The High Frequency Data GARCH Model

Eric Ghysels

UNC and Federal Reserve Bank of New York

June 1, 2009

Presentation based on:

- **The High Frequency Data GARCH Model**, *Xilong Chen, Eric Ghysels and Fangfang Wang*
- **Continuous record asymptotics for high frequency GARCH Filtering**, *Eric Ghysels, Per Mykland and Eric Renault*

Motivation

- Propose convenient parametric model for daily/weekly/monthly volatility using respectively intra-daily/daily/daily returns.
- Loosely speaking, this is a 'GARCH-version' of Ghysels, Santa-Clara and Valkanov (2005, *JFE*), without ARCH-in-Mean and allowing for intra-daily seasonals.
- We propose new estimation procedures, either based on semi-strong GARCH using QMLE, or weak GARCH using MIDAS-type regressions.

Related Literature

- Combine GARCH model and RV measurement.
  - Gallo and Engle (2006),
  - Frank, Shephard and Sheppard (2009).
- Direct versus iterated volatility forecasting.
  - Ghysels, Rubia and Valkanov (2009).

Related Literature

- Relates to Filtering and RV.
  - Andreou and Ghysels (2002),
  - Meddahi (2002),
  - Ghysels, Mykland and Renault (2008).
- Relates to Temporal Aggregation.
  - Drost and Nijman (1993),
  - Drost and Werker (1996),
  - Meddahi and Renault (2004).

The Model Dropping from the Sky

- The model:

$$\begin{aligned} V_{\tau+1|\tau} &= \tilde{a} + \tilde{b} \exp\left(\sum_{i=1}^{m} \beta_i(\theta)\right) V_{\tau|\tau-1} \\ &+ \tilde{c} \left[\sum_{j=0}^{m-1} \exp\left(\sum_{i=1}^{j} \beta_i(\theta) \right) r_{j\tau}^2 \right] \end{aligned}$$

  where $\beta_i(\theta) \equiv \theta_0 + \theta_1 i + \theta_2 i^2$ (or possibly higher order polynomial).

- The unknown parameters are $\left(\tilde{a}, \tilde{b}, \tilde{c}, \theta_0, \theta_1, \theta_2\right)$.
  1. when $\theta_1 = \theta_2 = 0, \tilde{b} = 1$, we get GARCH(1,1) for HF data;
  2. when $\theta_0 = \theta_1 = \theta_2 = 0$, we get RV-driven GARCH model;
  3. when $\tilde{b} = 1$, we get Periodic GARCH for HF data.
  4. when $m = 1$, we obtain 'regular' GARCH(1,1).

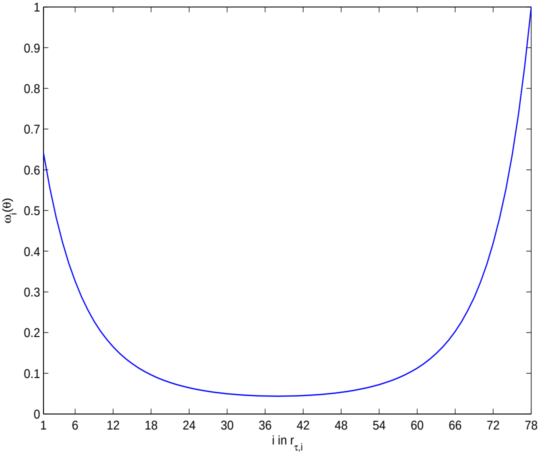
Empirical Evidence: Sample 1993 - 2006 with 5-min S&P 500 data

	S&P 500 Cash	S&P 500 Futures	Number parameters
		AIC Model Selection	
GARCH for HF data	13078.50	9451.35	3
P-GARCH for HF data	13082.50	9413.38	5
General Model	13070.30	9427.43	6
RV-driven GARCH	13083.80	9452.94	3

Empirical Evidence: Sample 1993 - 2006 with 5-min S&P 500 data

	S&P 500 Cash	S&P 500 Futures	Number parameters
	BIC Model Selection		
GARCH for HF data	13098.20	9471.58	3
P-GARCH for HF data	13115.30	9447.09	5
General Model	13109.60	9467.88	6
RV-driven GARCH	13103.40	9473.17	3

Intra-day Weighting Scheme



End of day gets most weight - beginning gets more weight than middle too - similar to Chen and Ghysels (2008).

Some First Impressions

- RV-driven models are not picked by BIC despite being more parsimonious

- Weighting scheme departs from equal weights.

Outline

- The Two-Period Example
- The High Frequency GARCH(1,1)
- Periodic HF-GARCH
- The Exponential Almon MIDAS Specification
- Comparison with HEAVY models
- Estimation Issues and Empirical Results
- Aggregation Bias Revisited
- Continuous Time Diffusions - Work in Progress

The Two-Period Example

- The traditional (weak) GARCH:

$$r_{t+1} = \sqrt{v_{t+1|t}}\varepsilon_{t+1}$$

$$v_{t+1|t} = a + bv_{t|t-1} + cr_t^2$$

- Suppose we construct a volatility for τ, equal to even t.
- Start with two-steps ahead volatility forecast:

$$v_{t+2|t} = a + bv_{t+1|t} + cE_t[r_{t+1}^2] = a + (b + c)v_{t+1|t}$$

and , define: $V_{\tau+1|\tau} \equiv V_{t+2|t} \equiv v_{t+1|t} + v_{t+2|t}$.

The Two-Period Example - Continued

- With some algebra one obtains:

$$V_{\tau+1|\tau} = a(2+c)(1+b) + b^2 V_{\tau|\tau-1} + c(1+(b+c))[r_\tau^2 + br_{\tau-1/2}^2]$$

where we use the convention $r_t = r_\tau$ and $r_{t-1} = r_{\tau-1/2}$.

- Which can be rewritten as:

$$V_{\tau+1|\tau} = \tilde{a} + \tilde{b} V_{\tau|\tau-1} + \tilde{c}[\tilde{\omega}_1 r_\tau^2 + \tilde{\omega}_2 r_{\tau-1/2}^2]$$

with $\sum_i \tilde{\omega}_i = 1$.

The Two-Period Example - Continued

- Note difference with a temporarily aggregated GARCH:

$$\bar{V}_{\tau+1|\tau} = \bar{a} + \bar{b}\bar{V}_{\tau|\tau-1} + \bar{c}[r_\tau + r_{\tau-1/2}]^2$$

- We can link the two via a MIDAS augmentation:

$$\begin{aligned} \bar{V}_{\tau+1|\tau} &= \bar{a} + \bar{b}\bar{V}_{\tau|\tau-1} + \bar{c}[r_\tau + r_{\tau-1/2}]^2 \\ &\quad + \bar{c}[(\tilde{\omega}_1 - 1)r_\tau^2 + (\tilde{\omega}_2 - 1)r_{\tau-1/2}^2 - r_{\tau-j}r_{\tau-1/2-j}] \end{aligned}$$

The Two-Period Example - Continued

- An intermediate case is 'realized volatility' - or RV-driven GARCH:

$$V_{\tau+1|\tau} = \underline{a} + \underline{b} V_{\tau|\tau-1} + \underline{c} RV_\tau \equiv \underline{a} + \underline{b} V_{\tau|\tau-1} + \underline{c}[r_\tau^2 + r_{\tau-1/2}^2]$$

and note that (with $RV_\tau \equiv [r_\tau^2 + r_{\tau-1/2}^2],$):

$$\begin{aligned} \bar{V}_{\tau+1|\tau} &= \bar{a} + \bar{b}\bar{V}_{\tau|\tau-1} + \bar{c}RV_\tau \\ &\quad + \bar{c}[(\tilde{\omega}_1 - 1)r_\tau^2 + (\tilde{\omega}_2 - 1)r_{\tau-1/2}^2] \end{aligned}$$

- Which amounts to a MIDAS correction of RV-driven GARCH.

High Frequency GARCH(1,1)

- Define the total volatility over the m steps:

$$V_{\tau+1|\tau} \equiv V_{t+m|t} \equiv v_{t+1|t} + v_{t+2|t} + \ldots + v_{t+m|t}$$

- This yields a HF-GARCH model with m sampling frequencies:

$$V_{\tau+1|\tau} = \tilde{a} + \tilde{b} V_{\tau|\tau-1} + \tilde{c} \sum_{j=0}^{m-1} \tilde{\omega}_j r_{\tau,j}^2$$

where $\tilde{\omega}_j \sim b^j, \forall j$, $\sum_{j=0}^{m-1} \tilde{\omega}_j = 1$, and $\tilde{b} = b^m$.

High Frequency GARCH(1,1) - Continued

- Back of the envelope calculation:
  - If daily $\tilde{b}$ is say .9 then intra-daily weights vary from .9986 to .9012.
  - If daily $\tilde{b}$ is say .9 then weekly forecast weights vary from .6561 to .9.

Periodic HF-GARCH(1,1)

- When using intra-daily data we have to deal with seasonal patterns.

- We will focus on the periodic GARCH(1,1) of Bollerslev and Ghysels (1996). In particular, consider m periods and the following volatility structure:

$$r_{i,\tau} = \sqrt{v_{i,\tau|i-1,\tau}}\varepsilon_{i,\tau}$$

  and where

$$v_{i,\tau|i-1,\tau} = a_i + b_i v_{i-1,\tau|i-2,\tau} + c_i r_{i-1,\tau}^2$$

Periodic HF-GARCH(1,1) - Continued

- This periodic structure yields the following HF-GARCH dynamics:

$$V_{\tau+1|\tau} = \tilde{a}_p(m) + \left(\prod_{i=1}^{m} b_i\right) V_{\tau|\tau-1} + \tilde{c}_p(m) \sum_{j=1}^{m} \left(\prod_{i=2+j}^{m+1} b_i\right) c_{j+1} r_{j,\tau}^2$$

where, letting $a_{m+1} \equiv a_1$, $b_{m+1} \equiv b_1$ and $c_{m+1} \equiv c_1$, $\tilde{a}_p(m)$ and $\tilde{c}_p(m)$ are as follows:

$$\tilde{a}_p(m) = \left[\sum_{j=1}^{m} \sum_{i=2}^{j} a_i \prod_{l=i+1}^{j} (b_l + c_l)\right] \left[1 - \left(\prod_{i=1}^{m} b_i\right)\right] + \left[\sum_{j=1}^{m} \prod_{i=2}^{j} (b_i + c_i)\right] \left[\sum_{j=1}^{m} a_j \prod_{i=j+1}^{m+1} b_i\right]$$

$$\tilde{c}_p(m) = \sum_{j=1}^{m} \prod_{i=2}^{j} (b_i + c_i)$$

Periodic HF-GARCH(1,1) - Continued

- Therefore, the periodic HF-GARCH model is of the form

$$V_{\tau+1|\tau} = \tilde{a}_p + \tilde{b}_p V_{\tau|\tau-1} + \tilde{c}_p \sum_{j=1}^{m} \omega_j r_{j,\tau}^2$$

where $\omega_j \geq 0$ and $\sum_j \omega_j = 1$.

- Practically, it would be difficult to estimate $m \times 3$ parameters in the periodic GARCH case, i.e. all a_i, b_i and c_i, for all i.

The Exponential Almon MIDAS specification

- There is a convenient shortcut, inspired by the Exponential Almon MIDAS filter proposed in Ghysels et al. (2005). Unconstrained model:

$$\begin{aligned} V_{\tau+1|\tau} &= \tilde{a} + \tilde{b} \exp\left(\sum_{i=1}^{m} \beta_i(\theta)\right) V_{\tau|\tau-1} \\ &\quad + \tilde{c} \left[\sum_{j=0}^{m-1} \exp\left(\sum_{i=1}^{j} \beta_i(\theta) \right) r_{j\tau}^2 \right] \end{aligned}$$

  where $\beta_i(\theta) \equiv \theta_0 + \theta_1 i + \theta_2 i^2$.

- The unknown parameters are $\left(\tilde{a}, \tilde{b}, \tilde{c}, \theta_0, \theta_1, \theta_2\right)$, and
  1. when $\theta_1 = \theta_2 = 0, \tilde{b} = 1$, we get ID GARCH(1,1);
  2. when $\theta_0 = \theta_1 = \theta_2 = 0$, we get RV-driven GARCH model;
  3. when $\tilde{b} = 1$, we get ID Periodic GARCH model.
  4. when $m = 1$, we obtain 'regular' GARCH(1,1).

Comparison with HEAVY model and multi-step ahead forecasting

- HEAVY models (Shephard et al. (2009)) amount to RV-GARCH plus an additional AR(1) equation for RV.

- Direct versus iterated volatility forecasting: similar to Ghysels, Rubia and Valkanov (2009) we formulate direct rather than iterated forecasts - using mixed frequencies which appears to be the best strategy.

Comparison with HEAVY model and multi-step ahead forecasting

- In the HF-GARCH model there is no need to add an extra equation. For two reasons:
  - Recall we can define the total volatility over over any *m* steps:

$$V_{\tau+1|\tau} \equiv V_{t+m|t} \equiv v_{t+1|t} + v_{t+2|t} + \ldots + v_{t+m|t}$$

  - More importantly, can we use parameter estimates from say a daily model with 5-min data to formulate say a weekly model (with the same 5-min data): in several cases the answer is *yes!*

Estimation Issues

- We propose two approaches: semi-strong GARCH inspired QMLE and weak GARCH MIDAS regression approach.
- Suppose now that we have intra-day returns $\{r_{\tau,j},\, j = 1, \ldots, m,\, \tau = 1, 2, \ldots, T\}$ and 'daily' squared returns $\{R_\tau^2,\, \tau = 1, 2, \ldots, T\}$.

Estimation Issues

- To estimate the parameters, consider the following quasi-MLE $\hat{\Theta}$ which is the minimizer of

$$-logl = \sum_{\tau=2}^{T} \left(\log V_{\tau|\tau-1}(\Theta) + \frac{R_\tau^2}{V_{\tau|\tau-1}(\Theta)} \right)$$

- This amounts to imposing a low frequency semi-strong GARCH assumption - say at the daily level.

Estimation Issues

- Preliminary theoretical and simulation results suggest that HF-GARCH specification with QMLE provides *consistent* (and asymptotically normal) estimates of underlying GARCH(1,1) under suitable regularity conditions.

- This is different from temporal aggregation results.

Temporal aggregation biases revisited

We use Meddahi and Renault simulation design. Instead of temporal aggregation we use a HF-GARCH model with m sampling frequencies based on GARCH(1,1) specification:

$$V_{\tau+1|\tau} = \tilde{a} + \tilde{b} V_{\tau|\tau-1} + \tilde{c} \sum_{j=0}^{m-1} \tilde{\omega}_j r_{\tau,j}^2$$

	MSE $\tilde{a}$	MSE $\tilde{b}$	MSE $\tilde{c}$
m = 5 / T = 500	0.000014	0.016862	0.017070
m = 5 / T = 1000	0.000000	0.000131	0.002232
m = 78 / T = 500	0.000347	0.000269	1.546709
m = 78 / T = 1000	0.000223	0.000107	0.801159
m = 288 / T = 500	0.015587	0.000438	36.078381
m = 288 / T = 1000	0.006146	0.000117	12.158628

Weak GARCH estimation

- We replace the ARMA approach of Francq and Zakoïan (2002) with the following MIDAS regressions:

$$
\begin{aligned}
R_\tau^2 &= \tilde{a} + \tilde{b} \exp\left(\sum_{i=1}^{m} \beta_i(\theta)\right) R_{\tau-1}^2 \\
&\quad + \tilde{c} \left[\sum_{j=0}^{m-1} \exp\left(\sum_{i=1}^{j} \beta_i(\theta) \right) r_{j\tau}^2 \right] + u_\tau
\end{aligned}
$$

  which is called a ADL-MIDAS regression in Andreou, Ghysels and Kourtellos (2009).

- Note that we can also write the above equations in terms of RV, yielding ADL-MIDAS in RV instead of daily squared returns.

- The X-MIDAS representations are very similar to the regressions estimated by Ghysels et al. (2005).

Empirical Results

- We analyze four data sets which consist of five-minute intra-day returns of respectively Dow Jones and S&P500 cash and futures markets.
- Since number of parameters is critical, we look at: Information Criterion:

$$AIC = -2ln(L) + 2k$$

$$BIC(orSBC) = -2ln(L) + kln(T)$$

$$HQC = -2ln(L) + 2kln(ln(T))$$

where L stands for log-likelihood; k for number of parameters; T for sample size.

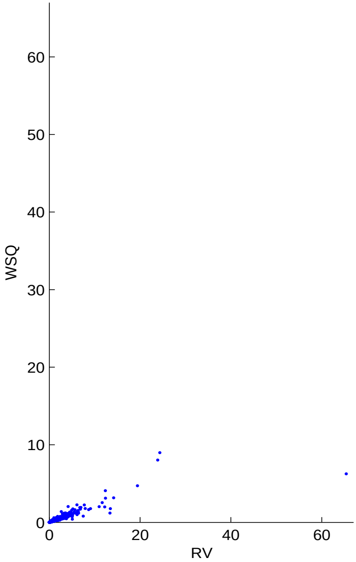
Empirical Evidence: Sample 1993 - 2006 with 5-min S&P 500 data

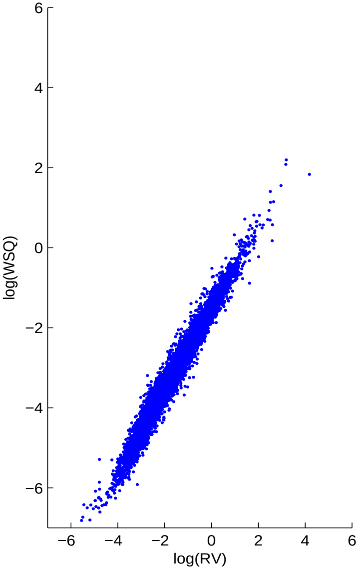
	S&P 500 Cash	S&P 500 Futures	Number parameters
		AIC Model Selection	
GARCH for HF data	13078.50	9451.35	3
P-GARCH for HF data	13082.50	9413.38	5
General Model	13070.30	9427.43	6
RV-driven GARCH	13083.80	9452.94	3

Empirical Evidence: Sample 1993 - 2006 with 5-min S&P 500 data

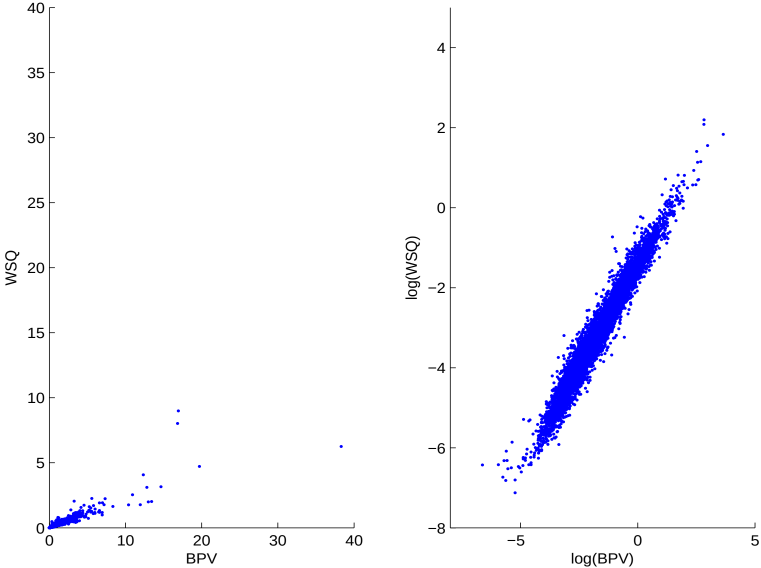
	S&P 500 Cash	S&P 500 Futures	Number parameters
		BIC Model Selection	
GARCH for HF data	13098.20	9471.58	3
P-GARCH for HF data	13115.30	9447.09	5
General Model	13109.60	9467.88	6
RV-driven GARCH	13103.40	9473.17	3

$\sum_j \tilde{\omega}_j r_{\tau,j}^2$ versus RV - S & P 500 Futures





$\sum_j \tilde{\omega}_j r_{\tau,j}^2$ versus BPV - S & P 500 Futures



Continuous Time Diffusions - Work in Progress

- Starting from Drost and Werker (1996) we can always 'close the GARCH gap'.
- More challenging is to view HF-GARCH is a mis-specified model/filter for underlying continuous time process. Work in progress, Ghysels, Mykland and Renault (2009).