Properties of Estimates of Daily GARCH Parameters Basaed on Intra-Day Observations

Article · July 2000

Source: CiteSeer

Victoria Zinde-Walsh
McGill University

18. 5. 2000

Properties of Estimates of Daily GARCH Parameters Based on Intra-day Observations

John W. Galbraith and Victoria Zinde-Walsh

Department of Economics
McGill University
855 Sherbrooke St. West
Montreal, Quebec H3A 2T7 CANADA

Preliminary and incomplete draft

Abstract

We consider estimates of the parameters of GARCH models of daily financial returns, obtained using intra-day (high-frequency) returns data to estimate the daily conditional volatility. Two potential bases for estimation are considered. One uses aggregation of high-frequency Quasi- ML estimates, using aggregation results of Drost and Nijman (1993). The other uses the integrated volatility of Andersen and Bollerslev (1998), and obtains coefficients from a model estimated by OLS or LAD, in the latter case providing consistency and asymptotic normality in cases where moments of the volatility estimation error may not exist. In particular, we consider estimation in this way of an ARCH approximation, and obtain GARCH parameters by a method related to that of of Galbraith and Zinde-Walsh (1997) for ARMA processes. We offer some simulation evidence on small-sample performance, and characterize the gains relative to standard quasi-ML estimates based on daily data alone.

Key words: GARCH, high-frequency data, integrated volatility, LAD

JEL Classification numbers: C12, C22

* The authors thank the *Fonds pour la formation de chercheurs et l'aide à la recherche* (Quebec) and the Social Sciences and Humanities Research Council of Canada for financial support of this research, and CIRANO (*Centre Inter-Universitaire de Recherche en Analyse des Organisations*) for research facilities. We are grateful to Nour Meddahi, Tom McCurdy and Eric Renault for valuable comments and discussions.

1. Introduction

GARCH models are widely used for forecasting and characterizing the conditional volatility of economic and (particularly) financial time series. Since the original contributions of Engle (1982) and Bollerslev (1986), the models have been estimated by Maximum Likelihood (or quasi-ML) methods on observations at the frequency of interest. In the case of asset returns, the frequency of interest is often the daily fluctuation.

Financial data are often recorded at frequencies much higher than the daily. Even where our interest lies in volatility at the daily frequency, these data contain information which may be used to improve our estimates of models at the daily frequency. Of course, following Andersen and Bollerslev (1998), higher-frequency data may also be used to estimate the daily volatility directly.

The present paper considers two possible strategies for estimation of daily GARCH models which use information about higher-frequency fluctuations. The first uses the known aggregation relations (Drost and Nijman, 1993) linking the parameters of GARCH models of high-frequency and corresponding low-frequency observations. When such estimates are based on QML estimates for the high-frequency data, however, relatively stringent conditions are required, which may not be met in (for example) asset-return data.

The second potential strategy is to use the observation of Andersen and Bollerslev (1998) that the volatility of low-frequency asset returns may be estimated by the sum of squared high-frequency returns. While the resulting estimate may be used directly to characterize the process as in Andersen and Bollerslev or Andersen et al. (1999), it is also possible to use the sequence of low- (daily-) frequency estimated volatilities to obtain estimates of conditional volatility models such as GARCH models, explicitly allowing for estimation error in the estimated daily volatility. The resulting models may be estimated by a variety of techniques (including LS); by using the Least Absolute Deviations (LAD) estimator, it is possible to obtain consistent and asymptotically normal estimates under quite general conditions (in particular, without requiring the existence of moments of the returns). We are then able to obtain estimates of GARCH parameters using an estimator related to that of Galbraith and Zinde-Walsh (1997) for ARMA models.

In section 2 we describe the models and estimators to be considered and give some relevant definitions and notation. Section 3 provides several asymptotic results, while section 4 presents simulation evidence on the finite- sample performance of regression estimators relative to that of standard GARCH estimates based on the daily observations alone.

2. GARCH model estimation using higher-frequency data

2.1 Processes and notation

We begin by establishing notation for the processes of interest. Consider a driftless diffusion process $\{X_t\}$ such that

$$X_t = X_0 + \int_0^t \sigma_s X_s dW_s,$$

where $\{W_t\}$ is a Brownian motion process and σ_s^2 is the instantaneous conditional variance. This is a special case of the structure used by, e.g., Nelson (1992), Nelson and Foster (1994).

The process is sampled discretely at an interval of time ℓ (e.g., each minute). We are interested in volatility at a lower-frequency sampling, with sampling interval $h\ell$ (e.g., daily), so that there are h high-frequency observations per low-frequency observation. Define one unit of time as a period of length ℓ.

We index the full set of observations by τ and the lower-frequency observations by t, so that $t = \{h, 2h, \ldots, hT\}$. The size of the sample of low-frequency observations is therefore T, and of the full set of high-frequency observations is hT. Following Andersen and Bollerslev (1998), estimate the conditional volatility at t as the estimated conditional variance

$$\hat{\sigma}_t^2 = \sum_{j=(i-1)h+1}^{ih} r_j^2, \tag{2.1.0}$$

with $r_j^2 = (x_j - x_{j-1})^2$. See Andersen and Bollerslev on convergence of $\hat{\sigma}_t^2$ to $\int_{t-1}^{t} \sigma_s^2 ds$.

Now consider ARCH and GARCH models at the lower-frequency observations:

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2, \tag{2.1.1}$$

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{i-1}^{p} \beta_i \sigma_{t-i}^2, \tag{2.1.2}$$

where $\varepsilon_t = y_t - \mu_t$ for a process y_t with conditional mean μ_t, or in the driftless case $\varepsilon_t = X_t$. So $E(\varepsilon_t^2|\psi_{t-i}) \equiv \sigma_t^2$. Models in the form (2.1.1), (2.1.2) are directly estimable if, as in Andersen and Bollerslev, we have measurements of σ_t^2. We return to this point in Section 2.3 below.

Finally, we will refer below to the definitions of Strong, Semi-strong and Weak GARCH given in Drost and Nijman (1993). In strong GARCH, $\{\varepsilon_t\}$ is such that $z_t \equiv \varepsilon_t/\sigma_t \sim IID(0,1)$; semi-strong GARCH holds where $\{\varepsilon_t\}$ is such that $E[\varepsilon_t|\varepsilon_{t-1},\ldots] = 0$ and $E[\varepsilon_t^2|\varepsilon_{t-1},\ldots] = \sigma_t^2$; weak GARCH holds where $\{\varepsilon_t\}$ is such that $P[\varepsilon_t|\varepsilon_{t-1},\ldots] = 0$ and $P[\varepsilon_t^2|\varepsilon_{t-1},\ldots] = \sigma_t^2$, where $P[\varepsilon_t^2|\varepsilon_{t-1},\ldots]$ denotes the best linear predictor of ε_t^2 given a constant and past values of both ε_t and ε_t^2.

2.2 Estimation by aggregation

Drost and Nijman (1993) showed that time-aggregated weak GARCH processes lead to processes of the same class, and gave deterministic relations between the coefficients (and the kurtosis) of the high frequency process and corresponding time-aggregated (low-frequency) process for the weak GARCH (1,1) case. As Drost and Nijman noted, such relations can in principle be used to obtain estimates of the parameters at one frequency from those at another. In this section we examine the strategy of low-frequency estimation based on prior high-frequency estimates. Time aggregation relations of course differ for stock and flow variables; here we treat flows, such as asset returns.

Consider the high-frequency GARCH(1,1) process

$$\sigma_\tau^2 = \omega + \alpha_1 \varepsilon_{\tau-1}^2 + \beta_1 \sigma_{\tau-1}^2; \tag{2.2.1}$$

if $\varepsilon_{(h)t} = \sum_{j=t(h-1)+1}^{th} \varepsilon_j$ is the aggregated flow variable, then its volatility at the low frequency follows the weak GARCH(1,1) process

$$\sigma_{(h)t}^2 = \overline{\alpha}_0 + \overline{\alpha}_1 \varepsilon_{(h)t}^2 + \overline{\beta}_1 \sigma_{(h)t-1}^2, \tag{2.2.2}$$

with $\overline{\alpha}_0, \overline{\alpha}_1, \overline{\beta}_1$ given by the corresponding formulae (13-15) for $\overline{\psi}, \overline{\alpha}, \overline{\beta}$ in Drost and Nijman (1993), adjusting for notation. To obtain consistent estimation by QML of the high-frequency model, it will be necessary that the process is semi-strong GARCH: the standard Quasi-ML estimator of the GARCH model will in general be inconsistent in weak GARCH models (as noted by Meddahi and Renault 1996, 2000 and Francq and Zakoïan 1998; see the latter reference for an example and M-R 2000 for a Monte Carlo example on samples of 80 000 – 150 000 simulated low frequency observations).

We will show that the mapping

$$\begin{pmatrix} \overline{\alpha}_0 \\ \overline{\alpha}_1 \\ \overline{\beta}_1 \end{pmatrix} = \psi \begin{pmatrix} \omega \\ \alpha_1 \\ \beta_1 \end{pmatrix} \tag{2.2.3}$$

provided by the Drost-Nijman formulae is a continuously differentiable mapping; it is also analytic over the region where the parameters are defined.

This implies that any consistent estimator of the high-frequency parameters $(\omega, \alpha_1, \beta_1)$ leads to a consistent estimator of the low-frequency parameters $(\overline{\alpha}_0, \overline{\alpha}_1, \overline{\beta}_1)$, and similarly that an asymptotically Normal estimator of the high-frequency parameters results in asymptotic Normality of the low-frequency parameters.

Denote the vector $\begin{pmatrix} \omega \\ \alpha_1 \\ \beta_1 \end{pmatrix}$ by η and, correspondingly, let $\overline{\eta} = \begin{pmatrix} \overline{\alpha}_0 \\ \overline{\alpha}_1 \\ \overline{\beta}_1 \end{pmatrix}$. Then $\psi(\eta) = \overline{\eta}$.

Now denote by $\Omega \in \mathcal{R}^3$ the region

$$\Omega = \{(\omega, \alpha_1, \beta_1) \in \mathcal{R}^3 |\ \omega > 0,\ \alpha_1 \geq 0,\ \beta_1 \geq 0;\ \alpha_1 + \beta_1 < 1\},$$

that is, the region for which the GARCH(1,1) process is defined (see, e.g., Bollerslev 1986).

Theorem 1. For any estimator $\hat{\eta}$ of η such that (i) $\hat{\eta} \xrightarrow{p} \eta$; (ii) $\hat{\eta} \overset{a}{\sim} N(\eta, V(\eta))$, the estimator $\hat{\overline{\eta}} = \psi(\hat{\eta})$ is such that for $\hat{\eta}$ satisfying (i),

$$\hat{\overline{\eta}} \xrightarrow{p} \overline{\eta},$$

and for $\hat{\eta}$ satisfying (ii),

$$\hat{\overline{\eta}} \overset{a}{\sim} N(\eta, V(\hat{\overline{\eta}})),$$

where the asymptotic covariance matrix is $V(\hat{\overline{\eta}}) = \frac{\partial \psi}{\partial \eta'} V(\eta) \frac{\partial \psi'}{\partial \eta}$.

Proof. It follows from (i) and consequently also from (ii) that since $\eta \in \Omega$,

$P(\hat{\eta} \in \Omega) \to 1$. Consider now the formulae for $\overline{\eta} = \psi(\eta)$ over Ω in Drost and Nijman (1993). From (15) of D-N we can obtain $\overline{\beta}_1$ from a solution to a quadratic equation of the form $Z^2 - cZ + 1 = 0$, where

$$c = c(\omega, \alpha_1, \beta_1, \kappa) \tag{2.2.4}$$

is obtained from the expression in (15) of D-N. For $\eta \in \Omega$ it follows that $c > 2$ and therefore $\overline{\beta}_1 = \frac{c}{2} - \left[(\frac{c}{2})^2 - 1\right]^{\frac{1}{2}}$ is such that $0 < \overline{\beta}_1 < 1$. Moreover, it can be shown that $\overline{\beta}_1 < (\alpha_1 + \beta_1)^h$ and so $\overline{\alpha}_1$ obtained from (13) in D-N also lies between 0 and 1.

The transformation $\psi \begin{pmatrix} \omega \\ \alpha_1 \\ \beta_1 \end{pmatrix}$ can be written as

$$\psi \begin{pmatrix} \omega \\ \alpha_1 \\ \beta_1 \end{pmatrix} = \begin{pmatrix} h\omega \frac{1-(\beta_1+\alpha_1)^h}{1-(\beta_1+\alpha_1)} \\ (\beta_1 + \alpha_1)^h - \frac{c}{2} + \left[(\frac{c}{2})^2 - 1\right]^{\frac{1}{2}} \\ -\frac{c}{2} + \left[(\frac{c}{2})^2 - 1\right]^{\frac{1}{2}} \end{pmatrix},$$

where c is given by (2.2.4); it is defined and differentiable everywhere in Ω. ∎.

Note that (as follows from Drost and Nijman 1993), even if $\beta_1 = 0$, $\overline{\beta}_1$ is non-zero as long as $\alpha > 0$. As h increases, $\overline{\alpha}_1$ and $\overline{\beta}_1$ decline; given α_1 and β_1, conditional heteroskedasticity vanishes for sufficiently large h. Therefore, for substantial conditional heteroskedasticity to be present in the low-frequency (aggregated) flow process, $\alpha_1 + \beta_1$ must be close to unity.

Suppose now that a standard quasi-Maximum Likelihood estimator is used with semi-strong GARCH high-frequency data to obtain estimates of η. Its asymptotic covariance matrix is $V[\hat{\eta}_{QML}] = [W'W]^{-1}B'B[W'W]^{-1}$, where

$$W'W = \sum_{\tau=1}^{h} T \left[\frac{g_\tau}{\sigma_\tau^2}\right] \left[\frac{g_\tau}{\sigma_\tau^2}\right]'$$

$$\text{and } B'B = \sum_{\tau=1}^{T} \left[\frac{\varepsilon_\tau^2}{\sigma_\tau^2} - 1\right]^2 \left[\frac{g_\tau}{\sigma_\tau^2}\right] \left[\frac{g_\tau}{\sigma_\tau^2}\right]',$$

with $g_\tau = \frac{\partial \sigma_\tau^2}{\partial \eta} = \begin{pmatrix} 1 \\ \varepsilon_{\tau-1}^2 \\ \sigma_{\tau-1}^2 \end{pmatrix}$. The asymptotic variance of the estimator $\hat{\hat{\eta}}$ based on flow aggregation is then

$$\frac{\partial \psi}{\partial \eta'}[W'W]^{-1}B'B[W'W]^{-1}\frac{\partial \psi'}{\partial \eta}. \tag{2.2.5}$$

If $\hat{\eta}_{QML}$ is the MLE this reduces to

$$\frac{\partial \psi}{\partial \eta'}[W'W]^{-1}\frac{\partial \psi'}{\partial \eta}. \tag{2.2.6}$$

Example 1. Let the high-frequency process be ARCH(1); aggregation then leads to a weak GARCH(1,1) process for the low-frequency data. However, the asymptotic covariance matrix for the estimator $\hat{\hat{\eta}}$ is of rank 2 rather than 3, since the middle part in (2.2.5) or (2.2.6) is of dimension 2×2. This indicates that there are cases where $\hat{\hat{\eta}}$ is clearly more efficient than $\overline{\eta}_{ML}$ (or $\overline{\eta}_{QML}$) based on low-frequency data alone, with covariance matrix of rank 3.

While estimation is feasible by this method, the requirements of this strategy, even for consistent estimation, are fairly severe. In particular, the potential inconsistency of QML estimation when only weak GARCH conditions apply means that we must assume semi-strong GARCH at the high frequency if estimation is by QML. This is, however, an arbitrary specification; if the high frequency data are themselves aggregates of yet higher frequencies, the semi-strong conditions do not follow. While consistent estimation of weak GARCH models is in principle possible (see Francq and Zakoïan 1998), the QML estimator does not accomplish this.

More generally, estimation based on aggregation presumes knowledge of the high-frequency structure, and requires the computation of different aggregation formulae for each model form to be estimated. For these reasons, we will proceed to investigate estimators based on the integrated volatility, which do not presume knowledge of the high-frequency process beyond the conditions necessary for consistency of the daily volatility estimate.

2.3 Estimation by regression using integrated volatility

As noted above, models of the form (2.1.1) and (2.1.2) are directly estimable if we have estimates of the conditional variance of the low-frequency observations, σ_t^2, for example from the daily integrated volatility, as in Andersen and Bollerslev.[1] However, we will not follow Andersen and Bollerslev in treating the observation as exact. Instead, we introduce into the model the measurement error arising in estimation of $\overline{\sigma}_t^2$ from the daily integrated volatility (2.1.0), a specification also employed by Maheu and McCurdy (2000). Let

$$\hat{\sigma}_t^2 = \sigma_t^2 + e_t; \tag{2.3.1}$$

properties of $\{e_t\}$ follow from (2.1.0) and will be considered below.

The ARCH and GARCH models become

$$\hat{\sigma}_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + e_t, \tag{2.3.2}$$

$$\hat{\sigma}_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i \hat{\sigma}_{t-i}^2 - \sum_{i=1}^{p} \beta_i e_{t-i} + e_t. \tag{2.3.3}$$

Both (2.3.2) and (2.3.3) are in principle estimable as regression models. The model (2.3.3) has an error term with an MA(s) form; the coefficients of this moving average

[1] Since we will be discussing low-frequency parameters hereafter, we no longer need to distinguish from high-frequency and will omit the 'bar' in symbols, referring to σ^2, α, β, etc. for the low-frequency values.

process are subject to the constraint embodied in (2.3.3) that they are the same (up to sign) as the coefficients on lagged values of $\hat{\sigma}_t^2$. Estimation of these models by LS or QML does however require relatively strong moment conditions to hold on the regressors and the volatility estimation errors $\{e_t\}$; note that this is unlike the standard GARCH model estimated by QML where conditions are usually applied to the rescaled squared innovations.

Maheu and McCurdy (2000) find good results using the constrained model, estimated by QML, on foreign exchange returns. Bollen and Inder (1998) estimate a model similar to (2.3.3) by standard QML methods (without accounting for the error autocorrelation structure), using intra-day data to obtain estimates of an unobservable sequence related to daily volatility. This approach requires consistency of the estimates of the unobservable sequence as the number of intra-day observations per day increases without bound, to obtain consistency of the estimator; Bollen and Inder find good results on a sample of S & P Index futures with a large number of observations per day.

Here we will consider estimation of models having the ARCH model (2.3.2), followed by computation of GARCH parameters from the ARCH approximation. This strategy also has the advantage of producing immediately an estimated model which is directly useable for forecasting, and of allowing computation of parameters of any GARCH(p, q) model from a given estimated ARCH representation. Sufficient conditions for consistent and asymptotically normal estimation are given in Section 3 below; it is not necessary that the number of intra-day observations per day (h) increase without bound.

To obtain estimates of GARCH parameters from the ARCH representation we pursue an estimation strategy related to that of Galbraith and Zinde-Walsh (1994, 1997), in which autoregressive models are used in estimation of MA or ARMA models. A high-order ARCH model is used, and estimates of GARCH (p, q) parameters are deduced from the patterns of ARCH coefficients.

Consider first a case of known conditional variance σ_t. The GARCH process (2.1.2) has a form analogous to the ARMA(p, q); using standard results on representation of an ARMA (p, q) process in MA form (see, e.g., Fuller 1976), we can express (2.1.2) in the form

$$\sigma_t^2 = \kappa + \sum_{\ell=1}^{\infty} \nu_\ell \varepsilon_{t-\ell}^2, \tag{2.3.4}$$

with $\nu_0 = 0$ and

$$\begin{aligned} \nu_1 &= \alpha_1 \\ \nu_2 &= \alpha_2 + \beta_1 \nu_1 \\ &\vdots \\ \nu_\ell &= \alpha_\ell + \sum_{i=1}^{min(\ell,p)} \beta_i \nu_{\ell-i}, \qquad \ell \leq q, \\ \nu_\ell &= \sum_{i=1}^{min(\ell,p)} \beta_i \nu_{\ell-i}, \qquad \ell > q; \end{aligned} \tag{2.3.5}$$

and finally

$$\kappa = (1 - \beta(1))^{-1}\omega = \left(1 - \sum_{i=1}^{p} \beta_i\right)^{-1} \omega. \tag{2.3.6}$$

Giraitis et al. (2000) give general conditions under which the ARCH(∞) representation is possible for the GARCH(p,q) case; only the existence of the first moment and summability of the coefficients ν_ℓ (in our notation) are required for the existence of a strictly stationary ARCH(∞) solution as given in (2.3.4).

To estimate the model using a truncated version of this ARCH(∞) representation, we use the estimated low-frequency conditional variance from (2.1.0), defining the estimation error as in (2.3.1) and substituting into (2.3.4) to obtain

$$\hat{\sigma}_t^2 = \kappa + \sum_{\ell=1}^{k} \nu_\ell \varepsilon_{t-\ell}^2 + e_t. \tag{2.3.7}$$

The truncation parameter k must be such that $k \to \infty$, $k/T \to 0$ for consistent estimation of the GARCH model.

This model may be estimated by LS or, to obtain results robust to less restrictive conditions on the volatility estimation errors, LAD. Asymptotic properties of the estimator are considered in Section 3. Estimation proceeds by first obtaining estimates of $\beta = (\beta_1, \beta_2, \ldots, \beta_p)$ from (2.3.5) for $\ell > q$, followed by estimation of the q parameters of α from the first q relations of (2.3.5), and of ω from (2.3.6).

Begin by defining

$$v(0) = \begin{bmatrix} \nu_{q+1} \\ \nu_{q+2} \\ \vdots \\ \nu_k \end{bmatrix}, \text{ and } v(-i) = \begin{bmatrix} \nu_{q+1-i} \\ \nu_{q+2-i} \\ \vdots \\ \nu_{k-i} \end{bmatrix}. \tag{2.3.7}$$

Next define the $(k-q) \times p$ matrix $V = [v(-1)v(-2)\ldots v(-p)] =$

$$\begin{bmatrix} \nu_q & \nu_{q-1} & \cdots & \nu_{q-p+1} \\ \nu_{q+1} & \nu_q & \cdots & \nu_{q-p+2} \\ \vdots & & & \vdots \\ \nu_{k-1} & \nu_{k-2} & \cdots & \nu_{k-p} \end{bmatrix}.$$

where $\nu_r = 0$ for $r \leq 0$. It follows from (2.3.5) that $v(0) = \beta' V$.

The $p \times 1$ vector of estimates $\hat{\beta}$ is defined by

$$\hat{\beta} = (\hat{V}'\hat{V})^{-1}\hat{V}'\hat{v}(0), \tag{2.3.9}$$

where the circumflex indicates replacement of ν_ℓ with the OLS-estimated values $\hat{\nu}_\ell$ in the definitions above. An estimate of α can then be obtained using the estimate of β and the relations (2.3.5): that is

$$
\begin{aligned}
\hat{\alpha}_1 &= \hat{\nu}_1 \\
\hat{\alpha}_2 &= \hat{\nu}_2 - \hat{\beta}_1 \hat{\nu}_1 \\
&\vdots \\
\hat{\alpha}_q &= \hat{\nu}_q - \sum_{i=1}^{min(q,p)} \hat{\beta}_i \hat{\nu}_{q-i}.
\end{aligned}
\tag{2.3.10}
$$

Finally,

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

The covariance matrix of the estimates can be obtained easily from the estimates of the representation (2.3.7) and the Jacobian of the transformation. Let the parameter vector be $\delta = (\omega, \beta, \alpha)$, and let ψ^2 be the variance of the noise e_t in (2.3.1). Then

$$\mathrm{var}(\delta) = J'(\mathrm{var}\hat{\nu})J,$$

where $\hat{\nu}$ is the vector of estimated ARCH parameters and J is the Jacobian of the transformation (2.3.9)-(2.3.10). Computation of the covariance matrix of the LAD parameter vector is discussed in section 3.

3. Asymptotic properties of the integrated volatility–regression estimates

In this section we discuss conditions for consistent and asymptotically Normal estimation of the integrated volatility–regression model of Section 2 by Least Absolute Deviations. Results for Quasi-Maximum Likelihood estimation of the ARCH model were established by Weiss (1986), using the assumption of finite fourth moments of the unnormalized data. Lumsdaine (1996) established consistency and asymptotic Normality of the QMLE for GARCH models by imposing conditions on the re-scaled data, $z_t = \varepsilon_t/\sigma_t$, including the

IID assumption and the existence of high-order moments. Lee and Hansen (1994) generalized these results to $\{z_t\}$ which are not IID, but simply strictly stationary and ergodic.

Consider the ARCH(k) model (2.3.7). Denote by V_T^2 the matrix with elements $\{V_T^2\}_{ij} = (\sum_{t=1}^T \varepsilon_{t-i}^2 \varepsilon_{t-j}^2)$, such that $V_T = (V_T^2)^{\frac{1}{2}}$.

Assumption 1. Suppose that there exists c such that

$$T^{-c}\left(\sum_{t=1}^{T} \varepsilon_{t-i}^2 \varepsilon_{t-j}^2\right) = O_p(1) \quad \text{and} \quad T^c V_T^{-1} = O_p(1) \tag{i}$$

(that is, the matrix is invertible with probability approaching 1), and

$$\max\ \varepsilon_t^2 = o_p(T^{c/2}). \tag{ii}$$

This assumption requires that $\max\ \varepsilon_t^2$ not grow too fast in probability relative to the sum $(\sum_{t=1}^T \varepsilon_{t-i}^2 \varepsilon_{t-j}^2)$.[2]

Next consider an assumption on the daily volatility estimation error,

$$e_t = \hat{\sigma}_t^2 - \sigma_t^2. \tag{3.1.1}$$

This assumption embodies both the Error Assumption of Pollard (1991, p.189) and the additional assumption of Pollard's Theorem 2 that the realizations of the process and the errors (here, volatility estimation errors) are assumed independent.

[2]As an example, consider a case where the eighth unconditional moment of ε_t exists. Then (i) is satisfied for $c = 1$ by the WLLN, $T^{-1}(\sum_{t=1}^T \varepsilon_{t-i}^2 \varepsilon_{t-j}^2) \xrightarrow{p} E(\varepsilon_{t-i}^2 \varepsilon_{t-j}^2)$. At the same time, if we re-write the ARCH(k) model (2.3.7) as a stationary AR(k) model by defining $w_t = \varepsilon_t^2 - \sigma_t^2$ (note that $E(w_t | \varepsilon_t^2, \varepsilon_{t-1}^2, \ldots) = 0$ and $\mathrm{var}(w_t) < \infty$), we obtain $\varepsilon_t^2 = \kappa + \sum_{\ell=1}^k \nu_\ell \varepsilon_{t-\ell}^2 + w_t$. Following example 2 of Pollard (1991) (generalizing to AR(k)), it follows that $\max\ \varepsilon_t^2 = o_p(T^{1/2})$. Of course, Assumption 1 can also hold in cases where moments do not exist.

Assumption 2. The volatility estimation errors $\{e_t\}$ are IID with median 0 and a continuous positive density $f(.)$ in the neighbourhood of zero. The sequences of errors $\{e_t\}$ and of realizations of the process $\{\varepsilon_t\}$ are independent.

Theorem 2. Consider the model (2.3.7),

$$\hat{\sigma}_t^2 = \kappa + \sum_{\ell=1}^{k} \nu_\ell \varepsilon_{t-\ell}^2 + e_t,$$

and suppose that Assumptions 1 and 2 are satisfied. Then if $\hat{\nu} = (\hat{\kappa}, \hat{\nu}_1, \ldots, \hat{\nu}_k)$ is the the LAD-estimated parameter vector,

$$2f(0)V_T(\hat{\nu} - \nu) \xrightarrow{D} N(0, I_{k+1}).$$

Proof. Follows from Pollard (1991, Theorem 2 and Example 1) for $c = 1$; can be extended to any c. Assumption 1 satisfies the conditions (ii)-(iv) of Theorem 2 of Pollard (1991), and combined with Assumption 2 provides all of the conditions (i)-(iv) for the asymptotic distribution to hold. ■

4. Simulation evidence

In this section we present evidence primarily on the finite-sample performance of the regression estimator of 2.3 using the daily integrated volatility, and for comparison the standard Quasi-ML estimator based on daily data alone. The Quasi-ML procedure described and implemented by Schoenberg (1998) is used for the standard estimates. The regression estimator uses (2.1.0) for a daily volatility estimate, followed by estimation of (2.3.7) by OLS or LAD, and transformation to GARCH parameter estimates via (2.3.9)-(2.3.10).

In the first set of simulation exercises, low-frequency (daily) data alone are simulated, because we need the high-frequency data only to generate estimates of $\hat{\sigma}_t^2$, which we accomplish instead by adding noise directly to σ_t^2, which is known in simulation. The variance of this noise controls directly the accuracy of our daily volatility estimate. The variance of the added noise is set equal to the unconditional variance of the simulated process–a very large noise element, corresponding to very weak information from the high-frequency data (see Andersen and Bollerslev 1998, Maheu and McCurdy 2000 for examples of noise variance expected in particular cases). The innovations are Normal, leading to a strong GARCH model estimated by a true ML estimator. These cases are therefore as favourable as possible for the QML estimator, but in the case of the regression estimators use noise variances which are very high (unfavourable) by the standards of typical cases. The fact that the process is strong GARCH is favourable, however, for the regression estimators, and in particular for OLS relative to LAD.

The low-frequency sample size T is set at $200, 600$ and the number of replications is 2000 for each experiment. The three sets of parameter values used result from the aggregation of high-frequency GARCH processes having parameters $\{\omega, \alpha, \beta\}$ equal to (.01, .05, .945), (.01, .08, .89) and (.01, .10, .85).

In a limited second set of simulations, for $T = 600$ only, the high-frequency GARCH process is simulated directly as in the first experiments, (strong GARCH, with normal errors) and aggregated to form a (weak GARCH) daily returns process. Estimates of the daily GARCH parameters on these daily data are obtained by QML, the regression estimators, and the aggregation estimator. The 'true' daily parameters are computed from the aggregation formula of Drost and Nijman (1993) for the GARCH (1,1) flow case, for comparison with the estimates from each method. Note that QML applied directly to the daily data is now technically inconsistent because only the weak GARCH conditions can be guaranteed to apply. By contrast, the aggregation estimator operates here under the most favourable possible circumstances: a high-frequency process which is strong GARCH and of known GARCH(1,1) form is estimated by true ML, yielding consistent estimates because of the conditions presumed for the high-frequency data. For the regression estimates, the a daily volatility estimate has a noise variance implied by the number of observations per day, h, which in these examples is set at 25 to keep overall sample sizes manageable. This number, unrealistically small in most circumstances, is again unfavourable for the regression estimates.

Results from the first set of simulations are contained in Figure 1xx. (results for estimates of ω are not plotted, but are similar in accuracy to those for β.). Results for the second set of simulations is reported in Figure 2xx.

Two general conclusions emerge from Figure 1xx. First, although the normally-distributed errors make OLS efficient relative to LAD in this case, the sacrifice in using the more generally-applicable LAD estimator is small; second, although a very noisy estimate of daily volatility is presumed, there are substantial gains in using the regression estimators[3] which take advantage of the higher-frequency data, relative to ML on the daily returns data alone.

From Figure 2xx, we note first that OLS and LAD again perform similarly, but OLS shows a longer upper or lower tail in most cases. Relative to the QML estimator, either of these regression estimators shows good performance in most of these examples. Finally, in these cases where conditions are ideal for the aggregation estimator, it markedly out-

[3]The regression estimators are presented here in unconstrained form, but can of course be estimated with the same constraints as QML.

performs any of the alternatives. While such circumstances may be unrealistic, the result is unsurprising in the sense that this is the only estimator that forms estimates using each one of the data points in the full high-frequency data set.

5. Concluding remarks

References

Andersen, T.G. and T. Bollerslev (1997) Intraday Periodicity and Volatility Persistence in Financial Markets. *Journal of Empirical Finance* 4, 115-158.

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 (1999) The Distribution of Exchange Rate Volatility. Working paper.

Baillie, R.T., T. Bollerslev, and H.O. Mikkelsen (1996) Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics* 74, 3-30.

Bollen, B. and B. Inder (1998) A General Volatility Framework and the Generalised Historical Volatility Estimator. Working paper, Monash University.

Bollerslev, T. (1986) Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics* 31, 307-327.

Drost, F.C. and B.J.M. Werker (1996) Closing the GARCH Gap: Continuous Time GARCH Modeling. *Journal of Econometrics* 74, 31-57.

Drost, F.C. and T.E. Nijman (1993) Temporal Aggregation of GARCH Processes. *Econometrica* 61(4), 909-927.

Engle, R.F. (1982) Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation. *Econometrica* 50, 987-1008

Francq, C. and J.-M. Zakoïan (1998) Estimating Weak GARCH Representations. Working paper, Université du Littoral–Cte d'Opale, Université de Lille I.

Fuller, W.A. (1976) *Introduction to Statistical Time Series.* Wiley, New York.

Galbraith, J.W. and V. Zinde-Walsh (1994) A Simple, Non-iterative Estimator for Moving-Average Models. *Biometrika* 81, 143–155.

Galbraith, J.W. and V. Zinde-Walsh (1997) On Some Simple, Autoregression-based Estimation and Identification Techniques for ARMA Models. *Biometrika* 84, 685–696.

Giraitis, L., P. Kokoszka and R. Leipus (2000) Stationary ARCH Models: Dependence

Structure and Central Limit Theorem. *Econometric Theory* 16, 3-22.

He, C. and T. Teräsvirta (1999) Fourth Moment Structure of the GARCH(p,q) process. *Econometric Theory* 15, 824-846.

Lee, S.-W. and B.E. Hansen (1994) Asymptotic Theory for the GARCH(1,1) Quasi-Maximum Likelihood Estimator. *Econometric Theory* 10, 29-52.

Lumsdaine, R.L. (1996) Consistency and Asymptotic Normality of the Quasi-Maximum Likelihood Estimator in IGARCH(1,1) and Covariance Stationary GARCH(1,1) Models. *Econometrica* 64, 575-596.

Maheu, J.M. and T.H. McCurdy (2000) Nonlinear Features of Realized FX Volatility. Working paper, University of Alberta/ University of Toronto.

Meddahi, N. and E. Renault (1996) Aggregation and Marginalization of GARCH and Stochastic Volatility Models. Working paper 3597, CRDE, Université de Montréal.

Meddahi, N. and E. Renault (2000) Temporal Aggregation of Volatility Models. Working paper, Université de Montréal.

Nelson, D. (1992) Filtering and Forecasting with Misspecified ARCH Models I: Getting the Right Variance with the Wrong Model. *Journal of Econometrics* 52, 61-90.

Nelson, D. and D.P Foster (1995) Filtering and Forecasting with Misspecified ARCH models II: Making the Right Forecast with the Wrong Model. *Journal of Econometrics* 67, 303-335.

Phillips, P.C.B. (1995) Robust Nonstationary Regression. *Econometric Theory* 11, 912-951.

Pollard, D. (1991) Asymptotics for Least Absolute Deviation Regression Estimators. *Econometric Theory* 7, 186-199.

Schoenberg, R. (1998) *FANPAC*. Aptech Systems, Maple Valley, WA.

Weiss, A.A. (1986) Asymptotic Theory for ARCH Models: Estimation and Testing. *Econometric Theory* 2, 107–131.