11^a Escolade Séries

Temporais e Econometria

Analysis of High Frequency Financial Data: Methods, Models and Software. Part II: Realized Variance

Eric Zivot

Department of Economics

University of Washington

August 1, 2005

Introduction

- Key problem in financial econometrics: modeling, estimation and forecasting of conditional return volatility and correlation.
  - Derivatives pricing, risk management, asset allocation

- Conditional volatility is highly persistent

- Inherent problem: conditional volatility is unobservable

- Traditional latent variable models: ARCH-GARCH, Stochastic volatility (SV) based on squared returns
  - difficult estimation
  - high frequency data not utilized
  - standardized returns not Gaussian
  - Imprecise forecasts
  - multivariate extensions are difficult

- New approach uses estimates of latent volatility based on high frequency data (realized variance measures)
  - Volatility is observable
  - Traditional time series models are applicable
  - High dimensional multivariate modeling is feasible

Construction of Realized Variance Measures

- $p_{i,t}$ = log-price of asset i at time t (aligned to common clock)
  - $\mathbf{p}_t = (p_{1,t}, \ldots, p_{n,t})' = n \times 1$ vector of log prices

- Δ = fraction of a trading session associated with the implied sampling frequency,

- $m = 1/\Delta$ = number of sampled observations per trading session

- T = number of days in the sample $\Rightarrow mT$ total observations

Example (FX market): Prices are sampled every 30 minutes and trading takes place 24 hours per day

- $m = 48$ 30-minute intervals per trading day

- $\Delta = 1/48 \approx 0.0208$.

Example (Equity market): Prices are sampled every 5 minutes and trading takes place 6.5 hours per day

- $m = 78$ 5-minutes intervals per trading day

- $\Delta = 1/78 \approx 0.0128$.

- Intra-day continuously compounded (cc) returns from time t to $t + \Delta$

$$\begin{aligned} r_{i,t+\Delta} &= p_{i,t+\Delta} - p_{i,t},\ i = 1, \ldots, n \\ \mathbf{r}_{t+\Delta} &= \mathbf{p}_{t+\Delta} - \mathbf{p}_t \end{aligned}$$

- Daily returns

$$\begin{aligned} r_{i,t} &= r_{i,t-1+\Delta} + r_{i,t-1+2\Delta} + \cdots + r_{i,t-1+m\Delta} \\ \mathbf{r}_t &= \mathbf{r}_{t-1+\Delta} + \mathbf{r}_{t-1+2\Delta} + \cdots + \mathbf{r}_{t-1+m\Delta} \end{aligned}$$

- *Realized variance* (RV) for asset i on day t

$$RV_{i,t} = \sum_{j=1}^{m} r_{i,t-1+j\Delta}^2,\ t = 1, \ldots, T$$

- *Realized volatility* (RVOL) for asset i on day t:

$$RVOL_{i,t} = \sqrt{RV_{i,t}}$$

- *Realized log-volatility* (RLVOL) :

$$RLVOL_{i,t} = \ln(RVOL_{i,t})$$

- The $n \times n$ *realized covariance* (RCOV) matrix on day t

$$RCOV_t = \sum_{j=1}^{m} \mathbf{r}_{t-1+\Delta} \mathbf{r}'_{t-1+\Delta}, \;\; t = 1, \ldots, T$$

  – The $n \times n$ matrix $RCOV_t$ will be positive definite provided $n < m$

- The realized correlation between asset i and asset j

$$\begin{aligned} RCOR_{i,j,t} &= \frac{[RCOV_t]_{i,j}}{\sqrt{[RCOV_t]_{i,i} \times [RCOV_t]_{j,j}}} \\ &= \frac{[RCOV_t]_{i,j}}{RVOL_{i,t} \times RVOL_{j,t}} \end{aligned}$$

Non-overlapping RV measures over h days:

$$
\begin{aligned}
RV_{i,t}^{h} &= \sum_{j=1}^{h} RV_{i,t}, \; t = h, 2h, \ldots, T/h \\
RCOV_{i,t}^{h} &= \sum_{j=1}^{h} RCOV_{t}, \; t = h, 2h, \ldots, T/h
\end{aligned}
$$

Quadratic Return Variation and Realized Variance

Two fundamental questions about RV are:

Q1 What does RV estimate?

Q2 Are RV estimates economically important?

Answers are provided in a number of important papers:

- Andersen, Bollerslev, Diebold, Labys (ABDL): "The Distribution of Realized Exchange Rate Volatility" JASA, 2001

- Andersen, Bollerslev, Diebold, Labys: "Modeling and Forecasting Realized Volatility" ECTA, 2003

- Barndorff-Nielsen and Shephard (BNS): "Estimating Quadratic Variation Using Realized Variance" JAE 2002

- Barndorff-Nielsen and Shephard: "Econometric Analysis of Realized Volatility and Its Use in Estimating Stochastic Volatility Models" JRSSB, 2002.

Continuous time arbitrage-free log-price process

- let $p(t)$ denote the univariate log-price process for a representative asset defined on a complete probability space (Ω, F, P), evolving in continuous time over the interval $[0, T]$.

- Let F_t be the $\sigma-$field reflecting information at time t such that $F_s \subseteq F_t$ for $0 \leq s \leq t \leq T$.

Result: If $p(t)$ is in the class of special semi-martingales then it has the representation

$$p(t) = p(0) + A(t) + M(t), \;\; A(0) = M(0) = 0$$

where $A(t)$ is a predictable drift component of finite variation, and $M(t)$ is a *local martingale*. Note: jumps are allowed in both $A(t)$ and $M(t)$.

- Let mT be a positive integer indicating the number of return observation obtained by sampling $m = 1/\Delta$ times per day for T days

- The cc return on asset i over the period $[t-\Delta, t]$ is

  $$r(t, t-\Delta) = p(t) - p(t-\Delta),\ t = \Delta, 2\Delta, \ldots, T$$

- The daily cc and cumulative returns are

$$\begin{aligned} r(t, t-1) &= p(t) - p(t-1) \\ r(t) &= p(t) - p(0) \end{aligned}$$

Definition: The *quadratic variation* (QV) of the return process at time t is

$$[r](t) = p\,\text{lim} \sum_{j=0}^{m-1} \{p(s_{j+1}) - p(s_j)\}^2$$

where $0 = s_0 < s_1 < \cdots < s_M = t$ and the limit is for the mesh size

$$\max_{1 \leq j < m} |s_j - s_{j-1}| \rightarrow 0 \text{ as } m \rightarrow \infty$$

- The QV process measures the realized sample path variation of the squared return process.

- QV is a unique and invariant ex-post realized volatility measure that is essentially model free.

The definition of QV implies the following convergence result for semi-martingales:

$$RV_t \xrightarrow{p} [r](t) - [r](t-1) \equiv QV_t, \text{ as } m \to \infty$$

That is, *daily RV converges in probability to the daily increment in QV*. This answers the first question Q1.

Remark:

- As noted by ABDL, QV_t is related to, but distinct from, the daily conditional return variance. That is, in general

$$QV_t \neq var(r(t,t-1)|F_{t-1})$$

Result (ABDL 2001): If

(i) the price process $p(t)$ is square integrable;

(ii) the mean process $A(t)$ is continuous;

(iii) the daily mean process, $\{A(s){-}A(t{-}1)\}_{s\in[t-1,t]}$, conditional on information at time t is independent of the return innovation process, $\{M(u)\}_{u\in[t-1,t]}$,

(iv) the daily mean process, $\{A(s){-}A(t{-}1)\}_{s\in[t-1,t]}$, is a predetermined function over $[t-1,t]$,

then for $0 \leq t-1 \leq t \leq T$

$$var(r(t,t-1)|F_{t-1}) = E[QV_t|F_{t-1}]$$

That is, *the conditional return variance equals the conditional expectation of the daily QV process*.

Note: the ex post value of RV_t is an unbiased estimator for the conditional return variance $var(r(t,t-1)|F_{t-1})$:

$$E[RV_t|F_{t-1}] = E[QV_t|F_{t-1}] = var(r(t,t-1)|F_{t-1})$$

Therefore, RV_t is economically important which answers the second question Q2.

Remark: The restrictions on the conditional mean process allow for realistic price processes.

- price process is allowed to exhibit deterministic intra-day seasonal variation.

- mean process can be stochastic as long as it remains a function, over the interval $[t-1, t]$, of variables in F_{t-1}.

- jumps are allowed in the return innovation process $M(t)$

- leverage effects caused by contemporaneous correlation between return innovations and innovations to the volatility process are allowed.

Results for *Itô processes*

- $p(t)$ is described by the stochastic differential equation

$$\begin{aligned} dp(t) &= \mu(t)dt + \sigma(t)dW(t) \\ W(t) &= \text{Wiener process} \end{aligned}$$

where $\mu(t)$ and $\sigma(t)$ may be random functions. Note: $\sigma(t)$ may exhibit jumps, dirurnal effects, long memory or be nonstationary.

- Daily return

$$r(t, t-1) = \int_{t-1}^{t} \mu(s)ds + \int_{t-1}^{t} \sigma(s)dW(s)$$

- There may be leverage effects. That is, $\sigma(t)$ may be correlated with $W(t)$. For example,

$$\begin{aligned} d\sigma(t) &= \tilde{\mu}(t)dt + \tilde{\sigma}(t)d\tilde{W}(t) \\ cov(dW(t), d\tilde{W}(t)) &\neq 0 \end{aligned}$$

- Daily increment to QV

$$QV_t = \int_{t-1}^{t} \sigma^2(s)ds = IV_t$$

where IV_t denotes daily *integrated variance* (IV).

Result: Since $RV_t \xrightarrow{p} QV_t$, it follows that

$$RV_t \xrightarrow{p} IV_t$$

Remark: IV_t plays a central in option pricing with stochastic volatility (e.g. Hull and White (1987))

Result (ABDL (2003)): If the mean process, $\mu(s)$, and volatility process, $\sigma(s)$, are independent of the Wiener process $W(s)$ over $[t-1, t]$ then

$$r(t, t-1)|\sigma\{\mu(s), \sigma(s)\}_{s\in[t-1,t]} \sim N\left(\int_{t-1}^{t} \mu(s)ds, IV_t\right)$$

where $\sigma\{\mu(s), \sigma(s)\}_{s\in[t-1,t]}$ denotes the $\sigma-$field generated by $(\mu(s), \sigma(s))_{s\in[t-1,t]}$.

- Since $\int_{t-1}^{t} \mu(s)ds$ is generally very small for daily returns and RV_t is a consistent estimator of IV_t, for *Itô processes* daily returns should follow a *normal mixture distribution* with RV_t as the mixing variable.

- If there are jumps in $dp(t)$, then $RV_t \xrightarrow{p} IV_t$ but returns are no longer conditionally normally distributed.

Asymptotic Distribution Theory for Realized Variance

- For a diffusion process, the consistency of RV_t for IV_t relies on the sampling frequency per day, Δ, going to zero.

- Convergence result is not attainable in practice as it is not possible to sample continuously.

  - Theory suggests sampling as often as possible to get the most accurate estimate of IV_t.

  - Market microstructure frictions eventually dominate the behavior of RV as $\Delta \rightarrow 0$, which implies a practical lower bound on Δ for observed data.

  - For $\Delta > 0$, RV_t will be a noisy estimate of IV_t.

Define the error in RV_t for a given Δ as

$$u_t(\Delta) = RV_t - IV_t$$

or

$$RV_t = IV_t + u_t(\Delta)$$

Result (Meddahi (2002) JAE):

- The mean of $u_t(\Delta)$ is non-zero when the drift $m(t)$ is non-zero

- $u_t(\Delta)$ is heteroskedastic

- Under leverage effect, $cov(IV_t, u_t(\Delta)) \neq 0$

- $corr(IV_t, u_t(\Delta)) = O(\Delta^{3/2})$ as $\Delta \to 0$

Result (BNS (2001)): For the Ito diffusion model under the assumption that mean and volatility processes are jointly independent of $W(t)$,

$$\sqrt{m}\frac{u_t(\Delta)}{\sqrt{2 \cdot IQ_t}} = \sqrt{m}\frac{(RV_t - IV_t)}{\sqrt{2 \cdot IQ_t}} \xrightarrow{d} N(0,1)$$

where

$$IQ_t = \int_{t-1}^{t} \sigma^4(s)ds$$

is the *integrated quarticity* (IQ).

Remarks:

- RV_t converges to IV_t at rate $\sqrt{m}$,

- The asymptotic distribution of RV_t is mixed-normal since IQ_t is random.

- IQ_t may be consistently estimated using the following scaled version of *realized quarticity* (RQ)

$$\frac{m}{3} RQ_t = \frac{m}{3} \sum_{j=1}^{m} r_{t+\Delta}^{4}$$

- The feasible asymptotic distribution for RV_t is

$$\frac{RV_t - IV_t}{\sqrt{\frac{2}{3} \cdot RQ_t}} \overset{A}{\sim} N(0,1)$$

which result suggests

$$\widehat{SE}(RV_t) = \sqrt{\frac{2}{3} \sum_{j=1}^{m} r_{t+\Delta}^4}$$

- Using the delta-method

$$\frac{RVOL_t - \sqrt{IV_t}}{\sqrt{\frac{2}{12} \cdot \frac{RQ_t}{RV_t}}} \overset{A}{\sim} N(0,1)$$

which suggests

$$\widehat{SE}(RV_t) = \sqrt{\frac{2}{12} \cdot \frac{RQ_t}{RV_t}}$$

- BNS find that the finite sample distribution of RV_t and $RVOL_t$ can be quite far from their respective asymptotic distributions for moderately sized m.

- BNS show that the asymptotic distribution of $RLVOL_t^2$,

$$\frac{RLVOL_t^2 - \ln(IV_t)}{\sqrt{\frac{2}{3} \cdot \frac{RQ_t}{RV_t^2}}} \overset{A}{\sim} N(0, 1)$$

is closer to its finite sample asymptotic distribution than the asymptotic distributions of RV_t and $RVOL_t$.

- BNS (2004) extend the above asymptotic results to cover the multivariate case, providing asymptotic distributions for $RCOV_t$ and $RCOR_{i,j,t}$, as well as realized regression estimates.

- These limiting distributions are much more complicated than the ones presented above, and the reader is referred to BNS (2004) for full details and examples.

Practical Problems in the Construction of RV

- The foremost problem is the choice of sampling frequency Δ (or number of observations per day m).
  - Bandi and Russell (2003) propose a data-based method for choosing Δ that minimizes the MSE of the measurement error.
  - Simulations and empirical examples suggest optimal sampling is around 1-3 minutes for equity returns.

- As discussed in Bai, Russell and Tiao (2000), various market microstructure effects (bid/ask bounce, infrequent trading, calendar effects etc.) induce serial correlation in the intra-day returns $r_{i,t+\Delta}$ which may induce biases in RV measures.
  - Filter the intra-day returns using simple MA or AR models prior to constructing RV measures.