31. 1. 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. We obtain asymptotic properties of the estimators and 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

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.

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 estimates 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. The second uses 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.

In section 2 we describe the models and estimators to be considered. Section 3 provides asymptotic results on each of the estimators, while section 4 presents simulation evidence on the finite- sample performance of the 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 t and the lower-frequency observations by τ, so that $\tau = h, 2h, \ldots$ or $\tau_i = ih,\ i = 1, 2, \ldots$. Following Andersen and Bolleslev (1998), estimate the conditional volatility at τ_i as the estimated conditional variance

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

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

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

$$\sigma_\tau^2 = \alpha_0 + \sum_{i=1}^{\tau} \alpha_i \varepsilon_{\tau-i}^2, \tag{2.1}$$

$$\sigma_\tau^2 = \alpha_0 + \sum_{i=1}^{\tau} \alpha_i \varepsilon_{\tau-i}^2 + \sum_{i-1}^{s} \beta_i \sigma_{\tau-i}^2, \tag{2.2}$$

where $\varepsilon_\tau = y_\tau - \mu_\tau$ for a process y_τ with conditional mean μ_τ, or in the driftless case $\varepsilon_\tau = X_\tau$. So $E(\varepsilon_t^2|\psi_{\tau-i} \equiv \sigma_\tau^2$.

Models in the form (2.1), (2.2) are directly estimable if, as in Andersen and Bollerslev, we have measurements of σ_τ^2. We return to this point in Section 2.3 below.

2.2 The aggregation estimator

Drost and Nijman (1993) showed that time aggregated 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 GARCH (1,1) case. As Drost and Nijman noted, such relations can be used to obtain estimates of the parameters at one frequency from those at another. In this section we examine some properties of a low-frequency estimator 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

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2; \tag{2.3}$$

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

$$\sigma_{(h)\tau}^2 = \overline{\alpha}_0 + \overline{\alpha}_1 \varepsilon_{(h)\tau}^2 + \overline{\beta}_1 \sigma_{(h)\tau-1}^2. \tag{2.4}$$

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.

We will show that the mapping

$$\begin{pmatrix} \overline{\alpha}_0 \\ \overline{\alpha}_1 \\ \overline{\beta}_1 \end{pmatrix} = \psi \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \beta_1 \end{pmatrix} \tag{2.5}$$

provided by these 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 $(\alpha_0, \alpha_1, \beta_1)$ 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 $\begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \beta_1 \end{pmatrix}$ 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 $\Omega \in R^3$ the region

$$\Omega = \{(\alpha_0, \alpha_1, \beta_1) \in R^3 |\ \alpha_0 > 0,\ \alpha_1 \geq 0,\ \beta_1 \geq 0;\ \alpha_1 + \beta_1 < 1\},$$

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

$$c = c(\alpha_0, \alpha_1, \beta_1, \kappa) \tag{2.6}$$

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 $\psi \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \beta_1 \end{pmatrix}$ can be written as

$$\psi \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \beta_1 \end{pmatrix} = \begin{pmatrix} h\alpha_0 \frac{1-(\beta_1+\alpha_1)^h}{1-(\beta_1+\alpha_1)} \\ (\beta_1 + \alpha_1)^h - \frac{c}{2} + \left[(\frac{c}{2})^2 - 1\right]^{\frac{1}{2}} \\ -\frac{c}{2} + \left[(\frac{c}{2})^2 - 1\right]^{\frac{1}{2}} \end{pmatrix},$$

where c is given by (2.6); 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 the high-frequency data to obtain estimators of η. Its asymptotic covariance matrix is $V[\hat{\eta}_{QML}] = [W'W]^{-1}B'B[W'W]^{-1}$, where

$$W'W = \sum_{t=1}^{T} \left[\frac{g_t}{\sigma_t^2}\right]\left[\frac{g_t}{\sigma_t^2}\right]'$$

$$\text{and } B'B = \sum_{t=1}^{T} \left[\frac{\varepsilon_t^2}{\sigma_t^2} - 1\right]^2 \left[\frac{g_t}{\sigma_t^2}\right]\left[\frac{g_t}{\sigma_t^2}\right]',$$

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

$$\frac{\partial \psi}{\partial \eta'}[W'W]^{-1}B'B[W'W]^{-1}\frac{\partial \psi'}{\partial \eta}. \tag{2.7}$$

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

$$\frac{\partial \psi}{\partial \eta'}[W'W]^{-1}\frac{\partial \psi'}{\partial \eta}. \tag{2.8}$$

Now consider for comparison the results of standard estimation of $\overline{\eta}$ directly from low-frequency data; the ML estimator $\overline{\eta}_{ML}$ has the asymptotic covariance matrix $[\overline{W}'\overline{W}]^{-1}$,

$$\text{where } [\overline{W}'\overline{W}]^{-1} = \sum_{\tau=1}^{T/h} \left[\frac{\overline{g}_\tau}{\overline{\sigma}^2_{(h)\tau}}\right]\left[\frac{\overline{g}_\tau}{\overline{\sigma}^2_{(h)\tau}}\right]', \tag{2.9}$$

$$\text{with } \overline{g}_\tau = \frac{\partial \overline{\sigma}^2_{(h)\tau}}{\partial \overline{\eta}} = \begin{pmatrix} 1 \\ \overline{\varepsilon}^2_{(h)\tau-1} \\ \overline{\sigma}^2_{(h)\tau-1} \end{pmatrix}.$$

These expressions, (2.9) and (2.7) or (2.8), can be compared to determine the relative asymptotic efficiencies of the aggregation based estimates $\hat{\overline{\eta}}$ and the conventional low-frequency estimates $\overline{\eta}_{ML}$.

Example 1. Let the high-frequency process be ARCH(1); aggregation then leads to a ARCH(1,1) process for the low-frequency data. However, the asymptotic covariance matrix for the estimator $\hat{\overline{\eta}}$ is of rank 2 rather than 3, since the middle part in (2.7) or (2.8) is of dimension 2×2. This indicates that there are cases where $\hat{\overline{\eta}}$ is clearly more efficient than $\overline{\eta}_{ML}$ (or $\overline{\eta}_{QML}$), with covariance matrix of rank 3.

[General results not available yet]

2.3 The regression estimator

As noted above, models of the form (xx1) and (xx2) are directly estimable if we have estimates of the conditional variance of the low-frequency observations, $\overline{\sigma}^2_\tau$, for example from the daily integrated volatility, as in Andersen and Bollerslev. 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 $\overline{\sigma}_\tau^2$. Let $e_t = \hat{\overline{\sigma}}_\tau^2 - \overline{\sigma}_\tau^2$; then the ARCH and GARCH models become

$$\hat{\overline{\sigma}}_\tau^2 = \alpha_0 + \sum_{i=1}^{r} \alpha_i \varepsilon_{\tau-i}^2 + e_t, \tag{2.10}$$

$$\hat{\overline{\sigma}}_\tau^2 = \alpha_0 + \sum_{i=1}^{r} \alpha_i \varepsilon_{\tau-i}^2 + \sum_{i=1}^{s} \beta_i \hat{\overline{\sigma}}_{\tau-i}^2 - \sum_{i=1}^{s} \beta_i e_{\tau-i} + e_t. \tag{2.11}$$

Both (2.10) and (2.11) are estimable as regression models. Nonetheless, there are several difficulties, particularly with respect to the GARCH model (2.11). First, this model has an error term with an MA(s) form; the coefficients of this moving average process, however, must obey the constraint embodied in (2.11) that the MA coefficients are the same as the coefficients on lagged values of $\hat{\overline{\sigma}}_\tau^2$. Estimation, whether by (quasi-) ML or otherwise, must therefore be constrained beyond the constraints imposed by the GARCH model. Second, the MA structure applying to $\{e_t\}$ induces an MA error term in the regression model if the sequence $\{e_t\}$ is itself uncorrelated. If this is not the case–that is, if there is autocorrelation in the errors of estimation of daily volatility–then the error term in (2.11) may follow a more general time series process.

We will first consider estimation of (2.11) by constrained quasi-ML, so that an asymptotic covariance matrix for the estimator follows from standard results.

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. *Monash University Working Paper* .

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

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.