Three Essays in Financial Econometrics

by
Serguei Zernov
Department of Economics
McGILL UNIVERSITY, Montreal
February 2004

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

Your file Votre référence
ISBN: 0-494-06357-2
Our file Notre référence
ISBN: 0-494-06357-2

Abstract

The three essays of this thesis touch a variety of topics in financial econometrics. The first is an empirical investigation of aspects of transactions dynamics of currency futures. Based on the analysis of years of transactions data, the author describes the seasonal dynamic component of trade and price durations of futures and then analyses dynamics of durations using the stochastic conditional duration model (SCD).

The second essay develops a model for the robust analysis of time series. Asymptotic properties of the parameter estimates of this model are established. The model is applied to the analysis of dynamics of conditional quantiles of the quadratic variation in currency exchange rates. Forecasting properties of the model are also evaluated.

Finally, the third essay uses recent advances in the theory of extremal events to analyse the effects of institutional changes in financial markets on the extremal behaviour of major stock indices, as far as this behaviour is reflected in the evolution of Hill's estimator of the tail index.

Résumé

Cette thèse est composée de trois chapitres qui ont trait à divers sujets en économétrie des marchés financiers. Le premier chapitre consiste en une recherche empirique sur des aspects de la dynamique des transactions pour les contrats futurs de devises. En se basent sur dix ans de données de transaction, l'auteur décrit la composante dynamique saisonnière qui caractérise les durées entre les transactions consécutives et ainsi que celles des changements de prix des contrats futurs. Il analyse la dynamique des durées en utilisant le modéle de durée stochastique conditionnel

Le deuxième chapitre développe un modèle pour l'analyse robuste des séries temporelles. Des propriétés asymptotiques pour les estimateurs sont établies. Le modèle est appliqué dans l'analyse de la dynamique des quantiles de la variance ralisée des taux de change. La qualité des prévisions du modèle est également évaluée.

Finalement, le troisième chapitre exploite des développements récents dans la théorie d'événements extrêmes afin d'analyser les effets de changements institutionnels dans les marchés financiers, sur le comportement extrême des principaux indices boursiers. Les statistiques basées sur l'estimateur de Hill de l'indice de queue sont utilisées pour déceler des changements dans ce comportement extrême.

Contents

Acknowledgements

This thesis would not be possible without my advisors, John Galbraith and Vicky Zinde-Walsh. Not only was their guidance absolutely essential in choosing and pursuing the research topics of this thesis, but also their care and support have been protecting me from the turbulence of the outside world allowing to concentrate on study and research. It has been a real pleasure to work with them professionally and to meet them personally. I also acknowledge their generous financial help.

The contribution to my personal development of Joseph Greenberg, who has been my friend and a holistic advisor, and a lifestyle consultant during my years at McGill University, can neither be overstated nor explained in any detail within the limited space available here.

An opportunity to meet the leading researchers in financial econometrics at seminars and conferences organised by CIRANO has benefited tremendously to my professional growth. I also gratefully acknowledge the financial support of this organisation.

I would like to thank the faculty and the administrative staff of the Department of Economics for their kind attention to my problems and their help.

Finally, I acknowledge the generous financial support of the Department of Graduate Studies of McGill University through Max Bell and McGill Major Scholarships and the financial support of the Department of Economics.

Chapter 1

Introduction

Financial econometrics has only recently evolved into an independent discipline or subfield within econometrics. It is not surprising that the literature and the arsenal of tools offered by this new and developing discipline (as well as the set of its sometimes undecipherable acronyms) is burgeoning. The three essays of my thesis touch, directly or indirectly, several areas of financial econometrics; in the introduction I would like to position the topics of my research within the main directions of financial econometrics, and to show connections between the essays.

1.1 The discipline of financial econometrics

As a new discipline, financial econometrics finds its foundations in time series analysis, time series econometrics, financial economics, business practice, and even stochastic physics. The majority of models and methods of financial econometrics, at least of the vintage of financial econometrics to which I have been exposed, deal with the analysis of financial time series or financial processes. The presence of time in the models of financial econometrics is not surprising: any problem in finance is ultimately about a return on investment, and the return is a flow variable defined on a per unit of time basis. A dedication to practical and empirical topics is yet another prominent feature of research in financial econometrics; it is a de facto standard for any publication in the field to present at least a small empirical example, even if the focus of the publication is mainly theoretical.

When an econometrician studies financial markets, his object of interest is formally the conditional distribution of returns of a set of assets at some point in the future, $P_{t+\delta}$, given the information at time t, $\Im_t$ (we denote this distribution $F(P_{t+\delta}|\Im_t)$), or some particular properties of this conditional distribution. Until the beginning of the 1980s, econometrics of financial markets was essentially limited to estimating the conditional mean of the distribution $F(P_{t+\delta}|\Im_t)$. This conditional mean is not, however, a very interesting object of research: under the hypothesis of market efficiency (in its simplest form) this expectation has to be equal to zero, and the market efficiency hypothesis in its various forms has become since the 1970s a dominant methodological paradigm in economics and finance, especially in academic research. The interesting side of the dynamics of financial series - the properties of conditional variances-covariances and of higher moments - was awaiting to be discovered.

Engle's (1982) article, which first introduced the Autoregressive Conditional heteroskedasticity (ARCH) model, may be considered among the first publications in financial econometrics[1]. An enormous family of models describing the dynamics of conditional variance - direct descendants of ARCH - followed soon after (GARCH, ARCH-M, EGARCH, IGARCH, FIGARCH, FIEGARCH and many others). Multivariate GARCH-type models have been developed too. A number of models in which the dynamics of the second moment are driven by latent variables, such as stochastic volatility or switching state models, in the univariate or multivariate context, and usually much more demanding computationally than their GARCH-type relatives, have also been developed in the 1990s.

The common property of the models mentioned above is that the observed time

[1]The Nobel Committee has recently recognised the importance of Engel's contribution to scientific progress, awarding him, together with Clive Granger, the 2003 Nobel Prize in Economics.

series (the logarithm of asset returns p_t, for example) are modelled as the product of a conditional scale variable σ_t and an innovation ϵ_t, where $\{\epsilon_t\}$ are often assumed to be i.i.d. normally distributed. A few papers have parted with the assumption of normality of re-scaled innovation (see, for example, Engle and Gonzalez-Rivera (1991), Drost and Klassen (1997) and Hafner and Rombouts (2003)), but the estimation of time-varying conditional higher moments in empirical applications is still in the very early stages of development.

An important class of models in financial econometrics encompasses the models that assume the data generating process (DGP) to be a continuous-time diffusion process, probably with jumps. An attractive feature of these models is that some of them allow for an analytically tractable derivative pricing theory, unlike the discrete time models. Regrettably, many continuous time models used today in empirical research and by practitioners do not fit the empirical data well. Newer models developed in recent years, including multi-factor stochastic volatility models with jumps and models based on non-Gaussian Ornstein-Uhlenbeck processes (this latter class has been popularised by Ole E. Barndorff-Nielsen, Neil Shephard and their co-authors) show greatly improved empirical performance. Feasible estimation of continuous time models also poses difficult econometric problems: the algorithms are often complex, computationally intensive, and sensitive to model misspecifications. Recent developments in using high-frequency data promise a significant improvement in our ability to estimate continuous time models; we shall return to the econometrics of high-frequency financial data below. These latest developments are very encouraging for the future of continuous-time models as a tool for empirical research and for solving applied problems of finance.

Continuous-time and discrete-time models offer the researcher two different ways to look at the same phenomenon - financial time series. While these two approaches are sometimes presented in the literature as if they were mutually exclusive, they, in my opinion, compliment one another. Seminal results in Drost and Nijman (1993) and in Drost and Werker (1996) pioneered the research towards bridging the gap between the continuous time and the discrete time approaches. A discrete time process is viewed by the authors as a discretisation of an underlying diffusion; thus, a diffusion process generates a connected family of discrete time processes sampled at different frequencies. A discrete time process can represent a sampling from any diffusion process belonging to a certain class of diffusion processes; the processes within this class may be not distinguishable based on a single discrete data series.

Intuitively, a discrete time framework is often more flexible and requires fewer prior assumptions about the data generating process. In empirical sciences like economics or finance any model is merely an approximation. It is legitimate to take into account specific research or practical objectives during the model selection process: one can use discrete time modelling in exploratory analysis of the data or in forecasting and rely on a continuous time model for pricing derivatives, or even use a collection of models, both discrete-time and continuous time, to describe an empirical object in greater detail and from different perspectives.

To say that the availability of high-frequency financial data and the ability to process such data have brought revolutionary changes into financial econometrics is not an overstatement. While twenty years ago daily or even weekly data had been sometimes termed *high frequency* in macroeconomics, in financial econometrics the term *high-frequency* means intra-day data up to ultimately transaction data. Use of intra-day data contributes to the ability of the econometrician to estimate financial and econometric models because of the additional information brought by these data, compared to the information contained in only the daily series; as a bonus, algorithms based on intra-day information often happen to be simpler, both conceptually and computationally, than algorithms based on the daily measurements.

As we have already pointed out, the majority of models of financial series are, in their essence, models of the conditional dynamics of the second moment. The estimation of these models using high-frequency data is premised by the fact that the statistic called *realised quadratic variation (RV)* is, under certain conditions, a consistent and asymptotically unbiased estimate of a quantity called *quadratic variation (QV)*[2] [3]. Realised quadratic variation computed using intra-day data can be used to estimate either continuous or discrete time models.

Estimation methods based on RV assume that the data is sampled at regular intervals, thus do not take into account the information contained in the timing of market events. The fundamental result upon which the use of RV is based is that it converges to the QV as the sampling frequency increases. Financial data, however, cannot be sampled at an arbitrarily high frequency - the natural limit to how often it can be sampled is the transactions data. As the sampling frequency increases, so does the effect on the estimate of the microstructure noise (the market microstructure effects contributing to the noise include the presence of the bid-ask spread and bounce, different information content of trades due to block trading or the strategic trading, discreteness of the price quotes and numerous others). If the microstructure noise is not modelled explicitly, there is usually the optimal sampling frequency of the intra-day data to be used to compute the RV, and this optimal frequency is lower than the highest available[4]. In practice, RV is always a noisy measure of QV, both due to the practical upper limit of the sampling frequency and the presence of the microstructure noise. It is important for the econometrician to be aware of this error and either to account for it in the model or to use a robust estimation technique.

An inherent property of the transactions data is that it is not recorded at regular intervals, and moreover, the timing between transactions is endogenous to the process: hence, the timing is informative with respect to the market dynamics. Theoretically, the

[2] Andersen and Bollerslev (1998) were the first to introduce realised quadratic variation, quadratic variation and to describe their properties; they used the term realised volatility rather than realised quadratic variation, but the latter term is becoming more widely used.

[3] The theory of quadratic variation will be introduced in detail in Chapter 3, which is why we keep the exposition here very concise.

[4] Aït-Sahalia and Mykland (2003) show that in the presence of microstructure noise RV may converge to the variance of the noise instead of QV. They offer also an algorithm to choose the sampling frequency optimal in the mean-square sense when the microstructure is not modelled as well as approaches to modelling the microstructure noise. The problem of optimal sampling in the presence of microstructure noise is also considered in Bandi and Russell (2003).

information contained in the timing of the transactions can be used for more accurate estimation of the market dynamics (for example, in the context of our previous discussion, for constructing more accurate estimates of QV). It is a difficult task, however, and there are yet very few studies which use transaction timing in modelling the dynamics of asset prices; I would like to mention here Gerhard and Hautsch (2002b) as an example of the use of transaction timing to estimate the volatility dynamics.

Investigation of the transactions data promises answers to many questions. Among them, for example: insights into the market microstructure and market interactions, and determinants of market liquidity and market performance under stress, with applications of the above to optimal market design. Transactions data is often described as a *point process*[5], and one of the possible ways to study point processes is through the dynamics of durations - intervals between consecutive occurrences of events of interest. Since the seminal paper of Engle and Russell (1998), financial econometrics has offered many models of durations (intertrade durations and other other types of durations). While the majority of the existing models of financial point processes are econometric models (i.e., they are not derived from our knowledge of market structure and operation), I believe that once researchers achieve a better understanding of the empirical properties of transaction data and develop the appropriate econometric tools, they will turn to the constructive design of transaction data models, i.e. to design based on knowledge of the structure of financial institutions and of the nature of interactions between the market participants.

The final comment in this section will be about the econometrics of extremal events. Models like GARCH and numerous incarnations of stochastic volatility models imply a complete specification of the distribution of a financial variable of interest. Practice shows, however, that models that adequately describe the dynamics of the centre of the distribution of asset returns often fail to accommodate the dynamics of rare events, i.e., of the tails of the distribution. We find ourselves once again in a situation typical to

[5]The definition of a point process is given in Section 2.2.

empirical research, and specifically to research in economics and finance, where there is no "correct" model, and the model selection can depend, among other factors, on the application for which the model is intended. Financial econometrics has responded to the needs of market regulators and institutional risk managers to study and forecast the behaviour of tails of the returns of financial assets: the theory of extremal events in applications to financial data has become a rapidly developing part of financial econometrics.

1.2 Contribution of this research

The three essays of this thesis touch in greater or smaller degree virtually all of the topics mentioned in the introductory review. The essay *Dynamics of Trade and Price Durations of Currency Futures* is an empirical study of the dynamics of transactions data. The subject of existing empirical studies of transactions durations was stocks; the empirical contribution of this research is that it investigates the dynamics of currency futures: neither currencies nor derivative contracts have been investigated from that perspective. Futures contracts have more complex seasonal dynamics of transactions than stocks because the former have an additional dimension to their seasonal component - time to expiration - and our study documents this complex seasonal behaviour. While existing studies of the dynamics of durations use data sets spanning a period of no more than several months, the data used in this research spans ten years. The large data set used allows us to estimate meaningfully the multi-dimensional seasonal component, and to study the evolution of the transaction dynamics over a period of ten years. Analysis of the data strongly suggests that transaction durations possesses long memory properties. In this essay, we introduce a new model capable of accounting for long memory in transaction durations, the *Fractionally Integrated Stochastic Conditional Duration* (FISCD) model, and suggest an estimation algorithm for it.

The second essay of the thesis - *Estimation of Conditional Quantiles of Variance*

Using Auxiliary Variance Information[6] - is an exercise in robust techniques of time series analysis. The developed techniques are used to estimate, using high-frequency data, the dynamics of conditional quantiles of QV in a stock index and in currency exchange rates.

The asymptotic properties of the estimator of GARCH parameters based on infinite robust regression are derived here for the case of an arbitrary quantile. I conduct an analysis of the dynamics of currency exchange rates, including the analysis of the forecasting performance of our modelling framework, by quantile regression.

This approach allows us to model dynamics of any conditional quantile of interest in the distribution of the second moment; therefore, it can be viewed as a non-parametric technique going beyond specifying the conditional dynamics of only the second moment. In the context of modelling the dynamics of QV, our technique can be either an alternative or a complimentary tool to the continuous time modelling framework. The advantages of the former are its robustness to misspecifications, its ability to focus on a specific part of the conditional distribution which is important in risk management applications, and its simplicity in implementation and computational frugality. In addition, our technique provides a flexible framework for including a variety of information auxiliary to the time series being modelled. Continuous time financial models, on the other hand, offer a transparent analytical treatment of option pricing and have a better predictive power provided that the model and the true DGP are close [7].

The third essay of the thesis - *Circuit Breakers and the Tail Index of Equity Returns*[8] - uses techniques based on Hill's estimator of the tail index to examine effects of institutional changes on the extremal behaviour of major stock indices.

The important advantage of the Hill's estimator is its robustness with respect to a wide range of the specifications of the DGP. Some of the statistics used in our research rely, nevertheless, on the rather restrictive assumption that the series under investigation

[6]This chapter is based on a joint article with John W. Galbraith and Vicky Zinde-Walsh.

[7]It is not surprising that continuous time models have better predictive power theoretically: because they are narrowly parameterised, they bring more *a priori* information to the estimation, and they fare better in comparison with methods using less prior information, provided that this information is correct.

[8]This chapter is based on Galbraith and Zernov (2004).

follows a GARCH process. To ensure robustness of our conclusions to model misspecification, we use several alternative approaches to test our working hypotheses.

In this essay, we study from a historic perspective how adoption of new market strategies by investors, amendments to trading rules, and institutional changes by market regulators have been reflected in the evolution of the tail index of major stock indices (DJIA and S&P 500). This research brings us back in time to a very interesting topic which has been almost in oblivion in the academic literature in the recent past - the market crash of 1987. New econometric techniques and available market data for the years after the crash allow us to find compelling evidence of a decrease in the tail index measure (increase in the frequency of extreme events) of major stock indices in the beginning of the 1980s - the period coinciding with the introduction of the synthetic portfolio insurance. We also find evidence that changes to the market structure introduced after the events of 1987, circuit breakers in particular, may have been successful in returning the extremal behaviour of the index almost to levels that existed before the introduction of portfolio insurance programs.

1.3 Contribution of the authors

Chapter 3 of this thesis - *Estimation of Conditional Quantiles of Variance Using Auxiliary Variance Information* - is based on joint research with John Galbraith and Vicky Zinde-Walsh. The theoretical results of this chapter - the asymptotic distribution of the estimated parameters of the time series quantile regression - were developed by the three authors with equal participation. I have carried out most of the empirical analysis of the study.

Chapter 4 - *Circuit Breakers and the Tail Index of Equity Returns* - is based on joint research with John Galbraith. The authors worked closely together on this project and made equal contribution to the results.

Chapter 2

Dynamics of Trade and Price Durations of Currency Futures

2.1 Introduction

Recent technological progress has made easily available the computing facilities necessary to process high-frequency financial data, as companies providing financial services and research have eased their grip on high-frequency trading records. There has been an academic response to these new opportunities, including a number of publications that make use of high-frequency financial data[1].

Trading data (bids and ask quotes, prices, trade volumes etc.) can be viewed as observable manifestations of the economic process that involves interaction of the market participants given the institutional structure and the flow of economic and financial information. These data are the empirical basis of modelling the economic process and its components. Financial models where asset prices are described by a stochastic process have become a *de facto* standard in the science of finance. Models featuring stochas-

[1]The financial industry had been studying high-frequency data long before the recent surge of interest by the academics. The industry had the resources and the data. Results of this research were usually available only in-house; in addition, the objectives and methods of research in the industry are somewhat different from those of academic research.

tic volatility, jumps and other refinements reflect how we see the reaction of market participants to heterogeneous flow of economic information to which they are exposed.

Prices in economics are realized only through transactions. Thus, if we consider a dynamic asset pricing model (continuous-time or discrete time, or mixed), price variables of such a model are not usually directly observable, in contrast with the textbook wisdom, but manifest themselves through transactions. Whatever inference we desire to make about the constituent parts of financial markets, it will be based on the transactions data, plus some exogenous data series.

Fitting a dynamic model to transactions data may serve two purposes: first, the dynamic model may give insights into the operation of financial markets, and second, it may be used as a black-box for forecasting purposes. The main subject of interest of this paper is the empirical study of trading dynamics (the dynamics of trade and price durations) of currency futures. We use ten years of transactions data on Japanese yen (JPY)/US dollar (USD) futures traded at the Chicago Mercantile Exchange (CME). Following the approach adopted by other researchers we represent the durations process as consisting of two components: the non-stochastic seasonal component and the stochastic component. The assets that have been most often analyzed in empirical studies of durations are stocks. The seasonal component of futures, compared to that of stocks, has a dimension related to the life-cycle of the contract; we model and describe the seasonal behaviour of JPY/USD futures with respect to their life cycle. We model the stochastic part of dynamics of currency futures using the *Stochastic Conditional Duration (SCD)* model of Bauwens and Veredas (2004), which we estimate using the QML approach and Kalman filter.

Standard specification diagnostics of the model with SCD dynamics, estimated on the futures data (we test how well the model fits the dependency properties of the data and we test parametric specifications of the distributions of the innovations), show that the SCD model with an $AR(1)$ latent process does not describe very well the dependency in the data. The analysis of the joint information structure of the model and of the estimation algorithm suggests that increasing the order of the latent process is not a feasible alternative. We suggest in the appendix an extension to the SCD model that allows for long memory in the latent process. We call this extended model Fractionally Integrated SCD, $FISCD(p, x, q)$[2] (the FISCD(p,x,q) model is mathematically equivalent to the LMSV model of Breidt, Crato, and de Lima (1998)). In terms of basic notions of the theory of signal processing, SCD FISCD and LMSV are, in their essence, models of a discrete signal measured with white, possibly non-Gaussian, noise. We estimate the $FISCD(1, x, 0)$ model for trade and price durations of currency futures using the spectral QML approach. Breidt, Crato, and de Lima (1998) have proved the strong consistency of such estimates but other properties of these estimates are not known and will be a subject of future research.

The data set being analyzed in this paper spans a much longer period of time than the data used in previous publications on a similar subject. Thanks to this fact we are able to look at the evolution of estimated model parameters over the years. We are also able to investigate whether the parameters of the SCD model vary over different periods of the life-cycle of the futures contracts. The analysis of such behaviour allows us to draw conclusions about the ability of the SCD (or FISCD) model to capture invariant dynamic properties of the trading process.

We also describe asymptotic properties of the QML estimates of the SCD model in a case in which the durations process is not seamless but is re-initialized in the beginning of every trading day.

Our exposition proceeds as follows. Section 2.2 describes and classifies existing approaches to modelling financial point processes. Section 2.3 describes the data and the transformations which have been applied to the data. Section 2.4 formulates the econometric model. Section 2.5 describes the estimation and specification diagnostic methods. Section 2.6 presents estimation results, their interpretation and discussion. Section 2.7 concludes. Some technical details as well as notes on specification and estimation of the

[2]The latent process of the FISCD follows ARFIMA(p,x,q).

FISCD model are provided in Appendix B.

2.2 Models of durations in finance

Transactions data are often modelled within the framework of so-called *point processes.* A (temporal) point process is a sequence of events (points) on a time line, with each point representing a random arrival time t_i, and a sequence of random variables $\{x_i\}$, called *marks*, associated with these arrival times. Instead of a sequence of arrival times t_i it is often convenient to use a sequence of *durations* D_i, $D_i = t_i - t_{i-1}$. Thus, a data sample generated by a point process can be described as $(D_i, X_i), i = 1, \ldots, N$. A point process with marks can serve as a generator for derived *thinned* processes, the arrival times of which are a subsequence of the arrival times of the original process; this subsequence is chosen based on a criteria which is a function of the arrival times and of the marks of the generator. The thinned processes associated with a point process are also point processes.

It is not difficult to describe a point process in general terms: the description will be given by the joint distribution of the durations and the marks conditional on their past values (see, for example, Engle (2000)):

$$(D_i, X_i)|\Im_{i-1} \sim f\left(D_i, X_i|\bar{D}_{i-1}, \bar{X}_{i-1}; \theta\right) \tag{2.1}$$

where $\bar{z}_i \equiv \{z_i, z_{i-1}, \ldots, z_i\}$ and θ is the vector of parameters.

In practice, it may be very difficult or even infeasible to specify the conditional distribution in the form (2.1). The researcher, however, may be interested only in characteristics of the process which do not require the model to be completely specified. Among typical questions of interest to an econometrician are the following:

- What is the marginal distribution of marks and durations; what is their joint marginal distribution?

- What it the distribution of the next mark conditional on the history of marks and/or durations?
- What is the distribution of future durations conditional on the past durations and/or marks?
- The questions above, asked with respect to a specific thinned process

If the researcher is interested only in the conditional dynamics of durations (whether of the generating process or of a thinned process), the *conditional hazard function* provides a convenient framework for model specification.

The hazard function is defined as the limit of the conditional probability of observing an arrival in a small interval, given that there has been no arrival up to the beginning of this interval:

$$\lambda_\tau(t) \equiv \lim_{\Delta \to 0} \frac{\mathbf{Prob}\left[t \leq \tau \leq t + \Delta | \tau \geq t\right]}{\Delta}, \tag{2.2}$$

There is a simple relationship between the hazard function and the distribution function of the arrival time $Q_\tau(s)$:

$$\lambda_\tau(t) = -\frac{\partial \log\left(1 - Q_\tau(t)\right)}{\partial t} \tag{2.3}$$

The hazard function can be specified in the conditional context given the information set $\Im_t$: one only needs to choose the appropriate conditioning set in (2.2). Also, knowledge of the conditional density immediately implies the knowledge of the hazard function, while modelling the latter directly is often simpler. Cox (1972) and Cox (1975) suggested presenting the hazard function as a product of the baseline hazard and a positive function of the conditioning set and model parameters. The models based on this approach are called proportional hazard models and are often used in the analysis of point processes in statistics, and specifically in the econometrics of ultra-high frequency data (see, as an example, Gerhard and Hautsch (2002a)).

Another approach to modelling the dynamics of point processes was suggested in Engle and Russell (1998). They assume that specification of the conditional density

(2.1) requires only a mean function. They define conditional duration as a function of past durations and marks and then assume that a duration is equal to the product of this conditional duration and an innovation - a random variable with a positive support:

$$\Psi_i \equiv \Psi\left(\bar{D}_{i-1}, \bar{X}_{i-1}\right) = \mathbf{E}_{i-1}\left(D_i | \bar{D}_{i-1}, \bar{X}_{i-1}\right)$$
$$D_i = \Psi_i \epsilon_i \tag{2.4}$$

where $\epsilon_i \sim i.i.d.$.

The authors consider the specific form of the function Ψ:

$$\Psi_i = \omega + \sum_{j=0}^{m} \alpha_j D_{i-j} + \sum_{j=0}^{q} \beta_j \Psi_{i-j}, \tag{2.5}$$

and they call the corresponding model the *Autoregressive Conditional Duration* model and denote it $ACD(m,q)$. One can see that if the distribution of ϵ is known, it is easy to write down the likelihood function for this model, i.e., the model is easy to estimate. Forms of the mean function Ψ different from the above are possible (a comprehensive review and classification of mean functions used in ACD-type models is presented in Hautsch (2002)).

Knowing the distribution function of the errors ϵ_i, one can cast the model of Engle and Russell in terms of the hazard function. Specifically, the hazard function of the model (2.4) takes the following form:

$$\lambda\left(t; \bar{D}_{i-1}, \bar{X}_{i-1}\right) = \lambda_0\left(\frac{t - t_{i-1}}{\Psi_i}\right)\frac{1}{\Psi_i},\ t_{i-1} \leq t \leq t_i. \tag{2.6}$$

where $\lambda_0(t)$ is the baseline hazard computed according to (2.3). The class of models defined in (2.4) is called *accelerated time models* because the function Ψ enters as the denominator the corresponding baseline hazard.

The model upon which a larger part of the empirical analysis of this study is based, the SCD, is also an accelerated time model like the ACD. It is a mixture model at the same time: the acceleration factor in the SCD model is not a deterministic function of the past values of durations and marks but, conditional on the history of durations and marks, is still a random variable. Bringing a parallel with the models of conditional heteroskedasticity, the relationship between ACD and SCD is the same as the relationship between GARCH models and stochastic volatility models: in the (strong) GARCH, the conditional variance σ_t^2 is a deterministic function of past variances and squared innovations ϵ_t^2:

$$\sigma_t^2 = \sigma_0 + \sum_{j=0}^{m} \alpha_j \sigma_{t-j} + \sum_{j=0}^{q} \beta_j \epsilon_{i-j}^2,$$

while, conditional on the current information set, the variance in stochastic volatility models is a random variable, same as the duration in the SCD.

The advantage of the SCD is that it is more flexible [3] and better fits the empirical data. Estimating the SCD model is, however, a more difficult problem that estimating ACD-type models because the likelihood function of the SCD cannot be written explicitly. We shall introduce the SCD model and discuss its estimation in detail, later in this chapter.

Finally, we would like to mention another model which is a mixture model like the SCD but that does not belong to the class of accelerated time models. This is the *Stochastic Volatility Duration (SVD)* model suggested by Ghysels, Gouriéroux, and Jasiak (1997). The authors noticed that modelling only the mean function of the duration process might not be enough to capture empirical properties of financial transaction data. Therefore, in the SVD model the conditional second moment of durations is also stochastic. As with the SCD, the likelihood function of the SVD cannot be written down explicitly but only computed using simulation. The SVD model, however, is set up in such a way that the simulated likelihood is relatively easy to compute.

This review of the models of financial point processes, and specifically - of durations, is not comprehensive. The objective of this review was to position the SCD model among the main existing approaches to modelling the dynamics of durations. We proceed next

[3]In particular, the SCD framework allows one to reproduce a wider range of shapes of hazard function than the ACD, including non-monotonous shapes. See Bauwens and Veredas (2004) for details.

to the introduction of the data used in this study.

2.3 Description of the data

We examine aspects of transaction dynamics of currency futures traded on the CME. Specifically, we study JPY/USD futures. CME currency futures contracts follow the usual March-June-September-December cycle. New contracts are listed six month before the expiration (on the day after the front month expires); the contracts expire on the second business day before the third Wednesday (in our data set, the expiration day is always Monday). For example, the first trading date of the March 2003 futures contract (tick symbol - JYH3) was the 18th of September, 2002 and the last trading date was the 17th of March, 2003. Trading opens at 7:20 and ends at 14:00 Central time.

The data set spans the period from January 2, 1991 to August 31, 2001 and consists of almost 4 million records. The records of trades of 47 contracts are present in the set, from a contract expiring in March of 1991 to one expiring in January of 2002. For the purposes of our analysis the data was filtered. The records in the data set represent either transactions, bid quotes or ask quotes. We remove records that are marked as ask or bid quotes and do not represent actual transactions. As well, we do not consider contracts with fewer than 130 trading days within the time span of the data and we exclude contracts with expiration dates after August 31, 2001. This leaves 37 contracts in the set, from a contract expiring in June of 1991 to a June, 2001 contract. After filtering, the data consist of 2,743,740 records.

Trade durations are defined as time intervals between consecutive trades; the last duration of a day precedes the first duration of the next day in the duration series. One of the problems that we had to resolve was the treatment of multiple trades that happened within one second (one second is the precision of the time stamp), i.e., when recorded trade durations were equal to zero. There are several possible ways to deal with this problem, some being more sophisticated than the others. We have employed a simple solution to this problem of censored measurements. We assign the value $\frac{1}{k-1}$ to every zero duration, where k is the number of trades happening within the current second, and we increase the subsequent duration by $\frac{1}{k-1}$ so that the sum of all durations is unchanged[4]. Another naive approach would be, following Bauwens and Veredas (2004), deleting all null durations. Bauwens and Veredas (2004) motivated the latter approach by arguing that the trades that happened within a very short period of time were likely from the same trader, who split a large block of shares.

A price duration is defined as the lapse of time that is required to observe a price change not less than a given threshold. It is in some sense natural to measure the change of the price in percentage points (or to measure the logarithm of the change of the price): dynamic models in finance are most often formulated with respect to the logarithm of the price. The matter is complicated by the fact that the transaction price is quoted with a given number of significant digits, i.e., the observed prices take their values on a discrete set. This is yet another illustration of the idea that transactions may be seen as a manifestation of the latent economic process; finite accuracy of the reported prices is a property of the "transmission mechanism" - the market. Russell and Engle (1998) develop a model of price durations where the prices are explicitly discrete-valued. We have chosen to use a change in the logarithm of price as the criterion for thinning. The empirical results presented here are for the case in which the change in the logarithm of the price is equal to or larger than 0.05%[5].

[4]We are aware that this method may introduce spurious correlation in the data; our analysis showed however that this effect is negligible.

[5]Stock prices are usually recorded with lower accuracy that those of currency futures, which is why accounting for discretisation error is more important in the former case.

2.4 Modelling trade and price durations of futures contracts

Let $\{D_i\}$ denote the recorded durations (trade durations or price durations). In what follows, if there is no ambiguity, we shall use small letters to denote logarithms of the values denoted by the corresponding capital letters i.e. $d_i \equiv \ln(D_i), .\psi_i \equiv \ln(\Psi_i)$ etc. The model being estimated is formally specified as follows:

$$D_i = \Phi(\kappa_i)\Psi_i\varepsilon_i \tag{2.7}$$

We assume that $\varepsilon_i|I_{i-1} \sim iid\ \mathrm{D}(\eta)$ where I_{i-1} denotes the information set at the beginning of the spell of the duration d_i and $\mathrm{D}(\eta)$ is a distribution with a positive support with a parameter η. The fourth moment of $\mathrm{D}(\eta)$ exists and is finite. Usual choices of the parametric form of the distribution $\mathrm{D}(\eta)$ in the context of duration studies are the Weibull distribution and the standard Gamma distribution. The process $\psi_i = \ln\Psi_i$ follows, in the general case, a stationary $ARMA(p,q)$ process with Gaussian innovations.

The function $\Phi(\kappa)$ is assumed to be non-stochastic and strictly positive for all admissible values of κ. Taking the logarithm of the equation above,

$$d_i = \phi(\kappa_i) + \mu(\eta) + \psi_i + \xi_i \tag{2.8}$$

where $(\xi_i + \mu(\eta))$ is distributed as the logarithm of ε_i and $\mathbf{E}[\xi_i|I_{i-1}] = 0$. Under the specifications above, log-durations are sums of the non-stochastic part $\phi(\kappa_i) + \mu(\eta)$, and the stochastic part $\psi_i + \xi$. If we define $\hat{d}_i = d_i - \phi(\kappa_i)$ and assume that ψ_i follows $AR(1)$, the model in terms of $\hat{d}_i$ will be the SCD model as it has been formulated and studied in Bauwens and Veredas (2004).

We shall argue later that it is not practical to consider SCD models with a latent process of order higher than $AR(1)$. The FISCD model introduced in Appendix (B.3) is a more flexible alternative than SCD. FISCD is a complex econometric object; many properties of the parameter estimates under the FISCD are unknown and are a subject of future research.

The seasonal component in the dynamics of durations

The literature presents strong empirical evidence of seasonality in trade and price durations (see, for example, Engle and Russell (1998), Gouriéroux, Jasiak, and Le Fol (1999), and Bauwens and Veredas (2004)), which is why the seasonal component $\Phi(\kappa)$ is present in equation (2.7). Unlike stocks that may be thought of as having an infinite time horizon, derivative contracts, bonds, and some other assets have a life cycle, from a contract's inception to its expiration. This life cycle is reflected in the "seasonal" behaviour of time series describing the dynamics of such contracts. This form of seasonal behaviour of trade and price durations of futures, due to their life cycle, has been given less attention in the empirical literature than diurnal or weekly seasonality[6].

Under the model adopted in this study, duration series have two components: the deterministic seasonal component and the stochastic component that follows the SCD dynamics. We can approach the estimation of the seasonal component parametrically, semi-parametrically or non-parametrically, and there exist several possibilities in each of these classes of estimation techniques. An attractive feature of non-parametric modelling is its flexibility. We shall model the seasonal dynamics of the durations in the non-parametric spirit (strictly speaking, the approach that we use is semi-parametric, as the reader will see from the exposition below), similar to the approach adopted in Veredas, Rodriguez-Poo, and Espasa (2002) and in Bauwens and Veredas (2004) with the difference that the futures considered in this paper have a more complex seasonal structure.

The multiplicative presentation (2.7) and the additive presentation in logarithms (2.8) are equivalent at the stage of modelling. When it comes to estimation, application of seasonal adjustment before taking logarithm or after leads to different results. In Bauwens

[6]Gerhard and Hautsch (2002b) describe the seasonality over the maturity of intra-day volatility for BUND futures. They estimate intra-day volatility based on price durations.

and Veredas (2004), the data are seasonally adjusted before taking the logarithms. The advantage of this approach is that the results are easier to interpret and easier to apply to forecasting (presumably, we are interested in durations and not in the logarithms of durations). We have chosen, however, to apply the seasonal adjustment after taking the logarithms of the data, and we shall name two reasons for this choice. First, we shall assume later in this study that $\phi(\kappa)$ follows the additive model with the logarithm as the link function[7]. Properties and estimation of additive models are known better than are properties of GAM, and we would like to build upon this knowledge. Second, the SCD is a model with dynamics that are linear in the logarithm of durations; estimation using the Kalman filter is based on the model's linearity. Seasonal adjustment of the dynamic variable of the model seems to be more transparent than adjustment of the non-linear transformation of this variable.

As we have mentioned just above, we impose additional structure on the seasonal component of the durations. Specifically, we assume that

$$\ln \Phi(\kappa) \equiv \phi(\kappa) = A_\delta + \chi(t) + \zeta(\tau), \tag{2.9}$$

where $\kappa = \{\delta, t, \tau\}$, A_δ ($\delta \in \{Monday, ..., Friday\}$) describes the weekly seasonality, $\chi(t)$ the seasonality due to contract life cycle, t is the time to expiration, $\zeta(\tau)$ corresponds to the diurnal seasonal component, and τ is the time elapsed from the beginning of the trading session. We test the assumption of orthogonality of the weekly, the life-cycle and the diurnal components later, in Section 2.6.4. The additive form of ϕ reduces the dimensionality of the non-parametric regression problem. Preliminary analysis shows that even with our large data set the curse of dimensionality cannot be escaped, especially

[7] $\Phi(\theta)$ follows in this case the *Generalized Additive Model (GAM)*. It is said that a non-parametric regression function follows the GAM if

$$f(m(\mathbf{X})) = \alpha + \sum_{j=1}^{d} \chi_j(X_j)$$

where $f(\cdot)$ is a known link function, $\chi_1, ..., \chi_j$ are unknown univariate functions, and $\mathbf{X} = (X_1, ..., X_j)$.

when we estimate the seasonal component at longer horizons to expiration t where the trading is sparse. Under the assumption of the additive model one can achieve, under certain conditions, the univariate rates of convergence of the non-parametric regression (Linton and Nielsen 1995).

2.5 Estimation Methods

2.5.1 Estimation of the seasonal component

We model the seasonal deterministic component of log durations in the following manner:

$$\mathbf{E}\left[d_i|\delta_i, t_i, \tau_i\right] = \alpha + A_{\delta_i} + \chi\left(t_i\right) + \zeta\left(\tau_i\right)$$

One approach to estimating additive models is the so-called *backfitting* algorithm (see Fan and Gijbels (1996), pp.266-267). In application to our problem, the algorithm can be described as follows:

1. Initialization: $\alpha = \frac{1}{N}\sum_{i=1}^{N} d_i$. We subtract the sample mean from $\{d_i\}$ and make initial guesses about all but one seasonal component (about A_δ and $\chi\left(t\right)$, for example). We force the sample expectation of each of these components to be equal to zero, so that the mean of the seasonally adjusted data is equal to zero.

2. $d^{(k+1)}_{\tau,t,v} = d^{(k)} - A^{(l)}_{\delta} - \chi^{(l)}$, where $A^{(l)}_{\delta}$, $\chi^{(l)}$ are the latest estimates of A and χ. $d^{(k+1)}$ is used to obtain an updated estimate of ζ via a univariate non-parametric smoother.

3. We repeat step 2 for each of the seasonal components until convergence is achieved. (In practice, 2 or 3 rounds are usually sufficient).

The term *backfitting* in reference to the action above was first used in Friedman and Stuetzle (1981). If the additive specification is not the true model, the algorithm is expected to give an estimate that is the best additive approximation to the regression surface in the mean squared error sense (Breiman and Friedman 1985).

We have to specify a univariate smoother that will be used at step 2 of the algorithm above. Our current choice is the kernel regression (other choices like smoothing splines, for example, will probably work equally well). The form of kernel regression used is known otherwise as the Nadaraya-Watson estimator. It is defined as

$$f(x) = \frac{\sum_{i=1}^{N} K\left(\frac{x-x_i}{h}\right) d_i}{\sum_{i=1}^{N} K\left(\frac{x-x_i}{h}\right)}$$

We have chosen the quartic kernel for our regression because it is fast to compute and has a compact support which also helps to reduce the complexity of the computations. For the results reported in the paper the values of the bandwidth parameters were chosen by visual inspection of their performance.

The component A_δ is somewhat different from χ and ζ. We estimate it as the mean duration for a given day of a week (formally, this is equivalent to setting $h = 1$, if we want to preserve the uniformity of exposition). Taking into account that we force the sample mean of each component of $\phi(\kappa)$ to zero, A_δ has four degrees of freedom (four parameters to be estimated), which is why we have mentioned that our approach can be called "semi-parametric".

The argument of $\chi(t)$, the life-cycle seasonal component, is the time to expiration in business days, $t \in \mathrm{N}$. We consider in our study the records with $1 \leq t \leq 130$. The argument of the diurnal component $\zeta(\tau)$ represents the time from the beginning of the trading day in seconds, $\tau \in [0, 24000)$.

We should point out that the whole data set is used to estimate the seasonal component of durations, $\phi(\kappa)$. This will allow us to capture invariant properties of the seasonal components across the years spanned by the data set. We estimate SCD parameters individually for each contract in the data set. Part of the variation of trading intensity will be also accounted for by the parameter ω of the SCD model which we keep free.

2.5.2 QML estimation of the SCD model

Once the seasonal component of the logarithm of durations has been estimated, we estimate the SCD model using the adjusted series of logarithms of trade and price durations. The sample mean of seasonally adjusted log durations and the average value of each of the adjustment factors are equal to zero over the whole sample of thirty seven contracts, but not for each individual contract. To account for this we allow the conditional duration to have a non-zero mean and we subtract from each of the seasonal components their average values over that specific sample. We assume that the seasonally adjusted log durations follow the model

$$\begin{aligned} d_i &= \mu(\gamma) + \psi_i + \xi_i \\ \psi_i &= \omega + \beta\psi_{i-1} + u_i, \ |\beta| < 1 \end{aligned}$$

where $\{\xi_i + \mu(\gamma)\}$ has log-Weibull distribution with parameter γ, $\mathbf{E}[\xi_i] = 0$[8]. The parameters are estimated by maximising the QML, computed with the help of Kalman filter. Note that

$$\mathbf{E}\left[\frac{1}{N}\sum_{i=1}^{N} d_i\right] = \mathbf{E}[d_i] = \mu(\gamma) + \frac{\omega}{1-\beta}.$$

Making the change of variable $\psi^* = \psi - \frac{\omega}{1-\beta}$, the model can be written as:

$$\begin{aligned} d_i^* &= \psi_i^* + \xi_i \\ \psi_i^* &= \beta\psi_{i-1}^* + u_i, \end{aligned} \tag{2.10}$$

where ω has been concentrated out of the likelihood, as suggested in Ruiz (1994) for estimating stochastic volatility models. The vector of parameters of the quasi-likelihood to be minimized is $\theta = \{\sigma_u^2, \gamma, \beta\}$. The asymptotic theory for the QML estimate of θ was

[8]Our analysis as well as the results reported in Bauwens and Veredas (2004) show that the Weibull distribution is preferred to the alternatives such as the exponential or the log-normal distributions; two-parameter forms such as the Burr or the generalised gamma distributions do not noticeably improve the fit while increase the number of the parameters to be estimated.

developed in Dunsmuir (1979) yielding $T^{\frac{1}{2}}\left(\theta-\hat{\theta}\right) \overset{d}{\sim} \mathrm{N}\left(0, C\left(\theta\right)\right)$. The expression for $C\left(\theta\right)$ and details of the computations are given in Appendix B.1.

The estimation results presented in this paper are for the case in which the data are treated as continuous series, i.e. the end of one day precedes directly the beginning of the next except that we have cut out the first 20 minutes in the beginning of each day. We use a Kalman filter initialized with the diffuse prior to compute the quasi-likelihood function of the model (2.10). An alternative to this approach would be to re-initialize the Kalman filter at the beginning of every trading day. The asymptotic theory for the estimates obtained when the filter is re-initialised daily with a diffuse prior are presented in Appendix B.2. It is also possible to initiate the Kalman filter with a meaningful prior computed using auxiliary information. In this study, we do not pursue further strategies with daily initialisation of Kalman iterations.

2.5.3 Diagnostic methods

As far as it concerns the nonstochastic part of the model, our primary concern is whether the additive form of the seasonal component is supported by the data. We use graphical methods to investigate this: we estimate the diurnal component for each day of the week and draw them on the same graph; in the same manner we draw the life cycle seasonal component for different days of the week and for different periods of the day. If there is no interdependence between seasonal variables, the lines on each of these graphs will not be far apart one from another.

To assess how the SCD model describes the dynamics of seasonally adjusted durations we want to investigate two issues: first, how well the model accounts for the dependency properties of the process; second, how good are the parametric assumptions about the distributions of the innovations. Since the SCD model is a model with one latent variable, we can compute two series of residuals resulting from the model estimation:

$$\hat{\xi}_i = d_i^* - \hat{\psi}_i^* \text{ and}$$
$$\hat{u}_i = \hat{\psi}_i^* - \beta\hat{\psi}_{i-1}^*.$$

The serial dependence structures of the series $\left\{\hat{\xi}_i\right\}$ and $\{\hat{u}_i\}$ are similar one to another: we shall focus here only on the dynamics of $\left\{\hat{\xi}_i\right\}$ because we can compare it directly with the dependence in the duration series.

Under the model, the residuals $\left\{\hat{\xi}_i\right\}$ are distributed as log-Weibull with parameters $(\gamma, 1)$, and $\{\hat{u}_i\}$ have a standard normal distribution.

We use traditional analysis of the ACF and the partial autocorrelation function (PACF) as well as Spearman's coefficient to investigate the serial dependence in the residuals. Bauwens and Veredas (2004) argue that the Ljung-Box statistic (or similar statistics based on the sample autocorrelations) would not be a correct measure of dependence in the context of irregularly spaced data, which is why Spearman's coefficient should be preferred. This argument has its merits. However, within the framework of a discrete time model, the Ljung-Box statistic will detect the presence of linear dependence in the series, regardless of whether the measurements are taken at equal or irregular physical time intervals.

In order to judge the compatibility of our parametric assumptions with the empirical observations we use p-value plots and p-value discrepancy plots, as well as commonly used non-parametric goodness-of-fit tests such as the Anderson-Darling test and the Cramèr-von Mises test[9].

[9]These statistics, their properties and their critical values for normal and log-Weibull distributions can be found in (Stephens 1976) and (Stephens 1977).

2.6 Estimation results: interpretation and discussion

2.6.1 Estimated seasonality in trade and price durations

Before proceeding with the analysis of the seasonal pattern of the dynamics of durations we would like to discuss briefly certain descriptive properties of the data. Figure 2-1 presents the standard deviation of the data series corresponding to each of the contracts (right scale) as well as analysis of how this variation is explained by the seasonal component and the SCD model (left scale). The fact that the results are rather uniform across the contracts is encouraging: our model assumes that data corresponding to different contracts have similar statistical properties. The seasonal component explains about 10% of the total variation of the logarithm of trade durations; the SCD explains 30% of its variation. For the price durations, the seasonal component explains 10%-20% of the total variation, and the SCD model explains 50% - 60% of the total variation [10]. Overall, our model explains $\sim 40\%$ of the variation of the logarithm of trade durations and $\sim 65\% - 75\%$ of the variation of the logarithm of price durations.

Figure 2-2 presents graphs of the seasonal components A_δ, $\chi(t)$, and $\zeta(\tau)$ of trade durations (left column) and of price durations (right column). We observe that over a week, trading is the least active on Mondays with its intensity increasing towards Friday. This is in line with the results reported in the financial literature. Over a day, the trading activity is high in the morning, then it decreases gradually to its lowest level between 12:30CT and 13:00CT, and increases again near the end of the trading day. The observed diurnal pattern is similar to those described for stocks by Bauwens and Veredas (2004) or by Gouriéroux, Jasiak, and Le Fol (1999). The lowest levels of daily activity for the stocks studied by the aforementioned authors, Boeing in the former case and Alcatel in the latter, have been observed between 13:30ET and 14:00ET (Central time is equal to

[10]Our model assumes that the the seasonal component and the SCD dynamics of the data are orthogonal.

Eastern time minus one). The curious conclusion that we can draw is that there is a synchronicity between trading in Chicago and in New York, despite the difference in the time zones. The increase in the level of trading activity at the end of a day is more pronounced for the CME currency futures than for the stocks studied in the articles just mentioned. An explanation for this may be that the trading at the CME closes at 14:00CT, just one hour after the lunch break responsible presumably for the trough in trading activity, while the NYSE trades until 16:00ET. Therefore, the traders of CME currency futures have less time, compared to the traders of stocks on the NYSE, in the afternoon to take their end-of-the-day positions [11].

The seasonal pattern resulting from the contract life cycle is in accordance with our expectations. Trading is the most active for the contract closest to expiration, and the level of trading activity is relatively flat from 65 business days to about 5 business days to expiration. The closest to expiration contract is traded less actively in the last few days of its existence because the traders switch to the next contract. When the nearest to expiration contract has six business days to expiration, the second to expiration contract has usually (not always, because of holidays) sixty-nine business days to expiration. The change in the life-cycle seasonal component of durations between 70 and 60 days to expiration correspond to a more than tenfold change in the expected trade duration and to a change of about four times in the expected price duration. Traders take positions in the contract that is going to become the nearest to expiration. We do not study the trading dynamics beyond 130 business days to expiration because the trading there is very sparse, typically just a few trades per day, if any.

[11]The strength of this argument is mitigated by the fact that the CME futures are traded on GLOBEX 24 hours a day. The differences in the shapes of diurnal components for stocks and CME currency futures may be also due to institutional differences and/or differences in the mechanism through which the trades are reported.

2.6.2 Estimated parameters of the SCD model

We discuss next the results of estimation of the SCD log-Weibull model. Figure 2-3 shows estimated values of the parameters β, σ^2, and γ, for each of the contracts in the calendar order with two asymptotic standard deviation error bars. The left column shows trade duration parameters, the right - the price duration parameters.

A prominent feature of estimated parameters of trade durations is that asymptotic standard errors are narrow. Given the sample sizes (the number of observations per contract is in the range from $53,000$ to $135,000$), we believe that the asymptotic standard errors would be close to the true standard errors under the correct model specification. This assertion is supported also by the results of Monte-Carlo experiments reported in Bauwens and Veredas (2004), where the size of the simulated samples is up to $N = 50,000$. In their study, the simulated standard errors are very close to the asymptotic values. If we allow for the possibility that the model is misspecified and we want to evaluate how far apart are the estimated model and the data-generating process, this distance (in some metric) will have two contributing factors: the statistical error of estimation and the error due to the model misspecification. Tight asymptotic standard errors of the estimated model parameters and the fact that the asymptotic standard errors are close to the finite sample standard errors suggest that the first contributing factor, the model estimation error, is small. The argument just above is not so relevant for the model parameters of price durations because the sample sizes of price durations are smaller, hence, asymptotic standard errors are wider.

The estimates of β of the latent process lie between 0.961 and 0.987 for trade durations series and between 0.882 and 0.981 for price durations: i.e. both processes are very persistent. The values of β are significantly less than 1 at conventional levels. For trade durations, the estimated values of the parameter γ of the Weibull distribution are significantly less than one for the contracts before June, 1995 and significantly greater than one for later contracts. The price durations $\hat{\gamma}$ are between 1.1 and 1.33 and are always significantly greater than one. Whether $\gamma \geq 1$ or $\gamma < 1$ has implications for the possible shapes of the hazard function: in the former case it is possible that the hazard function is non-monotonous (increasing first and then - decreasing, constant or, for large values of σ^2, increasing, - while in the latter case the hazard is always decreasing).

The estimated values of σ_u^2 vary from 0.011 to 0.040 for trade durations and from 0.0057 to 0.019 for price durations. One can observe that the values of the model parameters β, γ, and σ_u^2 are significantly different, based on the asymptotic standard errors, across the contracts considered, both for trade durations series and for price durations. Qualitatively, however, we may conclude that the behaviour of the durations process as described by the SCD model is similar for all of the contracts studied: we observe high persistence and low signal-to-noise ratios (SNR)[12], ranging from 1.18 to 1.36 for trade durations and from 1.49 to 2.94 for price durations. As we shall see in a moment, a low SNR will have interesting implications from the point of view of model identification.

2.6.3 SCD parameters and the horizon to expiration

It is interesting to investigate the question of stability of model parameters across different horizons to expiration. There is a natural split in the trading data for each contract: the records when the contract is the nearest to expiration and the records when the contract is the second nearest. Stability of the model parameters over the horizon to expiration would support our approach to modelling in general and in particular, our algorithm of seasonal adjustment.

Consider first trade durations series. The left column of Figure 2-4 shows the estimated values of the SCD model parameters for contracts with horizon to expiration ranging from 70 to 130 business days. The values of the corresponding parameters estimated using the whole sample are given on the same graphs as a reference. A futures contract is traded much less actively when it is the second closest to expiration than when it is the closest to expiration, which is why the subsample corresponding to the trades with time to expiration from 70 to 130 days comprises only a small fraction of all

[12]We define SNR here as the ratio of the unconditional variance of $\{d_i\}$ to that of $\{\psi_i\}$.

trading records of a contract (less than 5% for some of the contracts).

The estimates of β and σ_u^2 for trade durations based on records with 1 to 130 days to expiration and on records with 70 to 130 days to expiration are close. Informal analysis suggests that the estimates of β in the former case are higher than in the latter; the estimates of σ_u^2 are smaller in the former case than in the latter. This difference in the estimated parameter values can be explained, in part at least, by our treatment of the data as continuous series. The second to expiration contract, as has been mentioned, is traded much less actively than the closest to expiration; hence, the links between intra-day spells of trades constitute a larger proportion of the data in the former case. The conditional distribution of a duration will intuitively depend less on the previous measurement if this measurement has been taken at the end of the previous trading day, but we do not account for this in the model. That is why we can expect lower persistence for contracts with longer horizons of expiration within our modelling framework, and also a higher variance of innovations of the latent process.

The estimates of trade durations γ based on records with 70-130 days to expiration are lower than those based on contracts with 1-130 days to expiration.

The SCD parameters of price durations for different horizons to expiration differ less than those of trade durations. We still observe that the estimates of the persistence parameter are lower at longer expiration horizons, and we can use the same rationale as for the trade duration to explain why this is the case. The estimates of σ_u^2 and of γ of price durations based on records with longer expiration horizons are not statistically different from those based on all records.

Price durations change less with the expiration horizon than do trade durations. Thus, the proportion of links between trading days in the price durations data does not increase as much, when we consider only trades with 70 to 130 business days to expiration, as in the trade durations data. This may be one of the explanations why the estimated SCD parameters of price durations differ less across expiration horizon than do those of trade durations. Another explanation may be that the dynamics of trade durations depend more on the specifics of the trading system than the dynamics of price durations; the model used in this study accounts for the properties of the trading mechanism only in very general terms.

The dynamics of price durations do depend on the trading mechanism, however. A futures price follows very closely (virtually one-to-one) the price of the underlying asset. One could expect that price durations would not change as a function of the horizon to expiration, but we definitely observe life-cycle seasonality in the price durations, albeit weaker than in trade durations. A possible reason for the presence of this life-cycle seasonality in price durations is that the number of futures contracts in circulation is smaller when the contract is the second to expiration than when it is the closest to expiration. This latter property, the number of contracts in circulation, is more closely related to the transmission mechanism than to the information process determining the dynamics of the "latent" futures price.

2.6.4 Specification diagnostics

Figure 2-5 shows the estimated additive contributions of the life-cycle and diurnal components for each day of a week and the contributions of the life-cycle components for three two-hour periods of a trading day. We can conclude from the visual examination of these graphs that the assumption of additive form of the seasonal component is compatible with the data both for trade durations and price durations.

Figures 2-6 and 2-7 illustrate how the SCD model accommodates the dependence properties of the seasonally adjusted series of correspondingly trade and price durations[13]. The measures of dependence for the seasonally adjusted logarithms of durations are in the left columns of figures 2-6 and 2-7, the measures of dependence for the model residuals are in the right columns[14]. Both trade and price durations exhibit strong dependence, as

[13]We use the series corresponding to a specific contract to illustrate the dependence properties of the data, the model residuals, and the goodness of fit of the parametric distributional assumptions. The results are representative of all contracts considered.

[14]We use $\{\hat{u}_i\}$ residuals to compute the ACF, PACF, and Spearman's coefficient shown.

has been documented in the financial literature. The SCD model fails to account fully the dependence properties of the data: the residuals retain a degree of dependence.

Bauwens and Veredas (2004) found that the residuals of the SCD model with an $AR(1)$ latent process, estimated using trade durations of a stock, are not independent (the authors used Spearman's ρ statistic as a measure of dependence). They mentioned as a possible explanation, citing Jasiak (1998), that trade durations may be fractionally integrated. Our preliminary analysis also suggests the presence of long memory in the duration series (the estimates of the fractional integration parameter x of the FISCD model introduced in the appendix are between 0.4 and 0.6).

One may hope that a mechanical increase of the order of the latent process will allow the model to better accommodate the dependency properties of durations. This simple approach does not work as well as one might have expected. The reason lies in the structure of the model. The resolution of a system registering a mixture of signal and noise, i.e. its ability to distinguish between different signals (latent processes), as it is known in the theory of signal processing, depends among other factors on the geometry of the space of the solutions and on the signal-to-noise ratio. Given the parameter values typical for our data, the SNR is low, especially for the trade duration series. If we increase the order of the latent process to $AR(2)$, we increase the domain of the possible solutions. In the presence of white noise, the QML estimation algorithm used loses its resolution abilities primarily at higher frequencies, but this is exactly where the $AR(1)$ and $AR(2)$ differ from one another.

To illustrate the argument above we compute the inverse of the information matrix (see equation (B.1) in Appendix) of ML estimates of the parameters of a Gaussian $AR(2)$ process measured with Gaussian noise (we ignore the correction for non-normality of the measurement noise for the sake of transparency of the exposition). The values of the parameters used in this example are $\beta_1 = 0.95, \beta_2 = 0.02, \sigma_u^2 = 0.02$ and $\gamma = 1$ (γ is used to compute the variance of the measurement noise, $\sigma_\xi^2 = \frac{\pi^2}{6\gamma^2}$); β_i are the parameters of the $AR(2)$ latent process[15]. The inverse information matrix for the model parameterised as $\theta = \{\beta_1, \beta_2, \sigma_u^2, \gamma\}$ is:

$$\mathcal{IF}^{-1} \cong \begin{pmatrix} 5323356.3 & -5166653.3 & -208937.10 & -32404.133 \\ -5166653.3 & 5014563.3 & 202786.57 & 31450.200 \\ -208937.10 & 202786.57 & 8200.6707 & 1271.8867 \\ -32404.133 & 31450.200 & 1271.8867 & 197.83256 \end{pmatrix}.$$

We observe that the estimates of β_1 and β_2 have very high variance and the correlation between them is almost -1 (this will be especially true when β_1 is close to one and β_2 is relatively small). In practical terms this means that we cannot distinguish changes in β_1 from the changes in β_2 (we can estimate well the quantity $(\beta_1 + \beta_2)$, however).

Compare now the inverse information matrix above to that of the $AR(1)$ process measured with white noise. We assume that $\beta = 0.95$ and that the values of γ and σ_u^2 remain unchanged:

$$\mathcal{IF}^{-1} \cong \begin{pmatrix} 0.177 & -0.0885 & -0.0591 \\ -0.0885 & 0.0778 & 0.0555 \\ -0.0591 & 0.0555 & 0.584 \end{pmatrix}.$$

The difference is striking. Keeping in mind that the inverse of the Fisher matrix sets the lower bound to the norm of the variance-covariance matrix of the estimates, we see how much the uncertainty in parameter estimates is increased by extending the class of possible latent processes from $AR(1)$ to $AR(2)$.

The analysis above shows that the problem of identifying the structure of the latent process of the SCD model using the QML approach has properties which make it similar to an ill-posed problem. It is not an ill-posed problem in the strict sense of the definition because the unique solution exists, and for any given accuracy there is a sample size at

[15]It is possible to compute the components of the inverse Fisher matrix analytically but the expressions are very bulky. Therefore, we present a numerical illustration.

which this accuracy can be achieved. For practical purposes, however, acceptable variance of the parameter estimates can be achieved only by restricting the space of possible solutions to the class of $AR(1)$ models. Restricting the space of possible solutions is the standard approach to solving ill-posed problems.

It has been noticed above that the asymptotic standard errors of the SCD model when the latent process is $AR(1)$ are very narrow, and the goodness of fit is determined primarily by how well the model is specified. We see that the class of SCD models with the $AR(2)$ is too wide given the information available which results in very large asymptotic errors. Intuition suggests to look for a model in a class more flexible than the SCD with $AR(1)$ latent process but which would have a structure of the latent process different from $AR(2)$. The FISCD model introduced in Appendix B.3 is an attempt to find such class of models.

We turn next to investigating parametric assumptions about distributions of the innovations. Analysis of the p-value plots and of the p-value discrepancy plots of the empirical distribution of $\left\{\hat{\xi}_i\right\}$ against the log-Weibull distribution, shown in figures 2-8 and 2-10, does not indicate gross incompatibility of the adopted parametric form either for the trade durations data or for the price durations data. The shape of the p-plots is very similar to the shape observed in Bauwens and Veredas (2004) for trade durations of Boeing stocks.

Analysis of the p-value plot of the empirical distribution of estimated innovations $\{\hat{u}_i\}$ against the normal distribution suggests a distinct departure from normality both in the case of trade durations and of price durations (Figure 2-9). The empirical distribution has fatter tails than the normal distribution. Our observations with respect to the tails of the empirical distribution of $\{u_i\}$ are opposite to those reported in Bauwens and Veredas (2004) for the trade and price durations of the Boeing stocks; in the latter case, the empirical distribution had thinner than normal tails.

Formal goodness-of-fit tests based on the Anderson-Darling statistic and Cramèr-von Mises statistics reject the parametric distributional assumptions of the model at any conventional level. We expected that the parametric assumptions would be rejected by these tests because they would have a high power given a typical size of the sample and because our model was capable, by design, of capturing only the most general features of the data. However, a closer look at the behaviour of the goodness-of-fit statistics illuminates directions for improving the model and for further research.

Our model, which treats the data as continuous series, does not describe well transitions from one day to another. The values of the goodness-of-fit statistics by an order of magnitude if we censor from the samples the residuals corresponding to initial and final moments of trading in every day (the statistics still remain in the rejection region however). These quantitative results confirm the graphical analysis above: the rejection of the assumption of normality of $\{\hat{u}_i\}$ is overwhelmingly stronger than the rejection of the assumption about the parametric form of $\left\{\hat{\xi}_i\right\}$, the latter still having p-values of less than 1%.

2.6.5 Discussion

Our empirical analysis suggests several ways to improve the performance of the model. When estimating the seasonal components non-parametrically, we chose the bandwidth parameters based on visual analysis of the graphs. A more formal approach based, for example, on cross-validation, is possible. The difficulty with automated choice of the bandwidth parameters is related to the fact that the methods of automated selection break down if the errors are dependent. There are few methods that can handle the dependent data (for example, Francisco-Fernandez, Opsomer, and Vilar-Fernandez (2003)) but they are relatively complicated algorithmically and have been developed only for univariate nonparametric regression. Thus, even if we adopted an automated algorithm to choose the bandwidth, our technique would still remain ad hoc, but become much more complicated algorithmically.

Treating the data as continuous series may be inappropriate when we study the trade durations of contracts with longer horizons to expiration, where we typically have just a few records per day, or when we study price durations. The asymptotic theory developed in Dunsmuir (1979) is directly applicable to Kalman QML estimation provided that the filter is initialized with the diffuse prior. We have amended the asymptotic theory of QML estimation in such a way that it is applicable to the case in which the data consists of a set of independent subsamples (see Appendix B.2). We have investigated two alternative approaches: we initialized the Kalman filter at 8:00 each day using the average logarithm of trade durations (seasonally adjusted) between 7:40 and 8:00 and we initialized the Kalman filter at the beginning of each day with the sample mean. The estimation results were very close for all three methods when we analyze the whole data set, especially in the case of trade durations. This is because the records corresponding to the nearest to expiration contract constitute a larger part of the data, and the number of records per day is large, hence, the initialization of the Kalman filter affects only marginally the value of the quasi-likelihood function. We would like to observe in the end that the initialization of the Kalman filter at the beginning of every day provides us with a tool for introducing, in a non-trivial way, the information accumulated overnight.

We observed that the empirical distribution of $\{\hat{u}_i\}$ departs from the normal distribution, especially for the trade durations series. We expect that this problem will be mitigated if daily initialization of the Kalman filter is used: large overnight innovations may be responsible in part for fatter tails in the empirical distribution of $\{u_i\}$. Using a parametric distribution with fatter tails to model the innovations of the latent process may also help to improve the empirical behaviour of the model: Student's t distribution is a good candidate (the model with normal innovations of the latent process is nested into the Student t parametrisation).

2.7 Conclusions

We have mentioned in the introduction that the latent economic process can be viewed as manifesting itself through the transmission mechanism of the institutional structure of trading. The relatively simple model used in this study is far from describing the economic process and the transmission mechanism in detail; it captures dynamic properties of the data without revealing the structure of the data-generating mechanism. Nevertheless, if we adopt a constructive approach to modelling the generating mechanism for high-frequency financial series, a parsimonious description of the output signal given by the SCD model can be a valuable resource in synthesising the transmission function, which, given as an input a signal described by one of the existing models of asset price dynamics, would generate output with dynamics similar to that of the empirical point processes investigated.

We believe that the synthesis of models of trading mechanisms, which would bridge the gap between the dynamic financial models and the empirical models of high-frequency financial series, will be a promising area of research. Designing a realistic model of trading would be a very complex task and would require substantial resources. Even a simple stylized model, however, may provide further insights into the microstructure of financial markets. Imagine a latent price process following a stochastic volatility model and that a new transaction occurs when the latent price deviates from the last observed price by a given margin[16]. This model has a continuous-time process as an input and a point process as an output and is probably the simplest conceivable model of the trading mechanism. Empirical evidence suggests the presence of long memory in the volatility process and in the durations process. We conjecture that in the model just described long memory in the volatility of the latent price process will translate into long memory in the durations process.

Summarising the empirical findings of this study, we observed that while the estimated parameters of the SCD process are statistically different from contract to contract, qualitatively the process does not change much over the years studied. This is an indication that the model captures certain invariant properties of the economic process and the

[16]A framework with informed and uninformed traders can be used to explain the liquidity of this stylised market.

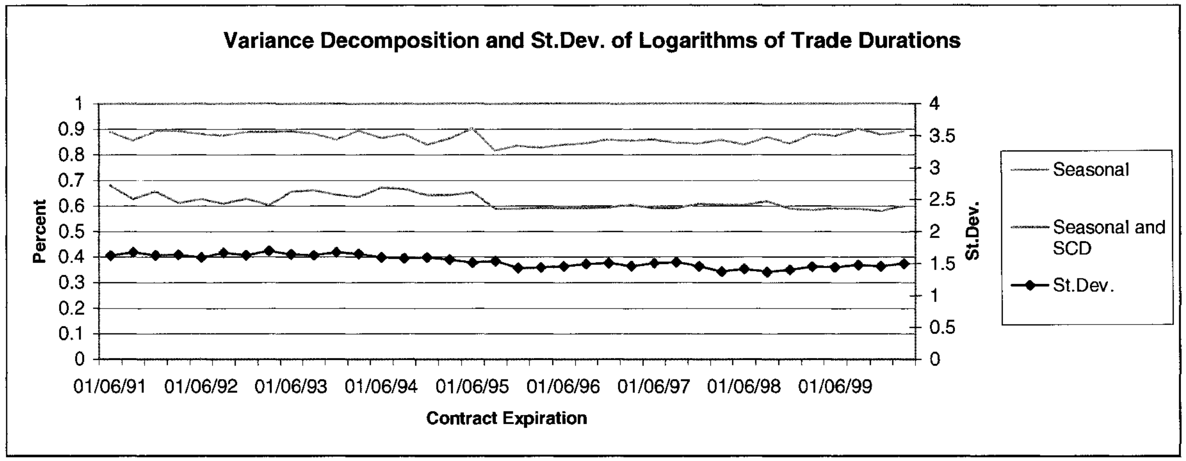
transmission mechanism. We observed that the SCD parameters, with a certain leeway, are also stable across the expiration horizon, the last statement being more accurate with respect to the price durations process. A price of a futures contract has almost a functional relationship with the price of the underlying asset. From this it follows that the life-cycle seasonality of price durations of futures contracts is determined to a great degree by characteristics of the trading mechanism rather than by only the properties of the latent economic process.

Preliminary analysis shows that the fit of the model improves noticeably if the overnight and the weekend interruptions in trading are taken into account. The simplest way to do this is to assume that in the beginning of every trading day the durations process is initialized with the unconditional mean. This approach has the advantage that the existing asymptotic theory of QML estimators can be applied with minor modifications (see Appendix B.2). Alternatively, the durations process can be initialised at the beginning of every trading day using auxiliary information available to the econometrician. Designing various initialization procedures is not a very interesting topic for academic research. However, the initialisation of the process will be crucial in any practical application of the model. The difficulty that one faces when using an informative prior to initialize the process in the beginning of every trading day is that the asymptotic theory for the estimates does not follow directly from the results of Dunsmuir (1979), since the essential assumption of ergodicity is violated, and the theory would have to be developed from scratch.

The science of signal processing has traditions of the analysis of the maximal achievable resolution of a system and of informational analysis of a transmission channel. We believe that the econometrics of high-frequency financial data can build on these traditions. The simple example of informational analysis of the SCD model given in this study is a modest contribution to this interesting direction of research. We illustrate practical limitations, given the model structure and the information available, of our ability to estimate and/or identify the signal (the latent process), and how these limitations can be discovered through the analysis of the informational structure of a system, comprising the model and the estimation algorithm (our analysis of the SCD model applies also to stochastic volatility models which have a similar mathematical structure). Our estimation and identification capabilities can be improved either by introducing new a priori information[17] or by using an estimation algorithm, if one exists, that makes better use of the existing information. In economics, choosing an alternative model of the transmission channel and of the signal itself can often be productive, in contrast to the natural sciences where due to the established methodological paradigm, the acceptable choice of models is restricted.

Because empirical evidence suggests long memory properties in the dynamics of durations, a model of durations should be able to accommodate such a possibility. It is not a very difficult task to estimate the long memory parameter; it is more difficult to estimate both the high-frequency dynamics and the long-memory of the process. The FISCD model provides a parametric framework which allows, in theory, the modelling of both the high frequency dynamics and the low frequency (long memory) properties. QML estimation of the FISCD model in the spectral domain is simple algorithmically and computationally. Investigation of the properties of the QML estimates of the parameters of the FISCD, beyond strong consistency, which is known to hold, remains, however, a challenging theoretical problem.

[17] An example of such a priori information is: "The latent process is an $AR(1)$.



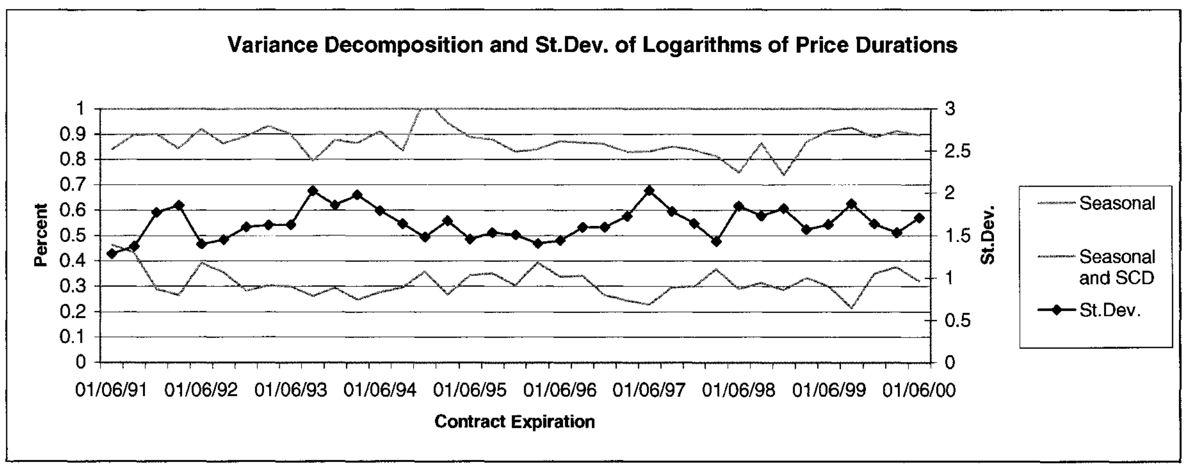


Figure 2-1: Variance Decomposition

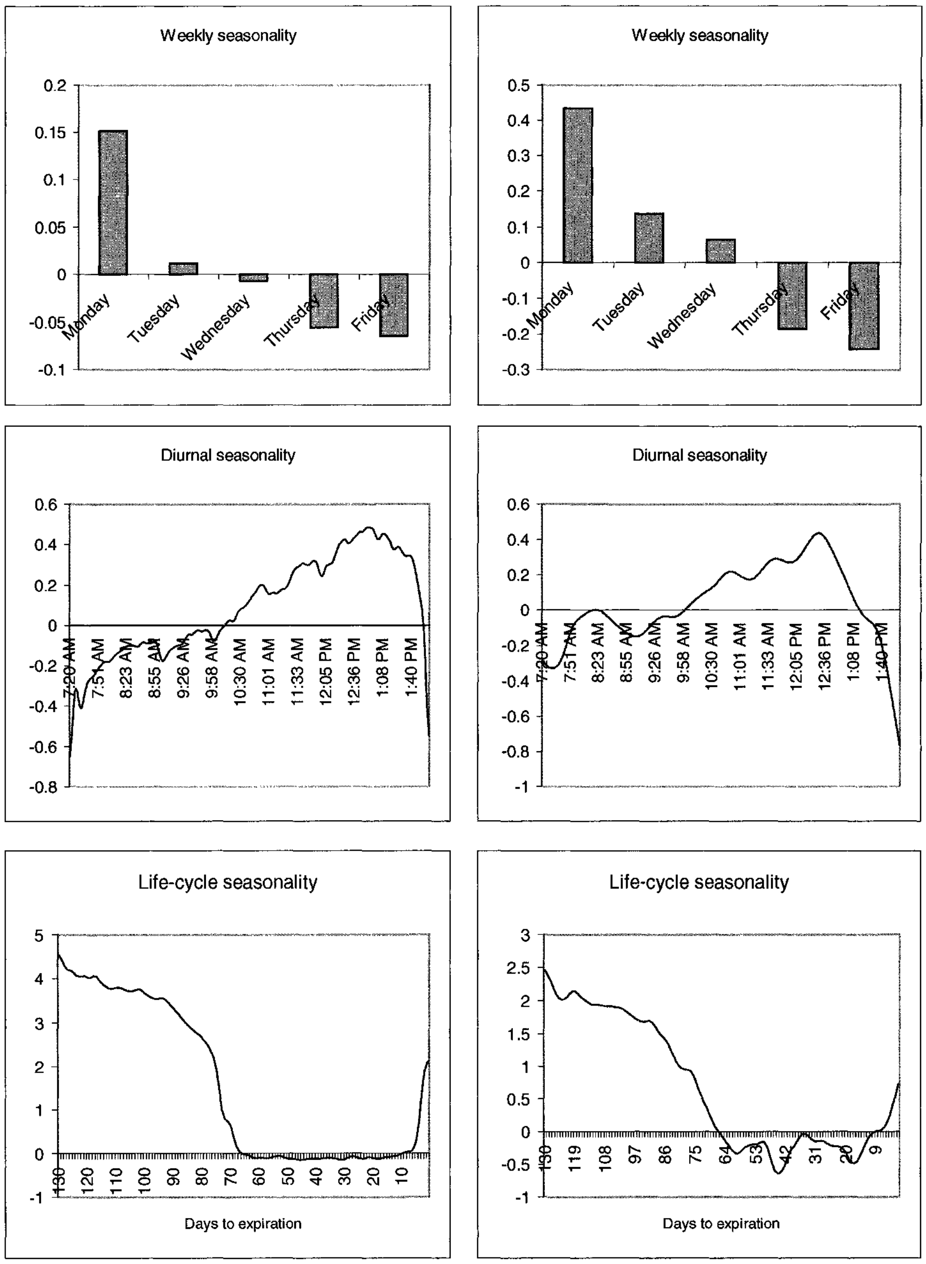


Figure 2-2: Estimates of the seasonal components $\hat{A}_{\delta}$, $\hat{\chi}(t)$, and $\hat{\zeta}(\tau)$ of trade durations (left column) and of price durations (right column)

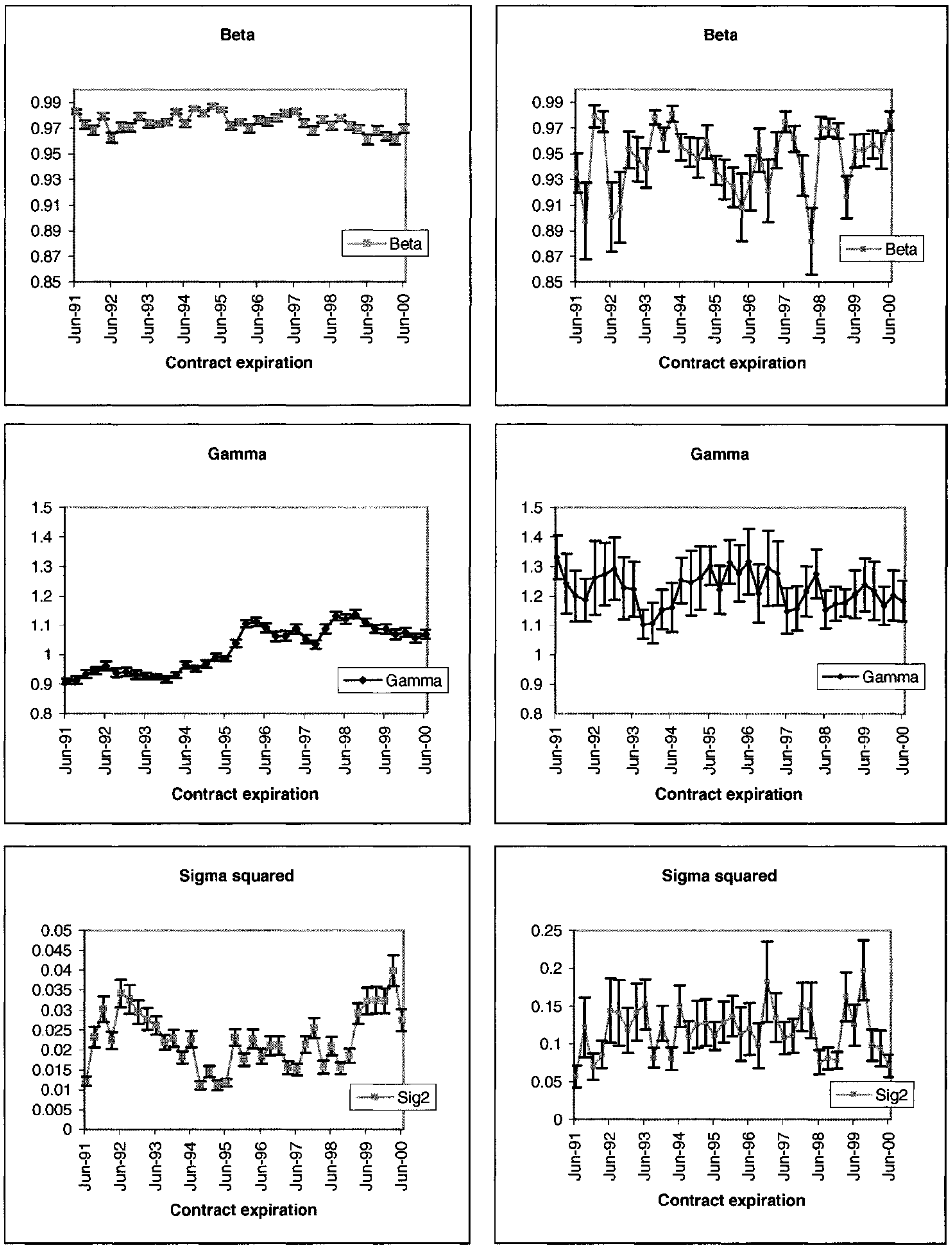


Figure 2-3: SCD parameters for contracts with different expiration dates. Left column - trade durations, right column - price durations. Error bars correspond to two SE.

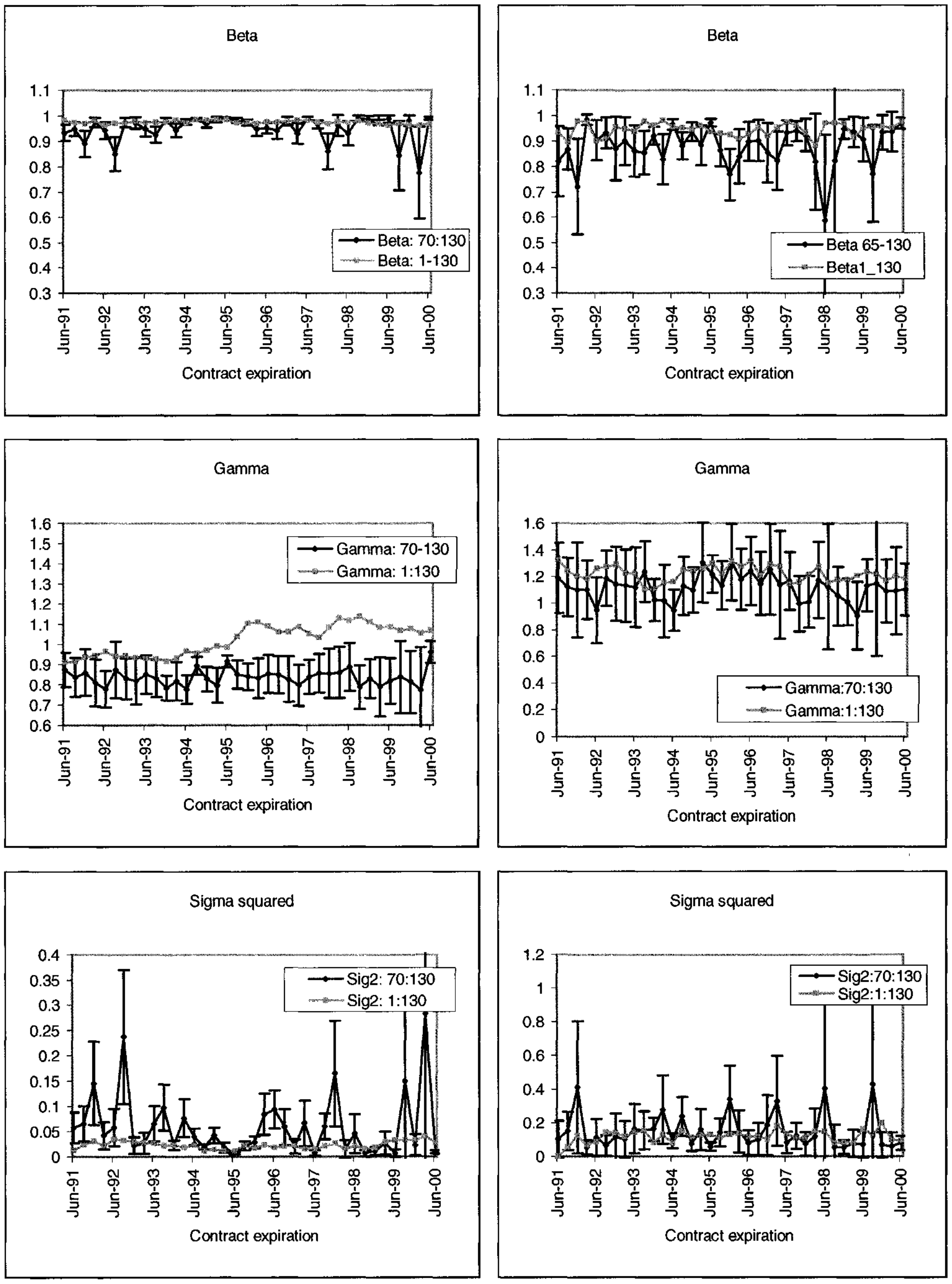


Figure 2-4: SCD parameters estimated using records with 70-130 business days to expiration. Left column - trade durations, right column - trade durations. Error bars correspond to two SE.

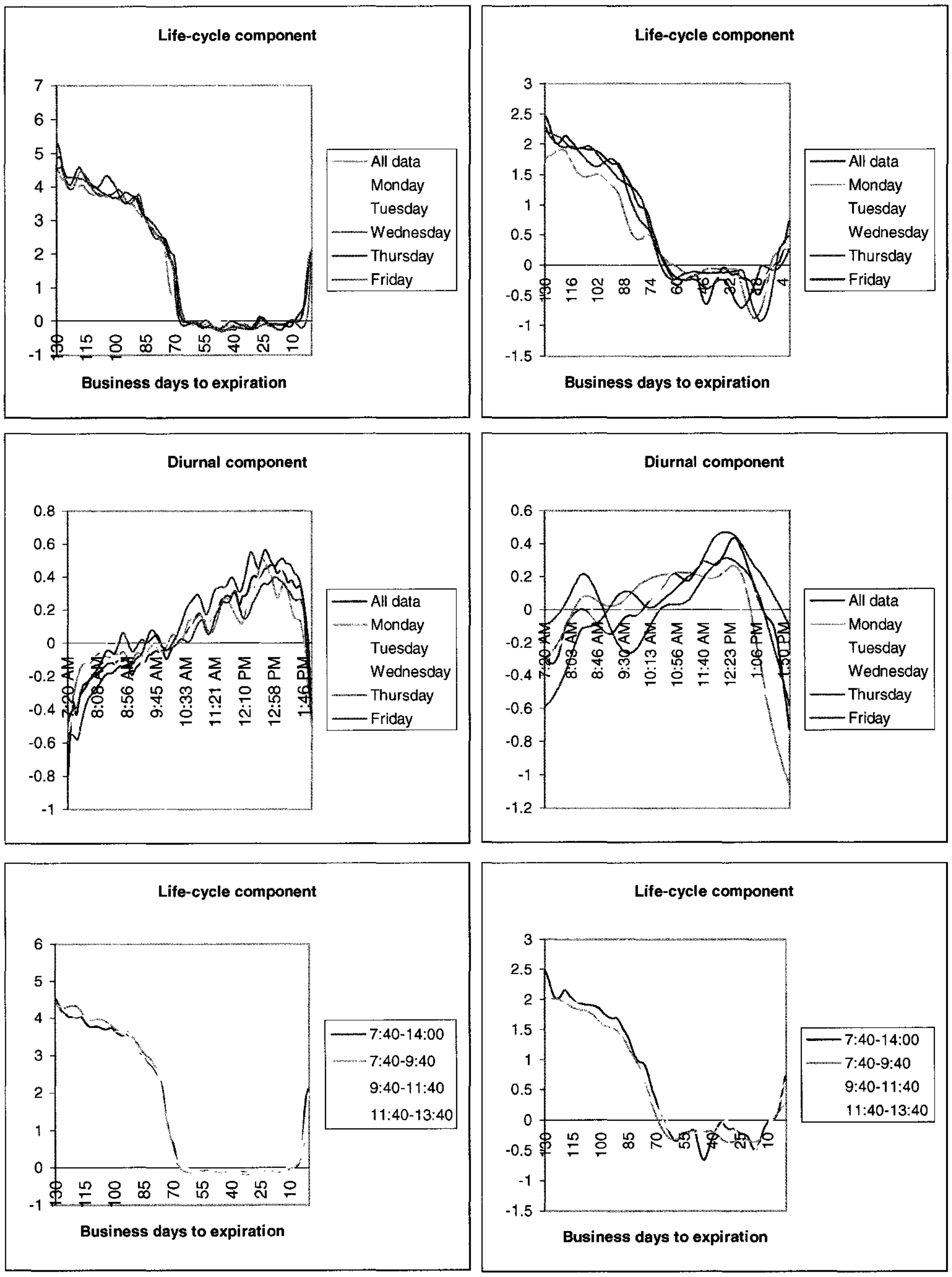


Figure 2-5: Graphic verification of the additivity property. Top two rows - life-cycle and diurnal contribution when the data is separated according to weekday. Bottom row - life-cycle contribution estimated for different periods of the trading day. Left column - trade durations, right - price durations.

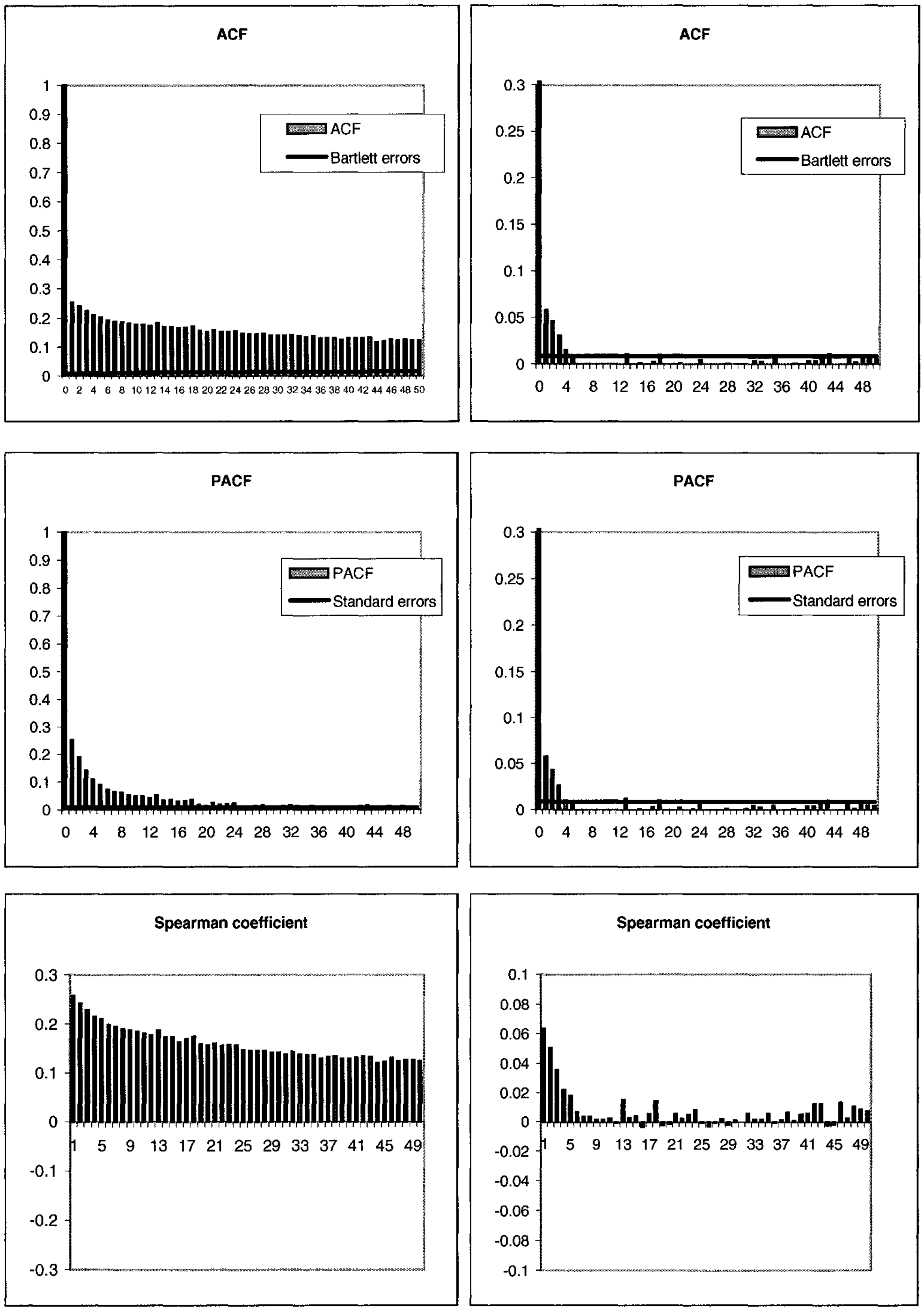


Figure 2-6: Measures of dependence for trade log-durations (left column) and for SCD residuals (right column).

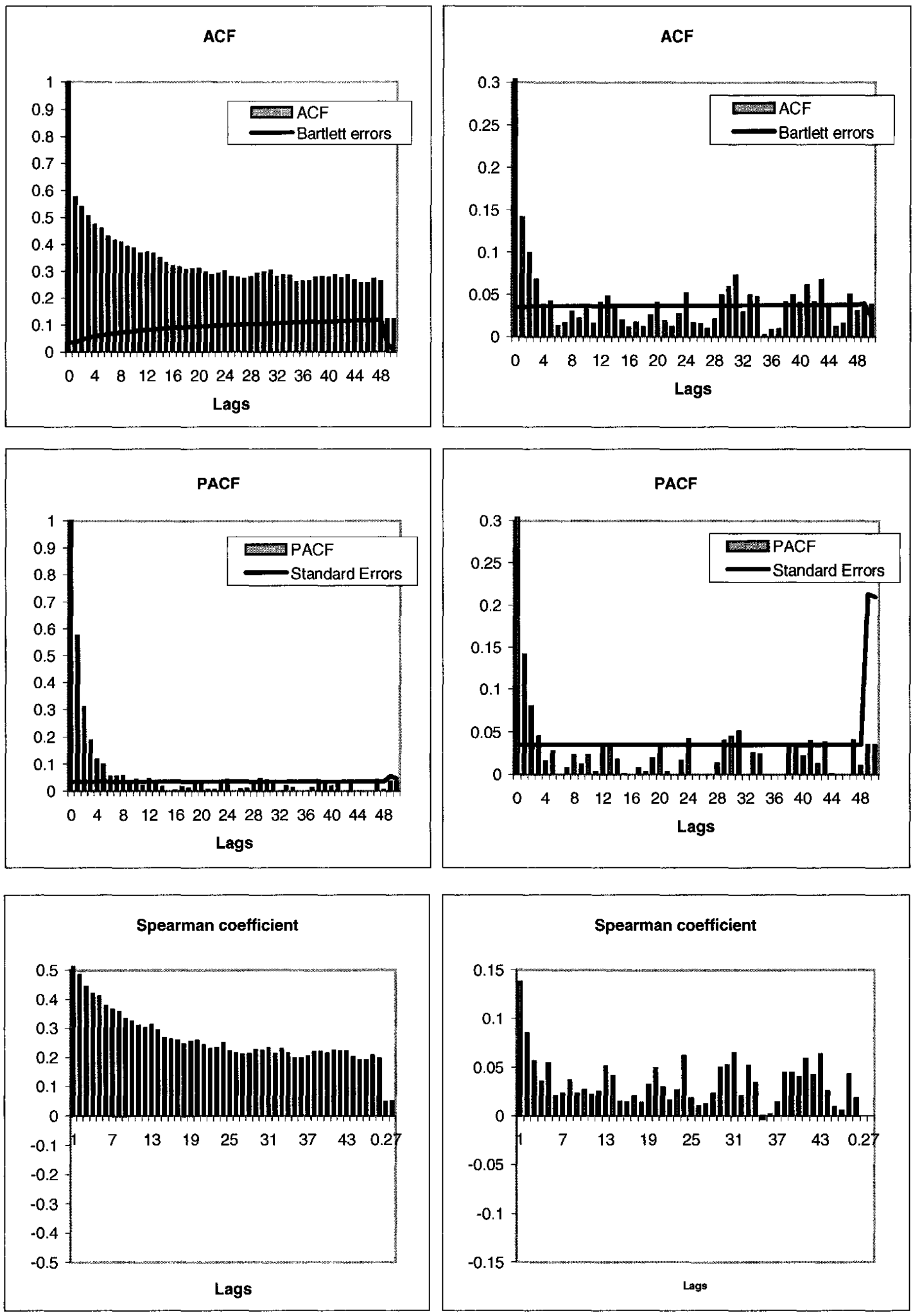


Figure 2-7: Measures of dependence for price log-durations (left column) and the SCD residuals (right column).

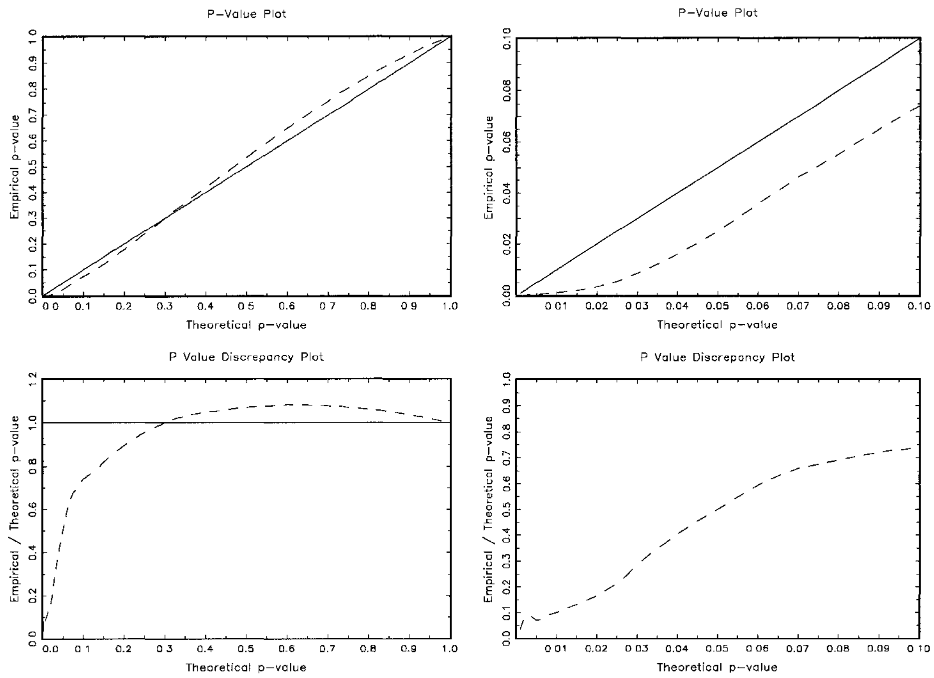


Figure 2-8: Trade durations. Goodness of fit of the empirical distribution of $\hat{\xi}$.

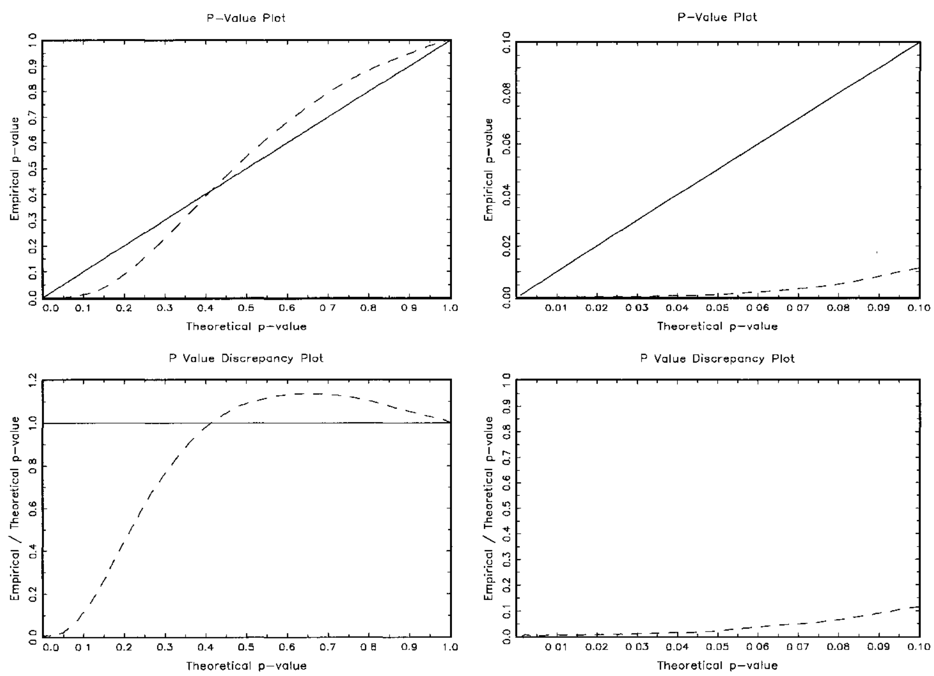


Figure 2-9: Trade durations. Goodness of fit of the empirical distribution of $\hat{u}$.

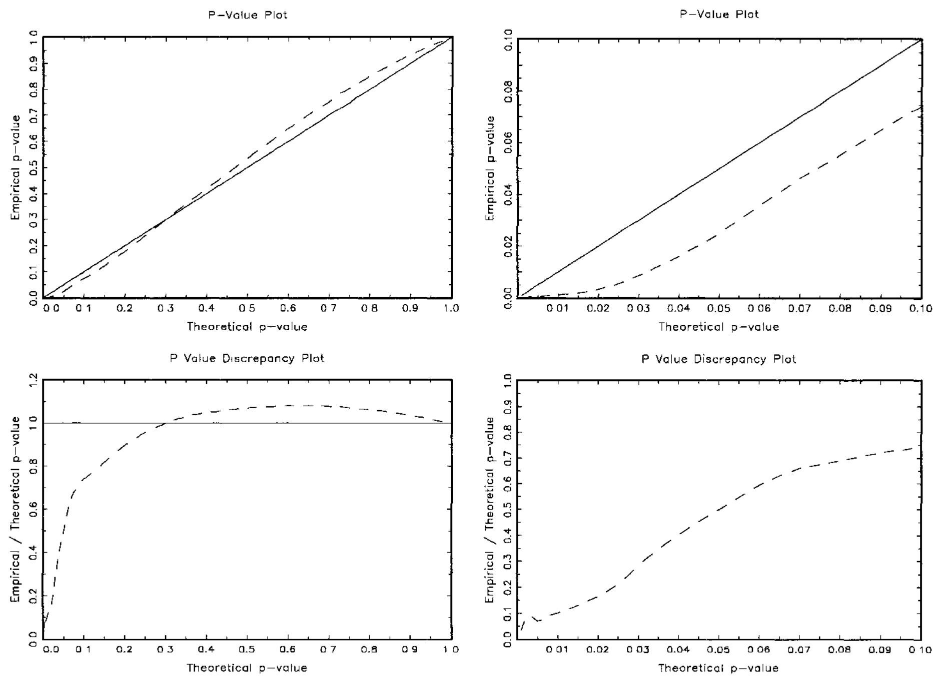


Figure 2-10: Price durations. Goodness of fit of empirical distribution of $\hat{\xi}$.

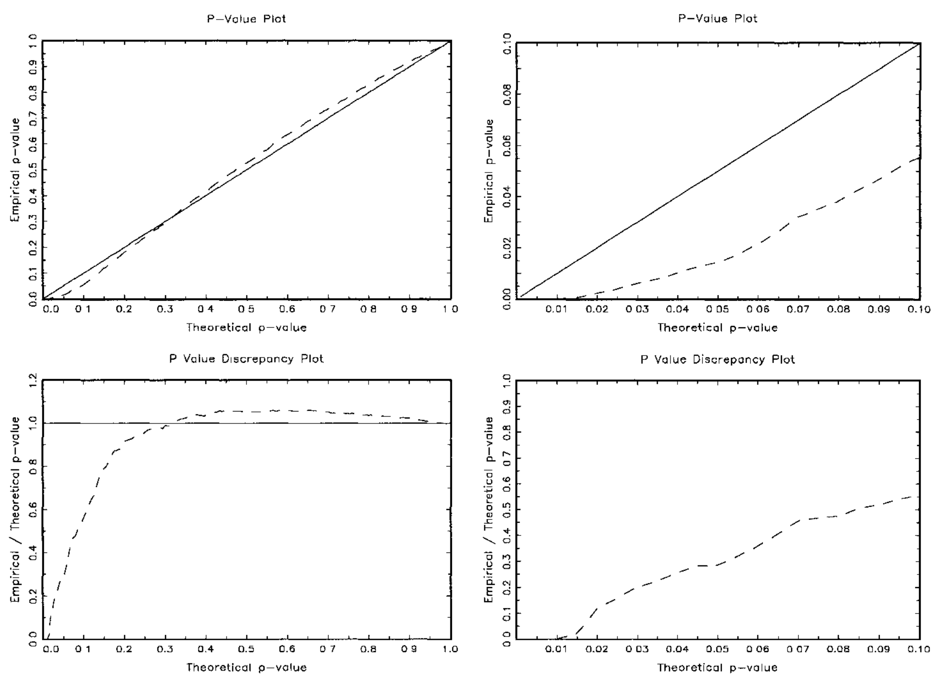


Figure 2-11: Price durations. Goodness of fit of empirical distribution of $\hat{u}$.

Chapter 3

Estimation of Conditional Quantiles of Variance Using Auxiliary Variance Information[1]

3.1 Introduction

A majority among researchers in finance and economics will agree that financial asset returns are not meaningfully predictable. It is an established empirical fact, however, that the volatility of the returns is very predictable. Finally, dynamic models of the higher moments of asset returns are still in their infancy. It is not surprising then that the bulk of dynamic models in financial econometrics are models of the dynamics of the second moment of asset returns[2].

Volatility is not observable directly, and until very recently our ability to investigate the dynamics of volatility was limited to estimating parametric models such as GARCH or stochastic volatility, where the volatility was a latent variable. Despite the great theo-

[1]This chapter is based on a joint article with John W. Galbraith and Vicky Zinde-Walsh.

[2]We shall use the terms *a model of volatility* and *a model of variance* interchangeably if this does not cause ambiguity. We reserve, however, the term *volatility*, to denote the square root of variance.

retical appeal of these models and the enthusiasm in the profession when they appeared, many deficiencies of these models were soon revealed. In particular, (G)ARCH-type models, while statistically highly significant, behaved poorly in predicting the variability of asset returns, measured by squared returns; stochastic volatility models, both in discrete and in continuous time, were notoriously difficult to estimate and did not offer any breakthrough improvement in predictive power either.

Merton (1980) was probably the first to mention in the economic literature that a precise estimate of the variance of Brownian motion over a period can be obtained if the sampling interval over this period approaches zero. This estimate is the sum of squared returns and is called *realised quadratic variation (RV)*. This knowledge had been virtually dormant until very recently, when Andersen and Bollerslev (1998) pointed out that intra-day data contains useful information about the variance of asset returns, and that this information can and should be used when gauging the empirical performance of dynamic models of variance[3]. It has also been recognised that intra-day data are useful for model estimation.

Andersen, Bollerslev, Diebold, and Labys (2001) (ABDL hereafter) have extended the result that RV converges to a certain measure of true variance - *quadratic variation* (QV) - from simple Brownian motion to the class of semi-martingale processes. RV computed using intra-day data was treated at first as if it were equal to QV (this includes the two articles by Andersen, Bollerslev and co-authors just mentioned). It is very appealing because under such an assumption QV becomes observable instead of being latent. The researchers have recognised, however, that RV is only an estimate of QV and that it is important to take into account its statistical properties when estimating models of volatility or gauging the performance of such models.

The error in RV as a measure of QV has two main components. First, even if the DGP satisfies the assumptions formulated in ABDL necessary for RV to be a consistent

[3]This discovery has coincided, incidentally or not, with the wider availability of intra-day market data and of the computing power necessary to process such data.

estimate when sampling frequency goes to infinity, for every finite sampling frequency there remains an *estimation error.* Second, any empirical high-frequency financial series are affected by market microstructure effects: bid-ask spread, discreteness of price quotes, irregular timing of quotes, and other factors, i.e., there is a *measurement error.*

Our knowledge about properties of the estimation error has improved significantly in the last few years; very important are the contributions of Barndorff-Nielsen, Shephard and co-authors, who have published a series of articles on the topic (see, for example, Barndorff-Nielsen and Shephard (2001), Barndorff-Nielsen, Nicolato, and Shephard (2002), Barndorff-Nielsen and Shephard (2002b), and Barndorff-Nielsen and Shephard (2002a)); another recent paper on the topic is Meddahi (2002b). The research by Barndoff-Nielsen and Shephard develops, in particular, the asymptotic theory of RV (and more generally, of realised power variation) and second-order properties of RV for certain classes of processes including stochastic volatility processes.

Our knowledge of properties of the measurement error is more limited than of those of the estimation error. These properties depend on the institutional structure of the market where the trading occurs and are difficult to formalise; in addition, many of these properties are market-specific, i.e., it is difficult if not impossible to develop a general theory of this measurement error.

The necessity to account for the measurement error has been recognised in the QV literature: ABDL, for example, notice that it may be suboptimal to use the highest available frequency of the data: while the estimation error decreases with an increase in the sampling frequency, the measurement error tends to "average out" at lower sampling frequencies. The recent article by Aït-Sahalia and Mykland (2003) considers the case in which the signal (the price process) is measured with additive i.i.d. noise, independent of the signal. They show that when the sampling frequency goes to infinity, RV may estimate consistently the variance of noise instead of the quantity of interest - the QV. The authors offer an algorithm to choose the optimal, in the MSE sense, sampling frequency in the case in which the measurement noise is not modelled; they argue also that modelling the microstructure noise often benefits the estimation even if the model is misspecified. The problem of optimal sampling in the presence of microstructure effects is also being solved in Bandi and Russell (2003).

The results of the research just mentionned promise a substantial improvement in the estimation of QV when the price process is measured with noise, but they will not eliminate the presence of the measurement error, which is why it is important to develop dynamic models and estimation methods robust to the specification of the error.

As we have noted above, use of RV facilitates and improves estimation of dynamic financial models, both in continuous and discrete time. It has been suggested in the recent literature to use the conditional moments of QV to estimate continuous-time stochastic volatility diffusion or jump-diffusion models (see, for example, Bollerslev and Zhou (2002) and Garcia, Lewis, and Renault (2001)). Care should be taken, however, to ensure that algorithms based on fitting asymptotic moments of QV are robust to the presence of measurement error in the empirical QV.

While continuous time stochastic volatility models are more convenient analytically for the purpose of derivative pricing, they are narrowly parameterised and often do not fit the empirical data well. Incorporation of jumps and of several latent variables improves the empirical fit but also complicates the estimation (it is recognised that identification and estimation of the jump component poses especially difficult econometric problems). For the purposes of asset pricing or risk management, if the price process is a continuous time stochastic volatility process, the object of interest of the econometrician is often not the instantaneous volatility but QV. Modelling directly the dynamics of QV as a discrete variable and the estimation of such models using realised variance series has been a subject of increasing interest in recent research. It has been shown (see, for example, Barndorff-Nielsen and Shephard (2002a), and Meddahi (2002a) among others) that certain classes of continuous time stochastic volatility processes imply ARMA dynamics of the QV and RV series, and that there is a correspondence between the number of latent volatility variables in the continuous time model and the order of the corresponding ARMA processes[4]. ARMA dynamics of QV may encompass a class of diffusion or jump-diffusion processes as possible DGPs, i.e., the discrete modelling provides a more flexible framework than that of the diffusion models which may serve as its generator[5] [6]. There is also a connection between ARMA dynamics of RV and weak GARCH processes: a weak GARCH process implies that the RV processes sampled at various frequencies are ARMA.

If it is assumed that the DGP is a well-behaved stochastic volatility or GARCH process, then the variance of the innovations in the ARMA presentation of RV exists (also the innovations are, of course, not Gaussian and moreover, they do not necessarily follow an m.d. sequence - see Meddahi (2002b) and references there) and it is a function of the parameters of the DGP. Then the ARMA can be estimated using QML or some other standard method, and the ARMA presentation can be used to estimate the parameters of the assumed continuous-time DGP. This approach is not legitimate, however, if there is measurement noise with unknown properties.

The objective of our research was to develop an estimation and forecasting methodology robust to the specification of the estimation noise and the measurement noise in RV. It is not possible, of course, to develop a model with no assumptions about the nature of the DGP, and the robustness does not come without costs - a loss of efficiency is one of them. We believe nevertheless that the technique we suggest complements the existing methods of modelling the dynamics of realised quadratic variation and quadratic variation. Our robust technique may offer superior performance in cases in which the gain in efficiency due to a more rigid model structure is more than offset by the misspecification error; our technique can set an empirical "reality check" benchmark for alternative approaches. We think also that thanks to robustness and computational simplicity our

[4]This research can be viewed as a continuation of the agenda started in Drost and Nijman (1993) and Drost and Werker (1996) on temporal aggregation of heteroskedastic processes and on bridging the gap between discrete time and continuous time modelling of financial processes.

[5]The adverse side of this flexibility is the problem of identifying the assumed diffusion or jump diffusion DGP based on the estimated discrete model of quadratic variation.

[6]See also Andersen, Bollerslev, and Meddahi (2002a) and Andersen, Bollerslev, and Meddahi (2002b) where the ability of dynamic models to forecast volatility is analysed.

approach offers researchers and practitioners a convenient tool for data exploration.

Our methodology builds on two main ideas: first, on using AR(∞) to estimate ARMA parameters and to identify ARMA models; this idea has been developed, in particular, in Galbraith and Zinde-Walsh (1997). Second, because of the measurement error on the l.h.s. we use robust least absolute deviation regression (LAD), or quantile regression, to estimate this infinite regression. Our model is a regression model in which the coefficients are those from the ARCH(∞) model (which can be seen as a representation of the underlying GARCH model). We apply our technique to studying the dynamics of realised quadratic variation of foreign exchange rates (DM/US$ and Yen/US$) and of the TSE 35 stock index. We also evaluate the forecasting performance of our estimator and compare it to the performance of alternative models of volatility dynamics.

The exposition will proceed as follows: Section 3.2 describes concisely the main results of the theory of QV and RV. Section 3.3 introduces the results on the asymptotic distribution of the infinite autoregressive quantile estimator and on LAD estimation and identification of ARMA processes using the AR(∞) representation. Empirical results of the study are reported in Section 3.4: we start the section with the description of the data, proceed with the results of LAD and quantile estimation of RV of the stock index and the exchange rates, and with the empirical analysis of the forecasting performance of our LAD estimators. Section 3.5 concludes the exposition. The proofs of the theorems are presented in the Appendix.

3.2 Semimartingales, Quadratic Variation and Realised Quadratic Variation

3.2.1 RV as a measure of QV

We write the log-price process as $y^*(t)$. Let $\hbar > 0$ denote a fixed interval; the return over this interval is defined as

$$y_i = y^*(i\hbar) - y^*((i-1)\hbar)$$

During the interval $\hbar$ we can compute M intra-interval returns which we denote

$$y_{j,i} = y^*\left((i-1)\hbar + \frac{\hbar j}{M}\right) - y^*\left((i-1)\hbar + \frac{\hbar(j-1)}{M}\right)$$

The RV is defined as

$$[y^*_M]_i = \sum_{j=1}^{M} y^2_{j,i}$$

A stochastic process $y^*(t)$ ($y^*(0) = 0$) is a *semimartingale,* if it can be presented as

$$y^*(t) = \alpha(t) + m(t), \ \alpha(0) = m(0) = 0$$

where $\alpha(t)$ is a process with locally bounded variation paths, and $m(t)$ is a local martingale. If $\alpha(t)$ is a predictable process, then $y^*(t)$ is called a *special semimartingale*; the processes of interest to us are always special semimartingales. A QV process is defined as

$$[y^*](t) = y^{*2}(t) - 2\int_0^t y^*(s-)\, dy^*(s)$$

For semimartingales, as it has been shown in ABDL,

$$[y^*]_i = p \lim_{M\to\infty} [y^*_M]_i$$

Let us consider several special cases. First,

$$[y^*](t) = [y^{*c}](t) + \sum_{0 \leq s \leq t} \{\Delta y_s^*\}^2$$

where y^{*c} is a continuous component of $y^*(t)$ and the rest is the jump component. In the context of special semimartingales we can write further

$$[y^*](t) = [m](t) + \sum_{0 \leq s \leq t} \{\Delta \alpha(s)\}^2 + 2 \sum_{0 \leq s \leq t} \Delta m(s) \Delta \alpha(s)$$

As one can see, if $\alpha(t)$ is continuous (even if $m(t)$ is not), then

$$[y^*](t) = [m](t).$$

From the above follows an important result about quadratic variation of semimartingales (noted first, probably, in ABDL) that

Claim 3.2.1 *For all semimartingales,* RV *converges* ***in probability*** *to* QV *as* $M \to \infty$. *The rate of convergence is unknown, as is the asymptotic distribution.*

A set of conditions, still in the general semimartingale setup, sufficient for:

$$Var(y^*(\hbar) | \mathfrak{F}_0) - \mathbf{E}[\{y_M^*\}_1 | \mathfrak{F}_0] = o(1)$$

is formulated in Barndorff-Nielsen and Shephard (2002b).

For the class of stochastic volatility processes for which the log-price follows

$$y^*(t) = \alpha(t) + \int_0^t \sigma(t) \, dw(t), \tag{3.1}$$

where the drift $\alpha(t)$ and $\sigma(t) > 0$ obey some mild regularity assumptions, Barndoff-Nielsen and Shephard developed (see Barndorff-Nielsen and Shephard (2002b) and other articles of these authors referrenced in the bibliography) the asymptotic theory of *RV* summarised by the expression below[7] [8]:

$$\frac{\sum_{j=1}^{M} y_{j,i}^2 - \int_{\hbar(i-1)}^{\hbar i} \sigma^2 (s)\, ds}{\sqrt{\frac{2}{3} \sum_{j=1}^{M} y_{j,i}^4}} \xrightarrow{L} N(0,1). \tag{3.2}$$

As one can see, the asymptotic distribution of the RV is mixed Gaussian since both the numerator and the denominator of (3.2) are random. The implications of this result are that while RV is a consistent estimator of QV under the stated assumptions, it is noisy, and that the approximation noise is higher at higher volatility levels. Barndorff-Nielsen and Shephard (2002b) illustrate this fact empirically using currency exchange rate series - the same data we use in our study. Our analysis using conditional quantiles detects qualitatively similar regularity in the data.

3.2.2 Second order properties of QV and RV

One can approach the modelling of the dynamics of QV from two perspectives: first, by describing the behaviour of this statistic implied by a DGP of practical or theoretical interest (in continuous-time or in discrete-time formulated at a higher frequency); second, by designing an econometric model of the QV process itself, assuming as little as possible about the true DGP in continuous time or at a higher frequency, or even not assuming existence of such a DGP at all. Our approach follows the latter paradigm, and as it is often done in statistics, we assume a certain linear structure in the dynamics of the statistics of interest - QV and RV. Remarkably, several widely used financial dynamic models also imply linear dynamics of these statistics - specifically that they follow an

---

[7]They extended this result to a more general case of *power variation.*

[8]The authors offer also an alternative formulation of the asymptotic theory of RV, which has, in their opinion, superior finite sample behaviour:

$$\frac{\sum_{j=1}^{M} \ln(y_{j,i}^2) - \ln(\int_{\hbar(i-1)}^{\hbar i} \sigma^2 (s)\, ds)}{\sqrt{\frac{2}{3} \frac{\sum_{j=1}^{M} y_{j,i}^4}{(\sum_{j=1}^{M} y_{j,i}^2)^2}}} \xrightarrow{L} N(0,1).$$

ARMA process.

It has been shown in Barndorff-Nielsen and Shephard (2002a) that if the spot volatility follows a linear combination of p constant elasticity of variance (CEV) or positive Ornstein-Uhlenbeck (OU)[9] processes and if $\alpha(t) = 0$ in (3.1), then RV follows an ARMA(p,p) process. Among other processes which imply ARMA dynamics of the QV and RV processes are the continuous SR-SARV process of Meddahi and Renault (2004), which encompasses some of the processes mentioned above, and the weak GARCH process; in all these cases the restriction that the drift of the process $\alpha(t) = 0$ stays[10].

Knowing the ARMA representation of QV is helpful for forecasting, filtering, and impulse response analysis. The relationship between the parameters and the hyperparameters of the ARMA and the parameters of the assumed DGP allows estimation and statistical inference with respect to the parameters of the DGP based on the estimated ARMA representation. Under the assumption that the DGP is one of the processes listed just above, estimation of the weak ARMA dynamics of QV poses little problem: the DGP guarantees the existence of moments of innovations necessary for applying traditional estimation techniques, for example - QML (with the quasi-likelihood function computed using the Kalman filter, as suggested in Barndorff-Nielsen and Shephard (2002a). Our robust approach comes to the rescue when the underlying DGP is not known, or when the measurements of QV are polluted with an error, the properties of which are not well known, and thus the convergence of traditional estimation techniques can not be assured.

[9]The use of positive OU processes in financial econometrics has been popularised by Barndorff-Nielsen and Shephard; see, for example, Barndorff-Nielsen and Shephard (2001).

[10]The requirement to set the drift to zero is a limitation, especially in financial applications, where the risk premium on an asset is a function of time, implying the presence of drift. In the article just mentioned and in other publications Barndorff-Nielsen and Shephard claim that for practical purposes the error resulting from setting the drift to zero is small. Our technique does not require assumptions with respect to the drift.

3.3 Asymptotic Theory of Robust Infinite Regression

3.3.1 Conditional Quantile Estimation

We consider a discrete stochastic process $\{y_t, X_t\}$, and an increasing sequence of σ-fields $\{\Im_t\}$, where X_t is measurable w.r.t. $\Im_t$. Denote $\chi_q(y_t) \equiv \chi_q(y_t|\Im_t)$ the q^{th} quantile of the conditional distribution of y_t. Define the *check function* as:

$$f_q(x) \equiv \left[\left(q - \frac{1}{2}\right) + \frac{1}{2}\mathrm{sgn}x\right] x. \tag{3.3}$$

We introduce some notational conventions. We denote the vector $X_t(k) \equiv (X_{0,t}, X_{1,t}, \ldots, X_{k,t})'$ with $X_{0,1} = 1$ (we allow $k = \infty$). For any $k < \infty$, $X_t(\infty)$ can be partitioned as $X_t(\infty) = (X_t(k), X_t(k+1,\infty))$; analogously, $\gamma_q(\infty)$, an infinite vector of coefficients of the linear representation of $\chi_q(y_t)$ considered below, can be partitioned as $(\gamma_q(k), \gamma_q(k+1,\infty))$.

Assumption 3.3.1

For a sequence of (possibly random) matrices $\{V_T(k)\}$,

(a) $V_T(k)^{-1}X_t(k)$ *is $\Im_t$-measurable for all T, k*[11]*;*

(b) $\chi_q(y_t|\Im_t) = X_t'(\infty)\gamma_q(\infty)$ *;*

(c) $e_t - \varpi_q = y_t - X_t'(\infty)\gamma_q(\infty)$, *where ϖ_q is a constant, is such that*

(i) $\{e_t, X_t\}$ *is a stationary ergodic sequence*

(ii) p.d.f. of $p_e(x)$ exists and is continuous and positive at ϖ_q

(iii) $\{f_q'(e_t - \varpi_q), \Im_t\}$ *is a m.d. sequence*[12]*;*

[11] T is an integer index; in practice, T can be thought as a sample size.

[12] The derivative of the check function, $f_q'(x)$, is defined and continuous everywhere except 0.

(d) $\sup_{1\le t\le T} \max \left|V_T^{-1}(k)X_t(k)\right| = o_p(1)$[13];

(e)

$$\max \left| \sum_{t=1}^{T} V_T(k)^{-1} X_t(k)' X_t(k) V_T(k)^{-1} - I_{k+1} \right| = o_p(1);$$

(f) *There exists a monotonically increasing function $\omega(x)$ such that $k = \omega(T) \to \infty$ as $T \to \infty$ and*

$$sup_{1\le t\le T} \left|X_t\left[k+1,\infty\right)' \gamma_q \left[k+1,\infty\right)\right| = o_p(T^{-\frac{1}{2}}).^{14}$$

We denote by $\hat{\gamma}_q(k)$ the quantile estimator of $\gamma_q(k)$:

$$\hat{\gamma}_q(k) = \arg\min_{\gamma(k)} \sum_{t=1}^{T} f_q\left(y_t - X_t(k)\gamma(k)\right).$$

For any fixed $k' < k$, define $\Omega_k \equiv \begin{bmatrix} I_{k'} & 0 \\ 0 & 0 \end{bmatrix}$. We are now ready to formulate the result pertaining the asymptotic distribution of the quantile estimator.

Theorem 3.3.1 *Under Assumption 3.3.1 as $T \to \infty$, $k = \omega\left(T\right)$,*

$$\Omega_k V_T(k)\left(\hat{\gamma}_q(k) - \gamma_q(k)\right) \Rightarrow \mathbf{N}\left(0, \frac{q(1-q)}{p_e(\varpi_q(e))}\Omega_k\right).$$

Note that we assume only that a specific quantile is represented as in Assumption 3.3.1b and that the approximation error satisfies the other conditions formulated in Assumption 3.3.1. We do not claim that these conditions are satisfied when the true DGP follows one of the processes considered in the previous section. We just offer this linear representation of conditional quantiles as a flexible framework for modelling the dynamics of QV.

[13]For any matrix X, $\max|X|$ denotes in this paper the absolute value of the largest component of the matrix.

[14]Note that if components of γ, γ_k decline exponentially in k and components of X_t grow at most at a polynomial rate then for $k = T^{\alpha}$, $\alpha > 0$ condition **(f)** is satisfied.

If $X_{l,t} = y_{t-l},\ l \geq 1$ then this estimation approach deals with an AR(∞) representation which is more general than any ARMA representation. In some cases it may be advantageous to use the ARMA model instead of the AR(∞), or its feasible counterpart - AR(p). For example, a more parsimonious representation like ARMA may yield a better forecasting performance, or we may want to incorporate into the estimation and forecasting prior knowledge about the process - that it is indeed an ARMA. To obtain estimates of ARMA parameters from the AR(∞) presentation we pursue the strategy developed in Galbraith and Zinde-Walsh (1994) and Galbraith and Zinde-Walsh (1997), where the autoregressive presentation and OLS are used for estimating an ARMA. The parameters of AR(∞) are estimated using a truncated AR(k) representation; the order of truncation k must be such that $k \to \infty$, $k/T \to 0$ for consistent estimation of the parameters of the ARMA. The ARMA parameters are evaluated using the deterministic relationship between them and the coefficients of AR(∞). Since the coefficients of an AR(∞) presentation are continuous functions of the corresponding ARMA parameters, obtaining the asymptotic distribution of the estimates of the latter poses no difficulties. We refer the reader to Galbraith and Zinde-Walsh (1997) for details of the method, and we shall illustrate its application in the example below.

Example: Robust Estimation of Weak GARCH using *RV*. Let's assume that the data generating process (DPG) at a higher frequency is a weak GARCH[15], i.e., that the linear projection of the conditional variance is expressed as:

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \sigma_{t-i}^2 + \sum_{i=1}^{q} \beta_i \epsilon_{t-i}^2. \tag{3.4}$$

A weak GARCH process, as has been shown in Francq and Zakoïan (2000) results in a stationary, invertible ARMA presentation of squared innovation ϵ_t^2 and thus allows for an infinite autoregressive representation of the variance $\sigma_t^2 \equiv \mathbf{Pr}\left(\epsilon_t^2 | \Im_t\right)$[16], $\Im_t =$

[15]Since weak GARCH is invariant with respect to aggregation, we do not need to worry if the choice of the high sampling frequency is appropriate.

[16]$\mathbf{Pr}\left(\cdot | \Im_t\right)$ denotes here a linear projection.

$\sigma\left\{\epsilon_{t-1}^2, \epsilon_{t-2}^2, \ldots\right\}$:

$$\sigma_t^2 = \kappa + \sum_{l=1}^{\infty} \nu_l \epsilon_{t-l}^2.$$

We have a measurement $\hat{\sigma}_t^2$ for σ_t^2; σ_t^2 is measured with error, and the value $\hat{\sigma}_t^2$ can be represented as $\hat{\sigma}_t^2 = \sigma_t^2 + \omega_t$. Then

$$\hat{\sigma}_t^2 = \kappa + \sum_{l=1}^{\infty} \nu_l \epsilon_{t-l}^2 + \omega_t \tag{3.5}$$

Using the notation introduced at the beginning of this section, we denote $X_{0,t} = 1$, $X_{l,t} = \epsilon_{t-l}^2$ and re-write equation (3.5) as

$$\hat{\sigma}_t^2 = X_t'(\infty)\gamma(\infty) + \omega_t = X_t'(k)\gamma(k) + X_t'[k+1,\infty)\gamma[k+1,\infty) + \omega_t.$$

To proceed, we need to specify a set of conditions.

Assumption 3.3.2

(a) $\{\epsilon_t^2, \omega_t\}$ *is a stationary ergodic sequence;*

(b) $\mathbf{E}\epsilon_t^4 < \infty$*;*

(c) $\max_{1\leq t\leq T} |\epsilon_t^2| = o_p(T^{\frac{1}{2}})$*;* $\sup_{-\infty<t<0} |\epsilon_t^2| = O_p(1)$*;*

(d) $\chi_{0.5}\left(\hat{\sigma}_t^2|\Im_t\right) = X_t'(\infty)\gamma(\infty)$ *or equivalently,* $\chi_{0.5}(\omega_t|\Im_t) = 0$*,*
where $\chi_{0.5}$ *denotes the median (50% quantile) of the distribution;*

(e) *the probability density function of* ω_t*,* $f_\omega(x)$*, exists, and is positive and continuous in a neighbourhood of* $\chi_{0.5}$*.*

We now define

$$\Sigma(k) = \mathbf{E}\left(T^{-1}X_t'(k)X_t(k)\right),$$

where k can be infinite; the matrix obtained from $\Sigma(\infty)$ by deleting the first row and the first column is the covariance matrix for $\{\epsilon_t^2\}_{t=1}^{\infty}$. Define $V_T = T^{\frac{1}{2}}\Sigma(k)^{\frac{1}{2}}$.

Theorem 3.3.2 *Under the Assumption 3.3.2 the least absolute deviation (LAD) estimator* $\hat{\gamma}(k)$, *for any matrix* $\Omega = \begin{pmatrix} I_{k'} & 0 \\ 0 & 0 \end{pmatrix}$, *has the following asymptotic distribution:*

$$\sqrt{(T)}\Omega\left(\gamma(\hat{k}) - \gamma(k)\right) \Rightarrow \mathbf{N}\left(0, \frac{1}{\left(2f_{\omega}\left(\chi_{0.5}\right)\right)^2}\left(\Omega\Sigma\left(\infty\right)\Omega\right)^{-1}\right)$$

where $\left(\Omega\Sigma\left(\infty\right)\Omega\right)^{-1}$ *can be consistently estimated by* $\Omega\left(\frac{1}{T}X'(k)X(k)\right)^{-1}\Omega$ *provided that as* $T \rightarrow \infty$, $k = \omega(T) = o\left(T^{\frac{1}{2}}\right)$; $\ln\omega(T) \rightarrow \infty$.

The proof of the theorem is in Appendix 3.3.2. Notice that the conditions of Assumptions 3.3.2 are not very restrictive: we require only the existence of the 4^{th} moment of the GARCH innovations, plus ergodicity; these conditions can be substituted by some alternative conditions which would allow the application of the law of large numbers (LLN) (see proof). We do not impose any assumptions about the existence of the moments of the measurement error of σ^2, as long as the error has zero median.

To estimate the parameters of the GARCH model (3.4) we notice that if the GARCH and the AR(∞) representations of conditional variance are non-stochastic, then the AR(∞) parameters can be expressed recursively as functions of the GARCH parameters:

$$\begin{aligned} \nu_1 &= \beta_1 \\ \nu_2 &= \beta_2 - \alpha_1\nu_1 \\ &\cdot \\ \nu_l &= \beta_l - \sum_{i=1}^{\min(l,p)} \alpha_i\nu_{l-i},\ l \leq q \\ \nu_l &= \sum_{i=1}^{\min(l,p)} \alpha_i\nu_{l-i},\ l > q \end{aligned} \tag{3.6}$$

and finally,

$$\kappa = \left(1 - \sum_{i=1}^{p} \alpha_1\right)^{-1} \omega.$$

With $\nu_0 \equiv 0$ we define

$$v(0) = \begin{bmatrix} \nu_{q+1} \\ \nu_{q+2} \\ \cdot \\ \cdot \\ \cdot \\ \nu_k \end{bmatrix}, \text{ and } v(-i) = \begin{bmatrix} \nu_{q+1-i} \\ \nu_{q+2-i} \\ \cdot \\ \cdot \\ \cdot \\ \nu_{k-i} \end{bmatrix}.$$

Next, define the $(k-q) \times p$ matrix $\mathbf{W} = [v(-1)v(-2)\ldots v(-p)]$ where $\nu_r \equiv 0$ for $r \leq 0$. It follows from (3.6) that $v(0) = \alpha'\mathbf{W}$. The $p \times 1$ vector of estimates of GARCH parameters $\hat{\alpha}$ is defined by

$$\hat{\alpha} = \left(\hat{\mathbf{W}}'\hat{\mathbf{W}}\right)^{-1} \hat{\mathbf{W}}'\hat{v}(0),$$

where the ν_l are substituted in the expression above by their estimates $\hat{\nu}_l$. The estimate of the vector of GARCH parameters β is easily obtained from (3.6), knowing $\hat{\alpha}$ and $\hat{\nu}_l$. The variance-covariance matrix of the estimated vector of GARCH parameters $(\hat{\omega}, \hat{\alpha}, \hat{\beta})$ is computed knowing the variance-covariance of the estimates of the AR(k) parameters given by the Theorem 3.3.2 and the Jacobian of the transformation, using the usual sandwich formula.

End of the example

We must point out a caveat here. The ARMA presentation of RV series considered in Section 3.2 describes the dynamics of the mean of RV (conditional on the information set consisting of past values of RV). The estimation procedure described in the example above yields the conditional median; the two are equal if the conditional distribution is symmetric. While we cannot claim that the conditional distribution is necessarily symmetric when we apply our model to RV series, i.e. our characterisation of the conditional location is different from that of the models of conditional mean, we do not see this as an obstacle to practical use.

In the example above and in the empirical part of our study we project the conditional quantile on the set of past squared innovations, but our theoretical result says nothing specific about the nature of the projection set: it may consist of other non-linear functions of the innovations or the set may include auxiliary variables. These variables in the linear presentation must only meet the requirements of Assumption 3.3.1.

3.4 Empirical results: dynamics of QV and RV of foreign exchange rates and a stock index

3.4.1 Description of the data

As we have noted, two types of data are used in this study: an equity-price index and a set of foreign exchange prices. The former is a short (one-year) span of very high-frequency data, spaced fifteen seconds apart, on the Toronto Stock Exchange index of thirty-five large-capitalization stocks (TSE 35), for the calendar year 1998[17][18]. The latter is a fourteen-year sequence of observations spaced at five minutes, and pertain to the DEM/USD and JPY/USD (German mark and Japanese yen respectively to US dollar) exchange rates. The fifteen-second intra-day data on the TSE 35 index value (as well as bid and ask) are available through the 9:30 to 16:00 trading day, for a total of approximately 1560 observations per day.

The data must be filtered to recognize the fact that trading does not take place throughout the 24-hour day, and that there are occasional anomalies near the beginning

[17]This index has since been superseded by the S&P/TSE 60 Index of large capitalisation stocks.

[18]It is well documented in the financial literature that stock prices exhibit a leverage effect. Our dynamic model does not take the leverage effect into account. However, for the class of SV models, for example, convergence in probability of RV to QV holds in the presence of leverage, i.e., projecting RV on past innovations may yield meaningful inference about QV even in the presence of leverage.

of the trading day. First, we treat the change in the index value between the 16:00 close and 9:30 open on the following day as a contribution to realised quadratic variation. The logic for such treatment is the following: one can assume that there is a latent information process which is manifested through trading; even though there are no trades between the close and the open, this latent process is not interrupted, and in the absence of trading the best available estimate of the QV overnight (or over the weekend) is probably the squared return over this period[19]. Second, the first few minutes of the trading day typically show the index value outside the bid/ask range; within the first two minutes of trading, the index value is usually again within the range. We therefore use the midpoint between bid and ask for the first two minutes of the trading day, by which point the two measures are almost invariably compatible. Finally, of course, we must decide on a level of time aggregation, or $\hbar$ in the notation of Section 2. Since the raw data are provided at 15-second intervals, summing four squared returns to obtain a single intra-day observation, that is aggregating to one-minute returns, implies $\hbar = 1560/4 = 390$ observations per day; aggregation to the five-minute interval corresponds to $\hbar = 1560/20 = 78$; the results presented in this paper correspond to the latter aggregation frequency. Each of the filtering operations applied to these TSE 35 data are described more fully in Galbraith and Kisinbay (2002).

The foreign exchange data used in this study are taken from the HFDF 2000 data set compiled and distributed by Olsen Group, Switzerland. Foreign exchange returns recorded every five minutes span the period from January 2, 1986, 00:00:00 GMT to January 1, 1999, 23:35:00. The returns are computed as the mid-quote price difference, expressed in basis points (i.e. multiplied by 10,000). The midquote price at the regular time point is estimated through a linear interpolation between the previous and following mid-price of the irregularly spaced tick-by-tick data. The average bid-ask spread over the last 5-minute interval is expressed in basis points. If there is no quote during an interval,

[19]I.e., we cannot feasibly construct RV series in the way required by the asymptotic theory, and we know very little about the properties of the approximation error. This is a case in which a robust approach becomes useful

the mean bid-ask spread is zero. Currencies are traded continuously throughout the day, seven days a week; thus the data set contains 1,262,016 5-minute returns, expressed in USD terms. In order to compute realised variances, it is necessary to perform filtering. We have followed Bollerslev and Domovitz (1993) and other researchers who have worked with these data, in defining the trading day t as the interval from 21:05 GMT of the previous calendar day to 21:00 GMT on the calendar day t. The estimate of daily realised quadratic variation, is computed by summing squares of currency's five-minutes returns over the day. Following ABDL, we have filtered out of the data days with low trading volume. The filters that we applied to the data eliminated weekends, fixed holidays (December 24-26, 31, January 1-2) as well as moving holidays (Good Friday, Easter Monday, Memorial Day, the Fourth of July, Labour Day, Thanksgiving (US) and the day after Thanksgiving. In addition, we have eliminated from the data the days for which the indicator variable (the bid-ask spread) had 144 or more zeroes, thus corresponding with technical "holes" in the recorded data. Application of all of these filters reduced the data sets to $876,096$ data points, or 3042 days, for the DEM/USD, and $877,248$ or 3046 days for the JPY/USD. Graphs of daily series of realised quadratic variation of DEM/USD and JPY/USD exchange rates are presented in Figure 3-1.

3.4.2 Projection of quantiles of QV on past squared innovations

In this section we use quantile regression to describe certain empirical dynamic properties of the conditional distribution of QV and RV.

Figure 3-2 depicts a one-period-ahead in-sample forecast of conditional 0.1, 0.3, 0.5, 0.7, and 0.9 quantiles of QV for the period from November 1995 to December 1998[20]. The forecast is computed using the AR(14)[21] representation.

Visual inspection of the graph confirms our *a priori* conjecture that the distribution

[20] We have chosen not to show all of the data, to improve the readability of the graph; the choice of the period is arbitrary.

[21] Here and below in this section AR(k) denotes LAD or quantile regression of RV on the k past squared daily returns.

Data Series	T	**Corr** $\left(RV, \frac{\hat{\chi}_{.8}}{\hat{\chi}_{.2}}\right)$	**Corr** $\left(RV, \frac{\hat{\chi}_{.9}}{\hat{\chi}_{.1}}\right)$
TSE 35	352	0.199	0.028
		(0.122)	(0.043)
DEM/USD	2970	0.288	0.326
		(0.030)	(0.031)
JPY/USD	2970	0.106	0.286
		(0.065)	(0.083)

Table 3.1: Correlation of realised quadratic variation with ratios of upper- to lower conditional quantiles

of the variance would be skewed to the right: almost uniformly the difference between the conditional median and the 0.7 quantile (or 0.9 quantile) exceeds the distance between the 0.3 (0.1) quantile and the median; the distance between the 0.7 and 0.9 quantiles exceeds the distance between the 0.1 and 0.3 quantiles.

As has been mentioned, the asymptotic theory developed by Shephard and co-authors implies that typically the variability of RV as an estimator of QV is higher when the level of volatility is higher. Our empirical analysis discovers qualitatively similar properties in the dynamics of the data. We analyse the correlation between the level of realised quadratic variation and the ratios of upper-to-lower quantiles of conditional forecast of QV; the quantitative results of this analysis - the correlations and the corresponding HAC standard errors are presented in Table 3.1. In all cases we observe that the correlation is positive; it is not statistically significant for the smaller TSE data set but for larger currency exchange series it is significant at the 1% level for three cases out of the four presented. In the fourth case the estimated **Corr** $\left(RV, \frac{\hat{\chi}_{.8}}{\hat{\chi}_{.2}}\right)$ for JPY/USD has the p-value of 10%.

3.4.3 ARMA Models of Conditional Quantiles of Variance

In the context of this section ARMA(p,q) denotes a parsimonious representation of the coefficients of the infinite-dimensional quantile regression. We now consider estimates of the ARMA parameters of the conditional quantiles of volatility using the methods introduced in Section 3.3[22]. We begin with the ARMA(1,1) model and then consider the possibility of using ARMA representations of higher order[23].

Table 3.2 presents estimates of the parameters of the ARMA(1,1) model for quantiles from 0.1 to 0.9 of daily QV for foreign exchange series. As an example of the interpretation of these numbers, consider the DEM/USD exchange rate. Conditional on a previous-period RV of 0.0004 and squared return of 0.00015, the 10th and 90th percentiles of the conditional distribution of quadratic variation lie at 0.00026 and 0.00033. That is, 10% of values of quadratic variation would be below the first number and above the second, given the conditions mentioned. This calculation therefore allows daily computation of the quantiles of the conditional distribution of the next day's volatility, given the conditions just observed.

If the DGP of the process were ARMA with homoscedastic innovations, then we would expect only the intercept of the quantile regression to change from one quantile to another, while the slope coefficient would remain invariant. We observe empirically, however, that while the intercept indeed increases from lower quantiles to higher, the slope coefficients change too, most notably - the moving average coefficient, which tends to be larger for higher quantiles than for lower lower quantiles. One of the implications of this observation is that it is not consistent with strong GARCH(1,1) as the true DGP.

On a smaller data set of equity returns, we present only 0.2, 0.5 and 0.8 conditional quantiles. We observe the same pattern: the conditional variance has a higher intercept and higher values of β at higher quantiles.

Next consider the specification of alternative ARMA models for the conditional quantiles. In Table 3.4 we report the results of ARMA model selection using the Bayesian

[22]As follows from the discussion in Section 3.3 and the Example on page 63, if the DGP is a weak GARCH(p,q), the QV and RV follow an ARMA(p,q), but other DGP may also generate ARMA dynamics of RV.

[23]Throughout this section the number of approximating AR terms used in the ARMA model estimation is k = 14 for the foreign exchange data and k = 8 for the smaller equity index data set. These orders are approximately those implied by the rule of thumb $k = 8 + \text{int}\left(2\ln\left(\frac{T}{100}\right)\right)$ stated in Galbraith and Zinde-Walsh (1997).

	DEM/USD			JPY/USD		
Quantile	ω	α	β	ω	α	β
0.1	1.0×10^{-5}	0.607	0.059	0.064	0.594	0.065
	(9.0×10^{-7})	(0.028)	(0.003)	(0.007)	(0.032)	(0.003)
0.2	1.2×10^{-5}	0.595	0.079	0.072	0.632	0.084
	(1.1×10^{-6})	(0.025)	(0.003)	(0.007)	(0.024)	(0.003)
0.3	1.2×10^{-5}	0.640	0.087	0.091	0.596	0.108
	(1.1×10^{-6})	(0.023)	(0.003)	(0.007)	(0.020)	(0.004)
0.4	1.5×10^{-5}	0.634	0.100	0.102	0.609	0.122
	(1.3×10^{-6})	(0.024)	(0.004)	(0.008)	(0.019)	(0.004)
0.5	1.5×10^{-5}	0.655	0.112	0.120	0.590	0.141
	(1.3×10^{-6})	(0.023)	(0.004)	(0.008)	(0.017)	(0.004)
0.6	1.6×10^{-5}	0.689	0.118	0.130	0.613	0.154
	(1.5×10^{-6})	(0.023)	(0.004)	(0.008)	(0.016)	(0.004)
0.7	1.7×10^{-5}	0.695	0.134	0.170	0.579	0.183
	(1.8×10^{-6})	(0.024)	(0.005)	(0.010)	(0.017)	(0.004)
0.8	2.0×10^{-5}	0.686	0.171	0.251	0.480	0.253
	(2.3×10^{-6})	(0.027)	(0.007)	(0.016)	(0.019)	(0.006)
0.9	2.5×10^{-5}	0.698	0.194	0.360	0.423	0.403
	(4.1×10^{-6})	(0.038)	(0.013)	(0.032)	(0.026)	(0.012)

Table 3.2: Foreign exchange rates: estimated ARMA(1,1) coefficients for specific quantiles of variance

Quantile	ω	α	β
0.2	$5. \times 10^{-6}$	0.70	0.05
	$(4. \times 10^{-6})$	(0.14)	(0.02)
0.5	$6. \times 10^{-6}$	0.76	0.10
	$(4. \times 10^{-6})$	(0.07)	(0.02)
0.8	$2. \times 10^{-5}$	0.55	0.16
	$(8. \times 10^{-6})$	(0.08)	(0.02)

Table 3.3: TSE 35 equity index returns: estimated GARCH(1,1) coefficients for specific quantiles

	DEM/USD			JPY/USD		
Quantile	$(p^\star, q^\star)$	$\sum \alpha_i$	$\sum \beta_i$	$(p^\star, q^\star)$	α	$\sum \beta_i$
0.1	(4, 1)	0.709	0.059	(4,1)	0.733	0.065
0.2	(4,2)	0.623	0.113	(3,4)	0.825	0.061
0.3	(3,2)	0.700	0.112	(2,2)	0.568	0.114
0.4	(4,1)	0.768	0.101	(2,2)	0.601	0.146
0.5	(4,1)	0.754	0.126	(2,2)	0.472	0.187
0.6	(3,2)	0.788	0.127	(2,2)	0.467	0.227
0.7	(4,3)	0.760	0.165	(2,2)	0.474	0.247
0.8	(3,4)	0.711	0.229	(3,2)	0.550	0.282
0.9	(2,2)	0.687	0.264	(4,1)	0.547	0.403

Table 3.4: Foreign exchange rates: ARMA model selection by quantile of variance

Information Criterion (BIC) and the Final Prediction Error (FPE), from the set of model orders (p,q), $p = 1, \ldots, 4; q = 1, \ldots, 4$. The chosen optimal order is the same for the two criteria in each case.

The information criteria favour models with at least four ARMA parameters in addition to the intercept. In these higher-order models as well we see the larger values of β associated with higher quantiles. The intercept values (not recorded in Table 3.4) also increase monotonically with the quantile, as in the results of Table 3.2, while there is no pattern of regular increase in the coefficients on lagged realised quadratic variation: that is, the higher quantiles do show similar persistence of volatility, and the higher values are reflected in higher weight on recent squared returns.

The observation that fitting empirical realised quadratic variation series requires an ARMA process of order higher than one is consistent with the analysis presented by other researchers (for example, Barndorff-Nielsen and Shephard (2002a), Meddahi (2002a), and Andersen, Bollerslev, and Meddahi (2002a)), who notice that continuous-time stochastic volatility models often require more than one latent factor driving the spot volatility, to achieve a satisfactory fit to empirical data. As has been mentioned, many of these models imply ARMA dynamics of QV of order equal to the number of independent latent factors.

3.4.4 Forecasting performance of AR(p) and ARMA estimators

Forecasting performance, especially out-of-sample forecasting performance, is traditionally an important criterium of model evaluation in empirical financial research. A comprehensive investigation of forecasting performance with respect to QV and RV of dynamic models of volatility, estimated using high-frequency data, has been conducted recently in Andersen, Bollerslev, and Meddahi (2002a) and Andersen, Bollerslev, and Meddahi (2002b). This research shows that if the DGP is one of the popular stochastic volatility models with parameters typical for financial series [24] then first, QV is highly predictable with an ideal (infeasible) one-period-ahead forecast R^2 exceeding in many cases 95% , and second, that the loss in the performance of the feasible forecast based on RV, compared with the ideal forecast, is not large, and that the feasible forecast R^2 is in many cases in the 80^{th} or even in the 90^{th} percentile. The study points out also that the theoretical predictive power of forecasts based on past daily squared innovations is substantially inferior to that of model-based forecasts.

In practice, unfortunately, we do not know the true dynamic model of volatility; the properties of both misspecification error and the estimation error are difficult to formalise. As one would expect, while the dynamic models of volatility show a good ability to forecast RV in empirical applications, they do not achieve their theoretical forecasting power.

We have anticipated that the flexibility and robustness of our estimation approach would translate into good forecasting performance. Table 3.5 presents results of the study of the forecasting performance of our models. For this analysis the data set is split into two parts: we use the first three quarters of the data to estimate the model and for in-sample analysis; we use the remaining quarter for the analysis of out-of-sample performance. In the table we report the ordinary R^2 of one-day-ahead forecasts and also the results (the intercept, the slope and the R^2) of the Mincer-Zarnovitz regression of

[24]Three models are considered: the GARCH diffusion model of Nelson (1990), the two-factor affine stochastic volatility model, and the log-normal diffusion.

realised values on the forecasted values. We also report the intercepts and the slopes of LAD Mincer-Zarnovitz regressions for the ARMA model (the intercept and the slope for the AR model are 0 and 1 by definition).

We consider two models: AR(12) and ARMA(1,1), and two estimation techniques, OLS and LAD. This gives for each currency (JPY and DEM) four sets of results in-sample and four out-of-sample sets. Both for DEM and JPY the in-sample R^2s are in the upper 20^{th} - lower 30^{th} percentiles for the models considered. We would not be surprised if there were a bias in the Mincer-Zarnovitz regression for LAD estimation of the AR model, since it is forecasting the conditional median and not the conditional mean, but the estimated bias is not statistically significant. The in-sample forecasts using the ARMA(1,1) model have a statistically significant bias; in addition, the slope of the Mincer-Zarnovitz regression is less than one, i.e., a more parsimonious model produces a smoother the forecast in this case. The bias and the slope are similar for the two estimation techniques. The in-sample predictive R^2 are very close to the values reported in Andersen, Bollerslev, and Meddahi (2002b), who studied the same data set [25]. The in-sample predictive power of the ARMA(1,1) model is consistently but not substantially lower than that of the AR(12) model, which is not surprising because the former is more parsimonious. The in-sample predictive R^2s of the models estimated using the LAD approach are somewhat lower than those estimated using OLS, which is again natural because the R^2 criterium puts LAD at a disadvantage in comparison with OLS.

We turn now to the analysis of out-of-sample forecasts. For the DEM series, the predictive R^2 of the AR(12) model, 20%, is almost identical to the value reported in Andersen, Bollerslev, and Meddahi (2002b). A more parsimonious GARCH(1,1) model outperforms the forecast obtained using the AR(12). The robust LAD estimator outperforms the OLS. The out-of-sample predictive R^2s achieved by our models applied to the

[25]The results reported in Andersen, Bollerslev, and Meddahi (2002b) are for an AR(5) model, but according to the authors, other models produce similar R^2s; the R^2s reported in this reference are adjusted for bias due to measurement error in the future RV. We do not adjust our R^2 because we do not want to introduce the additional assumptions required for such adjustment, i.e. our R^2s are biased downwards.

JPY/USD series ranges from the upper 30^{th} to the upper fiftieths percentiles; they are higher than than those resulting from the in-sample forecasting. While we attribute these unexpectedly high predictive R^2s to peculiarities of the data, we again notice that out-of-sample the ARMA(1,1) performs better than the AR(12), and that robust estimation yields better results than the OLS approach.

It was suggested (see, again, for example, Andersen, Bollerslev, and Meddahi (2002b) and references there) that a linear model be used to describe the dynamics of a non-linear transformation (square root or logarithm) of QV and RV instead of those quantities themselves. While these transformations bring the distribution of the error closer to the normal and thus improve the accuracy of the estimation and of the forecast, the drawback is that the transformation producing the best estimation or forecast results is not necessarily the quantity of interest to the researcher. We have a chance to emphasise again here the advantage of the robust approach, which allows the researcher to specify the model in terms of the quantity of interest.

3.5 Concluding remarks

We develop in this research a framework for modelling conditional quantiles for a class of processes, where these quantiles can be represented or approximated by linear combinations of infinite series of random variables. We impose very mild restrictions on these series which ensures the flexibility of our modelling framework. Since our approach allows us to model any quantile of the distribution, it can be viewed as a non-parametric tool permitting us to investigate the whole shape of the conditional distribution of the series of interest.

We apply our technique to investigate an object that has attracted a great deal of attention from researchers in the last several years - the dynamics of quadratic variation (QV) and realised quadratic variation (RV). The advantage of our estimator here is that it is robust - it retains consistency and asymptotic normality, without requiring that the

Model	Method	R^2	a	b	R^2	a_{LAD}	b_{LAD}
JPY/USD in-sample							
AR(12)	OLS	0.312	-	-	-	-	-
	LAD	0.302	1.14e-5	0.939	0.304	-	-
			(1.65e-6)	(0.030)			
ARMA(1,1)	OLS	0.288	1.82e-5	0.688	0.363	-	-
			(1.33e-6)	(0.019)			
	LAD	0.289	2.04e-5	0.697	0.356	1.05e-5	0.738
			(1.30e-6)	(0.020)		(9.24e-7)	(0.014)
JPY/USD out-of-sample							
AR(12)	OLS	0.399	-4.20e-5	1.943	0.522	-	-
			(5.50e-6)	(0.069)			
	LAD	0.440	-2.58e-5	1.953	0.578	5.75e-6	1.055
			(4.65e-6)	(0.062)		(3.61e-6)	(0.048)
ARMA(1,1)	OLS	0.396	2.03e-5	0.814	0.418	-	-
			(4.62e-6)	(0.036)			
	LAD	0.450	1.91e-5	0.898	0.456	1.85e-5	0.660
			(4.44e-6)	(0.036)		(3.24e-6)	(0.026)
DEM/USD in-sample							
AR(12)	OLS	0.333	-	-	-	-	-
	LAD	0.319	2.93e-6	1.116	0.322	-	-
			(1.80e-6)	(0.034)			
ARMA(1,1)	OLS	0.254	2.01e-5	0.663	0.342	-	-
			(1.33e-6)	(0.019)			
	LAD	0.286	2.04e-5	0.712	0.342	1.20e-5	0.722
			(1.30e-6)	(0.021)		(8.61e-7)	(0.014)
DEM/USD out-of-sample							
AR(12)	OLS	0.199	-1.024e-6	0.885	0.203	-	-
			(2.76e-6)	(0.065)			
	LAD	0.199	1.65e-6	0.989	0.199	9.15e-7	0.852
			(2.60e-6)	(0.074)		(1.84e-6)	(0.052)
ARMA (1,1)	OLS	0.287	9.68e-6	0.713	0.343	-	-
			(1.49e-6)	(0.037)			
	LAD	0.310	1.00e-5	0.764	0.343	8.47e-6	0.674
			(1.47e-6)	(0.039)		(1.074e-6)	(0.029)

Table 3.5: One-Day-Ahead Predictive Power

daily volatility estimate become arbitrarily accurate through a high frequency sampling interval converging to zero. This point is particularly important because in practical application, despite the very high frequencies of sampling sometimes available, market microstructure effects such as bid-ask bounce invalidate the diffusion approximation at very high frequencies. In practice, therefore, one must place some limit on the frequency of sampling, implying that consistency cannot be obtained via complete convergence of the daily volatility estimates to the true values.

Our analysis of conditional quantiles of the volatility of a stock index and of currency exchange series enabled us to discover certain regularities in the dynamics of the conditional distribution of realised quadratic variation; specifically, that the higher quantiles of conditional variance increase more than proportionally in periods of higher levels of variance. The latter observation is qualitatively consistent with the asymptotic theory of RV developed in recent research by Barndorff-Nielsen, Shephard, and co-authors for a class of stochastic volatility models.

Model selection based on information criteria (BIC or AIC) shows that an ARMA model of order higher than one is necessary to fit the data well (specifically, the information criteria favour ARMA(4,1) for the median of the DEM/USD exchange rate and ARMA(2,2) for the median of the JPY/USD exchange rate). This is consistent with recent empirical analysis of stochastic volatility models in application to financial series showing that at least two independent latent factors driving volatility are required to fit the data adequately; for the popular affine stochastic volatility diffusion models the number of independent latent factors translates into the autoregressive order of the ARMA presentation of the corresponding QV and RV series.

Finally, we evaluate the forecasting performance of our estimation framework. We show that while the long AR presentation exhibits higher predictive power in-sample (as measured by predictive R^2), the parsimonious ARMA(1,1) presentation achieves better results in out-of-sample tests. The robust estimation approach consistently improves out-of-sample prediction R^2s relative to OLS. In the empirical example considered in this study (DEM/USD and JPY/USD exchange rates), the predictive performance of our robust estimators is either comparable or exceeds the performance of alternative techniques (see Andersen, Bollerslev, and Meddahi (2002b) and Andersen, Bollerslev, Diebold, and Labys (2003)).

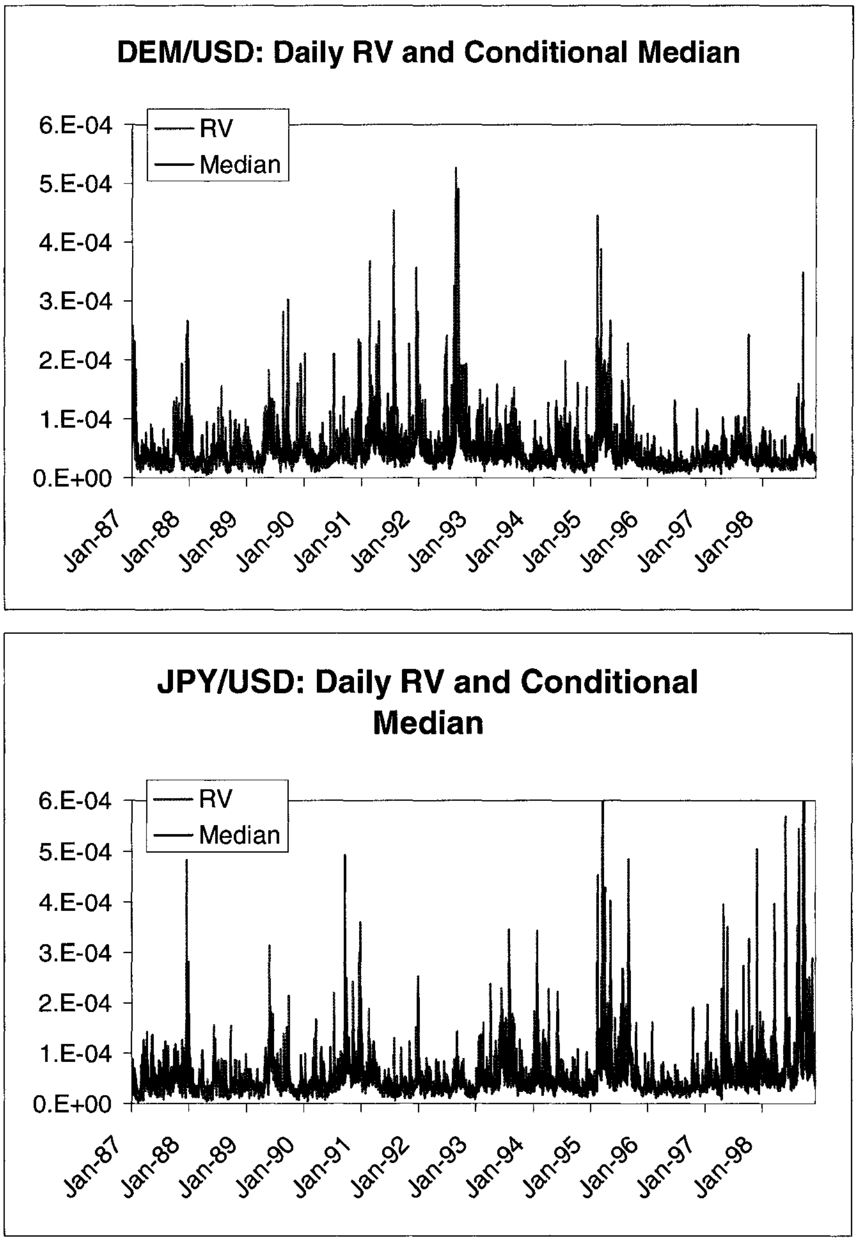
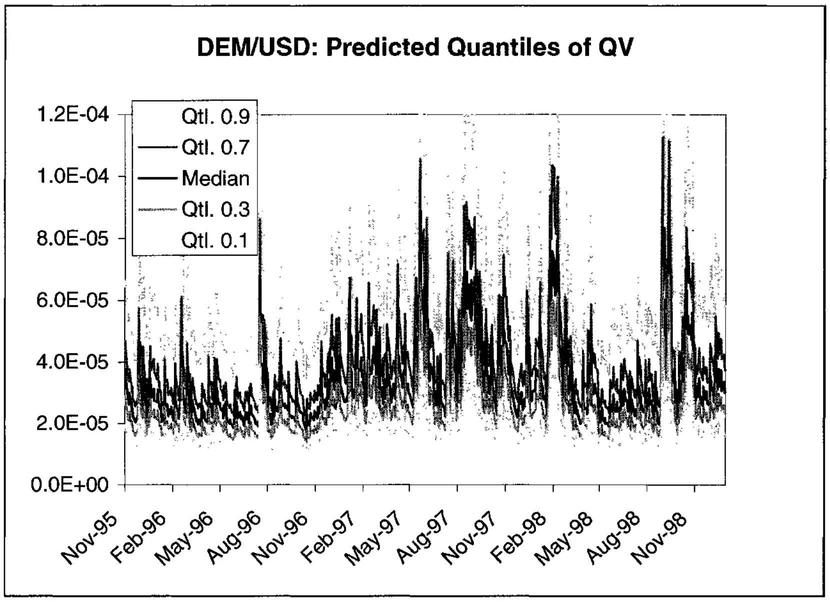


Figure 3-1: Daily Realised Quadratic Variation and In-Sample Conditional Median



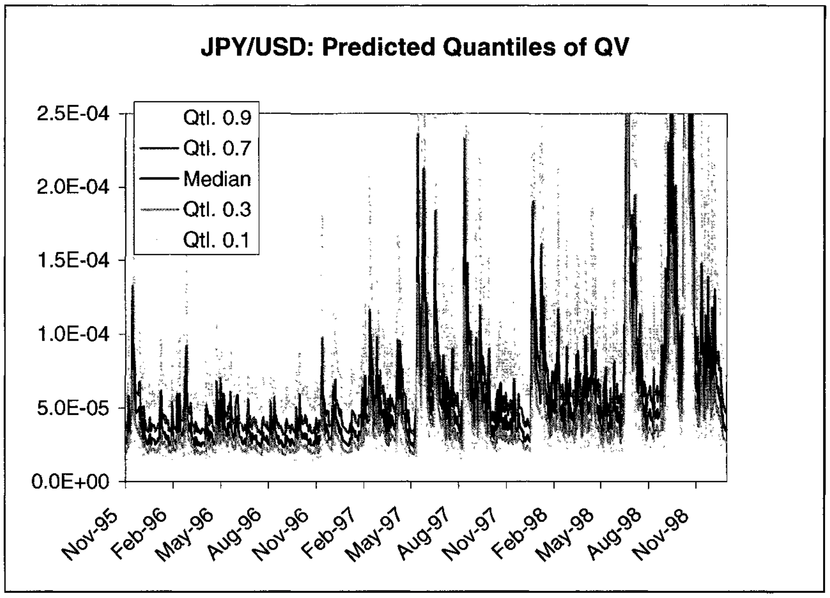


Figure 3-2: In-Sample Conditional Quantiles of Volatility

Chapter 4

Circuit Breakers and the Tail Index of Equity Returns [1]

4.1 Introduction

The idea of this short empirical investigation came in the spring of 2002. At the time, the stock market was continuing its steady fall from the stratospheric highs of the nineties (the key word here is *steady*). Indeed, although at the time Dow Jones Industrial Average (DJIA) had declined almost by 20% from its highest level in January of 2000 (and it had been even lower immediately after the events of September 11, 2001), we did not observe during this decline extreme fluctuations similar to those before the famous (or the infamous) crash of October 17, 1987. It is difficult to pinpoint a specific event which may have caused the market crash in October, 1987, i.e. one does not have reasons to treat it as an outlier when describing the dynamics of the stock market of that period. There may, however, have been a qualitative difference between the dynamics of the market in the recent past and in 1987. These observations came out in a discussion with John Galbraith, and we decided that it would be interesting to analyse them more formally.

A number of publications, both academic and professional, offered possible expla-

[1]This chapter is based in part on Galbraith and Zernov (2004)

nations for the events of October 17, 1987. Changes aimed, in particular, at limiting the impact of *program trading*(and especially of *portfolio insurance* programs) which was widely blamed for the crash, were introduced in the last years of the 1980s and at the beginning of the 1990s. There were also publications that analysed the effects of these changes. The interest in the topic, however, naturally faded away in the early nineties, and few people have touched it since. There are two main reasons why a new study may be able to provide further evidence regarding changes in market dynamics before and after the crash. First, there has been considerable progress in the theory and practice of the analysis of financial extreme events. We build our empirical study on these recent developments and focus our attention on the tails of distribution rather than on volatility measures, the methodology used by earlier researchers. By focusing on the tails, we can examine the evolution of the extreme behaviour over time, within a framework that does not put too restrictive conditions on the dynamics of the process. Second, a large sample is typically required for any study of extreme behaviour. This is due to the very nature of extreme events - they are rare. More than a dozen years of post-crash data allow us to obtain a more accurate description of the extreme dynamics of the time series of interest.

The rest of this chapter will proceed as follows: Section 4.2 presents historical facts regarding the emergence of program trading and of portfolio insurance techniques, introduces hypotheses about how program trading and portfolio insurance may have affected the dynamics of the stock market, describes the institutional changes which took place soon after the events of October, 1987, and reviews the previous literature on the effects of circuit breakers on the dynamics of the stock market; Section 4.3 reviews recent developments in the statistical theory of extreme events and describes the inferential methods which will be used in our study; Section 4.4 describes the data and empirical results; and Section 4.5 concludes the chapter.

4.2 Portfolio Insurance, Crash of October 1987, and Circuit Breakers

The events of October 1987 caught many academic economists and professionals in the the investment industry by surprise. It not only destroyed billions of dollars worth of wealth but also the comfortable belief shared by many professionals that the science of economics and finance had harnessed the wild forces of the market, that famous and spectacular market fads, bubbles and crashes, like the Dutch tulipmania of the 17^{th} century, or the South Sea bubble in Great Britain at the beginning of the 18^{th} century, or Black Monday in 1929 in the United States[2] were nothing more than historic facts, and that nothing similar would be possible again.

Elaborate market regulations and safeguards adopted since the crash of 1929 and the Great Depression, in the United States and elsewhere in the world, were among the sources of increased confidence. Other source was the ground-breaking developments in economics and finance that occurred at the end of 1960s and the beginning of 1970s: the rational expectations revolution in economics; the development, under the reigning hypothesis of market efficiency, of asset pricing and asset allocation theories; and the emergence of financial engineering. Ironically, in the postmortem analysis of the events of October 1987, many put the blame for the crash on investors' over-reliance on the financial technologies which were supposed to protect them from financial bad weather.

The buzz phrases mentioned almost every time in discussions about causes of the events of October 1987 are: *program trading, index arbitrage*, and *portfolio insurance.* Sometimes, however, authors are not very scrupulous in distinguishing between these terms. To avoid this, we define and discuss briefly each of these terms below.

[2]There is an abundant literature describing and analysing these famous historical market catastrophes. See, for example, Posthumus (1929), Garber (1990), and White (1990).

4.2.1 Program trading

The term *program trading* was probably born at the beginning of 1970s when investment professionals were forming groups of stocks with similar characteristics, in order to design portfolios with given properties. These groups were called *baskets.* When adjusting their portfolios, managers were trying to sell whole baskets of stocks as quickly as possible and to re-invest funds into other baskets. A basket of stocks to be bought or sold came to be referred to as a *buy* or *sell program.* This term had come into use even before index funds and portfolio insurance arrived on the scene (Miller 1989, Chapter 3). In general, program trading can be defined as a sequence of trades, usually automatically initiated, aimed at achieving a goal which is most often a pre-determined function of available market information. Program trading was a tool for implementing in practice the financial technologies developed at the end of 1960s and the beginning of 1970s; the emergence of affordable mainframe computers made program trading a fact of everyday life on Wall Street.

Although the trading on the exchanges was not computer-based at the time and sellers and buyers were matched either by specialists (for example, on the NYSE) or through the open outcry system (for example, on the CME), computers and computerised communications were becoming more and more important in the operation of the exchanges. In 1976, the AMEX and the NYSE introduced the Designated Order Turnaround system (DOT). Originally, DOT was developed as a system to batch together small orders to relieve specialists from some repetitive functions. After amendments and modifications, however, the capacity of DOT to handle larger trades had grown, and the system became a key component in the implementation of program trades. The performance of DOT would ultimately prove unsatisfactory during the crash of 1987, contributing to the turmoil on this day.

4.2.2 Index arbitrage

In the USA, the first derivative product based on an index was introduced in 1982: futures on the index of 1500 stocks published by Value Line, and traded on the Kansas City Futures Exchange. Later in the same year, the trading of futures on the Standard & Poor 500 (S&P 500) started on the CME (see Miller (1989) for details). Futures based on the S&P 500 proved to be very popular - it became the most actively traded futures contract. The high liquidity of index futures translated into lower transaction costs. This, together with lower margin requirements and an absence of restrictions on short sales, have soon made index futures the preferred vehicle for asset allocation, speculation, stock indexing, and portfolio insurance.

The idea of index arbitrage is very simple: to buy (sell) futures contracts and to sell (buy) stocks when futures are underpriced (over-priced) relative to the underlying basket of stocks. Since the implementation of index arbitrage requires the trading of many securities at the same time, the purchases and sales are likely to be executed through program trading. Index arbitrageurs consider, of course, transaction costs: futures prices, which lie within the transaction cost range from the value determined by the underlying index, do not present arbitrage opportunities.

Index arbitrageurs play an important role in transmitting information between the futures market and the stock market and in establishing the efficiency of the markets, i.e., if we believe that the existence of the market for index derivatives is beneficial, then the presence of arbitrageurs is necessary for this market to function. After the events of October 1987, index arbitrage programs were sometimes blamed for the crash, but it seems more likely that it was rather the inability of the arbitrageurs to perform their functions which contributed to the severity of the market failure of October 19, 1987.

4.2.3 Synthetic options and portfolio insurance

The discovery of the option-pricing formula by Fischer Black, Myron S. Scholes, and Robert C. Merton in 1973 was without a doubt a major event among the many other developments in economics and finance at the time. This discovery coincided with the establishment of the first US exchange for trading options on stocks, and once again, with commercial availability of the mainframe computers that were essential in implementing option valuation algorithms. To derive their option pricing result, Black and Scholes used the assumption of no arbitrage, and the equivalence of an option to a continuously adjusted position in a risky asset and a risk-free bond. This notion of an equivalent portfolio immediately suggested the possibility of creating *synthetic options*. The use of synthetic options as a portfolio insurance tool was first popularised, probably, by O'Brien (1982) and Rubinshein and Leland (1981)[3][4].

Synthetic portfolio insurance strategy is essentially a replication of a protective put option. A real protective put strategy is a combination of a put option on a stock or an index and a long position in this stock or index. Synthetic put replication requires buying the stocks as the market rises and selling as it falls. There are several important differences between the strategy using a real protective put and the strategy based on put replication. The strategy using a real protective put provides the investor with a minimum floor below which the portfolio value cannot fall. The synthetic portfolio insurance strategy promises such a floor theoretically; this promise, however, may not be fulfilled in practice due to imperfections of the market, as the events of October 1987 demonstrated. If the investor decides to follow the protective put strategy, he has to pay the price of the put up front, which is essentially the insurance premium. With the synthetic strategy, the investor does not have to pay anything up front. This does not mean, however, that the synthetic insurance strategy is free: its cost is the reduction in the achievable returns in periods in which the stock returns exceed cash returns. This reduction is called a *shortfall*. In addition to the shortfall, there are other costs, some of which were anticipated by those who offered or adopted synthetic insurance strategies in

[3]The authors later joined forces to form Leland O'Brien Rubinstein Associates which launched its first portfolio insurance products in 1982.

[4]Historical development of portfolio insurance strategies and their presumed role in the crash of 1987 is eloquently presented in Jacobs (1999). Our presentation of the topic draws from this source.

the period of their introduction in the beginning-middle of the 1980s, yet some came to many as an unpleasant surprise.

The market for listed options was only emerging at the time, and large institutional investors were not satisfied with the available menu of options listed (even including available OTC contracts). Synthetic puts seemed to offer a way to go beyond this menu. Originally, most of the synthetic insurance programs were implemented through the stock market; but soon after the introduction of stock index futures in 1982 the latter became the preferred instrument for portfolio insurance programs, due to lower transaction costs and more lax margin requirements on short sales (index futures could be used to create synthetic cash positions by selling them short).

4.2.4 Synthetic portfolio insurance and market dynamics

As we have already mentioned, as a result of the rational expectations revolution in economics, theoretical discoveries in finance and advancements in financial engineering, belief in the concepts of market efficiency and rationality of investors was strong among practitioners and even more so - among academic researchers. After the events of October 1987, some researchers looked for the causes of the crash within the paradigm of market efficiency and rationality, while others, in the spirit of John Maynard Keynes, tried to find an explanation for the crash in the animal spirits of the investors.

It is not our objective to introduce these theories in any detail here. We simply want to point out that many of these theories suggest that implementation of synthetic portfolio insurance by some investors would amplify market movements. For example, one of the approaches to explaining crashes of financial markets is termed *informational cascades.* An informational cascade occurs when a sequence of agents base their actions not only on their private information but on observed actions of other agents. Such behaviour is not necessarily irrational: obtaining adequate private information may be costly or time-consuming, and it may well be an optimising behaviour to incorporate into the decision signals contained in the actions of other market participants (see an extensive review of theories of rational herding in Chamley (2003) and in references contained therein). Portfolio insurers are essentially noise traders who trade in the direction of the movement of the market. Some of the investors may find it rational over a short horizon to act on signals contained in the price rather than on fundamentals (see Froot, Scharfstein, and Stein (1992)), and thus the presence of noise traders who coherently trade in the direction of market movements may, under this theory, amplify market fluctuations.

Shleifer and Vishny (1997) suggest a theory where arbitrageurs may decide not to trade against mispricing caused by the noise traders, even though the former believe that the mispricing will be corrected eventually. This is true because the arbitrageurs have to commit capital as long as the mispricing persists, and in the real world they act under capital constraints (for example, they may need external investors who are willing to capitalise them). Again, the presence of noise traders such as synthetic portfolio insurers, who coherently trade along with the market may lead to an increase in the length of mispricing spells, and the longer is the persistence of these spells, the less may arbitrageurs be willing to trade against this noise.

Generalising different theories for the purposes of our discussion, we can view financial markets as a dynamic system; stock prices and stock indices measure responses of this system to arriving information, and they are informational inputs to the system at the same time. We can view synthetic portfolio insurance programs as a positive feedback loop. Starting at the beginning of the 1980s, the number of market participants who adopted the strategy grew, as did the strength of the feedback, leading, as predicted by the dynamic control theory, to longer and more volatile responses of the market to arriving news. Ultimately, if the feedback is sufficiently strong, the system may begin to amplify very small input signals. If this stylised vision of financial markets and of the role of portfolio insurance is correct, we may observe changes in the dynamic behaviour of the market at some point prior October 1987.

4.2.5 Circuit Breakers

As a response to the events of October 1987, regulatory authorities launched a broad investigation into possible causes of the crash. Two major investigative efforts in this direction were undertaken by the Presidential Task Force and by the Securities and Exchange Commission. The finding of the former are known as the *Brady Report*. Among the recommendations for changes in the institutional structure of the markets which would either prevent such events in the future or at least mitigate their effects were circuit breakers. Originally, US stock exchanges put circuit breakers in place in October 1988.

Circuit breakers constitute a set of rules that may halt trading when a change during a trading day, or from the previous day close price of a stock price or a market indicator like a stock index exceeds a defined threshold. These rules may also impose restrictions on certain types of trading during periods of higher market volatility determined again by a stock index going outside a given collar. These latter restrictions include a requirement that arbitrage trades be *stabilising*, i.e., that they be implemented on the uptick during market declines or on the downtick during market advances, or that program trades be put in a *sidecar*[5], i.e., delaying their execution for a certain period of time. Rules 80A and 80B set circuit breaker rules for the NYSE; as has been mentioned, they were introduced originally in 1988 and modified many times later. Specifically, the triggering thresholds, originally set in absolute values, were relaxed in July 1996 and in January 1997 to account for an increase in the absolute values of the stock indices (DJIA and S&P 500); starting in February 1999, the trigger thresholds came to be defined as percentages of the indices, and their absolute values have subsequently been set quarterly using these percentage trigger levels. For more detailed descriptions of circuit breaker trading rules, see Lindsey and Pecora (1998) and official documents of the exchanges.

The designers of the circuit breakers envisioned that trading halts would provide "breathing room" for investors during periods of intensive trading, and improve dissemi-

[5]The sidecar rule was abolished in 1999.

nation of information among the investors, i.e., essentially help the markets to approach a fully revealing rational expectations equilibrium (the motivation behind the circuit breakers is not expressed, of course, exactly in these words in the reports and regulatory documents)[6]. The evidence with respect to the efficiency and usefulness of circuit breakers is mixed: there are theoretical investigations which find that circuit breakers may have even adverse effects on market volatility, while some empirical studies have found no evidence that the imposition of price limits has reduced volatility of the markets. Subrahmanyam (1994), for example, concludes that circuit breakers may increase price volatility in the framework of the one-market intertemporal model considered, while in the two-market situation analysed, the agents may migrate out of the market with the circuit breaker, thereby transferring price variability to the alternative market.

Santoni and Liu (1993) analyse daily closing values of the S&P 500 index from July 1962 to May 1991, fitting an ARCH model to the data and looking for structural breaks in the estimated model. They conclude that the data is not consistent with the hypothesis that adoption and revision of circuit breaker rules reduces the conditional variance of stock returns. The two studies of Ma, Rao, and Sears (1989a) suggest that the introduction of trading restrictions decreased volatility in the bond and commodity markets studied.

Studies on equity markets outside the U.S. include those of Roll (1989), Bertero and Mayer (1990), and Lauterbach and Ben-Zion (1993). The last paper examines the period surrounding the 1987 crash on the Tel Aviv market, where circuit breaker mechanisms were in place in 1987, and for which order imbalance data are available. These authors suggest that trading restrictions did not affect the overall degree of market decline, but did smooth returns around the crash date. Roll (1989) and Bertero and Mayer (1990) each compare 23 stock markets around the time of the crash, in monthly and daily data respectively; the latter study finds a substantial impact of circuit breakers on price

[6]Continuing the analogy between financial markets and a dynamic system, circuit breakers may be viewed as an attenuator or a phase shifter in the feedback loop.

declines, while the former does not. Nonetheless, as Lauterbach and Ben-Zion point out, trading restrictions differ substantially across international markets making the results of such exercises difficult to interpret.

To summarise, despite conflicting evidence on their effectiveness, circuit breakers have remained a part of the trading mechanism for more than 15 years now; circuit breaker mechanisms have been revised many times but they are unlikely to be abolished in the near future.

Empirical testing of effects and efficiency of the circuit breakers in the U.S. stock market is further complicated by the fact that many other changes were being introduced to the market structure and to market regulations concurrently. Examples are: increase in capacity and improvements in the electronic execution and information systems of the exchanges; and changes to the capital requirements of the market participants. When we state therefore in this study, that empirical evidence suggests that the introduction of circuit breakers in the late 1980s to the early 90s has affected extremal dynamics of market indices, we must offer the caveat that this change must be attributed to a whole complex of changes in the market structure and in investors' behaviour. Our analysis presents historical evidence that changes in the dynamics of stock indices coincide sometimes with known dates of changes in parameters of circuit breakers, but this evidence, in light of the discussion above, must be interpreted with caution.

4.3 Statistical Modelling of Extremal Events

In this section we briefly summarise some facts from extreme value theory (EVT) and statistical techniques relevant for the empirical analysis in our study. First, consider a sequence of i.i.d. r.vs. $\{X_i\}$. In the language of statistics, rare (or extremal) events in such a sequence are described by the behaviour of the tails of the distribution of X_i.

4.3.1 Estimation of the shape parameter of the generalised extreme value distribution

Statistical inferential methods for tail behaviour of a distribution function are most often based on the application of the Fisher-Tippett theorem which specifies limit laws for maxima of a sequence of i.i.d. r.vs. (see, for example, Embrechts, Klüppelberg, and Mikosch (1997, Ch.3)): properly normalised and scaled maxima converge in distribution to one of the *standard extreme value distributions*: *Fréchet*, *Gumbel*, or *(reversed) Weibull*[7]. These three distributions can be conveniently represented using the *generalised extreme value distribution (GEV)* with parameter ξ, namely

$$H_{\xi}(x) = \begin{cases} \exp\left\{-(1+\xi x)^{-1/\xi}\right\} & if \quad \xi \neq 0 \\ \exp\left\{-\exp\left\{-x\right\}\right\} & if \quad \xi = 0 \end{cases}, \text{ where } 1+\xi \text{x} > 0.$$

We say that the r.v. X belongs to *the maximum domain of attraction (MDA)* of the GEV H_{ξ} ($X \in MDA(H_{\xi})$) if there exist constants $c_n > 0$, $d_n \in \mathcal{R}$ such that

$$c_n^{-1}\left(\max_{i=1,n}(X_i) - d_n\right) \xrightarrow{d} H_{\xi}.$$

The primary case of interest of our study is when $\xi > 0$ (the *Fréchet* distribution) which characterises the MDA for heavy-tailed distributions: heavy tails is an empirical reality for financial series such as stock returns[8].

Fitting the GEV is central to the statistical analysis of extreme events. In many applications, including this study, the main interest is in estimating the shape parameter ξ of the limiting distribution. There are three main approaches to estimating the shape parameter of the GEV under the maximum domain of attraction condition: Pickhand's,

[7]In the extreme value theory, the reversed Weibull distribution with the location and the scale parameters omitted is traditionally referred simply is the Weibull distribution. The reverse Weibull distribution and the Weibull distribution are related one to another through the change of sign of the variable

[8]To provide a point of reference, the case $\xi = 0$ corresponds to the MDA for the distributions ranging from moderately heavy-tailed such as the log-normal distribution to light-tailed such as the normal distribution.

Hill's, and Dekkers-Einmahl-de Haan's estimators. While we use only Hill's estimator and its modifications in the empirical section, we give a brief description here of each of these approaches, following once again Embrechts, Klüppelberg, and Mikosch (1997, Ch. 6.4).

Method 1: Pickhand's Estimator. It can be shown that for $F \in MDA(H_\xi)$, $U(t) = F^{-1}(1 - t^{-1})$ satisfies

$$\lim_{t\to\infty} \frac{U(c(t)\,t) - U(t)}{U(t) - U(t/c(t))} = 2^\xi$$

where $c(t) \to 2$.

Pickhand's estimator is defined as

$$\hat{\xi}^P_{k,n} = \frac{1}{\ln 2} \ln \frac{X_{k,n} - X_{2k,n}}{X_{2k,n} - X_{4k,n}}$$

Under certain conditions, this estimator is weakly-consistent (or even strongly consistent) and asymptotically normal (see Dekkers and de Haan (1989)).

Method 2: Hill's Estimator. Suppose that $X_1, ..., X_T$ are i.i.d. with d.f. $F \in MDA(H_\xi)$, $\xi > 0$[9]. It can be shown that the above is equivalent to

$$\bar{F}(x) = x^{-\alpha} L(x)\,,\ x > 0^{10}, \tag{4.1}$$

[9] I.e. that the distribution belongs to the MDA of the *Fréchet* distribution.

[10] See Embrechts, Klüppelberg, and Mikosch (1997, Theorem 3.3.7).

where $\alpha \equiv 1/\xi$, for a slowly-varying L [11]. Let $X_{j,T}$ denote the j^{th} order statistic from the sample. Hill's estimator takes the form

$$\hat{\alpha} = \left(\frac{1}{m} \sum_{j=1}^{m} \ln X_{j,T} - \ln X_{m,T} \right)^{-1} \tag{4.2}$$

where $m = m(n) \to \infty$ in an appropriate way. Hill's estimator can be derived in a few different ways: through the maximum likelihood approach, the regular variation approach, or the mean excess function approach. The fact that different approaches yield the same result can be viewed as an indication that Hill's estimator is, in some ways, very "natural".

The theorem given in Embrechts, Klüppelberg, and Mikosch (1997, Theorem 6.4.6) establishes the weak consistency of Hill's estimator for weakly dependent sequences or linear processes, strong consistency for i.i.d sequences, and asymptotic normality for i.i.d. sequences. A penultimate approximation of Hill's estimator's distribution (an improvement in comparison with the asymptotic normal approximation) is given in Cheng and de Haan (2001) for the case of i.i.d. observations.

The consistency (and under additional assumption - even asymptotic normality) of the Hill's estimator for a wide class of strictly stationary processes satisfying certain mixing conditions was first shown in Hsing (1991); this class includes linear processes. The latter result forms the basis of the statistical technique used in our study of detecting a change in the tail index at an unknown date (Quintos, Fan, and Phillips (2001)).

Method 3: The Dekkers-Einmahl-de Haan estimator for $\xi \in \mathcal{R}$. Dekkers, Einmahl, and de Haan (1989) suggested a generalisation of Hill's estimator for cases in which $\xi \leqslant 0$. Since the latter case is not of interest in this research, we do not present further details here.

[11] A Lebesgue-measurable function L on $(0, \infty)$ is called *slowly varying* at ∞ (we write $L \in \Re_0$) if $\lim_{x\to\infty} \frac{L(tx)}{L(x)} = 1, \ t > 0$

Maximum likelihood estimation of the GEV parameters. The generalized extreme value distribution can also be written as

$$H_{\xi;\mu,\psi}(x) = \exp\left\{-\left(1-\xi\frac{x-\mu}{\psi}\right)^{-\frac{1}{\xi}}\right\}, \quad \left(1-\xi\frac{x-\mu}{\psi}\right) > 0,$$

if one includes explicitly the location and the scale parameters in its definition. We shall denote the vector of parameters $\theta = (\xi, \mu, \psi)$. We shall assume that the random variables under consideration have approximately a $H_\theta(x)$ distribution, meaning that the distribution of these variables belongs to the domain of attraction of the corresponding extreme value distribution.

In maximum value theory data is often represented in the following form:

$$\begin{aligned} X^{(1)} &= \left(X_1^{(1)}, ..., X_s^{(1)}\right) \\ X^{(2)} &= \left(X_1^{(2)}, ..., X_s^{(2)}\right) \\ &\dots \\ X^{(n)} &= \left(X_1^{(n)}, ..., X_s^{(n)}\right) \end{aligned}$$

where the vectors $X^{(i)}$ are assumed to be i.i.d. The basic i.i.d. sample from $H_{\xi;\mu,\psi}$, on which the statistical inference is performed, then consists of $X_i = \max\left(X_1^{(i)}, ..., X_s^{(i)}\right)$.

The set-up above corresponds to the standard parametric case and *in principle* can be solved by maximum likelihood. Problems typical to extreme value theory ususlly arise when the support of the underlying d.f. depends on the unknown parameters. In finance, most of the distributions have a support unbounded to the right; hence, the MLE technique offers a reliable procedure.

4.3.2 Detection of a break in the shape parameter

Phillips-Loretan test for equality of shape parameters of two samples. Loretan and Phillips (1994) propose, in the context of the discussion of methods of testing covariance stationarity of heavy-tailed time series, a test that is suitable for testing for equality of tail indices of two subsamples. Assuming that the sample $\{X_i\}_{i=1}^{T}$ is split into two subsamples of sizes m_1 and m_2, $m_1 + m_2 = T$, the Phillips - Loretan statistic is defined as:

$$Q_1 = \frac{m_1 \left(\hat{\alpha}_1 - \hat{\alpha}_2\right)^2}{\left(\hat{\alpha}_1^2 \eta_1 + \frac{m_1}{m_2} \hat{\alpha}_2^2 \eta_2 - 2\eta_{12}\right)} \xrightarrow{d} \chi_1^2 \tag{4.3}$$

where $\hat{\alpha}_1$ and $\hat{\alpha}_2$ denote respectively the Hill's estimates of the tail index for the first and the second parts of the sample. It has been shown originally in Hall (1982) that Hill's statistic has an asymptotically normal distribution provided that m goes to infinity at an appropriate rate (dependent on the parameters of the distribution). This leads to a χ_1^2 asymptotic distribution of the Phillips-Loretan statistic. In practice, the scaling factors η_1 and η_2, and the covariance factor η_{12} are substituted in (4.3) by their estimates which are computed in a way accounting for the dependency in the series (the estimator of η_1 and η_2 is given by formula (4.5); the estimator of the covariance factor η_{12} is constructed similarly)[12].

Test for an unknown breakpoint in the shape parameter. The structural change tests of Quintos, Fan, and Phillips (2001) are based on sequences of tail index estimates. The null hypothesis is that the tail index has the constant value α over the real interval $t \in [t_0, T - t_0]$, with an alternative of departure from α at some point in the interval, and is tested with sequences of estimates defined over different sets of samples. Recursive estimates produce a sequence of estimates $\hat{\alpha}_t$ using samples $1, \ldots t$, for $t = t_0, t_0{+}1, \ldots T{-}t_0$; rolling estimates use samples $1 + (t - t_0), \ldots t$, with t indexed over the same values, for a constant sample size $t_0 \equiv \gamma T$. The sequential tests use both a recursive set of estimates and a reverse recursive set labelled $\hat{\alpha}_{2t}$ defined over samples $1, \ldots t$ and $t + 1, \ldots T$ respectively, where once again t indexes the values $t_0, t_0 + 1, \ldots T - t_0$. The sequences of

[12]This form of the Phillips-Loretan statistic, which accounts for dependence in the series and between the subsamples, was suggested in Quintos, Fan, and Phillips (2001).

test statistics are based on the sequences:

$$\begin{aligned} Y_T(t) &= \left(\frac{tm_t}{T}\right)^{1/2}\left(\frac{\hat{\alpha}_t}{\hat{\alpha}_T}-1\right) \\ V_T(t) &= (\gamma m_\gamma)^{1/2}\left(\frac{\hat{\alpha}_t^*}{\hat{\alpha}_T}-1\right) \\ Z_T(t) &= \left(\frac{tm_t}{T}\right)^{1/2}\left(\frac{\hat{\alpha}_t}{\hat{\alpha}_{2t}}-1\right), \end{aligned} \tag{4.4}$$

where $\hat{\alpha}_t^*$ is Hill's estimator using the sample of size $w = [\gamma T]$ ending at t, which is rolled through the whole sample. m_t and m_γ denote the number of order statistics used to compute correspondingly $\hat{\alpha}_t$ and $\hat{\alpha}_t^*$. The sup's over t of $Y_T^2(t)$, $V_T^2(t)$, and $Z_T^2(t)$ were suggested in Quintos, Fan, and Phillips (2001) as the statistics for detecting a breakpoint; the critical values of these statistics are tabulated [13].

Processes with Conditional heteroskedasticity and the Tail Index. A prominent property of financial data is that it is not i.i.d. but exhibits dependence in the second moment (and possibly in the higher moments). Quintos, Fan, and Phillips (2001) notice that the dynamics of the second moment of a GARCH process can be described by a linear process, and thus the asymptotic theory for linearly dependent processes developed by Hsing (1991) applies to squared logarithmic returns. The required modifications to the statistics described above are based on a re-scaling of each of the equations of (4.4) to account for the different variances of the Hill estimates when the raw data are serially dependent; with the appropriate re-scaling the same asymptotic distribution holds. That is, serial dependence does not affect the consistency or asymptotic normality, with convergence at the rate $m^{\frac{1}{2}}$, of the Hill estimator. The effect of serial dependence is to increase the variance in the asymptotic distribution, and a valid test must embody an estimate of this higher variance. With an appropriately modified asymptotic variance estimate, inference on changes in tail behaviour can then be conducted, despite the

[13] We shall not use the recursive test in the empirical part of our research because it has power with respect to a decrease in the tail index only, not an increase.

clustering of large squared returns that arises in such data.

In these cases the statistics defined in Equation (4.4) are re-scaled by the factor η, which is equal to 1 in the i.i.d. case. Hsing (1991, Theorems 3.3, 3.5, Corollary 3.4) and Quintos, Fan, and Phillips (2001, Theorem 8) describe the asymptotic theory for the scaling factor η and also suggest how an estimate of the scaling factor $\hat{\eta}$ can be constructed from the data. Specifically, $\hat{\eta} = 1 + \hat{\chi} + \hat{\omega} - 2\hat{\psi}$, where

$$\begin{aligned} \hat{\chi} &= 2\hat{\alpha}^2 m_w^{-1} \sum_{j=1}^{w} c_{w,j} c_{w,j+1} \\ \hat{\psi} &= \hat{\alpha} m_w^{-1} \sum_{j=1}^{w} c_{w,j} d_{w,j+1} + c_{w,j+1} d_{w,j} \\ \hat{\omega} &= 2 m_w^{-1} \sum_{j=1}^{w} d_{w,j} d_{w,j+1}. \end{aligned} \tag{4.5}$$

with $c_{w,j} = (\ln(X^2)_i^w - \ln(X^2)_{(w-m_w+1)}^w)_+$, and $d_{w,j} = I(\ln(X^2)_i^w > \ln(X^2)_{(w-m_w+1)}^w)$, where $I(.)$ is the indicator function and where it is now understood that $\hat{\alpha}$ is estimated on the squares of the data. We have followed Quintos et al. in these definitions and in defining the additional notation w to indicate the window on which estimation takes place, for the full sample $w = T$, for a rolling sample $w = \gamma T$, and so on. The modified statistics based on (4.5) account for GARCH(1,1) dependence in the squared returns. These are the versions of the (rolling and sequential) tests that we employ below: all of the core inferential results reported in sections 4.1 and 4.2 use these tests which allow for dependence in the conditional variance. If $\{X_i\}$ has a tail index α, then $\{X_i^2\}$ has a tail index $\alpha/2$; in the figures below we re-scale the estimated tail index parameters by 2 to report the estimated tail index for the original series of logarithmic returns.

Two important notes are required here. First, the limit theory developed in Quintos, Fan, and Phillips (2001) assumes that the process under investigation is a strong GARCH process. The conditional distribution of innovations is assumed to be normal and the tail parameter α is in this case a function of GARCH (or ARCH) parameters. The asymptotic theory developed in Quintos et al. does not apply to cases in which the conditional distribution of innovations has fat tails.

Second, the tests described above are on the unconditional distribution of the variable of interest (logarithmic equity returns in our case: $r_t \equiv \ln(P_t/P_{t-1})$, where P_t denotes here the price of an asset or the value of the index). In investigating changes in market structure, we are interested in the question of whether we observe corresponding changes in the distribution function of the returns $\{r_t\}$, which is an important object of interest to market regulators in particular. In contrast, short-term risk management activity would typically be concerned with a standardized return of the form $\tilde{r}_t = \frac{r_t}{\hat{\sigma}_t}$ or $\tilde{r}_t = \frac{(r_t - \hat{\mu}_t)}{\hat{\sigma}_t}$, where $\hat{\mu}_t$ and $\hat{\sigma}_t$ are the estimated daily mean and the conditional standard deviation of returns; because an estimate of $\hat{\sigma}_t$ is available to the risk manager, actions at date t can be conditioned on this information. Tests for changes in the tail index of standardized returns can also be constructed (see Andreou and Ghysels (2002)).

Even if a process under investigation is a strong GARCH, the break in the tail index of this process may happen either due to a change in the GARCH parameters, or to a change in the distribution of innovations, or both. The statistics of Quintos et al. can be used only in the first case. The statistics of Phillips and Loretan or tests based on the bias-corrected estimators described below can be used in any of these three cases to test the equality of the tail indices in the pre-determined subsamples.

4.3.3 Choice of m and the bias-corrected Hill's estimator

The choice of m is a moot point when using Hill's estimator. As we have stated above, consistency and asymptotic normality of the Hill's estimator require that m go to infinity at an appropriate rate. However, depending on the choice of m and the properties of the slowly-varying function $L(x)$ there may be a trade-off between the bias and the variance of Hill's estimator. Usually, an increase in the rate of growth of m decreases the asymptotic variance but increases the bias.

Algorithms to choose the number of order statistics m. One has to provide a more detailed specification of the d.f. of X than the one in Equation (4.1) in order to develop feasible methods for choosing m. Assume that the *second order properties* of $F(x)$ are specified as follows:

$$\lim_{x\to\infty} \frac{\bar{F(tx)}/\bar{F(x)} - t^{-\alpha}}{a(x)} = t^{-\alpha}\frac{t^{\rho}-1}{\rho},\ t > 0, \tag{4.6}$$

where $a(x)$ is a measurable function of constant sign. If $\rho = 0$, the right-hand side of (4.6) is interpreted as $t^{\alpha}\ln t$.

Hall (1990) suggests a solution to the problem of the optimal choice of m for the special case of the presentation in (4.6):

$$\bar{F}(x) = ax^{-\alpha}[1 + bx^{-\beta} + o(x^{-\beta})], \tag{4.7}$$

where $\alpha > 0$, $b \in \mathbb{R}$, $\beta > 0$ (the case $\beta = 0$ corresponds to the expansion $\bar{F}(x) = ax^{-\alpha}[1 + b\ln x + o(\ln x)]$). It was shown in Hall's (1990) article that for the class of distributions which can be presented in the form of Equation (4.7), for a given m,

$$\mathbf{E}[\hat{\alpha}^{-1}] \approx \frac{1}{\alpha} - \frac{b\beta}{\alpha(\alpha+\beta)}a^{-\frac{\beta}{\alpha}}\left(\frac{m}{n}\right)^{\frac{\beta}{\alpha}}. \tag{4.8}$$

The equation above shows clearly that the bias increases in m. Hall makes a further assumption that $\alpha = \beta$ in the presentation in (4.7) and develops a subsample bootstrap technique for choosing m. In Danielsson and de Vries (1997) and Danielsson, de Haan, Peng, and de Vries (2001) the approach of Hall is extended to allow for a more general specification of the second order behaviour, as specified in Equation (4.6).

While the focus of our empirical study is on the change in the tail index and not its level *per se*, and thus one can expect that the importance of the bias is attenuated when we look at the difference between two estimators, we still want to account for the bias whenever this is possible. Feuerverger and Hall (1999) and Quintos, Fan, and

Phillips (2001) (the latter in the context of their methods for detecting a break in the tail index) suggested an adaptive algorithm for choosing the time-varying values of m. In our empirical computations of the statistics in (4.4), we compute m as a proportion of the sample size, following a recommendation in DuMouchel (1983); choosing m adaptively does not qualitatively change the test results. When we test the equality of the tail index across sub-periods, we opt not to use the bootstrapping techniques mentioned above but apply a somewhat different approach to correcting the bias of Hill's estimator of the tail exponent.

Bias-corrected estimation of the tail index. The method of bias correction used in some of the empirical computations of this study was suggested in Huisman, Koedijk, Kool, and Palm (2001). Their approach is also based on the decomposition (4.7) of the distribution function of the data. They noticed that if in this decomposition one imposed the restriction that $\alpha = \beta$, then the bias of Hill's estimator (see Equation 4.8) would be linear in m[14] .This observation allows us to run a regression of Hill's estimator on a constant and the number of order statistics m used in its computation. The unbiased estimate of the tail exponents will be thus the estimated intercept of this regression. Specifically, Equation (4.8) can be transformed into

$$\alpha^{-1}(m) = \mu_0 + \mu_1 m + \epsilon(m), \quad m = 1, ..., k. \tag{4.9}$$

It follows from the analysis above that an unbiased estimate of α^{-1} can be obtained only when m approaches zero: evaluation of Equation (4.9) for m approaching zero gives an unbiased estimate of α^{-1}, under the assumptions specified, and it is equal to the intercept term μ_0.

The parameters in (4.9) can be estimated using ordinary least squares but certain

[14]The assumption that $\alpha = \beta$ has been used by other authors (eg. Hall (1990). Huisman et al. present an argument based on the simulation evidence and also citing Dacorogna, Müller, Pictet, and de Vries (1995) that imposing such a restriction does not distort estimation results for a range of relevant distributions

econometric problems remain to be resolved: first, the error term in the regression is heteroskedastic; second, there is a problem of overlapping data due to the way the $\alpha^{-1}(k)$'s are estimated. Several approaches can be taken to resolving these problems; we use in our study the weighted least squares estimation of the regression (4.9) with the weighting matrix computed using asymptotic properties of order statistics (see Huisman, Koedijk, Kool, and Palm (2001, Appendix)).

The number of Hill's statistics k included in the regression (4.9) still has to be determined but one may expect the sensitivity of this bias-corrected estimate to a choice of k to be lower than the sensitivity of the Hill's procedure to the value of m - the number of order statistics used - as there is implicit averaging across the values of Hill's statistics corresponding to different ms. The assertion just made is confirmed by simulation results presented in Huisman, Koedijk, Kool, and Palm's (2001) paper. Based on these simulation results, the authors recommend choosing k in the range $(\frac{T}{3}, \frac{T}{2})$; it would be unacceptable, of course, to choose m for Hill's estimator in this range due to the increasing bias.

4.4 Empirical Dynamics of Extremal Behaviour of Stock Indices

4.4.1 Data

This study examines the two longest-standing indices of U.S. equity prices, the Dow Jones Industrial Average and the broader S&P 500 Index. Because the structure of circuit breakers is based on daily stock price changes–that is, the measure of price change which these devices use is implicitly set to zero at the beginning of each trading day–we examine daily changes in these index levels. Daily information is available for a long historical period for the Dow Jones index; we use data from October 1 1928, the date at which the index took its current 30-stock form. Our daily sample of the S&P 500 begins on January

3 1950; both series end in March of 2002, for total sample sizes of 18443 days (DJIA) and 13129 days (S&P). Each of these data series is transformed to daily logarithmic returns, $r_t = \ln(\frac{p_t}{p_{t-1}})$, where p_t is the index value at time t, reducing sample sizes by 1.

4.4.2 Evolution of the tail index on the full samples

The first set of empirical results concerns the hypothesis of constancy of the tail index on the full historical samples. The results for rolling tests using different rolling windows as proportions of the sample and for the sequential tests are presented in Table 4.1 (see relevant formulas in Equation (4.4)). For the results reported in the tables, we take m to be 10% of the relevant sample size [15].

Test	γ	**1%c.v. (5%c.v)**	**DJIA**	**Date**	**S&P 500**	**Date**
Rolling	0.15	1.9	3.39	5-Nov-73	3.01	16-Aug-82
	0.2	2.3	3.33	10-Feb-86	4.24	24-Apr-84
	0.25	2.55	3.7	9-Dec-86	5.83	6-Oct-87
	0.3	2.86 (2.12)	5.44	3-Jun-87	6.81	27-Mar-87
Sequential		28.82 (18.31)	52.9	5-Nov-98	79.51	11-Sep-98

Table 4.1: Full sample rolling and sequential tests for different sizes (as a proportion of the sample) of the moving window γ

These results are easily summarised. On all tests and on both stock indices, the null hypothesis of constancy of the tail index is rejected at a confidence level of 0.01 (the smallest tabulated); that is, each of the test statistics exceeds (by a substantial margin) the 99th percentile of the null distribution[16]. The dates of maxima of the statistics occur,

[15] Quintos, Fan, and Phillips (2001) use the same rule in the empirical part of their paper.

[16] As we mentioned above, we shall not use the recursive test because of the lack of consistency against an increase in the tail index (decrease in the tail thickness) over the sample. We note, however, that the recursive statistics corresponding to those in Table 1 also show strong rejection of the null of constant α on the full samples, and that these tests are consistent against a substantial decrease in the tail index as appears to have occurred in the neighbourhood of 1987.

in most of the cases, in the eighties: they precede the crash of 1987. Sensitivity to the number of included order statistics, m, is very low on these full-sample tests. All rolling statistics are in the 1% tail for a wide range of values of m (we examined values as low as 2% of the relevant sample size). The sequential tests are similarly robust, with the exception that for very small values of m such as 2% of the sample size, the sequential test on the S&P500 data falls outside the 5% tail (the p-value is approximately 0.06). The sequential tests are unaffected in this case by moderate changes in the initial sample size, since the maxima do not occur near the ends of the sample, where the trimming affects the included statistics.

A further form of sensitivity analysis addresses the possibility that changes in index composition may have affected these results, for example by including more relatively-risky technology firms in the index[17]. To explore this, we examine four firms that have been part of the DJIA from the beginning of our sample in 1928, or shortly thereafter: E.I. DuPont de Nemours (from 1935), Exxon Mobil (Formerly Exxon and Standard Oil of New Jersey), General Electric and General Motors. We repeat the tests just described on an equally-weighted index of the prices of these four large industrial firms, using data beginning in July of 1962 (note that these sample sizes are therefore smaller than for the Table 4.1 results). The results are qualitatively unchanged. Test statistics for this four-stock index, corresponding with those of Table 4.1 for the DJIA and the S&P 500, are as follows. In the rolling tests for $\gamma = 0.15, 0.20, 0.25, 0.30$, and the sequential test with $t_0 = 500$, the values are 2.46, 2.88, 4.35, 6.28 and in the sequential test, 169.0. Each of these statistics exceeds the 1% critical value. The estimated dates of structural change are again in the mid-1980's, with the exception of the rolling test for $\gamma = 0.15$ where a date in 1976 gives the maximal statistic.

For the DJIA data, the months in which the maximum statistics in the rolling tests occur vary from April 1986 ($\gamma = 0.15$) to October 1987 ($\gamma = 0.2$); in the sequential test,

[17]Note, however, that current Dow components Intel and Microsoft were added to the index only in November of 1999, affecting only a small part of the sample, and well past the estimated dates of structural changes.

the maximum occurs in September 1998. In the S&P 500 data, the maxima occur between August 1982 ($\gamma = 0.15$) and October/December 1987 ($\gamma = 0.25, 0.30$); note, however, from Figure 4-2, upper-left panel, that the early peak is isolated and only slightly exceeds the statistics from 1987. The sequential test again shows a late maximum in October 1998[18]. All of these dates should be interpreted cautiously, in part because the alternative hypothesis of a break at a particular date is not something that we wish to interpret literally in the present discussion. Moreover, we note from Quintos, Fan, and Phillips (2001) (see esp.Table 3 in this source) that estimates of breakpoint dates may be quite poor, and even in favourable cases show substantial variability around the true date. Nonetheless, we note that the strongest evidence of statistically significant change in the tail index does occur in the period of interest, i.e. weeks or months before the events of October 1987.

The graphs of the rolling test statistic and the tail index for varying sizes of the moving window are presented in Figure 4-1 for the DJIA and in Figure 4-2 for the S&P 500. These graphs indicate clearly that the change in the tail index which is identified by the tests is in the direction of a fall in α, that is, an increase in the relative frequency of extreme events.

Next, we consider constancy of the tail index on the post-1984 sample alone.

4.4.3 Circuit breakers and the evolution of the tail index

The results of the analysis in the previous sections suggest that there was a qualitative change in the extreme behaviour of the stock index series some time before the crash of October 1987. We are now interested in the possibility that the tail index may have increased (tail thickness decreased) later in the sample. Since the first circuit breaker took effect in October of 1988, we have an initial sample of almost one thousand trading days (January 1985 - October 1988) on which to base pre-circuit breaker estimates.

[18]Recall, that the sequential test compare samples before and after a hypothesised break date, whereas rolling and recursive tests use data up to the particular date.

Recall, however, from the discussion in Section 4.2, that first, the post-crash changes to the regulations and circuit-breaker mechanisms were introduced gradually, and second, that the severity with which the circuit breaker regulations bind has been changing over time as the trigger points as a percentage of index value have changed (see Booth and Broussard (1998) for a more detailed analysis of this issue).

We examine the possibility of a significant change in α in this later sample of data using the rolling and sequential tests which are consistent for a change in α in any direction. The results of the tests on the post-program-trading subsample are presented in Table 4.2.

Test	γ	**1%c.v. (5%c.v)**	**DJIA**	**Date**	**S&P 500**	**Date**
Rolling	0.15	1.9	3.13	5-Mar-97	2.22	6-Jul-01
	0.2	2.3	2.49	13-Mar-97	3.91	12-Mar-02
	0.25	2.55 (1.98)	2.17	24-Mar-97	2.16	1-Mar-02
	0.3	1.86	4.62	10-Mar-97	2.24	24-Jan-96
Sequential	28.82	28.82 (18.31)	62.9	5-Nov-98	18.7	11-Sep-98

Table 4.2: Sub-sample rolling and sequential test statistics (2-Jan-85 - 14-Mar-02) for different sizes (as a proportion of the sample) of the moving window γ

The evidence is weaker on this pair of samples, and the tests with unknown breakpoint do not find a statistically significant increase in the tail index for the end of the 1980s - beginning of the 1990s, which could have been associated with the original introduction of circuit breakers. There is still substantial evidence, however, against the null of constancy of the tail index. Five of eight rolling statistics have p-values below 0.01, the remaining three rolling tests, and the sequential test are in the upper 5% of the null distribution. For the DJIA for all four sizes of the moving window γ of the rolling test, the estimates of the date of the change for this shorter sample occur early in 1997. This coincides with the known changes to the circuit breaker trigger levels (see section 4.2.5). Inspection of the graphs of the tail index of the DJIA in Figure 4-3 (left panels) shows that there is a decrease in the tail index at the beginning of 1997, and this decrease is being picked up by the test as statistically significant. The estimated timing of the change in the S&P 500 index is different and does not coincide with any known changes in the trading regulations[19] [20] (see Figure 4-3, right panels).

The conclusion about the presence of the change in α in the post-program-trading period is quite robust with respect to the choice of m : if it is equal to 5% of the relevant sample, the rolling test statistic is still significant at the 1% level in 6 cases out of 8, and in one case it is significant at the 5% level. With $m = 5\%$, the rolling test change dates are estimated to be in March of 1997 for two choices of γ, both for the DJIA and for the S&P 500 series. The sequential test statistics are not significant. The sequential test, however, does not perform well in small samples, as noted in Quintos, Fan, and Phillips (2001, p.65), and we regard the rolling test as the preferred alternative here.

Thus, although the analysis of the shorter post-program-trading period detects that there may have been statistically significant breaks in the tail indices in the post-program trading period and that the timing of these breaks can be associated with known changes in trading regulations, analysis using the statistics of Quintos, Fan, and Phillips (2001) does not provide evidence that the introduction of circuit breakers and other institutional changes in the late 1980s and early 1990s led to an increase in the tail index. This analysis, however, shouldn't be viewed as evidence of the contrary. We shall further investigate this issue in section 4.4.4 using the bias-corrected Hill's estimator. Before closing this section, we note, however, a few relevant features of the second-order dependence of the data.

Consider the properties of this dependence across sub-samples. The second-moment

[19] The relevant trigger levels of the circuit breakers are formulated in terms of the DJIA. Therefore, one may have expected that introduced changes would have a stronger effect on the DJIA tham on the broader S&P 500.

[20] Booth and Broussard (1998) present an empirical argument that a change in the trigger levels of the circuit breakers at the beginning of 1997 did not affect the distribution of the series. Our analysis shows the evidence to the contrary.

conditional dependence of financial returns data is very commonly modelled with a GARCH(1,1) specification: $\sigma_t^2 = \omega + a\sigma_{t-1}^2 + br_{t-1}^2$; typical estimates show that $a + b \cong 1$; moreover, $a + b = 1$ corresponds with the integrated GARCH (IGARCH) model, and this condition defines the threshold at which the unconditional second moment no longer exists. As has been mentioned in the discussion above (see page 99), if a process has the GARCH form with a known distribution of standardised residuals, then the tail index is a function of the GARCH parameters; see, in particular Mikosch and Stărică (2000). We would therefore expect that differences in the tail index across sub-samples would correspond with differences in the estimated GARCH parameters, to the extent that the GARCH model provides a reasonable approximation in these data. Table 4.3 presents estimates of this model on each of the three sub-samples examined in this section.

DJIA		
Sample	$\hat{a}$ (s.e.)	$\hat{b}$ (s.e.)
01-Jan-50 - 31-Dec-84	0.93 (0.04)	0.06 (0.01)
02-Jan-85 - 31-Oct-88	0.85 (0.01)	0.15 (0.05)
01-Nov-88 - 14-Mar-02	0.93 (0.06)	0.06 (0.02)
S & P 500		
01-Jan-50 - 31-Dec-84	0.92 (0.03)	0.08 (0.01)
02-Jan-85 - 31-Oct-88	0.83 (0.10)	0.16 (0.05)
01-Nov-88 - 14-Mar-02	0.95 (0.05)	0.04 (0.02)

Table 4.3: GARCH(1,1) a and b parameter estimates

We note two points relevant to interpreting the statistical inference of Section 4.4.3. First, each of the sub-samples shows a sum of $\hat{a} + \hat{b}$ which is insignificantly different from unity; however, the point estimates are less than 1 in five of the six sub-samples, and $\simeq 1$ for the intermediate DJIA sub-sample. The condition for fourth moment existence in the GARCH(1,1) with (for example) gaussian innovations, $1 - 3a^2 - 2ab - b^2 > 0$, is far from being fulfilled on any sample; the implications for second moment existence - that results lie close to the threshold - are also similar across the samples.

Second, we see that the second order dynamics, as described by GARCH parameters, are different between these three subsamples (the difference being especially noticeable between the second subsample and the first and third subsamples). We have to point out as a caveat, however, that estimates of the tail index implied by estimated GARCH parameters (we used a non-parametric technique to account for the properties of the distribution of the re-scales innovations) do not exhibit any regular pattern across the sub-samples. That is, there is no prima facie indication that differences in the tail index across sub-samples arise through differences in the second-order dynamic structure as it is described by GARCH.

We want to emphasize again, however, that our primary interest is in the unconditional distribution of $\{r_t\}$ itself, and not in attributing any changes in this distribution to changes in parameters of a specific parametric dynamic model.

4.4.4 Point estimates of the tail index

Our analysis of the extremal behaviour of stock indices through the crash of October 1987 and thereafter, using the methods suggested by Quintos et al., has produced interesting observations but it is far from being complete and conclusive. In this section, we complement our analysis of the matter using an alternative technique.

We estimate in this section the tail exponents of the stock index returns for three subsamples: the period before January 1985 when portfolio insurance trading presumably had not yet introduced qualitative changes in the extremal dynamics of the markets; the period from January 1985 through October 1988 - the period covering the crash of the markets and before the introduction of the after-crash institutional changes; and finally, the period from November 1988 to March 2002 - the period after the introduction of the circuit breakers.

The main technique used in this section is the bias-corrected estimator of Huisman, Koedijk, Kool, and Palm (2001) (see Equation (4.9). Table 4.4 presents these bias-corrected estimates with $\kappa = T/3$, on three subsamples chosen to approximate, respectively, the pre-program-trading period, the program trading period before the introduction of circuit breakers, and the period after the introduction of circuit breakers.

Sample	$\hat{\alpha}$ (s.e.)	
	DJIA	*S & P 500*
01-Jan-50 - 31-Dec-84	4.91 (0.22)	4.60 (0.21)
02-Jan-85 - 31-Oct-88	2.65 (0.27)	3.53 (0.36)
01-Nov-88 - 14-Mar-02	3.78 (0.26)	4.69 (0.32)

Table 4.4: Bias-corrected tail index estimates

These results are obtained using raw logarithmic returns instead of the squared returns used in our previous analysis. This allows us to examine only the left tail of the distribution. The estimates suggest several observations. First, in common with Huisman, Koedijk, Kool, and Palm (2001), we find that the bias-corrected estimated tail index values tend to be substantially higher (a lower relative frequency of extreme events) than the raw Hill's estimates. Second, estimates on the post-circuit breaker sample show a tail index similar to those prevailing in the 1950-1984 sample, although whether the estimate of the index actually attains its pre-1985 values depends on the particular equity index considered. Finally, the difference across the sub-samples, and in particular - the distinction between 1985-1988 period and those earlier and later: is not only statistically significant (as established earlier) but quite substantial.

For comparison we also provide estimation results using two other techniques: the traditional Hill plot and the maximum likelihood estimation of α via the generalised extreme value distribution. The Hill plots corresponding to the periods of interest for the DJIA and the S&P 500 are presented in Figure 4-4. It is well known that it is difficult to extract a single value from the Hill plot; for the purpose of comparison, however, we have attempted to do so. The results are recorded in Table 4.5, and come from a proportion of the sample (also recorded in the table) chosen by visual inspection to correspond with a region of approximate stability of the estimates. We tend to interpret these results as consistent with other evidence that the value of the tail index was lower in the intermediate period, but we emphasize the difficulty in extracting a reliable point estimate.

Sample	$\hat{\alpha}$ (s.e.) [proportion of the sample used as order statistics]	
	DJIA	*S & P 500*
01-Jan-50 - 31-Dec-84	3.83 (0.22)[3.1%]	3.59 (0.21)[2.7%]
02-Jan-85 - 31-Oct-88	3.00 (0.27)[7.0%]	2.97 (0.36)[8.0%]
01-Nov-88 - 14-Mar-02	3.75 (0.26)[3.9%]	3.45 (0.32)[3.1%]

Table 4.5: Tail index estimates from Hill plot

Our findings based on the bias-corrected estimator of Huisman et al. are also corroborated by estimation results based on the GEV maximum likelihood technique. We estimate the shape parameter of the GEV via the method of fitting block maxima, where block length is chosen such that dependence across blocks can be treated as negligible, despite within-block dependence (see Embrechts, Klüppelberg, and Mikosch (1997) and Gençay, Selçuk, and Ugulülyagci (2001) for an exposition and implementation). The method requires a block size choice, which we take here to be either 80 or 125 trading days (approximately 4 or 6 months of daily trading data). Estimates of α from the GEV are given in Table 4.6.

Sample	$\hat{\xi}(s.e.); \hat{\alpha} = \hat{\xi}^{-1}$	
	DJIA	*S & P 500*
	Block size: 80	
01-Jan-50 - 31-Dec-84	0.149 (0.06); 6.69	0.180 (0.07); 5.57
02-Jan-85 - 31-Oct-88	1.217 (0.98); 0.82	1.189 (0.54); 0.84
01-Nov-88 - 14-Mar-02	0.259 (0.13); 3.86	0.281 (0.15); 3.56
	Block size: 125	
01-Jan-50 - 31-Dec-84	0.269 (0.11); 3.72	0.260 (0.10); 3.85
02-Jan-85 - 31-Oct-88	1.097 (0.61); 0.91	1.307 (0.73); 0.77
01-Nov-88 - 14-Mar-02	0.211 (0.16); 4.75	0.295 (0.21); 3.39

Table 4.6: Maximum likelihood estimation of the GEV parameter ξ using block maxima

The most important observation from this last exercise is that we again see indications of a substantially lower tail index in the intermediate period; the GEV estimates display yet higher variation across sub-samples than the Huisman, Koedijk, Kool, and Palm (2001) bias corrected estimates reported in Table 4.4. The standard errors of the estimates of GEV's ξ are substantially higher than those of the bias-corrected Hill's estimates[21].

4.5 Conclusion

Revolutionary innovations in economics and finance that occurred in the early 1970s, together with important technological changes (commercial availability of mainframe computers being probably the most important), soon found their way into the development and adoption of new financial technologies by market participants. One of these technologies - synthetic portfolio insurance - had been available starting from the beginning of the 1980s and gained popularity among investors by the middle of the 1980s. Synthetic portfolio insurance is often blamed as one of major factors contributing to the market crash of 1987.

The first goal of this study is to consider whether known changes in financial technologies were consistent with the observed patterns in the dynamics of equity markets, to the extent that these dynamics are summarised by the tail index of the distribution of logarithmic returns of several major stock indices. The statistical evidence is very clear and it indicates a strong rejection of the hypothesis of constant tail behaviour of the stock indices considered. The dating of significant statistics in the sequence is compatible with the known historical schedule of the penetration of synthetic portfolio insurance technology into the market: the equity markets began to exhibit significantly more extreme behaviour in the second half of the eighties, before the market crash of October 1987.

The second question of interest is whether we can detect an attenuating influence

[21]It is possible to increase the accuracy of GEV parameter estimates by taking into consideration more than one order statistic from each block when computing the likelihood. However, the GEV approach is not the focus of our present study.

on extremal market behaviour, of regulatory reforms which followed the 1987 crash, and whether the changes in the extremal behaviour were consistent with the known historical implementation schedule of these reforms. The statistical evidence on the second question is less strong than for the first and not without ambiguity. When applied to the period after the introduction of portfolio insurance (chosen to be 02-Jan-85 - 14-Feb-02), the hypothesis of the constancy of the tail index is marginally rejected. In some cases, however, the tests detect not an increase of the tail index but a decrease (an increase in the frequency of extreme events). The timing of this decrease (in early 1997) is consistent with the known review (widening) of the NYSE circuit breaker collars. This last fact may be interpreted as evidence that circuit breakers do affect the dynamics of markets, but, of course, it does not allow us to make conclusions about a causal relationship between circuit breakers' parameters and the extreme dynamics of the markets.

Point estimates of the tail index show that for the S&P 500, in the period from November 1988 to March 2002, it attains almost the same value as before the portfolio insurance era; the tail index of logarithmic returns of the DJIA remains significantly below the levels it had achieved prior 1985.

An important conclusion that one can draw from these results is that long series of equity returns, covering the period before and after the middle of the 1980s should not be treated as drawn from the same distribution, when extremal behaviour of these data is the object of interest. Some past studies of equity returns, however, treat the data this way. In these cases the series is drawn from a mixture of several distributions with different tail indices; it is known that the estimate of the tail index will be dominated by the subsample having the lowest tail index (highest frequency of extremes) among all the subsamples (see, for example, Quintos, Fan, and Phillips (2001, Theorem 3)). Thus, if the heterogeneity among the subsamples is not taken into account, the tail index estimate will give a misleading impression concerning the frequency of extremal events in the present market conditions, which may in turn lead to costly errors (for example, an excessive allocation of reserve capital or excessive insurance price).

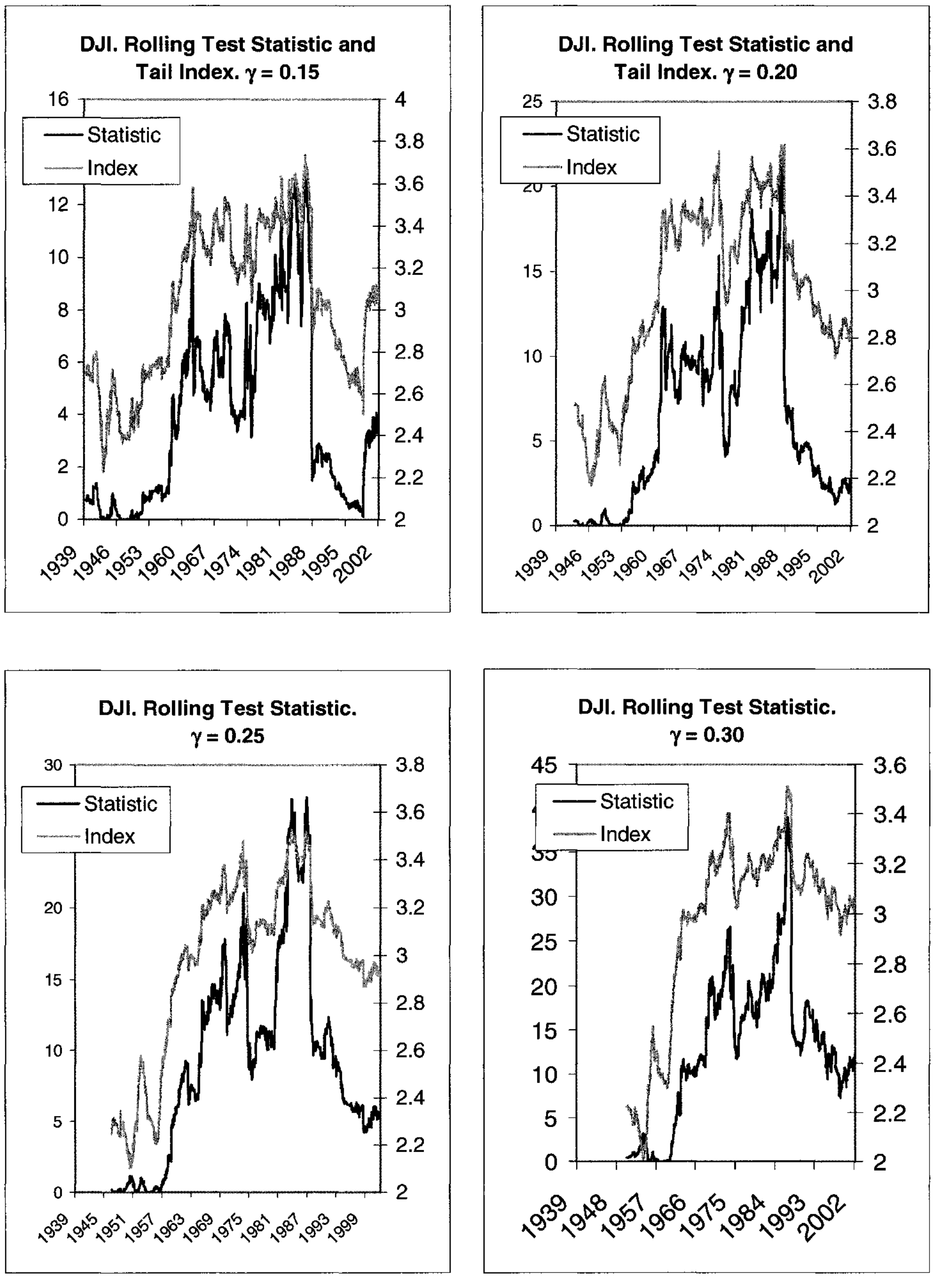


Figure 4-1: DJIA. Rolling Test Statistic (left scale) and Tail Index (right scale)

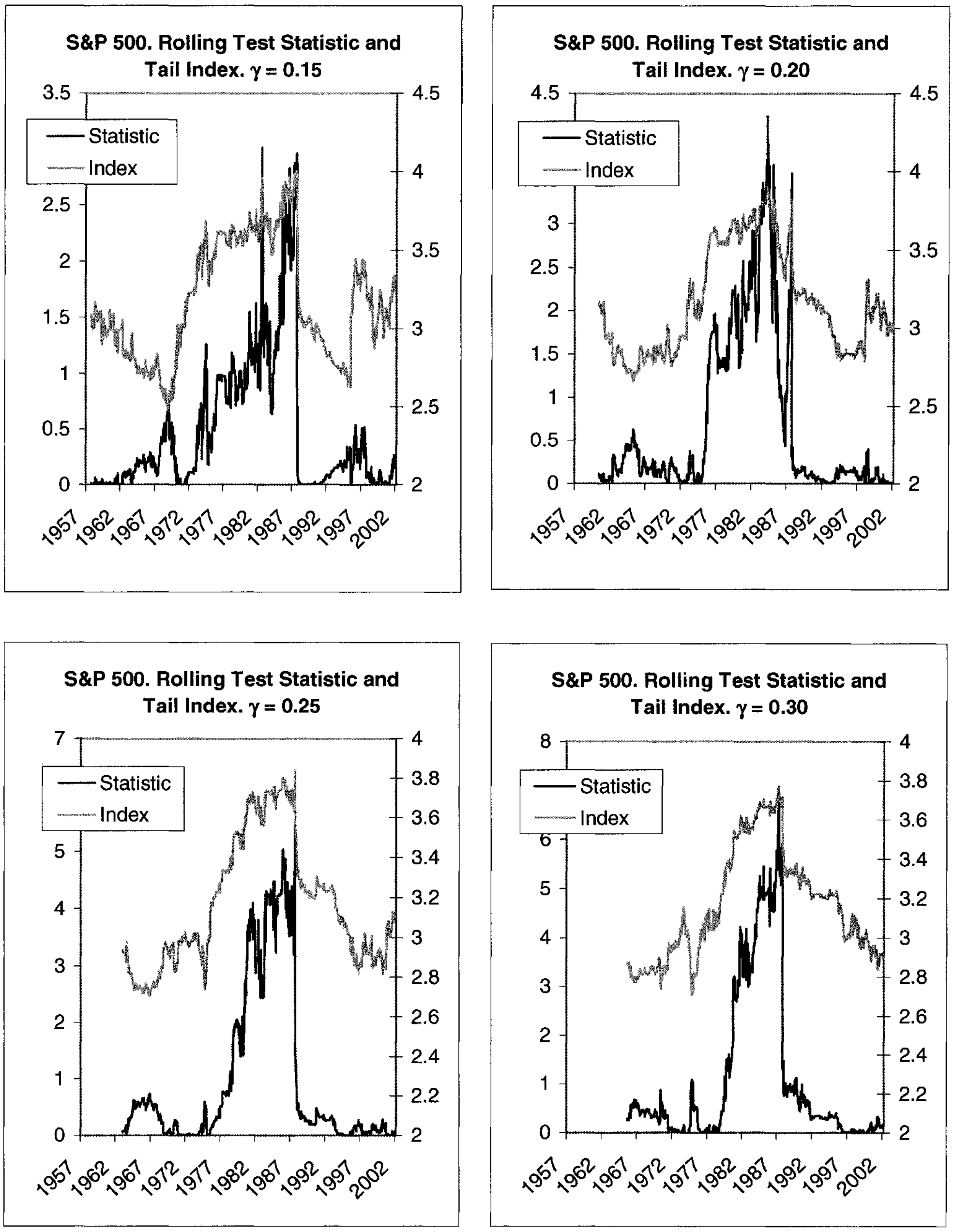


Figure 4-2: S&P 500. Rolling Test Statistic (left scale) and Tail Index (right scale)

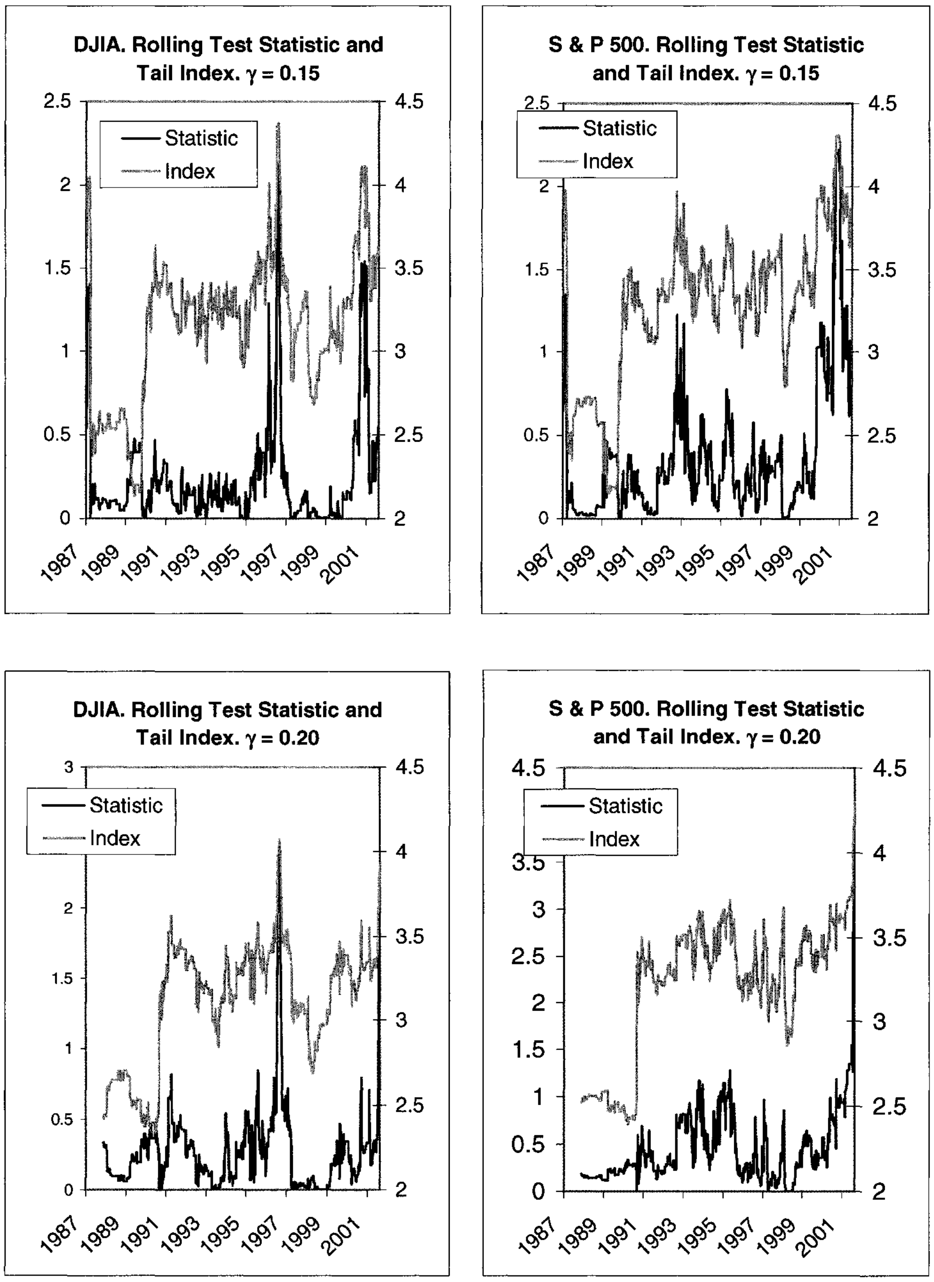


Figure 4-3: DJIA and S&P 500. Rolling Test Statistic (left scale) and Tail Index (right scale) in the post-program trading period (02-Jan-85 - 14-Mar-02)

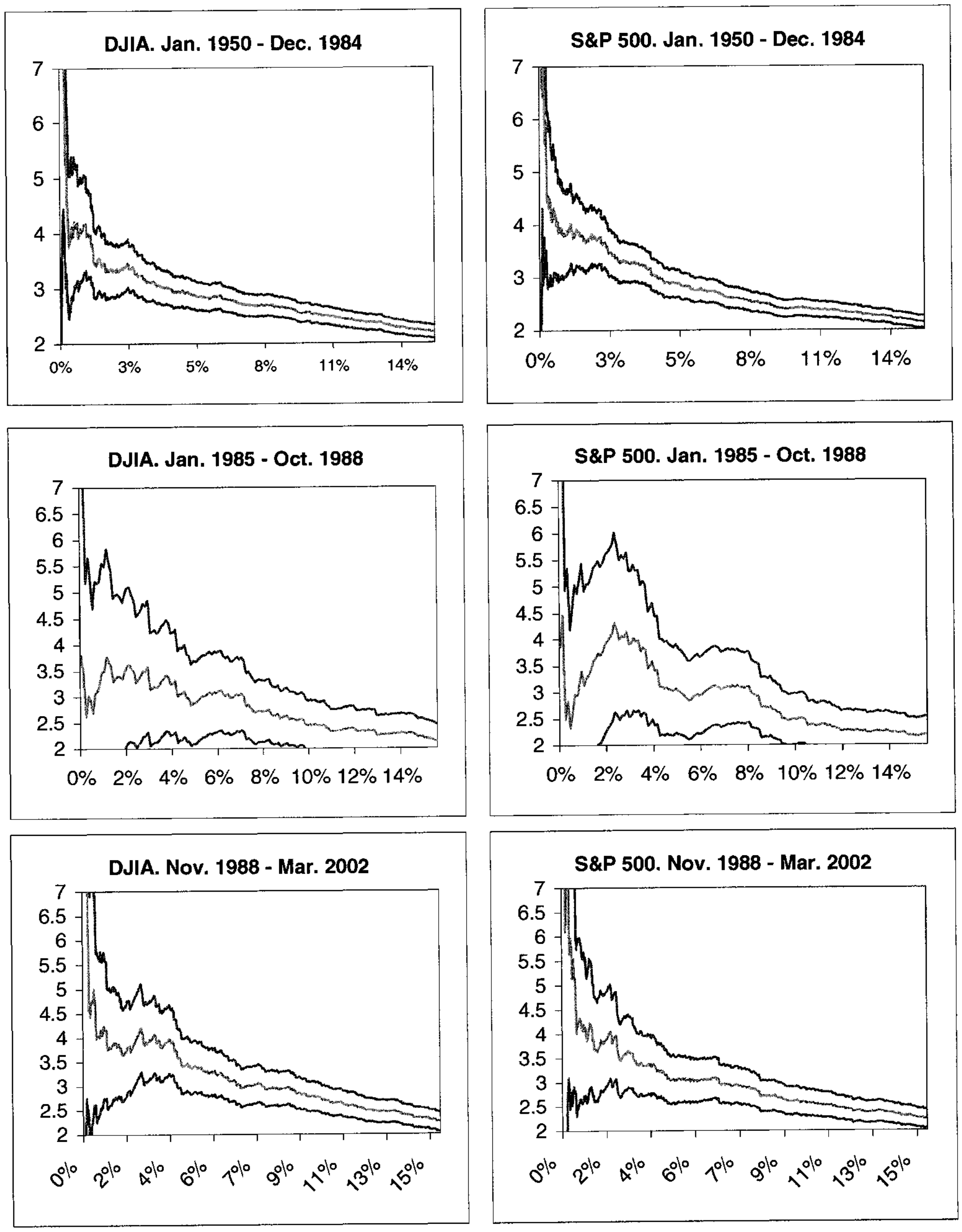


Figure 4-4: Hill plots with two standard deviation error bands

Appendix A

Quantile AR(∞) and ARMA Regression: Asymptotic Results

A.1 Proof of Theorem 3.3.1

Proof. Define

$$
\begin{aligned}
&Z_T\left(\gamma(k);q\right) \\
&\quad=\sum f_q\left(e_t-\varpi_q-X_t\left(k\right)\left(\gamma\left(k\right)-\gamma_q\left(k\right)\right)+X_t[k+1,\infty)\gamma_q[k+1,\infty)\right) \\
&\qquad\qquad-\sum f_q\left(e_t-\varpi_q+X_t[k+1,\infty)\gamma_q[k+1,\infty)\right)
\end{aligned}
$$

Here and below the summation is assumed to be over $t=1,\ldots,T$ if there is no ambiguity. Note that $\gamma(k)=\hat{\gamma}_q(k)$ is the minimiser of $Z_T\left(\gamma(k);q\right)$.

Following Phillips (1995), who considered generalised functions for LAD asymptotics, we choose a set of test functions Ψ, $\psi\left(x\right)\in\Psi$. It will be sufficient for our objectives to require that functions in Ψ are twice continuously differentiable and have a compact support in $[-1,1]$. We shall also normalise these functions so that $\int\psi\left(x\right)dx=1$. Using this set of test functions, a sequence of smoothed functions defined by

$$
f_q^m\left(x\right)=\int f_q\left(x-\frac{v}{m}\right)\psi\left(v\right)dv
$$

converges weakly, as integer $m\rightarrow\infty$, to the check function treated as a generalised function [1].

For a specific q, define now $g\left(k\right)=V_T\left(\hat{\gamma}_q\left(k\right)-\gamma_q\left(k\right)\right)$ and the smoothed process $Z_T^m\left(g(k);q\right)$:

$$
\begin{aligned}
&Z_T^m\left(g(k);q\right)= \\
&\quad\sum f_q^m\left(e_t-\varpi_q-X_t\left(k\right)V_T^{-1}g\left(k\right)+X_t[k+1,\infty)\gamma_q[k+1,\infty)\right) \\
&\qquad\qquad-\sum f_q^m\left(e_t-\varpi_q+X_t[k+1,\infty)\gamma_q[k+1,\infty)\right)
\end{aligned}
$$

[1] Note that $f_q^m(x)=f_q(x)$ when $|x|>m^{-1}$.

To simplify the notation, we introduce $\Lambda_{t,T}(k) \equiv X_t(k)V_T(k)^{-1}$ and use Λ_t instead of $\Lambda_{t,T}(k)$ whenever there is no ambiguity. Consider the Taylor expansion of $Z_T^m(g(k);q)$ around $g(k)=0$:

$$\begin{aligned} Z_T^m(\gamma_q, k; q) = \\ & -\sum f_q^{m\,\prime}(e_t - \varpi_q + X_t[k+1,\infty)\gamma_q[k+1,\infty))\,\Lambda_t g_q(k) \\ & +\frac{1}{2}\sum f_q^{m\,\prime\prime}(e_t^* - \varpi_q + X_t[k+1,\infty)\gamma_q[k+1,\infty))\,g_q'(k)\,\Lambda_t'\Lambda_t g_q(k) \end{aligned} \quad (A.1)$$

where $e_t^* = e_t + \Lambda_t \alpha g(k)$ with $0 \le \alpha \le 1$.

Note that $f_q'(x) = (q - \frac{1}{2}) + \frac{1}{2}sgn(x)$ and $f_q''(x) = \delta(x)$, where $\delta(x)$ denotes the Dirak delta-function and the equalities are defined for generalised functions. Moreover, $f_q^{m\prime} \to f_q'$ and $f_q^{m\prime\prime} \to f_{q\prime\prime}$, where convergence is understood as (weak) convergence of generalised functions (see, for example, Phillips (1995)).

We shall establish asymptotic limits of the terms in expansion (A.1) by showing that

$$\frac{1}{q(1-q)}\sum\left[\, f_q^{m\prime}(e_t - \varpi_q + X_t[k+1,\infty)\gamma_q[k+1,\infty))\right]\Lambda_t \quad \xrightarrow{D} \quad N(0,I) \quad (A.2)$$

and for

$$B_t = f_q^{m\prime\prime}(e_t^* - \varpi_q + X_t[k+1,\infty)\gamma_q[k+1,\infty))\,\Lambda_t'\Lambda_t$$

that

$$\sup_{\|g(k)\|<C_g}\left\|\left\{\sum B_t - p_e(\varpi_q)I_{k+1}\right\}\right\| \xrightarrow{p} 0 \quad (A.3)$$

for any constant $C_g > 0$.

To prove (A.2) notice first that by continuity of $f_q^{m\prime}$, stationarity of e_t and assumption **(f)**

$$\sup\left|f_q^{m\prime}(e_t - \varpi_q + X_t[k+1,\infty)\gamma_q[k+1,\infty)) - f_q^{m\prime}(e_t - \varpi_q)\right| = o_p\left(T^{-\frac{1}{2}}\right), \quad (A.4)$$

so that we can ignore $X_t[k+1,\infty)\gamma_q[k+1,\infty)$ in (A.2).

We have for any $\epsilon > 0$, e.g. $\epsilon = T^{-\frac{1}{2}}$,

$$\mathbf{Pr}\left(f_q^{m\prime}(e_t - \varpi_q)I\left(|e_t - \varpi_q| < m^{-1}\right)\right) < \mathbf{Pr}\left(I\left(|e_t - \varpi_q| < m^{-1}\right)\right)$$
$$= \int_{|e_t - \varpi_q| < m^{-1}} p_e(s)ds = p_e(\varpi_q)m^{-1} + \delta(m)m^{-1} = O(m^{-1})$$

where $\delta(m) \to 0$ as $m \to \infty$. Thus for $m \to \infty$, noting that f_q' is an ordinary function, we have:

$$f_q^{m\prime}(e_t - \varpi_q) = f_q^{m\prime}(e_t - \varpi_q)I\left(|e_t - \varpi_q| < m^{-1}\right)$$
$$+ f_q^{m\prime}(e_t - \varpi_q)I\left(|e_t - \varpi_q| > m^{-1}\right)$$
$$= o_p\left(T^{-\frac{1}{2}}\right) + f_q'(e_t - \varpi) \quad \text{(A.5)}$$

Since any $\sum_{t=1}^T |V_t| \leq T^{\frac{1}{2}}\left(\sum_{t=1}^T V_t^2\right)^{\frac{1}{2}}$, we have from (A.4), (A.5) and 3.3.1(**d**)

$$\max\left|\sum_{t=1}^T \left[f_q^{m\prime}\left(e_t - \varpi_q + X_t(k+1,\infty)\gamma(k+1,\infty)\right) - f_q'(e_t - \varpi_q)\right]\Lambda_t\right|$$
$$\leq \max\left|f_q^{m\prime}\left(e_t - \varpi_q + X_t(k+1,\infty)\gamma(k+1,\infty)\right) - f_q'(e_t - \varpi_q)\right| \cdot T^{\frac{1}{2}}\left(\operatorname{tr}\sum_{t=1}^T \Lambda_t\Lambda_t'\right)^{\frac{1}{2}}$$
$$= o_p(1).$$

Note that direct computation shows $\mathbf{E}\left(f_q'\left(e_t - \varpi_q\right)\right)^2 = q(1-q)$. Then consider

$$\xi_{T,t} = \left(\frac{1}{q(1-q)}\right)^{\frac{1}{2}} f_q'\left(e_t - \varpi_q\right)\Lambda_t$$

and for an arbitrary $\varepsilon > 0$

$$\zeta_{T,t} = \xi_{T,t}I\left(\sup_{1\leq t\leq T}\max|\Lambda_{t,T}(k)| < \varepsilon\right)$$

Since $f_q^{m\prime}$ is a bounded function and because of Assumption 3.3.1(**d**), we have that

$$\left|\sum \xi_{T,t}^l - \sum \zeta_{T,t}^l\right| \xrightarrow{p} 0 \ \texttt{for} \ l = 1, 2. \tag{A.6}$$

We show that the m.d. $\{\zeta_{T,t}, \Im\}$ array satisfies the conditions of the central limit theorem of McLeish (see, e.g., Bierens (1994, Theorem 6.1.6)). Indeed,

$$\sup_{T\geq 1} \mathbf{E}\left[\max_t \left(\zeta_{T,t}^2\right)\right] < \varepsilon^2 \mathbf{E}\left[\frac{1}{q(1-q)} \left(f_q'(e_t - \varpi_q)\right)^2\right] = \varepsilon^2 < \infty$$

and the condition (**a**) of the McLeish theorem is satisfied. Condition (**b**) follows from 3.3.1(**d**) and boundedness of f_q'. For condition (**c**) we need to show that $\sum_{t=1}^T \zeta_{T,t}^2 \to^p I$.

Consider

$$\eta_{T,t} = \zeta_{T,t}^2 - \Lambda_t'\Lambda_t = \left[\frac{1}{q(1-q)} \left(f_q'(e_t - \varpi_q)\right)^2 - 1\right] \Lambda_t'\Lambda_t.$$

Note that $\mathbf{E}\left[\eta_{T,t}\right] = 0$ and by Assumption 3.3.1 (**c**(**i**)) $\eta_{T,t}$ is a stationary ergodic sequence and thus $\sum \eta_{T,t} \xrightarrow{p} 0$. Since $\sum \Lambda'\Lambda \xrightarrow{p} I_{k+1}$ by 3.3.1(**e**) and from (A.6) we get the required condition (**c**). This concludes the proof of (A.2).

To prove (A.3) note that similarly to (A.4) we also can have

$$\begin{aligned} B_t = {} & \left[f_q^{m\prime\prime}\left(e_t - \varpi_q + X[k+1,\infty)\gamma_q[k+1,\infty)\right)\right] \Lambda_t'\Lambda_t \\ & = f_q^{m\prime\prime}(e_t^* - \varpi_q)\Lambda_t'\Lambda_t + o_p\left(T^{-\frac{1}{2}}\right) \end{aligned}$$

Similarly, over $\|g\| < C_q$ by differentiability of $f_q^{m\prime\prime}$ and 3.3.1(**d**),

$$\left|f_q^{m\prime\prime}(e_t^* - \varpi_q) - f_q^{m\prime\prime}(e_t - \varpi_q)\right| = o_p\left(T^{-\frac{1}{2}}\right),$$

thus $B_t = f_q^{m\prime\prime}(e_t - \varpi_q)\Lambda_t'\Lambda_t + o_p\left(T^{-\frac{1}{2}}\right)$. Also, as in (A.5) but for expectation[2],

$$\mathbf{E}\left[f_q^{m\prime\prime}(e_t - \varpi_q)\right] = \mathbf{E}\left[f_q''(e_t - \varpi_q)\right] + o\left(m^{-1}\right) = p_e(\varpi_q) + o\left(T^{-\frac{1}{2}}\right) \tag{A.7}$$

as long as $m^{-1} = o\left(T^{-\frac{1}{2}}\right)$.

Thus,

$$\sum B_t - p_e(\varpi_q)I_{k+1} = \sum \left(f_q^{m\prime\prime}(e_t - \varpi_q) - p_e(\varpi_q)\right)\Lambda_t'\Lambda_t + p_e(\varpi_q)\left[\sum \Lambda_t'\Lambda_t - I_{k+1}\right]$$

where the proof that the first sum goes to zero follows exactly the same steps as the similar proof for $\sum \eta_{T,t} \to 0$; the second term goes to zero by 3.3.1(**e**). This establishes (A.2) and (A.3).

Next, we follow the same argument as in Phillips (1995). The smoothed process $Z_T^m(\gamma_q; k, q)$ weakly (as a generalised random process) converges to the same limit process as $Z_T(\gamma_q; k, q)$ for each γ_q. The limit process is an ordinary random process (by (A.2) and (A.3)), it and $Z_T(\gamma_q; k, q)$ itself are convex as a result of (A.3) and (A.7); thus the sequence of minimisers of $Z_T(\gamma_q; k, q)$ has the same limit process as that of $g(k)$, the minimisers of $Z_T^m(\gamma_q; k, q)$, namely, the minimiser of the limit process. By (A.2) and (A.3) we have that

$$\Omega_k g(k) \xrightarrow{D} N\left(0, \frac{q(1-q)}{p_e^2(\varpi_q)}\Omega_k\right) \tag{A.8}$$

■

[2] f_q'' is a δ-function, but as in Phillips (1995) its expectation is an ordinary function as long as density at ϖ_q exists and is continuous.

A.2 Proof of the Theorem 3.3.2

Proof. The proof of this theorem is directly based on the results of the theorem 3.3.1. We verify the conditions of Assumption 3.3.1. $\Sigma(k) = \mathbf{E}\left(T^{-1}X(k)'X(k)\right) = T^{-1}V_T(k)'V_T(k)$ exists by Assumption 3.3.2(**a, b**). Assumption 3.3.2 (**a**) ensures that $T^{-1}\left(X(k)'X(k)\right)$ converges to $\mathbf{E}\left(T^{-1}X(k)'X(k)\right)$ a.s. by the ergodic theorem. Statements **a**, (**d**) and (**e**) of Assumption 3.3.2 provide (**b**) and (**c**) for Assumption 3.3.1. Part (**a**) of Assumption 3.3.1 follows from the definition of $\Im_t$ and from non-randomness of V_T. Part (**d**) follows from (**c**) of Assumption 3.3.2.

For (**e**) of Assumption 3.3.1 we have, by the ergodic theorem, that

$$T^{-1}X(k)'X(k) \overset{a.s}{\to} \Sigma, V_T(k) = T^{1/2}\Sigma^{1/2}.$$

Then (**e**) follows.

The (**e**) of Assumption 3.3.2 together with the fact that the $ARCH(\infty)$ coefficients for the $GARCH$ model are summable provides (**f**) of Assumption 3.3.1.

Note that while $T^{-1}X(k'X(k))$ as $T \to \infty$, $k \to \infty$ and $k/T \to 0$, consistently estimates Σ, by Berk (1974) we need extra conditions on the growth of k to ensure that as $T \to \infty$, $k \to \infty$, $\left(T^{-1}X(k)'X(k)\right)^{-1} \overset{p}{\to} \left(\Sigma(k)\right)^{-1}$. We assume here that the $ARCH(\infty)$ model is based here on a finite $GARCH$ and thus the coefficients of the $ARCH(\infty)$ presentation decline at an exponential rate. It follows that selecting k to be such that $k^{-2}T \to \infty$ and $\ln k \to \infty$ is sufficient to ensure the existence of the limit. ∎

Appendix B

Technical Notes to Chapter 2

B.1 Computation of asymptotic standard errors for the log-Weibull SCD model

Application of Dunsmuir's (1979) asymptotic theory to the case when the quasi-likelihood function is estimated using the Kalman filter is described in Harvey (1989, pp.220-221). The QML estimate of the parameter vector θ of the model is asymptotically normal, unbiased and has the variance-covariance matrix $\mathbf{C} = 2\mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}$ where

$$\mathbf{A} = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\partial \log\left(g\left(\lambda\right)\right)}{\partial \theta}\frac{\partial \log\left(g\left(\lambda\right)\right)}{\partial \theta'} d\lambda$$

is proportional to the information matrix of the process and

$$\mathbf{B} = \kappa\left[\frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\partial \log\left(g\left(\lambda\right)\right)}{\partial \theta} d\lambda\right]\left[\frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\partial \log\left(g\left(\lambda\right)\right)}{\partial \theta'} d\lambda\right].$$

Here, $g\left(\lambda\right)$ is the spectral-generating function of the process. It is easy to write down the spectral-generating function of the process $\hat{d}_i$: the spectrum of this process is different only by a constant form the spectrum of the AR process with normal innovations. The spectrum of the log-durations of the SCD model when the latent process is $AR\left(1\right)$ and the observation error has a log-Weibull distribution is

$$g\left(\lambda\right) = \frac{\pi^2}{6\gamma} + \frac{\sigma^2}{\left(1 - 2\beta\cos\lambda + \beta^2\right)^2}.$$

Analytical computation of the matrices $\mathbf{A}$ and $\mathbf{B}$ is a conceptually straightforward but tedious task. We opted to use approximate discrete representations of these integrals in our computations. Given our sample sizes, the discrete approximation is very accurate and much easier to implement.

As in Harvey (1989), we define discrete analogs of the matrices $\mathbf{A}$ and $\mathbf{B}$,

$$\mathbf{A}_T = \frac{1}{T}\sum_{m=0}^{T-1} \frac{\partial \log g\left(\frac{2\pi m}{T}\right)}{\partial \theta}\frac{\partial \log g\left(\frac{2\pi m}{T}\right)}{\partial \theta'}$$
$$\mathbf{B}_T = \kappa\left[\frac{1}{T}\sum_{m=0}^{T-1} \frac{\partial \log g\left(\frac{2\pi m}{T}\right)}{\partial \theta}\right]\left[\frac{1}{T}\sum_{m=0}^{T-1} \frac{\partial \log g\left(\frac{2\pi m}{T}\right)}{\partial \theta}\right],$$

$\lim_{T\to\infty}(\mathbf{A}_T) = \mathbf{A}$, $\lim_{T\to\infty}(\mathbf{B}_T) = \mathbf{B}$.

The information matrix that we are considering in the text of the paper is defined as (we omit the component which is due to the non-normality for the sake of simplicity):

$$\mathcal{IF} = \frac{1}{4\pi}\int_{-\pi}^{\pi} \frac{\partial \log (g(\lambda))}{\partial \theta}\frac{\partial \log (g(\lambda))}{\partial \theta'} d\lambda. \tag{B.1}$$

B.2 Asymptotic theory when the sample consists of T independent subsamples

Claim B.2.1 *Assume that the sample from the process $\{z_i(\theta_0)\}$ satisfying the conditions of Theorem 2.1 of (Dunsmuir 1979) consists of T independent subsamples $\{Z_i^j\}_{i=1}^{N_j}$, $N = \sum_{j=1}^{T} N_j$ (we consider only scalar processes to simplify the exposition). Assume that $N \to \infty$ in such way that $\frac{N_j}{N} \to \lambda_i$. Denote $g_\theta(\lambda)$ the spectral density of the process $z(\theta)$ and $I^j(\lambda)$ - the periodogram of the subsample. Then if $\bar{\theta}_N = \arg\max \bar{L}_N(\theta)$, $\bar{L}_N(\theta) = \sum_{j=1}^{T} L^j(\theta)$, where*

$$L^j(\theta) = \log\left(\frac{1}{2\pi}\int g_\theta(\lambda)\, d\lambda\right) + \frac{1}{2\pi}\int \frac{I^j(\lambda)}{g_\theta(\lambda)} d\lambda$$

then the quantity $N^{1/2}\left(\bar{\theta}_N - \theta_0\right)$ has an asymptotic normal distribution with zero mean and the covariance matrix

$$\left(T^{-2}\sum_{j=1}^{T}\lambda_j^{-1}\right)\Omega^{-1}(2\Omega + \Pi)\,\Omega^{-1}$$

where

$$\Omega = \frac{1}{2\pi}\int \frac{\partial \ln g_\theta(\lambda)}{\partial \theta'}\frac{\partial \ln g_\theta(\lambda)}{\partial \theta}d\lambda$$

and

$$\Pi = \kappa\left(\frac{1}{2\pi}\int \frac{\partial \ln g_\theta(\lambda)}{\partial \theta'}d\lambda\right)\left(\frac{1}{2\pi}\int \frac{\partial \ln g_\theta(\lambda)}{\partial \theta}d\lambda\right).$$

κ *is the forth cumulant of the innovations.*

Proof. The proof of the claim requires only a slight modification of the proof of the theorem 2.1 of (Dunsmuir 1979). We can see that under the conditions of the claim the quantity $\frac{\partial^2}{\partial\theta\partial\theta'}\bar{L}_N(\theta) \to_p T\Omega$ (each $L^j(\theta)$ converges to Ω), and the quantity $N^{1/2}\frac{\partial}{\partial\theta}\bar{L}_N(\theta)$ is asymptotically normal with the variance-covariance matrix

$$\sum_{j=1}^{T}\lambda_j^{-1}(2\Omega+\Pi)$$

The result of the claim immediately follows. ■

B.3 Long-memory in the dynamics of durations

FISCD model

The fractionally-integrated stochastic conditional duration model is specified as follows (we give here the general specification but we shall later consider only $FISCD(1,x,0)$):

$$\begin{aligned} d_i &= \mu(\gamma) + \psi_i + \xi_i \\ \phi(L)(1-L)^x\left(\psi_i - \bar{\psi}\right) &= \eta(L)\,u_i, \end{aligned} \tag{B.2}$$

where $\mathbf{E}[\psi_i] = \bar{\psi}$, and all roots of the polynomials $\phi(L)$ and $\eta(L)$ are outside the unit circle. As before, $\{\exp(\xi_i + \mu(\gamma))\}$ are i.i.d. with a distribution having a positive support, and $\{u_i\} \sim n.i.i.d.(0,\sigma_u^2)$. The spectral density of this process exists provided that $x < \frac{1}{2}$ and it has the following form:

$$g(\lambda) = \frac{1}{2\pi}\left(\sigma_\xi^2(\gamma) + \frac{\sigma_u^2\left|\eta\left(e^{-i\lambda}\right)\right|^2 2^{-x}(1-\cos\lambda)^{-x}}{\left|\phi\left(e^{-i\lambda}\right)\right|^2}\right),$$

where $\sigma_\xi^2(\gamma)$ is the variance of $\{\xi_i\}$. The spectrum has a singularity at $\lambda = 0$ which is integrable for the stationary range of the fractional integration parameter x. When the process is stationary one can easily compute the autocovariances of the process knowing the spectrum (they cannot be expressed in elementary functions but it is not important for our exposition):

$$\gamma_k = \int_{-\pi}^{\pi} g(\lambda)\exp(i\lambda k)\,d\lambda,\ \ k = 0, 1, ...$$

where γ_k denotes the k-th autocovariance. It is also possible to compute autocovariances of the process $\{D_i\}$, if necessary.

B.3.1 QML estimation of FISCD in the spectral domain

The asymptotic theory of Dunsmuir (1979) is not applicable to fractionally-integrated processes and the quasi-likelihood can not be computed using the state-space representation. Several approaches have been suggested for estimating models with latent variables that have a structure similar to that of the equation (B.2). These approaches, either based on the generalized method of moments or on computing the likelihood using simulations, have a common property: they are very computationally intensive. It is possible to compute the quasi-likelihood function of the $FISCD(p, x, q)$ process based on the sample spectrum of the process (this approach has been suggested in the context of stochastic volatility processes with long memory). The applicability of the spectral QML estimation technique (which gives essentially a Whittle-type estimator) to the problem at hand is based on results of Breidt, Crato, and de Lima (1998) who showed that the maximizer of the expression (B.3) is a strongly-consistent estimator of θ, provided that the parameter space is compact and that the parameter is uniquely identified at the true value. There is no asymptotic theory available for the spectral QML estimator.

The first steps in estimating the FISCD model are the same as those in the estimation of the SCD model. After seasonal adjustment and subtracting the mean, the model to be estimated has the following form:

$$d_i^* = \psi_i^* + \xi_i$$
$$\psi_i^* = (1 - \beta L)^{-1} (1 - L)^{-x} u_i$$

The logarithm of the spectral likelihood function of the process $\{d_i^*\}$ is

$$L(\theta) = -\frac{2\pi}{n} \sum_{k=1}^{[N/2]} \left(\log g_\theta (\lambda_k) + \frac{I_N (\lambda_k)}{g_\theta (\lambda_k)} \right) \tag{B.3}$$

where $g_\theta (\lambda)$ is the spectrum of the $FISCD(1, x, 0)$ process with the parameter vector $\theta = (\sigma_u^2, \gamma, x, \beta)$,

$$g(\lambda) = \frac{1}{2\pi} \left(\frac{\pi^2}{6\gamma^2} + \frac{\sigma_u^2 2^{-x} (1 - \cos \lambda)^{-x}}{\beta^2 - 2\beta \cos \lambda + 1} \right)$$

and $I_N (\lambda)$ is the sample periodogram. $\lambda_k = \frac{2\pi k}{N}$, $k = 0, 1, 2, ..., N$.

B.3.2 Estimation results and discussion

The estimated $FISCD(1, x, 0)$ parameters for trade and price durations are presented in Figure B-1 in a format similar to that used for depicting the estimates of the SCD parameters. The estimates of x vary between 0.42 and 0.52 for trade durations and between 0.23 and 0.65 for price durations. There are several contracts where the QML point estimates of x are greater than 0.5, i.e., are outside the stationarity region. We cannot draw definitive conclusions concerning the stationarity of the durations processes since the confidence intervals are not available.

The estimates of the parameter γ of the Weibull distribution lie between 0.95 and 1.26 for trade durations and between 1.2 and 2.65 for price durations. The estimates of σ_u^2 and β are very volatile and the data indicates that these two quantities have a strong negative correlation. It seems also that we observe two types of estimates: those with higher $\hat{\sigma}_u^2$ and $\hat{\beta}$ close to 0 and those with smaller $\hat{\sigma}_u^2$ and estimates of β in the range of $0.4 - 0.6$. Again, we cannot draw definitive conclusions because we don't know the distribution of the estimated parameters, even asymptotically. We can hypothesize that the instability of parameter estimates may be caused in part by the features of the data, such as overnight and weekend gaps in the trading, which are not accounted by the model. A study in a controlled environment through simulation will be helpful in discovering properties of QML parameter estimates of the FISCD model. We leave this as a topic for future research.

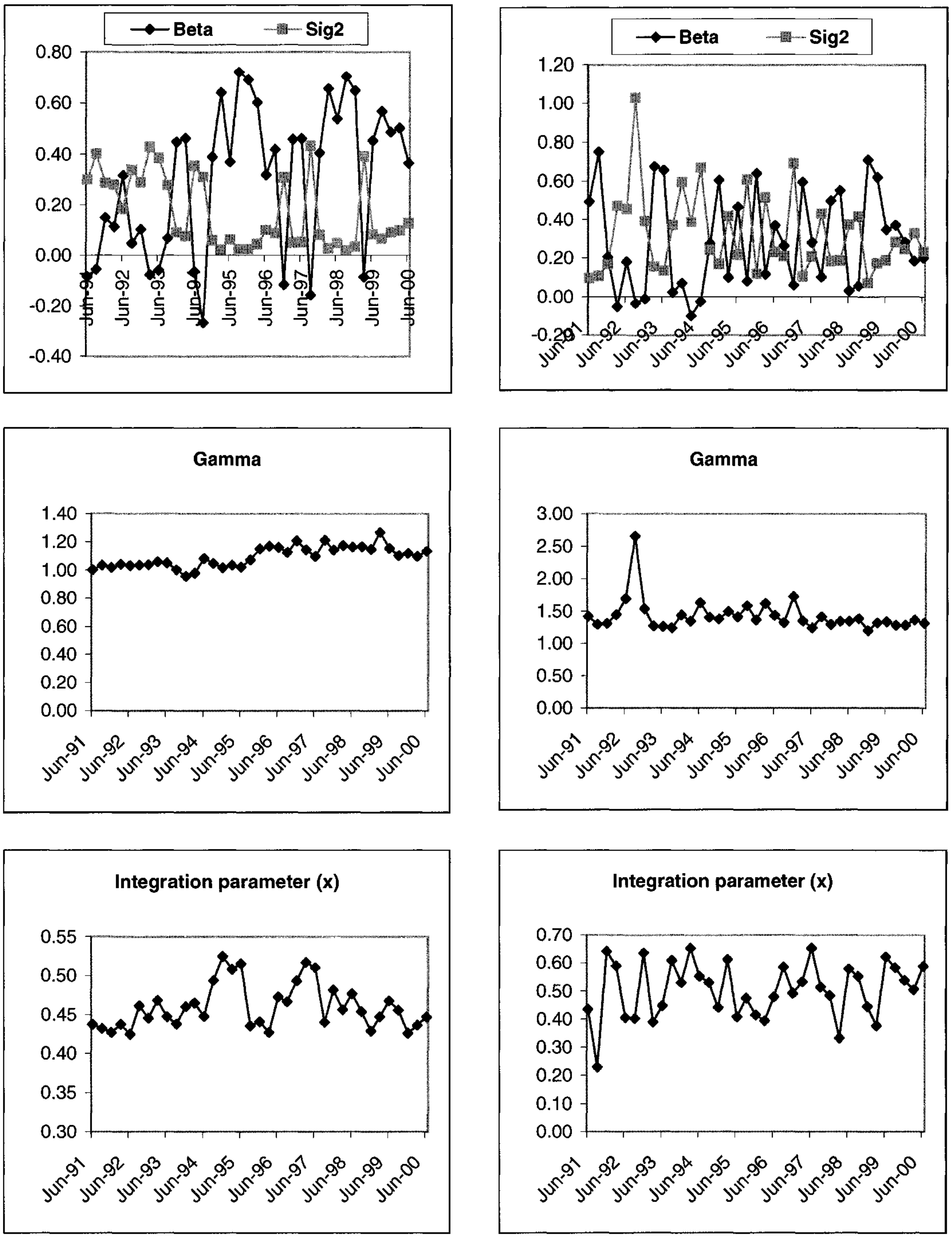


Figure B-1: FISCD(1,x,0) parameters for contracts with varying expiration dates. Trade durations are in the left column, price durations - in the right column.

Bibliography

Aït-Sahalia, Y., and P. A. Mykland (2003): "How Often to Sample a Continuous-Time Process in the Presence of Market Microstructure Noise," *NBER Working Paper No.9611.*

Andersen, T. G., and T. Bollerslev (1998): "Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts," *International Economic Review*, 39(4), 885–905.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2001): "The Distribution of Realized Exchange Rate Volatility," *Journal of American Statistical Association*, 96, 42–55.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2003): "Modelling and Forecasting Realised Volatility," *Econometrica*, 71, 579–625.

Andersen, T. G., T. Bollerslev, and N. Meddahi (2002a): "Analytic Evaluation of Volatility Forecasts," *CIRANO Scientific Series*, 2002s-90.

——— (2002b): "Correcting the Errors: A Note on Volatility Forecast Evaluation Based on High-Frequency Data and Realized Volatilities," *CIRANO Scientific Series*, 2002s-91.

Andreou, E., and E. Ghysels (2002): "Quality Control for Financial Risk Management: Monitoring Disruptions in the Distribution of Risk Exposure," *University of Cyprus Working Paper.*

Bandi, F. M., and J. R. Russell (2003): "Microstructure Noise, Realized Volatility, and Optimal Sampling," *Working Paper, Graduate School of Business, University of Chicago.*

Barndorff-Nielsen, O. E., E. Nicolato, and N. Shephard (2002): "Some Recent Developments in Stochastic Volatility Modelling," *Quantitative Finance*, 2, 11–23.

Barndorff-Nielsen, O. E., and N. Shephard (2001): "Non-Gaussian Ornstein-Uhlenbeck-Based Models and some of their Uses in Financial Economics," *Journal of Royal Statistical Society, Series B*, 63, 167–241.

——— (2002a): "Econometric Analysis of Realised Volatility and its Use in Estimating Stochastic Volatility Models," *Journal of the Royal Statistical Society, Series B*, 64, 253–280.

——— (2002b): "Estimating Quadratic Variation Using Realized Variance," *Journal of Applied Econometrics*, 17, 457–477.

Bauwens, L., and D. Veredas (2004): "The Stochastic Conditional Duration Model: A Latent Variable Model for the Analysis of Financial Durations," *Journal of Econometrics*, forthcoming.

Berk, K. N. (1974): "Consistent Autoregressive Spectral Estimates," *Annals of Statistics*, 2, 489–502.

Bertero, E., and C. Mayer (1990): "Structure and Performance: Global Interdependence of Stock Markets Around the Crash of October 1987," *European Economic Review*, 34, 1155–1180.

Bierens, H. J. (1994): *Topics in Advance Econometrics.* Cambridge University Press.

Bollerslev, T., and I. Domovitz (1993): "Trading Patterns and Prices in the Interbank Foreign Exchange Market," *Journal of Finance*, 48, 1423–1443.

Bollerslev, T., and H. Zhou (2002): "Estimating Stochastic Volatility Diffusion Using Conditional Moments of Integrated Volatility," *Journal of Econometrics*, 109, 33–65.

Booth, G. G., and J. P. Broussard (1998): "Setting NYSE Circuit Breaker Triggers," *Journal of Financial Services Research*, 13, 187–204.

Brady Comission (1988): "Report of the Presidential Task Force on Market Mechanism," *US government document.*

Breidt, F., N. Crato, and P. de Lima (1998): "The Detection and Estimation of Long Memory in Stochastic Volatility," *Journal of Econometrics*, 83, 325–348.

Breiman, L., and J. H. Friedman (1985): "Estimating Optimal Transformations for Multiple Regression and Correlation (with Discussion)," *Journal of the American Statistical Association*, 80, 580–619.

Chamley, C. (2003): "Rational Herding," *Boston University Working Paper.*

Cheng, S., and L. de Haan (2001): "Penultimate Approximation for Hill's Estimator," *Scandinavian Journal of Statistics*, 28, 569–575.

Cox, D. (1972): "Regression Models and Life Tables," *Journal of the Royal Statistical Society, Series B*, 34, 187–220.

——— (1975): "Partial Likelihood," *Biometrika*, 62, 269–276.

Dacorogna, M., U. Müller, O. Pictet, and C. de Vries (1995): *The Distribution of Extremal Foreign Exchange Data Returns in Extremely Large Data Sets.*, no. 95-70 in Discussion Papers. Tinbergen Institute.

Danielsson, J., L. de Haan, L. Peng, and C. de Vries (2001): "Using a Boostrap Method to Choose the Sample Fraction in Tail Index Estimaiton," *Journal of Multivatiate Analysis*, 76, 226–248.

DANIELSSON, J., AND C. G. DE VRIES (1997): "Tail Index and Quantile Estimation with Very High Frequency Data," *Journal of Empirical Finance*, 4, 241–257.

DEKKERS, A., J. EINMAHL, AND L. DE HAAN (1989): "A Moment Estimator for the Index of an Extreme-Value Distribution," *Annals of Statistics*, 17, 1833–1855.

DEKKERS, A. L. M., AND L. DE HAAN (1989): "On the Estimation of the Extreme-Value Index and Large Quantile Estimation," *Annals of Statistics*, 17, 1795–1832.

DROST, F. C., AND C. A. J. KLASSEN (1997): "Efficient Estimation in Semiparametric GARCH Models," *Journal of Econometrics*, 81, 193–221.

DROST, F. C., AND T. E. NIJMAN (1993): "Temporal Aggregation of GARCH Processes," *Econometrica*, 61(4), 909–927.

DROST, F. C., AND B. J. M. WERKER (1996): "Closing the GARCH gap: Continuous time GARCH modelling," *Journal of Econometrics*, 74, 31–57.

DUMOUCHEL, W. H. (1983): "Estimating the Stable Index α in Order to Measure Tail Thickness: A Critique," *Annals of Statistics*, 11, 1019–1031.

DUNSMUIR, W. (1979): "A Central Limit Theorem for Parameter Estimation in Stationary Vector Time Series and its Application to Models for a Signal Observed with Noise," *Annals of Statistics*, 7, 490–506.

EMBRECHTS, P., C. KLÜPPELBERG, AND T. MIKOSCH (1997): *Modelling Extremal Events*. Springer-Verlag, Berlin Heidelberg New York.

ENGLE, R., AND J. RUSSELL (1998): "Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data," *Econometrica*, 66(5), 1127–1162.

ENGLE, R. F. (1982): "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kindom Inflation," *Econometrica*, 50, 987–1007.

——— (2000): "The Econometrics of Ultra-High Frequency Data," *Econometrica*, 68, 1–22.

ENGLE, R. F., AND G. GONZALEZ-RIVERA (1991): "Semi-Parametrics ARCH Models," *Journal of Business and Economic Statistics*, 9, 345–359.

FAN, J., AND I. GIJBELS (1996): *Local Polynomial Modelling and its Applications.* Chapman and Hall, Suffolk.

FEUERVERGER, A., AND P. HALL (1999): "Estimating a Tail Exponent by Modelling Departure from a Pareto Distribution," *The Annals of Statistics*, 27, 760–781.

FRANCISCO-FERNANDEZ, M., J. OPSOMER, AND J. VILAR-FERNANDEZ (2003): "A Plug-in Bandwidth Selector for Local Polynomial Regression Estimator with Correlated Errors," *Journal of Nonparametric Statistics*, 16, 127–151.

FRANCQ, C., AND J.-M. ZAKOÏAN (2000): "Estimating Weak GARCH Representations," *Econometric Theory*, 16, 692–728.

FRIEDMAN, J. H., AND W. STUETZLE (1981): "Projection Pursuit Regression," *Journal of American Statistical Association*, 76, 817–823.

FROOT, K. A., D. S. SCHARFSTEIN, AND J. C. STEIN (1992): "Herd on the Street: Informational Inefficiencies in a Market with Short-Term Speculation," *Journal of Finance*, 47, 1461–1484.

GALBRAITH, J. W., AND T. KISINBAY (2002): "Volatility Forecast Horizons," *CIRANO Working Paper*, 2002s-21.

GALBRAITH, J. W., AND S. ZERNOV (2004): "Circuit Breakers and the Tail Index of Equity Returns," *Journal of Financial Econometrics*, 2, 109–129.

GALBRAITH, J. W., AND V. ZINDE-WALSH (1994): "A Simple Non-Iterative Estimator of Moving Average Models," *Biometrika*, 81, 143–155.

——— (1997): "On some Simple, Autoregression-Based Estimation and Identification Techniques for ARMA Models," *Biometrika*, 84, 685–696.

GARBER, P. M. (1990): "Famous First Bubbles," *Journal of Economic Perspectives*, 4, 35–54.

GARCIA, R., M.-A. LEWIS, AND É. RENAULT (2001): "Estimation of Objective and Risk-Neutral Distributions Based on Moments of Integrated Volatility," *University of Montreal Working paper.*

GENÇAY, R., F. SELÇUK, AND A. UGULÜLYAGCI (2001): "EVIM: A Software Package for Extreme Value Analysis in MATLAB," *Studies in Nonlinear Dynamics and Econometrics*, 5(3).

GERHARD, F., AND N. HAUTSCH (2002a): "Semiparametric Autoregressive Conditional Proportional Hazard Models," *Nuffield Economics Working Paper*, W2.

——— (2002b): "Volatility Estimation on the Basis of Price Intensities," *Journal of Empirical Finance*, 9, 57–89.

GHYSELS, E., C. GOURIÉROUX, AND J. JASIAK (1997): "Stochastic Volatility Duration Model," *CREST working paper 9746.*

GOURIÉROUX, C., J. JASIAK, AND G. LE FOL (1999): "Intra-Day Market Activity," *Journal of Financial Markets*, 2, 193–226.

HAFNER, C. M., AND J. V. K. ROMBOUTS (2003): "Semiparametric Multivatiate GARCH Models," CORE Discussion Paper.

HALL, P. (1982): "On some Simple Estimates of an Exponent of Regular Variation," *Journal of Royal Statistical Society. Series B.*, 44, 37–42.

——— (1990): "Using the Bootstrap to Estimate Mean Squared Error and Select Smoothing Parameters in Nonparametric Problems," *Journal of Multivariate Analysis*, 32, 177–203.

HARVEY, A. C. (1989): *Forecasting, Structural Time Series Models and the Kalman Filter.* Cambridge University Press, New York.

HAUTSCH, N. (2002): "Testing the Conditional Mean Function of Autoregressive Conditional Duration Models," *University of Konstantz Working Paper.*

HSING, T. (1991): "On Tail Index Estimation Using Dependent Data," *Annals of Statistics*, 19, 1547–1569.

HUISMAN, R., K. G. KOEDIJK, C. J. KOOL, AND F. PALM (2001): "Tail-Index Estimates in Small Samples," *Journal of Business and Economic Statistics*, 19, 208–216.

JACOBS, B. I. (1999): *Capital Ideas and Market Realities: Option Replication, Investor Behavior and Stock Market Crashes.* Blackwell Publishers Inc., Oxford, UK.

JASIAK, J. (1998): "Persistence in Intertrade Durations," *Finance*, 19, 166–195.

LAUTERBACH, B., AND U. BEN-ZION (1993): "Stock Market Crashes and the Performance of Circuit Breakers: Empirical Evidence," *The Journal of Finance*, 48, 1909–1925.

LINDSEY, R. R., AND A. P. PECORA (1998): "Ten Years After: Regulatory Developments in the Securities Markets Since the 1987 Market Break," *Journal of Financial Services Research*, 13, 283–314.

LINTON, O., AND J. P. NIELSEN (1995): "A Kernel Method of Estimating Structured Nonparametric Regression Based on Marginal Integration," *Biometrika*, 82, 93–100.

LORETAN, M., AND P. C. B. PHILLIPS (1994): "Testing the Covariance Stationarity of Heavy Tailed Time Series," *Journal of Empirical Finance*, 1, 211–248.

MA, C. K., R. P. RAO, AND R. SEARS (1989a): "Limit Moves and Price Resolution: The Case of the Treasury Bond Futures Market," *The Journal of Futures Market*, 9, 321–335.

MA, C. K., R. P. RAO, AND R. S. SEARS (1989b): "Volatility, Price Resolution, and the Effectiveness of Price Limits," *Journal of Financial Services Research*, 3, 165–199.

MEDDAHI, N. (2002a): "ARMA Representation of Integrated and Realised Variances," *CIRANO Scientific Series*, 2002s-93.

——— (2002b): "A Theoretical Comparison Between Integrated and Realized Volatilities," *Journal of Applied Econometrics*, 17, 479–508.

MEDDAHI, N., AND E. RENAULT (2004): "Temporal Aggregation of Volatility Models," *Journal of Econometrics*, in press.

MERTON, R. (1980): "On Estimating the Expected Return on the Market," *Journal of Financial Economics*, 8, 323–361.

MIKOSCH, T., AND C. STĂRICĂ (2000): "Limit Theory for the Sample Autocorrelations and Extremes of a GARCH(1,1) Process," *Ann.Statist.*, 28, 1427–1451.

MILLER, J. D. (1989): *Program Trading: The New Age of Investing.* J.K. Lasser Institute, New York, NY, USA.

NELSON, D. B. (1990): "ARCH Models as Duffusion Approximations," *Journal of Econometrics*, 45, 7–38.

O'BRIEN, J. W. (1982): "Dynamic Hedging: A New 'Option'," *Pensions and Investment Age*, p. 1982.

PHILLIPS, P. C. (1995): "Robust Nonstationary Regression," *Econometric Theory*, 11, 912–951.

POSTHUMUS, N. W. (1929): "The Tulip Mania in Holland in the Years 1636 and 1637," *Journal of Economic and Business History*, 1, 434–455.

QUINTOS, C., Z. FAN, AND P. C. PHILLIPS (2001): "Structural Change Test in Tail Behavior and the Asian Crisis," *Review of Economics Studies*, 68, 633–663.

ROLL, R. (1989): "Price Volatility, International Market Links, and their Implications for Regulatory Policies," *Journal of Financial Services Research*, 3, 211–246.

RUBINSHEIN, M., AND H. E. LELAND (1981): "Replicating Options with Positions in Stock and Cash," *Financial Analysts Journal*, 37, 63–72.

RUIZ, E. (1994): "Quasi-Maximum Likelihood Estimation of Stochastic Volatility Models," *Journal of Econometrics*, 63, 289–306.

RUSSELL, J., AND R. ENGLE (1998): "Econometric Analysis of Discrete-Valued Irregularly-Spaced Financial Transactions Data Using a New Autoregressive Conditional Multinomial Model," *University of California (San-Diego) Working Paper*, 98-10.

SANTONI, G., AND T. LIU (1993): "Circuit Breakers and Stock Market Volatility," *The Journal of Futures Market*, 13, 261–277.

SECURITIES AND EXCHANGE COMISSION (SEC) (1988): "The October 1987 Market Break," *Report.*

SHLEIFER, A., AND R. W. VISHNY (1997): "The Limits of Arbitrage," *Journal of Finance*, 52, 35–55.

STEPHENS, M. A. (1976): "Asymptotic Results for Goodness-of-Fit Statistics with Unknown Parameters," *The Annals of Statistics*, 4, 357–369.

——— (1977): "Goodness of Fit for the Extreme Value Distribution," *Biometrika*, 64, 583–588.

SUBRAHMANYAM, A. (1994): "Circuit Breakers and Market Volatility: A Theoretical Perspective," *The Journal of Finance*, 48, 237–254.

VEREDAS, D., J. RODRIGUEZ-POO, AND A. ESPASA (2002): "On the (Intraday) Seasonality and Dynamics of a Financial Point Process: A Semiparametric Approach," *CORE DP 2002/23.*

WHITE, E. N. (1990): "The Stock Market Boom and Crash of 1929 Revisited," *Journal of Economic Perspectives*, 4, 67–83.