.3	27.3	30.7	99.9
	SS	0.0	0.0	0.0	0.0	0.0	1.4	0.1	0.1	16.1	0.1	0.1	23.3
	SS^*	3.0	2.9	4.2	5.0	5.2	31.7	7.8	8.0	63.2	10.7	10.8	59.5

Table 4.7. Size and power comparison of joint tests under $M3$ with low-frequency (past lags) instruments, nominal level: 5%

Size

				$T=100$												$T=200$											
				$l=1$				$l=5$				$l=10$				$l=1$				$l=5$				$l=10$			
ρ	ϕ	λ	σ_v	AR	AR^*	SS	SS^*	AR	AR^*	SS	SS^*	AR	AR^*	SS	SS^*	AR	AR^*	SS	SS^*	AR	AR^*	SS	SS^*	AR	AR^*	SS	SS^*
0.25	0.50	1.33	1.92	4.8	2.8	4.6	2.9	4.6	2.8	0.1	3.3	3.5	2.3	0.0	3.0	5.0	2.6	5.0	2.6	4.7	3.1	0.1	3.1	4.0	2.8	0.0	3.1
	0.60	0.96	2.27	5.1	2.5	4.9	2.6	4.6	2.8	0.1	2.7	3.6	2.3	0.0	2.6	5.0	2.3	4.8	2.5	4.7	3.0	0.1	2.9	4.1	2.6	0.0	2.9
	0.70	0.76	2.54	5.2	2.4	5.2	2.4	4.8	2.6	0.1	2.5	3.6	2.3	0.0	2.4	5.1	2.0	5.3	2.4	4.8	2.8	0.1	2.4	4.1	2.5	0.0	2.5
	0.80	0.64	2.77	5.6	2.3	6.0	2.6	4.8	2.5	0.1	2.4	4.0	2.3	0.0	2.2	5.5	1.9	5.8	2.2	5.0	2.8	0.2	2.1	4.2	2.5	0.0	2.0
	0.90	0.56	2.97	7.1	2.8	7.6	3.2	5.4	2.6	0.1	2.5	4.4	2.4	0.0	2.3	6.3	2.1	6.8	2.4	5.4	2.9	0.1	2.5	4.4	2.5	0.0	1.9
	1.00	0.50	3.14	5.9	1.6	6.5	1.6	5.2	2.3	0.2	1.7	5.0	2.7	0.0	1.7	5.9	4.2	6.3	4.3	4.7	4.7	0.1	3.8	4.5	4.9	0.0	3.8
0.3	0.50	2.40	1.43	4.8	2.3	4.6	2.4	4.3	2.4	0.0	2.9	3.6	1.8	0.0	2.7	4.8	2.3	4.9	2.3	4.5	2.5	0.1	3.0	3.9	2.3	0.0	3.0
	0.60	1.56	1.78	4.9	2.0	4.7	2.2	4.7	2.2	0.1	2.5	3.4	1.8	0.0	2.3	4.8	2.0	4.9	2.1	4.9	2.4	0.1	2.5	4.1	2.1	0.0	2.4
	0.70	1.19	2.04	5.2	1.9	5.3	2.1	4.9	2.1	0.1	2.1	3.7	1.8	0.0	2.0	5.2	1.7	5.4	1.8	4.9	2.3	0.1	2.1	4.1	2.2	0.0	2.0
	0.80	0.97	2.25	5.7	1.9	6.1	2.1	4.9	1.9	0.1	1.8	4.0	1.8	0.0	1.8	5.7	1.6	5.9	1.6	5.0	2.2	0.1	1.8	4.3	2.0	0.0	1.6
	0.90	0.84	2.42	7.1	2.2	7.7	2.3	5.4	2.0	0.1	2.0	4.4	2.0	0.0	1.8	6.4	1.5	6.7	1.7	5.4	2.2	0.2	1.8	4.3	1.8	0.0	1.5
	1.00	0.75	2.57	6.0	1.0	6.5	1.0	5.4	1.8	0.2	1.1	4.9	2.0	0.0	1.2	6.2	2.0	6.7	2.0	4.7	2.5	0.1	1.8	4.7	2.8	0.0	1.9
0.35	0.50	5.60	0.94	4.6	2.1	4.4	2.0	4.1	1.9	0.1	2.5	3.2	1.4	0.0	2.0	4.8	1.9	5.2	1.9	4.5	1.9	0.1	2.9	3.9	1.6	0.0	2.4
	0.60	2.82	1.32	4.9	1.6	4.6	1.9	4.3	1.7	0.0	2.5	3.5	1.3	0.0	2.0	4.8	1.6	5.0	1.7	4.7	1.8	0.1	2.1	4.0	1.6	0.0	2.1
	0.70	1.96	1.59	5.2	1.5	5.2	1.8	4.8	1.6	0.0	1.9	3.5	1.2	0.0	1.7	5.2	1.3	5.5	1.4	5.0	1.8	0.1	1.8	4.1	1.5	0.0	1.5
	0.80	1.55	1.79	5.9	1.5	6.1	1.7	4.9	1.4	0.1	1.5	3.9	1.3	0.0	1.4	5.7	1.1	5.8	1.2	4.9	1.6	0.1	1.3	4.3	1.3	0.0	1.2
	0.90	1.31	1.94	7.4	1.5	7.8	1.6	5.4	1.3	0.1	1.4	4.4	1.4	0.0	1.4	6.6	1.0	6.8	1.1	5.3	1.4	0.2	1.1	4.3	1.2	0.0	1.1
	1.00	1.17	2.06	6.6	0.7	6.9	0.5	5.7	1.2	0.2	0.6	5.0	1.3	0.0	0.7	7.0	0.8	7.3	0.9	4.9	1.2	0.2	0.7	4.7	1.4	0.0	0.7

Power for $H_0: \phi = 0.50, \rho = 0.35$

				$T=100$												$T=200$											
				$l=1$				$l=5$				$l=10$				$l=1$				$l=5$				$l=10$			
ρ	ϕ	λ	σ_v	AR	AR^*	SS	SS^*	AR	AR^*	SS	SS^*	AR	AR^*	SS	SS^*	AR	AR^*	SS	SS^*	AR	AR^*	SS	SS^*	AR	AR^*	SS	SS^*
0.25	0.50	1.33	1.92	7.4	0.1	6.8	0.1	6.1	0.2	0.2	0.2	4.6	0.3	0.0	0.3	11.5	0.0	9.9	0.0	7.6	0.1	0.3	0.0	6.3	0.0	0.0	0.0
	0.60	0.96	2.27	22.4	2.1	17.8	2.0	13.8	2.0	0.9	1.7	9.3	1.5	0.1	1.3	45.3	1.7	36.8	2.0	28.2	1.3	3.5	1.2	19.7	1.2	0.2	1.0
	0.70	0.76	2.54	57.4	21.0	46.4	16.6	38.5	15.8	5.8	13.0	26.4	10.9	0.3	8.9	88.8	36.1	80.7	31.5	74.9	30.4	29.2	25.6	62.4	25.8	6.5	22.1
	0.80	0.64	2.77	89.9	68.8	80.8	58.0	77.3	57.8	28.5	49.0	62.9	45.4	3.8	37.7	99.6	92.8	98.4	87.0	98.3	89.8	80.6	83.8	96.4	85.4	48.0	80.1
	0.90	0.56	2.97	99.0	96.3	96.8	91.0	97.1	92.6	68.2	86.0	92.4	86.9	24.4	77.6	100.0	99.9	100.0	99.7	100.0	99.9	98.4	99.4	100.0	99.8	92.2	99.4
	1.00	0.50	3.14	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.6	100.0	97.2	100.0	97.2	100.0	97.5	100.0	97.5	100.0	97.4	100.0	97.4	100.0
0.3	0.50	2.40	1.43	5.6	0.1	5.2	0.2	4.7	0.4	0.1	0.4	3.9	0.4	0.0	0.3	6.6	0.0	5.8	0.1	5.3	0.1	0.1	0.0	4.4	0.1	0.0	0.0
	0.60	1.56	1.78	15.5	1.5	12.1	1.4	9.3	1.5	0.4	1.4	6.6	1.1	0.0	1.0	31.0	1.5	24.3	1.6	16.9	1.1	1.3	0.9	12.2	1.1	0.0	0.8
	0.70	1.19	2.04	46.4	16.3	36.9	12.5	28.8	10.9	3.3	8.3	18.8	7.1	0.1	5.4	80.0	31.5	69.7	26.2	60.4	23.4	16.9	18.8	47.5	17.8	2.5	15.2
	0.80	0.97	2.25	84.6	61.3	74.1	49.3	68.3	47.3	20.3	36.1	52.5	34.7	2.1	23.1	99.1	90.2	96.7	83.3	96.2	84.4	69.7	76.1	92.2	77.5	33.9	68.9
	0.90	0.84	2.42	98.4	94.4	95.2	87.3	95.0	88.9	60.4	77.0	88.9	80.5	19.0	60.2	100.0	99.9	100.0	99.5	100.0	99.7	97.1	98.9	99.9	99.4	87.4	98.3
	1.00	0.75	2.57	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.8	100.0	99.0	100.0	99.0	100.0	99.1	100.0	99.1	100.0	99.0	100.0	99.0	100.0
0.35	0.50	5.60	0.94	4.6	2.1	4.4	2.0	4.1	1.9	0.1	2.5	3.2	1.4	0.0	2.0	4.8	1.9	5.2	1.9	4.5	1.9	0.1	2.9	3.9	1.6	0.0	2.4
	0.60	2.82	1.32	8.8	2.6	7.6	2.1	5.8	2.3	0.2	2.8	4.2	1.8	0.0	2.5	15.2	5.3	12.6	4.0	8.6	3.6	0.3	3.7	6.5	2.7	0.0	3.4
	0.70	1.96	1.59	32.4	14.8	25.5	9.9	18.2	8.3	1.4	6.5	11.8	5.3	0.0	4.3	62.6	39.3	51.8	29.2	39.9	24.5	7.0	18.5	29.1	16.5	0.6	14.2
	0.80	1.55	1.79	75.4	54.7	63.6	41.0	54.4	37.8	12.3	24.9	39.5	25.0	1.0	13.7	97.0	90.7	92.3	81.9	89.6	80.9	52.1	67.2	81.9	70.7	18.4	55.7
	0.90	1.31	1.94	97.1	91.9	92.3	82.1	90.7	83.4	50.0	62.5	82.1	71.5	13.2	39.4	100.0	99.9	99.8	99.4	99.8	99.6	94.3	96.5	99.6	99.0	78.2	92.0
	1.00	1.17	2.06	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.9	99.9	99.8	100.0	99.8	100.0	99.8	100.0	99.8	100.0	99.8	100.0	99.8	100.0

Table 4.8. Size and power comparison of joint tests under $M3$ with $\epsilon_t \sim N(0,\pi^2/2)$ and low-frequency (past lags) instruments, nominal level: 5%

Size																											
				$T=100$												$T=200$											
				$l=1$				$l=5$				$l=10$				$l=1$				$l=5$				$l=10$			
ρ	ϕ	λ	σ_v	AR	AR^*	SS	SS^*	AR	AR^*	SS	SS^*	AR	AR^*	SS	SS^*	AR	AR^*	SS	SS^*	AR	AR^*	SS	SS^*	AR	AR^*	SS	SS^*
0.25	0.50	1.33	1.92	5.3	3.0	5.0	2.8	4.3	2.6	0.1	3.3	3.4	2.3	0.0	3.0	4.6	2.6	4.8	2.8	5.0	3.4	0.1	3.3	4.5	3.0	0.0	3.1
	0.60	0.96	2.27	5.1	2.7	5.0	2.5	4.2	2.6	0.1	2.8	3.5	2.1	0.0	2.8	4.8	2.2	4.7	2.6	5.2	3.3	0.1	2.8	4.4	2.9	0.0	2.8
	0.70	0.76	2.54	5.2	2.4	5.2	2.4	4.5	2.4	0.0	2.5	3.7	2.0	0.0	2.4	4.9	2.1	5.0	2.3	5.4	3.1	0.1	2.3	4.4	2.8	0.0	2.5
	0.80	0.64	2.77	5.7	2.3	5.9	2.4	4.6	2.3	0.0	2.4	3.8	2.1	0.0	2.3	5.2	2.1	5.3	2.1	5.5	2.9	0.1	2.1	4.5	2.7	0.0	2.1
	0.90	0.56	2.97	7.1	2.6	7.3	3.0	5.1	2.5	0.1	2.6	4.2	2.1	0.0	2.5	6.4	2.2	6.5	2.3	5.6	2.9	0.2	2.2	4.7	2.7	0.0	2.0
	1.00	0.50	3.14	5.9	1.4	6.4	1.5	5.5	2.3	0.1	1.7	4.5	2.1	0.0	1.6	5.8	4.4	6.1	4.5	4.4	4.7	0.1	3.9	4.8	5.5	0.0	4.1
0.3	0.50	2.40	1.43	5.1	2.6	5.1	2.6	4.5	2.3	0.1	3.1	3.4	1.8	0.0	2.5	4.9	2.3	5.0	2.4	4.9	2.9	0.1	3.2	4.2	2.4	0.0	2.9
	0.60	1.56	1.78	5.2	2.3	5.2	2.3	4.3	2.1	0.1	2.7	3.5	1.8	0.0	2.2	4.8	1.8	4.9	2.1	5.1	2.7	0.1	2.5	4.3	2.3	0.0	2.4
	0.70	1.19	2.04	5.3	2.0	5.4	2.0	4.4	1.9	0.0	2.4	3.7	1.7	0.0	2.0	4.9	1.6	4.9	1.9	5.3	2.3	0.1	1.9	4.3	2.1	0.0	2.1
	0.80	0.97	2.25	5.7	1.8	5.9	1.9	4.7	1.8	0.1	2.0	3.8	1.5	0.0	1.8	5.2	1.6	5.4	1.8	5.5	2.2	0.1	1.6	4.3	2.0	0.0	1.6
	0.90	0.84	2.42	7.1	2.0	7.4	2.3	5.3	1.9	0.1	2.1	4.2	1.6	0.0	1.8	6.4	1.6	6.6	1.8	5.6	2.2	0.2	1.6	4.6	2.1	0.0	1.5
	1.00	0.75	2.57	6.0	0.8	6.3	1.0	5.3	1.6	0.2	1.2	4.6	1.6	0.0	1.1	6.1	2.0	6.5	2.0	4.7	2.5	0.1	1.7	4.9	3.1	0.0	1.9
0.35	0.50	5.60	0.94	5.5	2.3	5.2	2.3	4.5	2.1	0.0	2.9	3.6	1.4	0.0	2.3	5.0	1.9	5.2	2.2	4.9	2.2	0.1	3.0	4.4	1.9	0.0	2.3
	0.60	2.82	1.32	5.3	2.0	5.3	1.9	4.6	1.8	0.1	2.7	3.5	1.3	0.0	1.9	4.9	1.4	5.1	1.6	5.2	2.1	0.1	2.4	4.3	1.7	0.0	2.1
	0.70	1.96	1.59	5.4	1.7	5.7	1.6	4.4	1.4	0.1	2.2	3.7	1.2	0.0	1.6	4.7	1.2	5.0	1.4	5.3	1.8	0.1	1.7	4.3	1.6	0.0	1.8
	0.80	1.55	1.79	6.1	1.4	6.1	1.4	4.7	1.3	0.1	1.8	3.8	1.0	0.0	1.3	5.4	1.1	5.3	1.3	5.5	1.6	0.1	1.2	4.4	1.5	0.0	1.1
	0.90	1.31	1.94	7.2	1.5	7.5	1.5	5.3	1.3	0.1	1.7	4.4	1.1	0.0	1.3	6.5	1.1	6.7	1.3	5.6	1.5	0.2	1.0	4.5	1.3	0.0	1.0
	1.00	1.17	2.06	6.4	0.5	6.8	0.4	5.4	1.0	0.2	0.6	4.7	1.1	0.0	0.7	6.8	0.8	6.9	0.9	5.0	1.1	0.2	0.8	5.0	1.4	0.0	0.7
Power for $H_0: \phi = 0.50, \rho = 0.35$																											
				$T=100$												$T=200$											
				$l=1$				$l=5$				$l=10$				$l=1$				$l=5$				$l=10$			
ρ	ϕ	λ	σ_v	AR	AR^*	SS	SS^*	AR	AR^*	SS	SS^*	AR	AR^*	SS	SS^*	AR	AR^*	SS	SS^*	AR	AR^*	SS	SS^*	AR	AR^*	SS	SS^*
0.25	0.50	1.33	1.92	7.6	0.1	6.5	0.1	6.0	0.2	0.2	0.1	4.6	0.2	0.0	0.2	11.5	0.0	10.1	0.0	8.6	0.0	0.3	0.0	6.6	0.0	0.0	0.0
	0.60	0.96	2.27	22.6	2.0	17.2	1.9	13.8	1.7	0.8	1.6	8.9	1.2	0.0	1.3	46.0	1.4	37.0	1.6	28.3	1.4	3.4	1.0	20.0	1.1	0.3	1.0
	0.70	0.76	2.54	57.4	21.1	46.3	16.5	38.5	15.4	5.8	12.1	25.9	10.0	0.3	8.8	89.3	36.0	80.8	30.5	75.3	29.8	29.4	24.9	62.6	25.3	6.3	21.4
	0.80	0.64	2.77	89.7	68.7	80.8	57.9	77.2	57.7	27.8	47.9	62.8	45.6	3.5	36.9	99.7	93.1	98.5	87.2	98.5	90.3	80.7	84.1	96.1	86.0	47.1	80.4
	0.90	0.56	2.97	99.2	96.1	96.8	90.9	97.2	93.0	67.9	85.9	92.4	87.0	25.0	77.4	100.0	99.9	100.0	99.7	100.0	99.9	98.6	99.5	100.0	99.7	92.1	99.6
	1.00	0.50	3.14	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.6	100.0	97.0	100.0	97.0	100.0	97.4	100.0	97.4	100.0	97.2	100.0	97.2	100.0
0.3	0.50	2.40	1.43	5.7	0.1	5.6	0.2	4.9	0.4	0.1	0.3	3.7	0.3	0.0	0.4	6.7	0.0	6.1	0.0	5.9	0.0	0.1	0.0	5.0	0.1	0.0	0.0
	0.60	1.56	1.78	15.0	1.4	11.4	1.4	9.4	1.2	0.4	1.2	6.2	1.0	0.0	1.1	31.3	1.3	24.9	1.3	17.7	1.1	1.2	0.7	12.3	1.0	0.1	0.7
	0.70	1.19	2.04	46.7	16.4	36.5	11.7	28.6	10.8	3.2	7.9	18.2	6.3	0.1	5.0	80.1	31.2	70.0	25.4	60.6	22.9	16.8	17.9	47.2	17.1	2.7	14.0
	0.80	0.97	2.25	84.5	60.8	74.1	49.0	67.6	46.7	19.2	35.0	53.0	34.2	1.9	22.8	99.1	90.6	97.1	83.3	96.3	84.9	69.2	76.1	91.9	77.5	32.8	68.4
	0.90	0.84	2.42	98.5	94.5	95.4	87.4	95.1	89.0	60.2	76.6	89.0	80.3	19.0	59.8	100.0	99.9	99.9	99.5	100.0	99.7	97.4	99.0	99.9	99.5	87.4	98.5
	1.00	0.75	2.57	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.8	100.0	98.9	100.0	98.9	100.0	99.0	100.0	99.0	100.0	98.9	100.0	98.9	100.0
0.35	0.50	5.60	0.94	5.5	2.3	5.2	2.3	4.5	2.1	0.0	2.9	3.6	1.4	0.0	2.3	5.0	1.9	5.2	2.2	4.9	2.2	0.1	3.0	4.4	1.9	0.0	2.3
	0.60	2.82	1.32	8.8	2.7	7.3	2.1	6.0	2.0	0.1	3.0	4.3	1.6	0.0	2.5	15.1	5.2	12.6	3.8	9.3	3.9	0.3	3.5	6.6	2.9	0.0	3.3
	0.70	1.96	1.59	32.5	14.4	24.7	9.2	18.2	8.3	1.2	6.6	10.9	4.6	0.1	4.3	62.2	39.3	51.7	29.2	40.1	25.2	6.4	18.4	28.7	16.7	0.6	13.5
	0.80	1.55	1.79	74.7	54.2	63.6	40.4	53.6	36.9	11.4	23.9	39.5	23.9	0.8	13.2	97.1	91.2	92.8	82.1	89.8	81.1	51.3	67.0	81.8	70.4	17.6	54.2
	0.90	1.31	1.94	97.0	92.2	92.2	82.5	90.9	83.6	48.7	61.8	82.3	71.6	12.9	39.6	100.0	99.9	99.9	99.4	99.9	99.5	94.6	96.6	99.6	99.0	78.0	91.9
	1.00	1.17	2.06	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.9	100.0	99.7	100.0	99.7	100.0	99.8	100.0	99.8	100.0	99.7	100.0	99.7	100.0

Table 4.9. Size comparison of tests ($H_0 : \phi_{lf} = \phi_0$) under $M4$ with high-frequency instruments, nominal level: 5%

				Size																							
				$T=100$												$T=200$											
				$l=1$				$l=5$				$l=10$				$l=1$				$l=5$				$l=10$			
Frequency	ϕ_{lf}	ϕ_{hf}	$\sigma_{v,hf}$	AR	AR^*	SS	SS^*	AR	AR^*	SS	SS^*	AR	AR^*	SS	SS^*	AR	AR^*	SS	SS^*	AR	AR^*	SS	SS^*	AR	AR^*	SS	SS^*
30-sec	0.000	0.000000	0.00537	5.0	5.1	4.9	5.0	4.8	4.8	0.2	5.2	5.1	5.0	0.0	4.9	5.0	5.0	4.8	5.1	5.1	5.0	0.1	5.1	4.9	4.8	0.0	4.9
	0.100	0.997052		5.1	5.0	5.0	5.2	4.8	4.8	0.1	4.9	5.2	4.8	0.0	4.9	5.0	4.8	4.8	4.7	5.0	4.9	0.1	5.0	4.7	4.9	0.0	5.0
	0.300	0.998458		4.9	4.9	4.8	4.8	5.1	5.2	0.1	4.6	5.1	4.7	0.0	4.6	4.8	4.7	4.5	4.6	5.1	5.0	0.1	4.6	4.9	5.2	0.0	4.3
	0.500	0.999112		4.7	4.7	4.6	4.6	5.1	5.1	0.1	4.2	5.3	5.0	0.0	4.2	4.6	4.5	4.6	4.5	5.1	4.8	0.0	4.2	5.3	5.2	0.0	3.7
	0.700	0.999543		4.7	4.7	4.5	4.5	5.0	5.0	0.0	3.7	5.3	5.3	0.0	3.5	4.4	4.4	4.6	4.5	5.2	4.9	0.0	3.3	5.4	5.3	0.0	3.2
	0.800	0.999714		4.5	4.4	4.4	4.3	5.2	5.1	0.0	3.3	5.4	5.1	0.0	3.1	4.3	4.3	4.4	4.4	5.2	4.9	0.0	2.9	5.4	5.3	0.0	2.6
	0.900	0.999865		4.3	4.2	4.2	4.1	5.2	4.9	0.0	3.0	5.3	5.1	0.0	2.7	4.1	3.9	4.1	4.1	5.2	4.8	0.0	2.3	5.3	5.3	0.0	1.9
	0.950	0.999934		4.1	4.1	4.1	4.1	5.2	4.9	0.0	2.8	5.6	5.4	0.0	2.8	3.7	3.6	3.8	3.9	5.2	4.8	0.0	1.9	5.4	5.3	0.0	2.0
	0.999	0.999999		5.4	5.3	5.1	4.9	6.2	6.0	0.1	4.1	6.7	6.1	0.0	3.4	6.1	5.8	5.5	5.6	6.9	6.5	0.0	4.0	6.7	6.5	0.0	3.6
1-min	0.000	0.000000	0.00760	5.6	5.6	5.4	5.3	4.7	4.9	0.1	5.1	5.4	5.1	0.0	5.1	5.0	5.1	4.8	4.8	5.4	5.4	0.1	5.1	4.7	4.5	0.0	4.9
	0.100	0.994113		5.7	5.5	5.3	5.3	4.8	4.7	0.1	4.9	5.5	5.2	0.0	5.1	5.0	4.8	5.1	5.0	5.6	5.5	0.1	5.2	4.7	4.6	0.0	4.7
	0.300	0.996918		5.6	5.3	5.3	5.3	5.1	5.0	0.1	4.8	5.8	5.4	0.0	4.7	5.0	4.8	4.8	5.1	5.7	5.8	0.1	5.2	5.1	5.0	0.0	4.6
	0.500	0.998224		5.2	5.0	5.1	5.0	5.4	5.2	0.1	4.3	6.1	5.7	0.0	4.3	5.0	4.9	4.9	4.8	6.0	6.1	0.1	4.3	5.3	5.2	0.0	4.2
	0.700	0.999086		4.8	4.6	5.0	4.8	5.6	5.4	0.1	3.7	6.3	5.8	0.0	3.6	4.9	4.9	4.5	4.5	6.1	6.2	0.1	3.5	5.3	5.3	0.0	3.3
	0.800	0.999428		4.7	4.4	4.8	4.8	5.6	5.3	0.1	3.6	6.4	5.8	0.0	3.3	4.8	4.6	4.4	4.4	6.2	6.2	0.1	2.7	5.3	5.4	0.0	2.5
	0.900	0.999730		4.3	4.3	4.4	4.4	5.6	5.3	0.1	3.2	6.4	5.8	0.0	2.7	4.2	4.2	4.0	4.0	6.2	6.0	0.0	2.1	5.2	5.3	0.0	1.9
	0.950	0.999868		4.0	4.0	4.2	4.2	5.6	5.5	0.1	2.8	6.4	5.9	0.0	2.4	3.6	3.5	3.7	3.7	6.2	6.0	0.0	1.5	5.3	5.2	0.0	1.4
	0.999	0.999997		4.6	4.4	4.5	4.3	6.2	5.8	0.0	2.6	6.8	6.5	0.0	2.6	4.6	3.9	4.1	4.0	7.1	6.7	0.0	2.0	6.2	6.0	0.0	1.4
5-min	0.000	0.000000	0.01698	4.9	4.9	4.9	5.0	4.8	4.8	0.1	5.1	5.3	5.2	0.0	5.2	5.0	5.0	5.1	5.1	4.8	5.0	0.1	4.9	5.2	5.0	0.0	5.0
	0.100	0.970911		5.2	5.0	5.0	4.8	5.0	5.0	0.1	4.9	5.4	5.3	0.0	5.0	5.3	5.1	4.8	5.0	4.9	4.8	0.1	5.1	5.3	5.0	0.0	5.0
	0.300	0.984683		5.0	4.8	5.0	4.8	5.7	5.8	0.1	4.7	5.8	5.6	0.0	4.6	5.2	5.1	4.9	4.8	5.5	5.4	0.1	5.0	5.9	5.8	0.0	4.7
	0.500	0.991153		5.1	4.9	4.8	4.7	6.3	6.1	0.2	4.6	6.6	6.2	0.0	4.3	5.3	5.2	5.0	4.8	6.0	5.8	0.1	4.9	6.5	6.3	0.0	4.8
	0.700	0.995438		4.9	4.6	4.8	4.7	6.5	6.3	0.2	4.5	6.7	6.2	0.0	4.0	4.9	4.9	5.0	4.8	6.0	5.9	0.2	4.6	6.9	6.6	0.0	4.4
	0.800	0.997143		4.4	4.1	4.8	4.7	6.3	6.1	0.2	4.2	6.7	6.1	0.0	3.9	4.7	4.5	4.8	4.6	5.9	5.5	0.2	4.2	6.9	6.4	0.0	3.9
	0.900	0.998650		4.2	3.7	4.5	4.5	6.1	5.7	0.1	4.1	6.3	5.7	0.0	3.5	4.6	4.2	4.5	4.3	5.7	5.2	0.1	3.2	6.6	5.9	0.0	3.2
	0.950	0.999343		4.1	3.5	4.2	4.3	6.1	5.4	0.1	3.7	6.1	5.2	0.0	3.1	4.7	3.6	4.1	3.8	5.7	5.0	0.1	2.5	6.5	5.7	0.0	2.5
	0.999	0.999987		4.0	3.1	4.0	3.7	6.2	5.3	0.1	3.2	6.2	5.4	0.0	2.8	4.9	3.3	3.6	3.2	6.0	4.8	0.1	1.6	6.9	5.7	0.0	1.6
10-min	0.000	0.000000	0.02402	5.1	5.1	5.3	4.8	4.9	5.0	0.1	4.9	5.1	5.0	0.0	5.1	4.7	4.9	4.8	4.8	5.3	5.5	0.1	5.2	4.6	5.1	0.0	4.5
	0.100	0.942668		5.3	5.1	5.4	4.9	5.5	5.4	0.1	5.0	5.6	5.3	0.0	5.1	5.1	5.0	5.0	4.8	5.6	5.6	0.1	5.1	5.1	5.1	0.0	4.6
	0.300	0.969601		5.4	5.1	5.3	5.1	6.1	5.6	0.1	4.8	6.4	5.7	0.0	4.9	5.2	5.0	5.3	4.9	6.1	5.8	0.2	5.1	5.9	5.6	0.0	4.6
	0.500	0.982384		5.2	5.1	5.4	5.1	6.6	6.4	0.1	4.8	7.1	6.4	0.0	4.9	5.3	5.3	5.3	5.1	6.7	6.5	0.2	5.2	6.6	6.3	0.0	4.6
	0.700	0.990896		5.3	5.0	5.3	5.0	6.8	6.6	0.2	4.9	7.5	6.9	0.0	4.6	5.1	5.0	5.4	5.1	6.8	6.5	0.1	5.0	6.9	6.5	0.0	4.3
	0.800	0.994295		5.0	4.8	5.3	4.8	6.8	6.7	0.2	4.8	7.5	6.7	0.0	4.5	4.9	4.6	5.2	5.0	6.7	6.3	0.1	4.6	6.7	6.4	0.0	4.0
	0.900	0.997302		4.5	4.1	5.1	4.7	6.6	6.3	0.1	4.4	7.1	6.2	0.0	4.0	5.6	4.8	5.0	4.9	6.7	5.9	0.1	3.9	6.7	6.0	0.0	3.3
	0.950	0.998686		4.4	3.9	4.9	4.5	6.5	5.7	0.1	4.3	6.9	5.8	0.0	3.9	6.5	4.8	4.8	4.5	6.9	5.6	0.1	3.4	6.7	5.5	0.0	2.7
	0.999	0.999974		4.9	3.4	4.3	4.3	6.4	5.2	0.1	3.9	6.7	5.3	0.0	3.7	8.3	4.3	4.2	3.6	7.7	5.4	0.1	2.7	7.2	5.2	0.0	2.0

Table 4.10. Power comparison of tests ($H_0 : \phi_{lf} = 0$) under $M4$ with high-frequency instruments, nominal level: 5%

				Power for $H_0 : \phi_{lf} = 0$																							
				$T = 100$												$T = 200$											
				$l = 1$				$l = 5$				$l = 10$				$l = 1$				$l = 5$				$l = 10$			
Frequency	ϕ_{lf}	ϕ_{hf}	$\sigma_{v,hf}$	AR	AR^*	SS	SS^*	AR	AR^*	SS	SS^*	AR	AR^*	SS	SS^*	AR	AR^*	SS	SS^*	AR	AR^*	SS	SS^*	AR	AR^*	SS	SS^*
30-sec	0.000	0.000000	0.00537	5.0	5.1	4.9	5.0	4.8	4.8	0.2	5.2	5.1	5.0	0.0	4.9	5.0	5.0	4.8	5.1	5.1	5.0	0.1	5.1	4.9	4.8	0.0	4.9
	0.100	0.997052		5.1	5.0	5.1	5.2	4.9	4.8	0.1	5.2	5.1	4.8	0.0	5.0	5.0	4.9	4.9	4.8	5.1	5.1	0.1	5.1	4.8	4.8	0.0	5.0
	0.300	0.998458		5.0	5.0	5.0	5.0	5.2	5.2	0.2	5.4	5.3	5.2	0.0	6.0	5.2	5.4	4.9	4.9	5.4	5.4	0.1	5.4	5.1	5.3	0.0	5.3
	0.500	0.999112		5.7	6.2	5.8	5.8	6.2	6.1	0.2	6.6	6.3	6.3	0.1	7.5	6.6	6.9	6.4	6.2	6.8	6.9	0.2	7.0	7.0	7.0	0.0	6.8
	0.700	0.999543		13.1	13.9	12.0	11.6	12.7	12.8	1.3	11.6	12.4	12.4	0.2	12.3	23.5	24.8	20.4	19.6	20.4	20.7	2.6	17.0	17.0	17.2	0.4	15.2
	0.800	0.999714		28.9	30.2	24.4	23.6	26.7	27.5	4.8	21.2	23.7	23.6	1.0	19.4	55.5	57.0	48.0	46.9	49.9	50.3	14.4	38.9	41.7	42.2	3.3	32.3
	0.900	0.999865		65.3	66.3	55.6	54.2	62.5	62.6	20.8	45.6	53.6	53.6	6.5	36.6	93.2	93.8	87.3	86.6	91.7	91.8	58.7	79.2	86.2	86.4	31.1	69.4
	0.950	0.999934		85.1	85.5	75.0	74.0	82.3	82.3	40.1	63.0	74.9	74.6	17.4	51.8	99.3	99.3	97.6	97.3	99.0	99.0	84.1	92.6	98.1	98.0	63.2	86.4
	0.999	0.999999		94.7	94.9	88.3	87.6	93.8	93.8	62.6	78.4	90.1	89.9	38.7	68.9	100.0	100.0	99.7	99.7	100.0	100.0	95.2	97.5	99.9	99.9	87.6	95.1
1-min	0.000	0.000000	0.00760	5.6	5.6	5.4	5.3	4.7	4.9	0.1	5.1	5.4	5.1	0.0	5.1	5.0	5.1	4.8	4.8	5.4	5.4	0.1	5.1	4.7	4.5	0.0	4.9
	0.100	0.994113		5.7	5.6	5.3	5.2	4.7	4.7	0.1	5.0	5.6	5.3	0.0	5.2	5.1	5.0	5.0	5.0	5.4	5.5	0.0	5.4	4.8	4.5	0.0	4.8
	0.300	0.996918		5.6	5.4	5.1	5.1	5.0	4.9	0.2	5.1	5.8	5.4	0.0	5.5	5.2	5.2	4.9	4.9	5.4	5.6	0.1	5.7	4.8	4.7	0.0	5.1
	0.500	0.998224		5.9	6.0	5.6	5.4	5.8	5.8	0.1	5.9	6.5	6.2	0.0	6.1	6.1	6.3	5.8	5.7	6.1	6.4	0.1	5.9	5.4	5.7	0.0	5.9
	0.700	0.999086		10.3	11.2	9.3	8.9	9.7	10.2	0.6	8.9	10.1	10.1	0.0	9.1	15.1	16.9	13.9	13.5	14.2	14.8	1.1	11.6	12.3	12.8	0.1	11.0
	0.800	0.999428		20.3	22.8	18.1	17.4	19.7	20.6	2.5	15.3	17.2	17.7	0.4	13.9	37.5	41.0	34.0	32.5	36.8	38.4	7.0	27.2	30.7	32.2	1.5	23.3
	0.900	0.999730		51.3	54.6	43.6	42.3	52.0	52.9	13.5	35.4	43.4	43.9	3.7	28.9	83.3	85.8	77.5	76.0	84.2	85.2	43.9	67.6	78.0	78.8	20.8	58.6
	0.950	0.999868		74.7	76.8	64.7	63.3	75.3	75.5	30.6	54.3	67.8	68.1	12.4	43.6	97.3	97.8	94.3	93.8	97.6	97.7	75.1	87.2	95.8	95.9	53.0	80.2
	0.999	0.999997		90.8	91.6	82.9	81.8	90.7	90.9	53.5	71.8	86.0	86.0	31.6	62.6	99.8	99.8	99.1	99.1	99.8	99.8	92.7	96.3	99.6	99.7	82.7	93.3
5-min	0.000	0.000000	0.01698	4.9	4.9	4.9	5.0	4.8	4.8	0.1	5.1	5.3	5.2	0.0	5.2	5.0	5.0	5.1	5.1	4.8	5.0	0.1	4.9	5.2	5.0	0.0	5.0
	0.100	0.970911		4.9	5.0	5.0	5.0	4.8	5.0	0.1	5.0	5.1	5.2	0.0	5.2	5.1	5.2	4.9	5.0	4.9	4.8	0.1	5.1	5.2	5.0	0.0	5.2
	0.300	0.984683		4.9	5.0	5.0	5.1	4.8	4.9	0.1	4.7	5.2	5.2	0.0	5.2	5.0	5.1	4.9	5.0	4.9	5.0	0.1	5.2	5.1	5.1	0.0	5.0
	0.500	0.991153		5.0	5.0	4.9	5.1	4.8	4.9	0.1	4.8	5.2	5.2	0.0	5.3	5.1	5.2	5.0	5.0	4.9	5.1	0.1	5.3	5.4	5.4	0.0	5.1
	0.700	0.995438		5.6	6.1	5.9	5.6	5.4	5.6	0.1	5.4	6.1	6.2	0.0	6.0	6.7	7.5	6.6	6.3	6.1	7.0	0.1	5.9	6.1	6.7	0.0	6.2
	0.800	0.997143		7.8	9.7	8.0	7.6	7.7	8.5	0.2	6.9	7.7	8.2	0.0	7.0	11.6	15.8	11.6	11.1	12.3	14.6	0.7	9.6	10.3	12.2	0.1	8.8
	0.900	0.998650		19.9	27.0	17.5	16.6	22.6	26.0	2.5	14.7	18.7	21.3	0.4	13.1	40.4	53.6	37.5	35.9	47.3	53.8	10.4	31.1	42.1	48.0	2.9	26.0
	0.950	0.999343		39.7	49.7	32.9	31.6	46.1	49.8	8.7	25.8	40.3	43.7	3.0	23.0	75.4	85.5	71.2	69.3	82.7	86.2	37.0	59.3	78.6	81.8	19.1	52.0
	0.999	0.999987		69.4	76.8	58.5	57.4	74.3	76.8	24.6	45.5	68.8	71.1	13.2	40.2	96.9	98.4	94.3	93.6	98.1	98.5	74.8	85.5	97.4	97.9	59.7	81.3
10-min	0.000	0.000000	0.02402	5.1	5.1	5.3	4.8	4.9	5.0	0.1	4.9	5.1	5.0	0.0	5.1	4.7	4.9	4.8	4.8	5.3	5.5	0.1	5.2	4.6	5.1	0.0	4.5
	0.100	0.942668		5.0	5.1	5.2	4.7	4.8	5.3	0.1	5.0	5.1	5.0	0.0	5.2	4.8	4.9	4.9	4.8	5.0	5.4	0.1	5.0	4.7	4.9	0.0	4.5
	0.300	0.969601		4.9	5.1	5.3	4.9	4.9	5.2	0.1	4.8	5.0	4.9	0.0	5.2	4.9	4.9	5.0	5.0	5.2	5.3	0.1	5.0	4.7	5.0	0.0	4.9
	0.500	0.982384		4.8	5.1	5.4	5.0	4.9	5.3	0.1	4.8	5.1	4.9	0.0	5.3	4.8	4.9	5.1	5.0	5.0	5.3	0.1	4.9	4.9	5.3	0.0	5.0
	0.700	0.990896		5.2	5.9	5.5	5.2	5.1	5.7	0.0	5.1	5.1	5.2	0.0	5.5	5.4	5.9	6.0	5.6	5.7	6.1	0.2	5.3	5.3	5.9	0.0	5.3
	0.800	0.994295		6.2	7.6	6.7	6.3	5.9	6.8	0.1	5.9	5.6	6.0	0.0	6.0	7.5	9.9	7.9	7.5	7.9	9.6	0.3	6.6	6.9	8.4	0.0	5.9
	0.900	0.997302		12.6	17.9	11.8	10.9	13.1	16.4	0.8	9.4	10.8	13.3	0.1	9.2	22.0	35.9	22.0	20.7	28.2	35.9	3.4	16.5	24.5	31.2	0.7	14.3
	0.950	0.998686		25.8	37.8	21.8	20.7	30.8	36.5	3.6	16.9	25.5	29.9	0.8	15.4	53.6	72.7	50.3	48.1	65.6	73.8	17.9	39.1	61.8	68.7	7.6	35.1
	0.999	0.999974		55.5	66.7	44.6	43.3	61.6	66.7	15.1	33.0	56.2	60.1	6.5	29.3	91.4	96.2	86.0	85.4	95.0	96.5	57.8	73.8	93.9	95.5	44.2	70.5

Table 4.11. Description of instruments

No	Classes of instruments		Subclasses
		HF realized measures not robust to jumps	
1-13	RV	Realized volatility	1s, 5s, 30s, 1m, 5m, 10m, 1t, 5t, 10t, 20t, 50t, 5m-ss, 10m-ss
14-24	RVbr	Realized volatility with optimal sampling	1s, 5s, 30s, 1m, 5m, 10m, 1t, 5t, 10t, 20t, 50t
25-35	MSRV	Multi-scales realized volatility	1s, 5s, 30s, 1m, 5m, 10m, 1t, 5t, 10t, 20t, 50t
36-40	Rkcub	Realized Kernel with fat-top cubic kernel	1t, 5t, 10t, 20t, 50t
41-45	Rkbart	Realized Kernel with fat-top Bartlett kernel	1t, 5t, 10t, 20t, 50t
46-50	RKth2	Realized Kernel with fat-top Tukey-Hanning kernel (power 2)	1t, 5t, 10t, 20t, 50t
51-55	RKnfp	Realized Kernel with non-fat-top Parzen kernel	1t, 5t, 10t, 20t, 50t
56-58	RRV	Realized range volatility	1m, 5m, 10m
		HF realized measures robust to jumps	
59-71	BV	Bipower variation	1s, 5s, 30s, 1m, 5m, 10m, 1t, 5t, 10t, 20t, 50t, 5m-ss, 10m-ss
72-77	MedRV	Nearest neighbor truncated median RV	1s, 5s, 30s, 1m, 5m, 10m
78-83	MinRV	Nearest neighbor truncated minimum RV	1s, 5s, 30s, 1m, 5m, 10m
		Aditional HF measures and jump variations	
84-96	RSVN	Realized semivariance due to negative returns	1s, 5s, 30s, 1m, 5m, 10m, 1t, 5t, 10t, 20t, 50t, 5m-ss, 10m-ss
97-109	RSVP	Realized semivariance due to positive returns	1s, 5s, 30s, 1m, 5m, 10m, 1t, 5t, 10t, 20t, 50t, 5m-ss, 10m-ss
110-120	JV	Jump variation	1s, 5s, 30s, 1m, 5m, 10m, 1t, 5t, 10t, 20t, 50t
121-131	SJV	Signed jump variation	1s, 5s, 30s, 1m, 5m, 10m, 1t, 5t, 10t, 20t, 50t
132-142	LJV	Log squared jump variation	1s, 5s, 30s, 1m, 5m, 10m, 1t, 5t, 10t, 20t, 50t
143-153	LSJV	Log squared signed jump variation	1s, 5s, 30s, 1m, 5m, 10m, 1t, 5t, 10t, 20t, 50t
154-156	PCF	HF principal component factor	1, 2, 3
		Other instruments	
157-162	ImV-C	Implied volatility (call option)	mean, min, max, q1, q2, q3
163-168	ImV-P	Implied volatility (put option)	mean, min, max, q1, q2, q3
169-174	ImV-A	Implied volatility (both call and put option)	mean, min, max, q1, q2, q3
175	1-day	Daily realized volatility	

Notes:

1. Sampling frequencies are tick, second and minute, *e.g.*, 1t stands for 1-tick, 1s stands for 1-second and 1m stands for 1-minute.
2. The use of 1-minute subsamples in the calculation of realized measures is denoted by ss.
3. Three principal component factors are extracted from HF instruments (1-109). PCF-1 stands for the largest factor.
4. Implied volatilities (ImV) are calculated from American options. We consider three classes: (1) only call options, (2) only put options, and (3) both call and put options. For each class, we use all implied volatilities at a given date to construct six ImV subclasses, which are mean, min, max, and three quantiles (q1, q2, q3).

Table 4.12. Strength comparison with daily past lags as instruments
(F-statistics from first-stage regression)

January 2009 - December 2013, $T = 1258$							
	# of instruments						
Ticker	1	2	3	4	5	6	7
GE	23.64	25.00	21.10	20.73	18.46	16.85	14.81
IBM	9.22	10.08	10.08	9.72	8.63	7.73	6.87
JPM	41.08	38.42	34.99	28.34	24.71	23.22	20.79
KO	6.19	10.24	8.82	9.00	8.31	7.08	7.00
PFE	14.99	11.17	7.53	7.43	7.41	7.45	7.06
PG	3.57	4.28	5.38	4.88	5.76	5.14	6.56
T	5.36	13.65	9.62	7.04	6.76	6.07	5.37
WMT	15.24	11.01	7.71	6.10	5.45	5.36	5.63
XOM	9.48	7.80	7.87	6.97	5.86	6.08	5.69
$CV\_Size(0.10)$	16.38	19.93	22.30	24.58	26.87	29.18	31.50

Notes:

1. The critical value (CV) is a function of one endogenous regressor, the number of instrumental variables, and the desired 10% maximal size of a 5% Wald test of $\phi = \phi_0$, for further details, see Table 5.2 of Stock and Yogo (2005).
2. Instruments are deemed weak if the first-stage F-statistic is less than the CV associated with the corresponding column.

Table 4.13. Strength comparison of all IV's
(F-statistics from first-stage regression)
Ticker: IBM, January 2009 - December 2013, $T = 1258$

No	Instruments	$l=1$	$l=3$	$l=5$	No	Instruments	$l=1$	$l=3$	$l=5$	No	Instruments	$l=1$	$l=3$	$l=5$	No	Instruments	$l=1$	$l=3$	$l=5$
1	RV-1s	70.4	29.2	17.5	45	RKbart-50t	139.0	46.3	27.7	89	RSVN-10m	79.3	31.7	19.3	133	LJV-5s	46.2	23.4	13.9
2	RV-5s	69.3	29.9	17.7	46	RKth2-1t	132.5	46.6	27.9	90	RSVN-1t	99.3	34.2	20.7	134	LJV-30s	9.4	8.3	6.0
3	RV-30s	95.7	34.1	20.5	47	RKth2-5t	139.7	46.3	27.9	91	RSVN-5t	103.6	34.9	21.4	135	LJV-1m	24.2	13.4	9.0
4	RV-1m	99.4	35.0	21.1	48	RKth2-10t	142.7	47.3	28.3	92	RSVN-10t	106.1	37.2	22.7	136	LJV-5m	16.2	12.5	8.3
5	RV-5m	96.8	34.5	21.6	49	RKth2-20t	143.5	47.7	28.6	93	RSVN-20t	122.0	42.4	26.3	137	LJV-10m	23.2	11.8	8.8
6	RV-10m	92.0	33.0	20.4	50	RKth2-50t	138.8	46.1	27.7	94	RSVN-50t	110.3	40.0	24.1	138	LJV-1t	92.5	32.1	19.2
7	RV-1t	99.9	34.5	20.8	51	RKnfp-1t	142.3	47.4	28.3	95	RSVN-5m-ss	93.6	35.2	21.2	139	LJV-5t	62.8	22.0	13.8
8	RV-5t	106.3	35.6	21.9	52	RKnfp-5t	139.7	47.0	28.0	96	RSVN-10m-ss	89.6	34.3	20.6	140	LJV-10t	71.8	27.5	17.7
9	RV-10t	110.7	38.5	23.5	53	RKnfp-10t	136.7	46.0	27.4	97	RSVP-1s	70.0	28.9	17.3	141	LJV-20t	44.1	19.6	14.4
10	RV-20t	128.3	43.6	27.3	54	RKnfp-20t	139.6	46.4	27.7	98	RSVP-5s	68.7	29.3	17.4	142	LJV-50t	90.1	30.6	19.9
11	RV-50t	117.5	40.9	24.7	55	RKnfp-50t	135.4	45.1	27.0	99	RSVP-30s	93.4	33.0	19.8	143	LSJV-1s	24.0	13.8	11.6
12	RV-5m-ss	104.8	36.4	22.0	56	RRV-1m	96.6	34.4	20.7	100	RSVP-1m	95.4	33.2	19.9	144	LSJV-5s	10.4	16.2	11.7
13	RV-10m-ss	101.4	35.4	21.3	57	RRV-5m	85.3	33.5	20.5	101	RSVP-5m	81.5	29.8	18.6	145	LSJV-30s	19.7	17.7	13.7
14	RVbr-1s	84.5	31.0	19.1	58	RRV-10m	80.5	32.6	20.1	102	RSVP-10m	69.1	26.4	16.6	146	LSJV-1m	16.5	13.7	9.4
15	RVbr-5s	81.0	29.8	18.5	59	BV-1s	80.2	30.2	18.3	103	RSVP-1t	99.7	34.6	20.8	147	LSJV-5m	22.6	14.3	10.1
16	RVbr-30s	71.5	27.4	17.5	60	BV-5s	71.6	29.2	17.8	104	RSVP-5t	106.1	35.6	22.0	148	LSJV-10m	13.4	12.3	9.7
17	RVbr-1m	76.5	29.5	18.4	61	BV-30s	97.6	34.5	20.8	105	RSVP-10t	109.6	38.2	23.4	149	LSJV-1t	40.1	17.0	11.8
18	RVbr-5m	87.7	35.1	21.9	62	BV-1m	100.5	35.1	21.1	106	RSVP-20t	125.3	42.5	26.9	150	LSJV-5t	35.4	14.6	10.1
19	RVbr-10m	61.8	27.7	17.3	63	BV-5m	95.5	34.5	21.2	107	RSVP-50t	111.9	38.6	23.5	151	LSJV-10t	38.1	20.6	13.6
20	RVbr-1t	99.4	36.4	21.7	64	BV-10m	87.6	31.3	19.3	108	RSVP-5m-ss	94.2	33.5	20.3	152	LSJV-20t	33.5	12.7	8.3
21	RVbr-5t	93.0	33.8	20.2	65	BV-1t	99.8	34.4	20.9	109	RSVP-10m-ss	82.5	30.6	18.5	153	LSJV-50t	37.3	16.5	10.9
22	RVbr-10t	93.8	34.2	21.5	66	BV-5t	106.8	35.8	22.1	110	JV-1s	0.6	0.5	0.8	154	PCF-1	102.7	35.3	21.4
23	RVbr-20t	95.0	34.1	20.6	67	BV-10t	105.3	36.9	22.4	111	JV-5s	0.7	0.5	0.7	155	PCF-2	98.5	34.1	20.6
24	RVbr-50t	92.3	33.2	20.7	68	BV-20t	129.0	43.8	27.4	112	JV-30s	0.0	2.3	2.2	156	PCF-3	67.4	24.8	15.8
25	MSRV-1s	99.6	34.6	21.2	69	BV-50t	120.6	42.1	25.6	113	JV-1m	2.9	5.7	4.2	157	ImV-C-mean	23.4	18.3	12.3
26	MSRV-5s	92.9	32.3	20.4	70	BV-5m-ss	95.5	34.5	21.2	114	JV-5m	9.1	9.0	7.2	158	ImV-C-min	84.8	29.3	17.4
27	MSRV-30s	94.1	34.2	21.7	71	BV-10m-ss	95.5	34.5	21.2	115	JV-10m	15.4	10.8	6.9	159	ImV-C-max	1.3	1.3	0.9
28	MSRV-1m	98.0	36.1	22.4	72	MedRV-1s	72.5	29.6	17.9	116	JV-1t	0.5	0.9	1.2	160	ImV-C-q1	87.5	29.2	17.8
29	MSRV-5m	83.2	33.2	20.9	73	MedRV-5s	62.9	28.4	16.9	117	JV-5t	0.6	1.2	1.3	161	ImV-C-q2	80.5	29.6	17.6
30	MSRV-10m	81.4	30.6	18.6	74	MedRV-30s	94.0	33.6	20.2	118	JV-10t	0.3	0.7	1.0	162	ImV-C-q3	25.1	18.1	12.1
31	MSRV-1t	123.9	43.2	25.9	75	MedRV-1m	97.6	34.3	20.8	119	JV-20t	0.1	3.6	2.5	163	ImV-P-mean	27.5	12.4	9.2
32	MSRV-5t	123.2	43.7	26.0	76	MedRV-5m	95.9	34.6	21.1	120	JV-50t	0.6	1.3	1.3	164	ImV-P-min	63.1	21.1	13.1
33	MSRV-10t	128.3	44.1	26.3	77	MedRV-10m	91.3	32.5	20.1	121	SJV-1s	0.9	1.3	0.8	165	ImV-P-max	0.2	1.0	1.0
34	MSRV-20t	126.0	42.8	26.3	78	MinRV-1s	74.3	29.1	17.8	122	SJV-5s	0.2	0.7	1.4	166	ImV-P-q1	72.4	27.4	16.6
35	MSRV-50t	142.3	47.3	28.9	79	MinRV-5s	62.1	26.8	16.3	123	SJV-30s	0.8	1.6	3.4	167	ImV-P-q2	71.4	25.4	15.4
36	RKcub-1t	102.8	40.2	24.4	80	MinRV-30s	93.9	33.6	20.2	124	SJV-1m	0.5	2.0	2.5	168	ImV-P-q3	44.0	15.9	10.6
37	RKcub-5t	127.7	42.7	25.6	81	MinRV-1m	97.2	34.1	20.6	125	SJV-5m	0.2	1.9	2.8	169	ImV-A-mean	35.1	17.9	12.0
38	RKcub-10t	145.2	48.2	28.9	82	MinRV-5m	92.1	34.2	20.9	126	SJV-10m	0.4	1.8	2.0	170	ImV-A-min	68.8	22.7	13.7
39	RKcub-20t	136.4	45.5	27.2	83	MinRV-10m	79.7	29.2	18.0	127	SJV-1t	0.7	11.6	7.1	171	ImV-A-max	1.1	1.6	1.1
40	RKcub-50t	134.3	44.8	26.8	84	RSVN-1s	70.5	29.5	17.6	128	SJV-5t	0.0	1.4	1.3	172	ImV-A-q1	83.8	31.3	19.1
41	RKbart-1t	133.9	45.2	27.0	85	RSVN-5s	69.3	30.3	18.0	129	SJV-10t	0.4	0.7	0.9	173	ImV-A-q2	82.3	28.0	17.0
42	RKbart-5t	139.9	46.3	27.9	86	RSVN-30s	92.9	34.2	20.5	130	SJV-20t	0.0	0.7	0.7	174	ImV-A-q3	51.7	21.8	13.5
43	RKbart-10t	141.9	47.0	28.2	87	RSVN-1m	95.0	35.2	21.2	131	SJV-50t	0.0	0.5	0.6	175	1-day	9.2	10.1	8.6
44	RKbart-20t	143.8	47.8	28.6	88	RSVN-5m	88.1	35.1	21.6	132	LJV-1s	56.7	26.6	15.9	$CV_{Size,0.10}$		16.4	22.3	26.9

Notes:

1. The critical value (CV) is a function of one endogenous regressor, the number of instrumental variables, and the desired 10% maximal size of a 5% Wald test of $\phi = \phi_0$, for further details, see Table 5.2 of Stock and Yogo (2005).
2. We use logarithms of RV-RSVP and PCF classes of instruments given in Table 4.11.
3. Instruments are deemed weak if the first-stage F-statistic is less than the CV associated with the corresponding column.

Table 4.14. Projection-based 90% confidence intervals for the volatility persistence parameter ϕ (Strong instruments)
Ticker: IBM, January 2009 - December 2013, $T = 1258$

Panel A

No	Instruments	$\bar{d}_{i,s}$	AR	AR^*	SS	SS^*
1	RSVN-5m-ss	0.8860	[0.950, 1.000]	[0.866, 1.000]	[0.932, 1.000]	[0.796, 1.000]
2	RSVN-5m	0.8855	[0.948, 1.000]	[0.864, 1.000]	[0.931, 1.000]	[0.799, 1.000]
3	RSVN-1m	0.8848	[0.947, 1.000]	[0.856, 1.000]	[0.929, 1.000]	[0.807, 1.000]
4	ImV-C-mean	0.8830	[0.964, 1.000]	[0.852, 1.000]	[0.937, 1.000]	[0.779, 1.000]
5	MinRV-5m	0.8828	[0.945, 1.000]	[0.867, 1.000]	[0.925, 1.000]	[0.794, 1.000]
6	RV-5m-ss	0.8825	[0.946, 1.000]	[0.863, 1.000]	[0.926, 1.000]	[0.795, 1.000]
7	BV-5m	0.8823	[0.945, 1.000]	[0.865, 1.000]	[0.926, 1.000]	[0.793, 1.000]
8	BV-5m-ss	0.8823	[0.945, 1.000]	[0.865, 1.000]	[0.926, 1.000]	[0.793, 1.000]
9	BV-10m-ss	0.8823	[0.945, 1.000]	[0.865, 1.000]	[0.926, 1.000]	[0.793, 1.000]
10	MedRV-5m	0.8823	[0.945, 1.000]	[0.866, 1.000]	[0.925, 1.000]	[0.793, 1.000]

Panel B

No	Instruments	$\bar{d}_{i,s}$	AR	AR^*	SS	SS^*
11	ImV-C-q3	0.8805	[0.964, 1.000]	[0.843, 1.000]	[0.940, 1.000]	[0.775, 1.000]
12	RV-1m	0.8800	[0.944, 1.000]	[0.857, 1.000]	[0.925, 1.000]	[0.794, 1.000]
13	ImV-C-q2	0.8795	[0.958, 1.000]	[0.860, 1.000]	[0.940, 1.000]	[0.760, 1.000]
14	RRV-1m	0.8790	[0.945, 1.000]	[0.858, 1.000]	[0.926, 1.000]	[0.787, 1.000]
15	MedRV-1m	0.8785	[0.944, 1.000]	[0.857, 1.000]	[0.926, 1.000]	[0.787, 1.000]
16	RV-5m	0.8783	[0.943, 1.000]	[0.858, 1.000]	[0.923, 1.000]	[0.789, 1.000]
17	BV-1m	0.8775	[0.944, 1.000]	[0.857, 1.000]	[0.925, 1.000]	[0.784, 1.000]
18	RSVN-10m-ss	0.8775	[0.949, 1.000]	[0.858, 1.000]	[0.931, 1.000]	[0.772, 1.000]
19	RSVN-10m	0.8760	[0.946, 1.000]	[0.861, 1.000]	[0.927, 1.000]	[0.770, 1.000]
20	RV-10m-ss	0.8758	[0.944, 1.000]	[0.857, 1.000]	[0.924, 1.000]	[0.778, 1.000]

Panel C

No	Instruments	$\bar{d}_{i,s}$	AR	AR^*	SS	SS^*
21	RSVN-30s	0.8753	[0.944, 1.000]	[0.848, 1.000]	[0.924, 1.000]	[0.785, 1.000]
22	RRV-5m	0.8750	[0.946, 1.000]	[0.855, 1.000]	[0.927, 1.000]	[0.772, 1.000]
23	MinRV-1m	0.8745	[0.943, 1.000]	[0.855, 1.000]	[0.924, 1.000]	[0.776, 1.000]
24	ImV-C-min	0.8743	[0.952, 1.000]	[0.834, 1.000]	[0.930, 1.000]	[0.781, 1.000]
25	MSRV-1m	0.8723	[0.939, 1.000]	[0.862, 1.000]	[0.920, 1.000]	[0.768, 1.000]
26	RSVP-1m	0.8715	[0.942, 1.000]	[0.852, 1.000]	[0.921, 1.000]	[0.771, 1.000]
27	RV-30s	0.8713	[0.942, 1.000]	[0.847, 1.000]	[0.922, 1.000]	[0.774, 1.000]
28	BV-30s	0.8713	[0.943, 1.000]	[0.847, 1.000]	[0.923, 1.000]	[0.772, 1.000]
29	ImV-C-q1	0.8710	[0.952, 1.000]	[0.840, 1.000]	[0.931, 1.000]	[0.761, 1.000]
30	MSRV-30s	0.8698	[0.935, 1.000]	[0.858, 1.000]	[0.917, 1.000]	[0.769, 1.000]

Notes:

1. The instrument set consists of a constant and a lag of an instrument, $l = 1$.
2. We use logarithms of RV-RSVP and PCF classes of instruments given in Table 4.11.
3. The inference procedures [AR, AR^*, SS, SS^*] are proposed in Sections 4.3.1-4.3.4 and corresponding test statistics are given in equations (4.3.22), (4.3.25), (4.3.31) and (4.3.35).
4. The confidence intervals are constructed by projection technique described in Section 4.3.5. The corresponding 95% confidence interval for the nuisance parameter λ is [33.943, 61.154] with $\hat{\lambda} = 47.548$ and $\mathrm{SE}(\hat{\lambda}) = 6.935$.
5. We use $\tau = 0.2$ for SS-type tests and employ 99 Monte Carlo replications for point-optimal type procedures.
6. The average precision of an instrument set i over the proposed inference methods is measured by $\bar{d}_{i,s} := S^{-1}\sum_{i=1}^{S} d_i$, where $s \in S$ and S is the set of identification-robust inference methods.

Table 4.15. Projection-based 90% confidence intervals for the volatility persistence parameter ϕ (Weak instruments)
Ticker: IBM, January 2009 - December 2013, $T = 1258$

No	Instruments	$\bar{d}_{i,s}$	AR	AR^*	SS	SS^*
Panel A						
1	JV-1s	0.0000	[0.000, 1.000]	[0.000, 1.000]	[0.000, 1.000]	[0.000, 1.000]
2	JV-5s	0.0000	[0.000, 1.000]	[0.000, 1.000]	[0.000, 1.000]	[0.000, 1.000]
3	JV-30s	0.0000	[0.000, 1.000]	[0.000, 1.000]	[0.000, 1.000]	[0.000, 1.000]
4	JV-1t	0.0000	[0.000, 1.000]	[0.000, 1.000]	[0.000, 1.000]	[0.000, 1.000]
5	SJV-1s	0.0000	[0.000, 1.000]	[0.000, 1.000]	[0.000, 1.000]	[0.000, 1.000]
6	SJV-5s	0.0000	[0.000, 1.000]	[0.000, 1.000]	[0.000, 1.000]	[0.000, 1.000]
7	SJV-10t	0.0000	[0.000, 1.000]	[0.000, 1.000]	[0.000, 1.000]	[0.000, 1.000]
8	SJV-20t	0.0000	[0.000, 1.000]	[0.000, 1.000]	[0.000, 1.000]	[0.000, 1.000]
9	SJV-50t	0.0000	[0.000, 1.000]	[0.000, 1.000]	[0.000, 1.000]	[0.000, 1.000]
10	ImV-C-max	0.0000	[0.000, 1.000]	[0.000, 1.000]	[0.000, 1.000]	[0.000, 1.000]
11	ImV-P-max	0.0000	[0.000, 1.000]	[0.000, 1.000]	[0.000, 1.000]	[0.000, 1.000]
12	ImV-A-max	0.0000	[0.000, 1.000]	[0.000, 1.000]	[0.000, 1.000]	[0.000, 1.000]
Panel B						
No	Instruments	$\bar{d}_{i,s}$	AR	AR^*	SS	SS^*
13	JV-20t	0.0038	[0.000, 1.000]	[0.000, 1.000]	[0.000, 0.985]	[0.000, 1.000]
14	SJV-5t	0.1875	[0.500, 1.000]	[0.250, 1.000]	[0.000, 1.000]	[0.000, 1.000]
Panel C						
No	Instruments	$\bar{d}_{i,s}$	AR	AR^*	SS	SS^*
15	JV-10t	0.3130	[0.000, 1.000]	[0.000, 1.000]	[0.745, 0.993]	[0.500, 1.000]
16	JV-50t	0.3283	[0.000, 1.000]	[0.000, 1.000]	[0.763, 1.000]	[0.550, 1.000]
17	JV-5t	0.3308	[0.000, 1.000]	[0.000, 1.000]	[0.753, 1.000]	[0.570, 1.000]
18	SJV-30s	0.3400	[0.000, 1.000]	[0.000, 1.000]	[0.860, 1.000]	[0.500, 1.000]
19	SJV-1m	0.3845	[0.000, 1.000]	[0.000, 1.000]	[0.898, 1.000]	[0.640, 1.000]
20	SJV-10m	0.3975	[0.000, 1.000]	[0.000, 1.000]	[0.930, 1.000]	[0.660, 1.000]
21	SJV-5m	0.3998	[0.000, 1.000]	[0.000, 1.000]	[0.919, 1.000]	[0.680, 1.000]
22	JV-1m	0.4028	[0.911, 1.000]	[0.700, 1.000]	[0.000, 1.000]	[0.000, 1.000]
23	JV-10m	0.4075	[0.890, 1.000]	[0.740, 1.000]	[0.000, 1.000]	[0.000, 1.000]
24	LSJV-20t	0.4108	[0.883, 0.992]	[0.750, 1.000]	[0.000, 0.998]	[0.000, 1.000]
25	LSJV-5t	0.4208	[0.913, 1.000]	[0.770, 1.000]	[0.000, 1.000]	[0.000, 1.000]
26	1-day	0.4255	[0.870, 0.965]	[0.750, 1.000]	[0.000, 0.953]	[0.000, 1.000]
27	LJV-30s	0.4268	[0.927, 1.000]	[0.780, 1.000]	[0.000, 1.000]	[0.000, 1.000]
28	LJV-1m	0.4373	[0.933, 1.000]	[0.816, 1.000]	[0.000, 1.000]	[0.000, 1.000]
29	SJV-1t	0.6795	[0.810, 0.992]	[0.700, 1.000]	[0.700, 1.000]	[0.500, 1.000]
30	LJV-10m	0.7465	[0.905, 0.998]	[0.750, 1.000]	[0.829, 1.000]	[0.500, 1.000]

Notes:

1. The instrument set consists of a constant and a lag of an instrument, $l = 1$.
2. We use logarithms of RV-RSVP and PCF classes of instruments given in Table 4.11.
3. The inference procedures [AR, AR^*, SS, SS^*] are proposed in Sections 4.3.1-4.3.4 and corresponding test statistics are given in equations (4.3.22), (4.3.25), (4.3.31) and (4.3.35).
4. The confidence intervals are constructed by projection technique described in Section 4.3.5. The corresponding 95% confidence interval for the nuisance parameter λ is [33.943, 61.154] with $\hat{\lambda} = 47.548$ and $\mathrm{SE}(\hat{\lambda}) = 6.935$.
5. We use $\tau = 0.2$ for SS-type tests and employ 99 Monte Carlo replications for point-optimal type procedures.
6. The average precision of an instrument set i over the proposed inference methods is measured by $\bar{d}_{i,s} := S^{-1}\sum_{i=1}^{S} d_i$, where $s \in S$ and S is the set of identification-robust inference methods.

Table 4.16. Projection-based 90% confidence intervals for the volatility persistence parameter ϕ
Strong instruments (Several lags)
Ticker: IBM, January 2009 - December 2013, $T = 1258$

	$l=1$					$l=3$					$l=5$				
Instruments	$\bar{d}_{i,s}$	AR	AR^*	SS	SS^*	$\bar{d}_{i,s}$	AR	AR^*	SS	SS^*	$\bar{d}_{i,s}$	AR	AR^*	SS	SS^*
RSVN-5m-ss	0.8860	[0.950, 1.0]	[0.866, 1.0]	[0.932, 1.0]	[0.796, 1.0]	0.8618	[0.949, 1.0]	[0.838, 1.0]	[0.895, 1.0]	[0.765, 1.0]	0.8603	[0.941, 1.0]	[0.845, 1.0]	[0.874, 1.0]	[0.781, 1.0]
RSVN-5m	0.8855	[0.948, 1.0]	[0.864, 1.0]	[0.931, 1.0]	[0.799, 1.0]	0.8668	[0.955, 1.0]	[0.845, 1.0]	[0.892, 1.0]	[0.775, 1.0]	0.8695	[0.960, 1.0]	[0.856, 1.0]	[0.871, 1.0]	[0.791, 1.0]
RSVN-1m	0.8848	[0.947, 1.0]	[0.856, 1.0]	[0.929, 1.0]	[0.807, 1.0]	0.8570	[0.951, 1.0]	[0.851, 1.0]	[0.887, 1.0]	[0.739, 1.0]	0.8555	[0.946, 1.0]	[0.845, 1.0]	[0.861, 1.0]	[0.770, 1.0]
ImV-C-mean	0.8830	[0.964, 1.0]	[0.852, 1.0]	[0.937, 1.0]	[0.779, 1.0]	0.8218	[0.970, 1.0]	[0.828, 1.0]	[0.849, 1.0]	[0.640, 1.0]	0.8078	[0.950, 1.0]	[0.813, 1.0]	[0.757, 1.0]	[0.711, 1.0]
MinRV-5m	0.8828	[0.945, 1.0]	[0.867, 1.0]	[0.925, 1.0]	[0.794, 1.0]	0.8493	[0.935, 1.0]	[0.837, 1.0]	[0.886, 1.0]	[0.739, 1.0]	0.8465	[0.928, 1.0]	[0.825, 1.0]	[0.856, 1.0]	[0.777, 1.0]
RV-5m-ss	0.8825	[0.946, 1.0]	[0.863, 1.0]	[0.926, 1.0]	[0.795, 1.0]	0.8560	[0.939, 1.0]	[0.837, 1.0]	[0.889, 1.0]	[0.759, 1.0]	0.8523	[0.932, 1.0]	[0.840, 1.0]	[0.866, 1.0]	[0.771, 1.0]
BV-5m	0.8823	[0.945, 1.0]	[0.865, 1.0]	[0.926, 1.0]	[0.793, 1.0]	0.8508	[0.935, 1.0]	[0.835, 1.0]	[0.887, 1.0]	[0.746, 1.0]	0.8500	[0.932, 1.0]	[0.827, 1.0]	[0.861, 1.0]	[0.780, 1.0]
BV-5m-ss	0.8823	[0.945, 1.0]	[0.865, 1.0]	[0.926, 1.0]	[0.793, 1.0]	0.8508	[0.935, 1.0]	[0.835, 1.0]	[0.887, 1.0]	[0.746, 1.0]	0.8500	[0.932, 1.0]	[0.827, 1.0]	[0.861, 1.0]	[0.780, 1.0]
BV-10m-ss	0.8823	[0.945, 1.0]	[0.865, 1.0]	[0.926, 1.0]	[0.793, 1.0]	0.8508	[0.935, 1.0]	[0.835, 1.0]	[0.887, 1.0]	[0.746, 1.0]	0.8500	[0.932, 1.0]	[0.827, 1.0]	[0.861, 1.0]	[0.780, 1.0]
MedRV-5m	0.8823	[0.945, 1.0]	[0.866, 1.0]	[0.925, 1.0]	[0.793, 1.0]	0.8493	[0.936, 1.0]	[0.833, 1.0]	[0.886, 1.0]	[0.742, 1.0]	0.8515	[0.930, 1.0]	[0.826, 1.0]	[0.859, 1.0]	[0.791, 1.0]
1-day	0.4255	[0.870, 0.965]	[0.750, 1.0]	[0.000, 0.953]	[0.000, 1.0]	0.7490	[0.855, 0.974]	[0.760, 1.0]	[0.782, 0.977]	[0.550, 1.0]	0.7485	[0.838, 0.979]	[0.770, 1.0]	[0.741, 0.986]	[0.610, 1.0]

Notes:

1. The instrument set consists of a constant and lags of an instrument, $l = 1, 3, 5$.
2. We use logarithms of RV-RSVP and PCF classes of instruments given in Table 4.11.
3. The inference procedures [AR, AR^*, SS, SS^*] are proposed in Sections 4.3.1-4.3.4 and corresponding test statistics are given in equations (4.3.22), (4.3.25), (4.3.31) and (4.3.35).
4. The confidence intervals are constructed by projection technique described in Section 4.3.5. The corresponding 95% confidence interval for the nuisance parameter λ is [33.943, 61.154] with $\hat{\lambda} = 47.548$ and $\mathrm{SE}(\hat{\lambda}) = 6.935$.
5. We use $\tau = 0.2$ for SS-type tests and employ 99 Monte Carlo replications for point-optimal type procedures.
6. The average precision of an instrument set i over the proposed inference methods is measured by $\bar{d}_{i,s} := S^{-1}\sum_{i=1}^{S} d_i$, where $s \in S$ and S is the set of identification-robust inference methods.

Table 4.17. Projection-based 90% confidence intervals for the volatility persistence parameter ϕ
Different combinations of strong instruments
Ticker: IBM, January 2009 - December 2013, $T = 1258$

Instrument set	$\bar{d}_{i,s}$	# of Instruments	AR	AR^*	SS	SS^*
RV-5m-ss, ImV-C-q3	0.8748	2	[0.954, 1.000]	[0.848, 1.000]	[0.910, 1.000]	[0.787, 1.000]
BV-5m-ss, ImV-C-q3	0.8775	2	[0.954, 1.000]	[0.850, 1.000]	[0.912, 1.000]	[0.794, 1.000]
RSVN-5m, ImV-C-q3	0.8820	2	[0.953, 1.000]	[0.857, 1.000]	[0.916, 1.000]	[0.802, 1.000]
RKcub-10t, ImV-C-q3	0.8583	2	[0.958, 0.995]	[0.854, 1.000]	[0.892, 1.000]	[0.724, 1.000]
BV-5m-ss, LJV-5s	0.8650	2	[0.936, 1.000]	[0.841, 1.000]	[0.908, 1.000]	[0.775, 1.000]
RKcub-10t, PCF-1, ImV-C-q3	0.8555	3	[0.967, 0.991]	[0.842, 1.000]	[0.878, 1.000]	[0.726, 1.000]
BV-5m-ss, LJV-5s, ImV-C-q3	0.8645	3	[0.946, 1.000]	[0.838, 1.000]	[0.898, 1.000]	[0.776, 1.000]
BV-5m-ss, LJV-5s, PCF-1	0.8568	3	[0.960, 0.999]	[0.844, 1.000]	[0.888, 1.000]	[0.734, 1.000]
BV-5m-ss, LJV-5s, PCF-1, ImV-C-q3	0.8553	4	[0.966, 0.996]	[0.829, 1.000]	[0.881, 1.000]	[0.741, 1.000]
BV-5m-ss, LJV-5s, LSJV-10t, PCF-1, ImV-C-q3	0.8493	5	[0.959, 0.997]	[0.820, 1.000]	[0.872, 1.000]	[0.743, 1.000]

Notes:

1. The instrument set includes a constant and a lag of each instrument given in column 1.
2. We use logarithms of RV-RSVP and PCF classes of instruments given in Table 4.11.
3. The inference procedures [AR, AR^*, SS, SS^*] are proposed in Sections 4.3.1-4.3.4 and corresponding test statistics are given in equations (4.3.22), (4.3.25), (4.3.31) and (4.3.35).
4. The confidence intervals are constructed by projection technique described in Section 4.3.5. The corresponding 95% confidence interval for the nuisance parameter λ is [33.943, 61.154] with $\hat{\lambda} = 47.548$ and $\mathrm{SE}(\hat{\lambda}) = 6.935$.
5. We use $\tau = 0.2$ for SS-type tests and employ 99 Monte Carlo replications for point-optimal type procedures.
6. The average precision of an instrument set i over the proposed inference methods is measured by $\bar{d}_{i,s} := S^{-1} \sum_{i=1}^{S} d_i$, where $s \in S$ and S is the set of identification-robust inference methods.

Chapter 5

Conclusion and future work

This thesis studied and contributed to the SV literature from the estimation, inference, and volatility forecasting viewpoints. On the whole, the thesis developed easily applicable statistical methods for stochastic volatility models. We conclude our study with suggestions for future work.

The simple estimation methods, which are developed in chapters 2-3, can be extended for several SV specifications: SV with conditional heavy-tailed distributions (especially Student's t-distribution and the generalized error distribution), multivariate SV models (higher-order, cross leverage, non-Gaussian distributions), asymmetric SV specification (which allows for leverage effects), and multi-factor SV models. These are the objects of ongoing research.

The inference methods proposed in chapter 4 can be adapted to other situations, *e.g.*, measurement error in ARMA-type models, or noisy realized measures in HAR volatility modeling. The extension to multivariate models and parameter estimation in GSV framework are topics of ongoing research.

Bibliography

Abramowitz, M. and Stegun, I. A. (1970), *Handbook of Mathematical Functions*, Vol. 55, Dover Publications Inc., New York.

Ahsan, M. N. and Dufour, J.-M. (2018), Simple estimators for higher-order stochastic volatility models and forecasting, Technical report, McGill University.

Ahsan, M. N. and Dufour, J.-M. (2019), A simple efficient moment-based estimator for the stochastic volatility model, *in* I. Jeliazkov and J. Tobias, eds, 'Topics in Identification, Limited Dependent Variables, Partial Observability, Experimentation, and Flexible Modeling', Vol. 40 of *Advances in Econometrics*, Emerald, Bingley, U.K., pp. 157–201.

Ahsanullah, M. and Wei, W. W. S. (1984), Effects of temporal aggregation on forecasts in an ARMA(1,1) process, *in* 'American Statistical Association, Proceedings of the Business and Economics Statistics Section', pp. 297–302.

Aït-Sahalia, Y., Mykland, P. A. and Zhang, L. (2005), 'How often to sample a continuous-time process in the presence of market microstructure noise', *The Review of Financial studies* **18**(2), 351–416.

Aït-Sahalia, Y., Mykland, P. A. and Zhang, L. (2011), 'Ultra high frequency volatility estimation with dependent microstructure noise', *Journal of Econometrics* **160**(1), 160–175.

Alizadeh, S., Brandt, M. W. and Diebold, F. X. (2002), 'Range-based estimation of stochastic volatility models', *The Journal of Finance* **57**(3), 1047–1091.

Andersen, T. G. and Bollerslev, T. (1998), 'Answering the skeptics: Yes, standard volatility models do provide accurate forecasts', *International Economic Review* pp. 885–905.

Andersen, T. G., Bollerslev, T., Diebold, F. X. and Ebens, H. (2001), 'The distribution of realized stock return volatility', *Journal of Financial Economics* **61**(1), 43–76.

Andersen, T. G., Bollerslev, T., Diebold, F. X. and Labys, P. (2001), 'The distribution of realized exchange rate volatility', *Journal of the American statistical association* **96**(453), 42–55.

Andersen, T. G., Chung, H.-J. and Sørensen, B. E. (1999), 'Efficient method of moments estimation of a stochastic volatility model: A Monte Carlo study', *Journal of Econometrics* **91**, 61–87.

Andersen, T. G., Dobrev, D. and Schaumburg, E. (2012), 'Jump-robust volatility estimation using nearest neighbor truncation', *Journal of Econometrics* **169**(1), 75–93.

Andersen, T. G. and Sørensen, B. E. (1996), 'GMM estimation of a stochastic volatility model: A Monte Carlo study', *Journal of Business and Economic Statistics* **14**(3), 328–352.

Andrews, D. W. (1991), 'Heteroskedasticity and autocorrelation consistent covariance matrix estimation', *Econometrica* **59**(3), 817–858.

Andrews, D. W., Moreira, M. J. and Stock, J. H. (2006), 'Optimal two-sided invariant similar tests for instrumental variables regression', *Econometrica* **74**(3), 715–752.

Asai, M. (2008), 'Autoregressive stochastic volatility models with heavy-tailed distributions: A comparison with multifactor volatility models', *Journal of Empirical Finance* **15**(2), 332–341.

Baillie, R. T. and Bollerslev, T. (1992), 'Prediction in dynamic models with time-dependent conditional variances', *Journal of Econometrics* **52**(1-2), 91–113.

Bandi, F. M. and Russell, J. R. (2006), 'Separating microstructure noise from volatility', *Journal of Financial Economics* **79**(3), 655–692.

Bandi, F. M. and Russell, J. R. (2008), 'Microstructure noise, realized variance, and optimal sampling', *The Review of Economic Studies* **75**(2), 339–369.

Barnard, G. A. (1963), 'Comment on 'The spectral analysis of point processes' by M. S. Bartlett', *Journal of the Royal Statistical Society, Series B* **25**, 294.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A. and Shephard, N. (2008), 'Designing realized kernels to measure the ex post variation of equity prices in the presence of noise', *Econometrica* **76**(6), 1481–1536.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A. and Shephard, N. (2011), 'Multivariate realised kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading', *Journal of Econometrics* **162**(2), 149–169.

Barndorff-Nielsen, O. E., Kinnebrock, S. and Shephard, N. (2010), Measuring downside riskrealised semivariance, *in* T. Bollerslev, J. Russell and M. Watson, eds, 'Volatility and Time Series Econometrics: Essays in Honour of Robert F. Engle', Oxford University Press, Oxford, U.K., pp. 117–136.

Barndorff-Nielsen, O. E., Nicolato, E. and Shephard, N. (2002), 'Some recent developments in stochastic volatility modelling', *Quantitative Finance* **2**(1), 11–23.

Barndorff-Nielsen, O. E. and Shephard, N. (2001), 'Non-Gaussian Ornstein–Uhlenbeck-based models and some of their uses in financial economics', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **63**(2), 167–241.

Barndorff-Nielsen, O. E. and Shephard, N. (2004), 'Power and bipower variation with stochastic volatility and jumps', *Journal of Financial Econometrics* **2**(1), 1–37.

Barndorff-Nielsen, O. E. and Shephard, N. (2006), 'Econometrics of testing for jumps in financial economics using bipower variation', *Journal of Financial Econometrics* **4**(1), 1–30.

Bates, D. M. and Watts, D. G. (1988), *Nonlinear Regression Analysis and its Applications*, John Wiley & Sons, New York.

Bekker, P. (1994), 'Alternative approximations to the distributions of instrumental variable estimators', *Econometrica* **62**, 657–681.

Benati, L. (2008), 'The great moderation in the United Kingdom', *Journal of Money, Credit and Banking* **40**(1), 121–147.

Blanchard, O. and Simon, J. (2001), 'The long and large decline in US output volatility', *Brookings Papers on Economic Activity* **2001**(1), 135–174.

Bollerslev, T. (1986), 'Generalized autoregressive conditional heteroskedasticity', *Journal of Econometrics* **31**, 307–327.

Bollerslev, T. (1987), 'A conditionally heteroskedastic time series model for speculative prices and rates of return', *The Review of Economics and Statistics* **69**(3), 542–547.

Bollerslev, T. (2010), Glossary to ARCH (GARCH), *in* T. Bollerslev, J. Russell and M. Watson, eds, 'Volatility and Time Series Econometrics: Essays in Honour of Robert F. Engle', Oxford University Press, Oxford, U.K.

Bollerslev, T. and Zhou, H. (2002), 'Estimating stochastic volatility diffusion using conditional moments of integrated volatility', *Journal of Econometrics* **109**(1), 33–65.

Bos, C. S. (2012), Relating stochastic volatility estimation methods, *in* L. Bauwens, C. Hafner and S. Laurent, eds, 'Handbook of Volatility Models and Their Applications', John Wiley & Sons, Hoboken, NJ, chapter 6, pp. 147–174.

Bougerol, P. and Picard, N. (1992), 'Stationarity of GARCH processes and of some nonnegative time series', *Journal of Econometrics* **52**(1), 115–127.

Breidt, F. J. and Carriquiry, A. L. (1996), Improved quasi-maximum likelihood estimation for stochastic volatility models, *in* W. O. Johnson, J. C. Lee and A. Zellner, eds, 'Modelling and Prediction: Honouring Seymour Geisel', Springer-Verlag, New York, pp. 228–247.

Broto, C. and Ruiz, E. (2004), 'Estimation methods for stochastic volatility models: a survey', *Journal of Economic Surveys* **18**(5), 613–649.

Buse, A. (1992), 'The bias of instrumental variables estimators', *Econometrica* **60**, 173–180.

Busetti, F. and Taylor, A. (2003), 'Testing against stochastic trend and seasonality in the presence of unattended breaks and unit roots', *Journal of Econometrics* **117**(1), 21–53.

Carnero, M. A., Peña, D. and Ruiz, E. (2004), 'Persistence and kurtosis in GARCH and stochastic volatility models', *Journal of Financial Econometrics* **2**(2), 319–342.

Carrasco, M. and Chen, X. (2002), 'Mixing and moment properties of various GARCH and stochastic volatility models', *Econometric Theory* **18**(1), 17–39.

Cavaliere, G. and Taylor, A. R. (2007), 'Testing for unit roots in time series models with non-stationary volatility', *Journal of Econometrics* **140**(2), 919–947.

Cavanagh, C. L., Elliott, G. and Stock, J. H. (1995), 'Inference in models with nearly integrated regressors', *Econometric Theory* **11**(5), 1131–1147.

Chan, J. C. and Grant, A. L. (2016), 'Modeling energy price dynamics: GARCH versus stochastic volatility', *Energy Economics* **54**, 182–189.

Chao, J. and Swanson, N. R. (2007), 'Alternative approximations of the bias and MSE of the IV estimator under weak identification with an application to bias correction', *Journal of Econometrics* **137**(2), 515–555.

Chaudhuri, S., Richardson, T., Robins, J. and Zivot, E. (2010), 'A new projection-type split-sample score test in linear instrumental variables regression', *Econometric Theory* **26**(6), 1820–1837.

Chaudhuri, S. and Zivot, E. (2011), 'A new method of projection-based inference in GMM with weakly identified nuisance parameters', *Journal of Econometrics* **164**(2), 239–251.

Chen, J.-S. and Jennrich, R. I. (1996), 'The signed root deviance profile and confidence intervals in maximum likelihood analysis', *Journal of the American Statistical Association* **91**, 993–999.

Chernov, M., Gallant, A. R., Ghysels, E. and Tauchen, G. (2003), 'Alternative models for stock price dynamics', *Journal of Econometrics* **116**(1), 225–257.

Chib, S., Nardari, F. and Shephard, N. (2002), 'Markov chain Monte Carlo methods for stochastic volatility models', *Journal of Econometrics* **108**, 281–316.

Christensen, B. J. and Prabhala, N. R. (1998), 'The relation between implied and realized volatility', *Journal of Financial Economics* **50**(2), 125–150.

Christensen, K. and Podolskij, M. (2007), 'Realized range-based estimation of integrated variance', *Journal of Econometrics* **141**(2), 323–349.

Cogley, T. and Sargent, T. J. (2005), 'Drifts and volatilities: monetary policies and outcomes in the post WWII US', *Review of Economic Dynamics* **8**(2), 262–302.

Corsi, F. (2009), 'A simple approximate long-memory model of realized volatility', *Journal of Financial Econometrics* **7**(2), 174–196.

Danielsson, J. (1994), 'Stochastic volatility in asset prices estimation with simulated maximum likelihood', *Journal of Econometrics* **61**, 375–400.

Danielsson, J. and Richard, J.-F. (1993), 'Accelerated Gaussian importance sampler with application to dynamic latent variable models', *Journal of Applied Econometrics* **8**, S153–S173.

Davidson, J. (1994), *Stochastic Limit Theory: An Introduction for Econometricians*, first edn, Oxford University Press, Oxford, U.K.

Davis, R. A. and Mikosch, T. (2009), Probabilistic properties of stochastic volatility models, *in* T. Mikosch, J.-P. Kreiß, A. R. Davis and G. T. Andersen, eds, 'Handbook of Financial Time Series', Springer-Verlag, Berlin, Heidelberg, pp. 255–267.

Den Haan, W. J. and Levin, A. (1997), A practitioner's guide to robust covariance matrix estimation, *in* G. S. Maddala and C. R. Rao, eds, 'Handbook of Statistics: Robust Inference', Vol. 15, Elsevier, pp. 299–342.

Diebold, F. X. and Mariano, R. S. (2002), 'Comparing predictive accuracy', *Journal of Business and Economic Statistics* **20**(1), 134–144.

Dufour, J.-M. (1989), 'Nonlinear hypotheses, inequality restrictions, and non-nested hypotheses: Exact simultaneous tests in linear regressions', *Econometrica* **57**, 335–355.

Dufour, J.-M. (1990), 'Exact tests and confidence sets in linear regressions with autocorrelated errors', *Econometrica* **58**, 475–494.

Dufour, J.-M. (1997), 'Some impossibility theorems in econometrics, with applications to structural and dynamic models', *Econometrica* **65**, 1365–1389.

Dufour, J.-M. (2003), 'Identification, weak instruments and statistical inference in econometrics', *Canadian Journal of Economics* **36**(4), 767–808.

Dufour, J.-M. (2006), 'Monte Carlo tests with nuisance parameters: A general approach to finite-sample inference and nonstandard asymptotics in econometrics', *Journal of Econometrics* **133**(2), 443–477.

Dufour, J.-M. and Jasiak, J. (2001), 'Finite sample limited information inference methods for structural equations and models with generated regressors', *International Economic Review* **42**, 815–843.

Dufour, J.-M. and Jouini, T. (2006), 'Finite-sample simulation-based tests in VAR models with applications to Granger causality testing', *Journal of Econometrics* **135**(1-2), 229–254.

Dufour, J.-M. and Jouini, T. (2014), 'Asymptotic distributions for quasi-efficient estimators in echelon VARMA models', *Computational Statistics and Data Analysis* **73**, 69 – 86.

Dufour, J.-M. and Khalaf, L. (2001), Monte Carlo test methods in econometrics, *in* B. Baltagi, ed., 'A Companion to Theoretical Econometrics', Blackwell Companions to Contemporary Economics, Basil Blackwell, Oxford, U.K., chapter 23, pp. 494–519.

Dufour, J.-M. and Khalaf, L. (2002), 'Simulation based finite and large sample tests in multivariate regressions', *Journal of Econometrics* **111**(2), 303–322.

Dufour, J.-M. and King, M. L. (1991), 'Optimal invariant tests for the autocorrelation coefficient in linear regressions with stationary or nonstationary AR(1) errors', *Journal of Econometrics* **47**, 115–143.

Dufour, J.-M. and Neves, J. (2018), Maximized Monte Carlo: an R package, Technical report, McGill University.

Dufour, J.-M. and Pelletier, D. (2005), 'Practical methods for modelling weak VARMA processes: identification, estimation and specification with a macroeconomic application', *Preprint* .

Dufour, J.-M. and Taamouti, M. (2003), Point-optimal instruments and generalized Anderson-Rubin procedures for nonlinear models, Technical report, C.R.D.E., Université de Montréal.

Dufour, J.-M. and Taamouti, M. (2005), 'Projection-based statistical inference in linear structural models with possibly weak instruments', *Econometrica* **73**(4), 1351–1365.

Dufour, J.-M. and Taamouti, M. (2007), 'Further results on projection-based inference in IV regressions with weak, collinear or missing instruments', *Journal of Econometrics* **139**(1), 133–153.

Dufour, J.-M., Trognon, A. and Tuvaandorj, P. (2017), 'Invariant tests based on M-estimators, estimating functions, and the generalized method of moments', *Econometric Reviews* **36**(1-3), 182–204.

Dufour, J.-M. and Valéry, P. (2006), 'On a simple two-stage closed-form estimator for a stochastic volatility in a general linear regression', *Advances in Econometrics* **20**, 259–288.

Dufour, J.-M. and Valéry, P. (2009), 'Exact and asymptotic tests for possibly non-regular hypotheses on stochastic volatility models', *Journal of Econometrics* **150**, 193–206.

Durbin, J. (1960), 'The fitting of time series models', *Revue de l'Institut International de Statistique* **28**, 233–244.

Durham, G. B. (2006), 'Monte Carlo methods for estimating, smoothing, and filtering one- and two-factor stochastic volatility models', *Journal of Econometrics* **133**, 273–305.

Durham, G. B. (2007), 'SV mixture models with application to S&P 500 index returns', *Journal of Financial Economics* **85**(3), 822–856.

Dwass, M. (1957), 'Modified randomization tests for nonparametric hypotheses', *Annals of Mathematical Statistics* **28**, 181–187.

Eberhart, R. and Kennedy, J. (1995), A new optimizer using particle swarm theory, *in* 'Proceedings of the Sixth International Symposium on Micro Machine and Human Science', IEEE, pp. 39–43.

Engle, R. F. (1982), 'Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation', *Econometrica* **50**(4), 987–1008.

Engle, R. F. and Gallo, G. M. (2006), 'A multiple indicators model for volatility using intra-daily data', *Journal of Econometrics* **131**(1), 3–27.

Fama, E. F. (1965), 'The behavior of stock-market prices', *The Journal of Business* **38**(1), 34–105.

Fan, J. and Wang, Y. (2007), 'Multi-scale jump and volatility analysis for high-frequency financial data', *Journal of the American Statistical Association* **102**(480), 1349–1362.

Fiorentini, G., Sentana, E. and Shephard, N. (2004), 'Likelihood-based estimation of latent generalized ARCH structures', *Econometrica* **72**(5), 1481–1517.

Flury, T. and Shephard, N. (2011), 'Bayesian inference based only on simulated likelihood: particle filter analysis of dynamic economic models', *Econometric Theory* **27**(05), 933–956.

Francq, C. and Zakoïan, J.-M. (2006), 'Linear-representation based estimation of stochastic volatility models', *Scandinavian Journal of Statistics* **33**(4), 785–806.

Francq, C. and Zakoïan, J.-M. (2010), *GARCH models: structure, statistical inference and financial applications*, Wiley, Chichester, U.K.

Fuller, W. A. (1996), *Introduction to Statistical Time Series*, second edn, John Wiley & Sons, New York.

Galbraith, J. W. and Zinde-Walsh, V. (1997), 'On some simple, autoregression-based estimation and identification techniques for ARMA models', *Biometrika* **84**(3), 685–696.

Gallant, A. R. (1987), *Nonlinear Statistical Models*, John Wiley & Sons, New York.

Gallant, A. R., Hsieh, D. and Tauchen, G. (1997), 'Estimation of stochastic volatility models with diagnostics', *Journal of Econometrics* **81**(1), 159–192.

Gallant, A. R., Hsu, C.-T. and Tauchen, G. (1999), 'Using daily range data to calibrate volatility diffusions and extract the forward integrated variance', *The Review of Economics and Statistics* **81**(4), 617–631.

Gallant, A. R. and Tauchen, G. (1996), 'Which moments to match?', *Econometric Theory* **12**, 657 – 681.

Gatheral, J. and Oomen, R. C. (2010), 'Zero-intelligence realized variance estimation', *Finance and Stochastics* **14**(2), 249–283.

Genon-Catalot, V., Jeantheau, T. and Laredo, C. (2000), 'Stochastic volatility models as hidden Markov models and statistical applications', *Bernoulli* **6**(6), 1051–1079.

Gentle, J. E. (2007), *Matrix algebra: theory, computations, and applications in statistics*, Springer Science & Business Media.

Ghysels, E., Harvey, A. and Renault, E. (1996), Stochastic volatility, *in* G. S. Maddala and C. R. Rao, eds, 'Handbook of Statistics: Statistical Methods in Finance', Vol. 14, North-Holland, Amsterdam, pp. 119–191.

Glosten, L. R., Jagannathan, R. and Runkle, D. E. (1993), 'On the relation between the expected value and the volatility of the nominal excess return on stocks', *The Journal of Finance* **48**(5), 1779–1801.

Goffe, W. L., Ferrier, G. D. and Rogers, J. (1994), 'Global optimization of statistical functions with simulated annealing', *Journal of Econometrics* **60**, 65–99.

Gouriéroux, C. and Monfort, A. (1995), *Statistics and Econometric Models, Volumes One and Two*, Cambridge University Press, Cambridge, U.K. Translated by Quang Vuong.

Granger, C. W. and Morris, M. J. (1976), 'Time series modelling and interpretation', *Journal of the Royal Statistical Society. Series A (General)* pp. 246–257.

Hafner, C. M. and Linton, O. (2017), 'An almost closed form estimator for the EGARCH model', *Econometric Theory* **33**(4), 1013–1038.

Hamilton, J. D. (1994), *Time Series Analysis*, Princeton University Press, Princeton, New Jersey.

Hannan, E. J. and Rissanen, J. (1982), 'Recursive estimation of mixed autoregressive-moving average order', *Biometrika* **69**(1), 81–94.

Hansen, B. E. (1994), 'Autoregressive conditional density estimation', *International Economic Review* **35**, 705–730.

Hansen, B. E. (1995), 'Regression with nonstationary volatility', *Econometrica* **63**(5), 1113–1132.

Hansen, L. (1982), 'Large sample properties of generalized method of moments estimators', *Econometrica* **50**, 1029–1054.

Hansen, P. R. (2005), 'A test for superior predictive ability', *Journal of Business and Economic Statistics* **23**(4), 365–380.

Hansen, P. R., Huang, Z. and Shek, H. H. (2012), 'Realized GARCH: a joint model for returns and realized measures of volatility', *Journal of Applied Econometrics* **27**(6), 877–906.

Hansen, P. R. and Lunde, A. (2006), 'Realized variance and market microstructure noise', *Journal of Business and Economic Statistics* **24**(2), 127–161.

Hansen, P. R., Lunde, A. and Nason, J. M. (2003), 'Choosing the best volatility models: the model confidence set approach', *Oxford Bulletin of Economics and Statistics* **65**, 839–861.

Hansen, P. R., Lunde, A. and Nason, J. M. (2011), 'The model confidence set', *Econometrica* **79**(2), 453–497.

Harvey, A. C. (1989), *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press, Cambridge, UK.

Harvey, A. C. and Shephard, N. (1996), 'Estimation of an asymmetric stochastic volatility model for asset returns', *Journal of Business and Economic Statistics* **14**(4), 429–434.

Harvey, A., Ruiz, E. and Shephard, N. (1994), 'Multivariate stochastic variance models', *The Review of Economic Studies* **61**, 247–264.

Heber, G., Lunde, A., Shephard, N. and Sheppard, K. (2009), 'Oxford-man institute's realized library, version 0.3'.

Hogg, R. V. and Craig, A. T. (1958), 'On the decomposition of certain χ^2 variables', *The Annals of Mathematical Statistics* **29**(2), 608–610.

Hull, J. and White, A. (1987), 'The pricing of options on assets with stochastic volatilities', *The Journal of Finance* **42**(2), 281–300.

Ikeda, S. S. (2013), 'Two-scale realized kernels: A univariate case', *Journal of Financial Econometrics* **13**(1), 126–165.

Jacod, J., Li, Y., Mykland, P. A., Podolskij, M. and Vetter, M. (2009), 'Microstructure noise in the continuous case: the pre-averaging approach', *Stochastic Processes and their Applications* **119**(7), 2249–2276.

Jacquier, E., Polson, N. and Rossi, P. (1994), 'Bayesian analysis of stochastic volatility models (with discussion)', *Journal of Economics and Business Statistics* **12**, 371–417.

Jurado, K., Ludvigson, S. C. and Ng, S. (2015), 'Measuring uncertainty', *American Economic Review* **105**(3), 1177–1216.

Kalnina, I. and Linton, O. (2008), 'Estimating quadratic variation consistently in the presence of endogenous and diurnal measurement error', *Journal of Econometrics* **147**(1), 47–59.

Kambouroudis, D. S. and McMillan, D. G. (2015), 'Is there an ideal in-sample length for forecasting volatility?', *Journal of International Financial Markets, Institutions and Money* **37**, 114–137.

Kapetanios, G., Khalaf, L. and Marcellino, M. (2016), 'Factor-based identification-robust interference in IV regressions', *Journal of Applied Econometrics* **31**(5), 821–842.

Kastner, G. (2016), 'Dealing with stochastic volatility in time series using the R package stochvol', *Journal of Statistical Software* **69**(5), 1–30.

Kim, S., Shephard, N. and Chib, S. (1998), 'Stochastic volatility: Likelihood inference and comparison with ARCH models', *The Review of Economic Studies* **65**, 361–393.

King, M. L. (1980), 'Robust tests for spherical symmetry and their application to least squares regression', *The Annals of Statistics* **8**, 1265–1271.

King, M. L. (1987), 'Towards a theory of point optimal testing', *Econometric Reviews* **6**(2), 169–218.

Kleibergen, F. (2002), 'Pivotal statistics for testing structural parameters in instrumental variables regression', *Econometrica* **70**(5), 1781–1803.

Knight, J. L., Satchell, S. E. and Yu, J. (2002), 'Estimation of the stochastic volatility model by the empirical characteristic function method', *Australian and New Zealand Journal of Statistics* **44**(3), 319–335.

Koop, G. and Korobilis, D. (2013), 'Large time-varying parameter VARs', *Journal of Econometrics* **177**(2), 185–198.

Koop, G., Leon-Gonzalez, R. and Strachan, R. W. (2009), 'On the evolution of the monetary policy transmission mechanism', *Journal of Economic Dynamics and Control* **33**(4), 997–1017.

Koopman, S. J., Jungbacker, B. and Hol, E. (2005), 'Forecasting daily variability of the S&P 100 stock index using historical, realised and implied volatility measurements', *Journal of Empirical Finance* **12**(3), 445–475.

Koopman, S. J. and Scharth, M. (2012), 'The analysis of stochastic volatility in the presence of daily realized measures', *Journal of Financial Econometrics* **11**(1), 76–115.

Koreisha, S. and Pukkila, T. (1990), 'A generalized least squares approach for estimation of autoregressive moving average models', *Journal of Time Series Analysis* **11**(2), 139–151.

Kristensen, D. and Linton, O. (2006), 'A closed-form estimator for the GARCH(1,1) model', *Econometric Theory* **22**, 323–337.

Li, Y. and Mykland, P. A. (2007), 'Are volatility estimators robust with respect to modeling assumptions?', *Bernoulli* pp. 601–622.

Liesenfeld, R. and Jung, R. C. (2000), 'Stochastic volatility models: conditional normality versus heavy-tailed distributions', *Journal of Applied Econometrics* **15**(2), 137–160.

Lindner, A. M. (2009), Stationarity, mixing, distributional properties and moments of GARCH(p, q)–processes, *in* T. Mikosch, J.-P. Kreiß, A. R. Davis and G. T. Andersen, eds, 'Handbook of Financial Time Series', Springer-Verlag, Berlin, Heidelberg, pp. 43–69.

Liu, L. Y., Patton, A. J. and Sheppard, K. (2015), 'Does anything beat 5-minute RV? a comparison of realized measures across multiple asset classes', *Journal of Econometrics* **187**(1), 293–311.

Liu, Y. and Morley, J. (2014), 'Structural evolution of the postwar US economy', *Journal of Economic Dynamics and Control* **42**, 50–68.

Loretan, M. and Phillips, P. C. (1994), 'Testing the covariance stationarity of heavy-tailed time series: An overview of the theory with applications to several financial datasets', *Journal of Empirical Finance* **1**(2), 211–248.

Mandelbrot, B. (1963), 'The variation of certain speculative prices', *The Journal of Business* **36**, 394–419.

McConnell, M. M. and Perez-Quiros, G. (2000), 'Output fluctuations in the United States: What has changed since the early 1980's?', *American Economic Review* **90**(5), 1464–1476.

Meddahi, N. (2001), An eigenfunction approach for volatility modeling, Technical report, CRDE, Université de Montréal, and CIRANO.

Meddahi, N. (2003), 'ARMA representation of integrated and realized variances', *The Econometrics Journal* **6**(2), 335–356.

Meddahi, N. and Renault, E. (2004), 'Temporal aggregation of volatility models', *Journal of Econometrics* **119**(2), 355–379.

Meeker, W. Q. and Escobar, L. A. (1995), 'Teaching about approximate confidence regions based on maximum likelihood estimation', *The American Statistician* **49**, 48–53.

Melino, A. and Turnbull, S. M. (1990), 'Pricing foreign currency options with stochastic volatility', *Journal of Econometrics* **45**(1), 239–265.

Miller, Jr., R. G. (1981), *Simultaneous Statistical Inference*, second edn, Springer-Verlag, New York.

Monfardini, C. (1998), 'Estimating stochastic volatility models through indirect inference', *The Econometrics Journal* **1**(1), C113–C128.

Moreira, M. J. (2003), 'A conditional likelihood ratio test for structural models', *Econometrica* **71**(4), 1027–1048.

Müller, U. A., Dacorogna, M. M., Davé, R. D., Olsen, R. B., Pictet, O. V. and von Weizsäcker, J. E. (1997), 'Volatilities of different time resolutions - analyzing the dynamics of market components', *Journal of Empirical Finance* **4**(2), 213–239.

Nelson, D. (1988), Times Series Behavior of Stock Market Volatility and Returns, PhD thesis, Department of Economics, Massachusetts Institute of Technology, Cambridge, Massachusetts.

Nelson, D. B. (1990), 'Stationarity and persistence in the GARCH(1,1) model', *Econometric Theory* **6**(03), 318–334.

Nelson, D. B. (1991), 'Conditional heteroskedasticity in asset returns: A new approach', *Econometrica* pp. 347–370.

Newey, W. K. and McFadden, D. (1994), 'Large sample estimation and hypothesis testing', *Handbook of Econometrics* **4**, 2111–2245.

Newey, W. K. and West, K. D. (1987), 'Hypothesis testing with efficient method of moments estimators', *International Economic Review* **28**, 777–787.

Newey, W. K. and West, K. D. (1994), 'Automatic lag selection in covariance matrix estimation', *The Review of Economic Studies* **61**(4), 631–653.

Noureldin, D., Shephard, N. and Sheppard, K. (2012), 'Multivariate high-frequency-based volatility (HEAVY) models', *Journal of Applied Econometrics* **27**(6), 907–933.

Omori, Y., Chib, S., Shephard, N. and Nakajima, J. (2007), 'Stochastic volatility with leverage: Fast and efficient likelihood inference', *Journal of Econometrics* **140**(2), 425–449.

Pagan, A. R. and Schwert, G. W. (1990*a*), 'Alternative models for conditional stock volatility', *Journal of Econometrics* **45**(1-2), 267–290.

Pagan, A. R. and Schwert, G. W. (1990*b*), 'Testing for covariance stationarity in stock market data', *Economics Letters* **33**(2), 165–170.

Perron, P. and Ng, S. (1996), 'Useful modifications to some unit root tests with dependent errors and their local asymptotic properties', *The Review of Economic Studies* **63**(3), 435–463.

Phillips, P. C. B. (1989), 'Partially identified econometric models', *Econometric Theory* **5**, 181–240.

Poon, S.-H. and Granger, C. W. (2003), 'Forecasting volatility in financial markets: A review', *Journal of Economic Literature* **41**(2), 478–539.

Primiceri, G. E. (2005), 'Time varying structural vector autoregressions and monetary policy', *The Review of Economic Studies* **72**(3), 821–852.

Richard, J.-F. and Zhang, W. (2007), 'Efficient high-dimensional importance sampling', *Journal of Econometrics* **141**(2), 1385–1411.

Robinson, P. M. and Velasco, C. (1997), Autocorrelation-robust inference, *in* G. S. Maddala and C. R. Rao, eds, 'Handbook of Statistics: Robust Inference', Vol. 15, Elsevier, pp. 267–298.

Rudin, W. (1976), *Principles of Mathematical Analysis*, third edn, McGraw-Hill, New York.

Ruiz, E. (1994), 'Quasi-maximum likelihood estimation of stochastic variance models', *Journal of Econometrics* **63**, 284–306.

Saikkonen, P. (1986), 'Asymptotic properties of some preliminary estimators for autoregressive moving average time series models', *Journal of Time Series Analysis* **7**(2), 133–155.

Sandmann, G. and Koopman, S. J. (1998), 'Estimation of stochastic volatility models via Monte Carlo maximum likelihood', *Journal of Econometrics* **87**(2), 271–301.

Savin, N. E. (1984), Multiple hypothesis testing, *in* Z. Griliches and M. D. Intriligator, eds, 'Handbook of Econometrics, Volume 2', North-Holland, Amsterdam, chapter 14, pp. 827–879.

Sbrana, G. and Poloni, F. (2013), 'A closed-form estimator for the multivariate GARCH(1,1) model', *Journal of Multivariate Analysis* **120**, 152–162.

Sensier, M. and Dijk, D. v. (2004), 'Testing for volatility changes in US macroeconomic time series', *The Review of Economics and Statistics* **86**(3), 833–839.

Shephard, N. (1994), 'Partial non-Gaussian state space', *Biometrika* **81**(1), 115–131.

Shephard, N. (1996), 'Statistical aspects of ARCH and stochastic volatility', *Monographs on Statistics and Applied Probability* **65**, 1–68.

Shephard, N. (2005), *Stochastic volatility: selected readings*, Oxford University Press, Oxford, U.K.

Shephard, N. and Andersen, T. G. (2009), Stochastic volatility: Origins and overview, *in* T. Mikosch, J.-P. Kreiß, A. R. Davis and G. T. Andersen, eds, 'Handbook of Financial Time Series', Springer-Verlag, Berlin, Heidelberg, pp. 233–254.

Shephard, N. and Sheppard, K. (2010), 'Realising the future: forecasting with high-frequency-based volatility (HEAVY) models', *Journal of Applied Econometrics* **25**(2), 197–231.

Sheppard, K. (2013), 'Oxford MFE Toolbox'.
URL: *http://www.kevinsheppard.com/MFE_Toolbox*

Shi, Y. and Eberhart, R. (1998), A modified particle swarm optimizer, *in* '1998 IEEE international conference on evolutionary computation proceedings', IEEE, Piscataway, NJ, USA, pp. 69–73.

Staiger, D. and Stock, J. H. (1997), 'Instrumental variables regression with weak instruments', *Econometrica* **65**, 557–586.

Steel, M. F. (1998), 'Bayesian analysis of stochastic volatility models with flexible tails', *Econometric Reviews* **17**(2), 109–143.

Stock, J. H. and Yogo, M. (2005), Testing for weak instruments in linear IV regression, *in* D. W. Andrews and J. H. Stock, eds, 'Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg', Cambridge University Press, pp. 109–120.

Straumann, D. (2005), *Estimation in conditionally heteroscedastic time series models*, Vol. 181 of *Lecture Notes in Statistics*, Springer, Berlin Heidelberg.

Takahashi, M., Omori, Y. and Watanabe, T. (2009), 'Estimating stochastic volatility models using daily returns and realized volatility simultaneously', *Computational Statistics and Data Analysis* **53**(6), 2404–2426.

Taylor, J. W. (1999), 'Evaluating volatility and interval forecasts', *Journal of Forecasting* **18**(2), 111–128.

Taylor, S. J. (1982), Financial returns modelled by the product of two stochastic processes – a study of daily sugar prices, *in* O. D. Anderson, ed., 'Time Series Analysis: Theory and Practice', Vol. 1, North-Holland, Amsterdam, pp. 203–226.

Taylor, S. J. (1986), *Modelling Financial Time Series*, John Wiley & Sons, New York.

Taylor, S. J. (1994), 'Modeling stochastic volatility: A review and comparative study', *Mathematical Finance* **4**(2), 183–204.

Tsay, R. S. and Tiao, G. C. (1984), 'Consistent estimates of autoregressive parameters and extended sample autocorrelation function for stationary and nonstationary ARMA models', *Journal of the American Statistical Association* **79**(385), 84–96.

Wang, J. and Zivot, E. (1998), 'Inference on structural parameters in instrumental variables regression with weak instruments', *Econometrica* **66**(6), 1389–1404.

West, K. D. (1996), 'Asymptotic inference about predictive ability', *Econometrica* **64**(5), 1067–1084.

White, H. and Domowitz, I. (1984), 'Nonlinear regression with dependent observations', *Econometrica* **52**, 143–161.

Wishart, J. (1947), 'The cumulants of the z and of the logarithmic χ^2 and t distributions', *Biometrika* **34**(1/2), 170–178.

Wright, J. H. (1999), 'Testing for a unit root in the volatility of asset returns', *Journal of Applied Econometrics* **14**(3), 309–318.

Xiu, D. (2010), 'Quasi-maximum likelihood estimation of volatility with high frequency data', *Journal of Econometrics* **159**(1), 235–250.

Yu, J. (2002), 'Forecasting volatility in the New Zealand stock market', *Applied Financial Economics* **12**(3), 193–202.

Zhang, L. (2006), 'Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach', *Bernoulli* **12**(6), 1019–1043.

Zhang, L., Mykland, P. A. and Aït-Sahalia, Y. (2005), 'A tale of two time scales: Determining integrated volatility with noisy high-frequency data', *Journal of the American Statistical Association* **100**(472), 1394–1411.