The Use of High-Frequency Data in Financial Econometrics: Recent Developments

Peter Reinhard Hansen

Department of Economics, Stanford University

Stanford Conference in Quantitative Finance, 2010
The Society for Financial Econometrics (SoFiE) Founded 2008
Oxford-Man Institute @ Oxford University (2007)
Volatility Institute @ NYU Stern (Robert Engle) (2009)

VLAB
Part 1: Accurate Measures of Volatility
Realized Measures of Volatility Computed from High-Frequency Data
- Ideal Case: Realized Variance
- Noisy Data (Market Microstructure)
  - Empirical Properties of Noise
- Robust Estimators:
  - Realized Kernel & Markov Chain Estimator

Part 2: Applications
Utilizing Realized Measures for Volatility Modeling and Forecasting
- Realized GARCH Models
Realized Measures of Volatility
High-Frequency Prices

Mid-quote (05-04-2007, 14588 obs.)

9:30 10:15 11 12 13 14 15 16
A Measure of Variation of Asset Prices

Suppose that \( Y_t = \log P_t \) is a semi-martingale

\[
Y_t = \int_0^t a_u \, du + \int_0^t \sigma_u \, dB_u + J_t,
\]

where

- \( a \) is a predictable locally bounded drift,
- \( \sigma \) is a cadlag volatility process,
- \( B \) is a standard Brownian motion, and
- \( J_t = \sum_{i=1}^{N_t} D_i \) is a finite activity jump process.

Quadratic Variation (over \([0, 1]\)) is

\[
QV = \int_0^1 \sigma_u^2 \, du + \sum_{i=1}^{N_1} D_i^2.
\]
Intraday Returns

- Divide day into $n$ subintervals

\[ 0 = T_0 < T_1 < \cdots < T_{n-1} < T_n = 1, \]

(e.g. equidistant $T_j - T_{j-1} = \frac{1}{n}$).

- Intraday returns

\[ y_{j,n} = Y_{T_j} - Y_{T_{j-1}}, \quad j = 1, \ldots, n. \]

- Daily return is the sum of intraday returns,

\[ Y_1 - Y_0 = y_{1,n} + \cdots + y_{n,n}. \]
Realized Variance (Empirical QV)

- Realized variance is defined by
  \[ RV = \sum_{j=1}^{n} y_{j,n}^2. \]

- Properties (ideal case)
  - \( RV \xrightarrow{p} QV \) as \( \sup_{i=1,\ldots,n} |T_i - T_{i-1}| \to 0. \)
  - No jumps (i.e. \( J_t = 0 \)) then
    \[ QV = IV := \int_0^1 \sigma_u^2 du, \quad \text{(integrated variance)} \]
  - and
    \[ \sqrt{n}(RV - IV) \overset{d}{\to} N(0, V). \]
Literature Invigorated by
Andersen and Bollerslev (1998)
Realized Variance useful for evaluation of GARCH models

Statistical Properties of Realized Variance
Andersen et al. (2001)
Barndorff-Nielsen and Shephard (2002)
The Great Tragedy of Science:

Thomas H. Huxley (1825-1895).
The Great Tragedy of Science:
The Slaying of a Beautiful Hypothesis by an Ugly Fact

*Thomas H. Huxley (1825-1895).*
High-Frequency Prices: AA 2007-05-04

Graphs showing time series data with price changes and autocorrelation functions (ACF) for different time periods. The ACF values are displayed for the lag lengths specified, with confidence intervals indicated in red.
High-Frequency Prices: AA 2007-01-16

Graphs showing price changes and ACF (Autocorrelation Function) for different time periods:
- Price changes from 9:30 to 16:00 on 01-16-2007 with 4631 observations.
- Price changes from 9:30 to 16:00 on 01-16-2007 with 13588 observations.
- ACF for lag length from 01-10-2007 with ACF(1) = -0.20.
- ACF for lag length from 01-10-2007 with ACF(1) = 0.09.
High-Frequency Prices: AA 2007-01-26

The graphs illustrate price changes and autocorrelation functions (ACFs) for two different sets of high-frequency data. The left pair of graphs shows price changes and ACFs for 5,325 observations, while the right pair shows price changes and ACFs for 14,271 observations. The plots indicate the presence of volatility and autocorrelation patterns in the data.
High-Frequency Prices: AA 2007-01-26 (zoom)
High-Frequency Returns are Autocorrelated

- Autocorrelation causes RV to be biased/inconsistent.
  - Computing RV with tick-by-tick returns says more about “noise” than “volatility”.
- Ad-hoc resolution: Sample sparsely.
  - Compute RV using 5-minute returns.
Sample Sparsely
Volatility Signature Plot Reveals Bias Problem
Price Without Noise

- Price

\[ Y_t \]
Price cloaked with noise

\[ X_t = Y_t + U_t \]
Properties of the Noise

- Hansen and Lunde (2006)
  Journal of Business and Economic Statistics
  Invited paper with Comments and Rejoinder.

- Noise $U_t$ is...
  - ... serial dependent
  - ... endogenous (not independent of $Y_t$)
  - ... has changed over time (tick size)
  - ... is “small”.
Robust
Realized Measures
Realized Kernel

- Realized Autocovariances

\[ \gamma_h = \sum_{i=1}^{n} y_{i,n} y_{i-h,n}. \]

- So \( RV = \gamma_0 \).

\[ RK = \sum_{h=-\infty}^{\infty} k\left(\frac{h}{H}\right) \gamma_h, \]

where
- \( k(\cdot) \) is a kernel function and
- \( H \) is a bandwidth parameter.
Figure 2.—Kernel functions, $k(x/c^*)$, scaled by their respective $c^*$ to make them comparable.
Realized Volatility Measure (Annualized Realized Kernel)

- Russian Crisis & LTCM
- 9/11
- Mini-Crash (Oct. 27, 1997)
- Fed Surprise Cut 50bp
- Asian Crisis
- Dot-Com Burst
- WorldCom
- Lehman Brothers
- Bear Stearns Collapse
- Liquidity Crunch

High = 161.70 (Oct. 10, 2008)
Multivariate Problem

- Asynchronous Trading (quoting)
- Refresh Time “aligns” observations, and induces additional noise
Realized Beta from Realized Kernel

- **Assets: AA-SPY**
  - Slope = 1.088 (0.013)
  - const = -0.082 (0.016)

- **Assets: C-SPY**
  - Slope = 1.059 (0.012)
  - const = -0.038 (0.011)

- **Assets: AA-SPY**
  - Slope = 0.961 (0.009)
  - const = 0.042 (0.012)

- **Assets: C-SPY**
  - Slope = 0.986 (0.012)
  - const = 0.008 (0.012)
Univariate Realized Kernel
Barndorff-Nielsen et al. (2008)

Multivariate Realized Kernel
Barndorff-Nielsen et al. (2010a)

Realized Kernels in Practice
Barndorff-Nielsen et al. (2009)

Relation to other estimators
Barndorff-Nielsen et al. (2010b)
Markov Chain Estimator
(Personal Favorite)

Hansen and Horel (2009)
Prices On A Grid

GE -- transaction prices

2004-11-02
The Markov Chain Estimator

- Exploits **discreteness** of price changes.
- **Simple** to compute
  - Estimate MC model (requires counting)
  - Estimator computed with basic matrix operations
- **Inference**...
  - Standard errors given in closed-form
  - Asymptotics is reliable
Empirical Example. GE 2004-11-01

\[
x = \begin{pmatrix}
-3 \\
-2 \\
-1 \\
1 \\
2 \\
3
\end{pmatrix} \quad \hat{P} = \begin{pmatrix}
0.17 & 0.00 & 0.33 & 0.50 & 0.00 & 0.00 \\
0.00 & 0.06 & 0.23 & 0.51 & 0.17 & 0.03 \\
0.00 & 0.01 & 0.25 & 0.71 & 0.02 & 0.00 \\
0.00 & 0.02 & 0.72 & 0.25 & 0.01 & 0.00 \\
0.00 & 0.14 & 0.64 & 0.19 & 0.03 & 0.00 \\
0.00 & 0.50 & 0.50 & 0.00 & 0.00 & 0.00
\end{pmatrix} \quad \hat{\pi} = \begin{pmatrix}
0.00 \\
0.02 \\
0.48 \\
0.48 \\
0.02 \\
0.00
\end{pmatrix}.
\]

Compute

\[
\Lambda_{\hat{\pi}} = \text{diag}(\hat{\pi}), \quad \hat{Z} = (I - \hat{P} + \hat{\pi})^{-1} \quad \hat{N} = \iota\hat{\pi}'.
\]

Markov Chain Estimator is:

\[
MC = x' \left[ \frac{n}{q_{\log}} \Lambda_{\hat{\pi}} (2\hat{Z} - I) \right] x
\]

\[
= x' \begin{pmatrix}
0.007 & -0.000 & -0.001 & 0.000 & -0.000 & -0.000 \\
-0.000 & 0.035 & -0.013 & -0.003 & 0.010 & 0.002 \\
-0.000 & -0.012 & 0.533 & 0.256 & -0.004 & -0.002 \\
-0.001 & -0.004 & 0.250 & 0.532 & -0.008 & -0.000 \\
-0.000 & 0.008 & 0.004 & -0.014 & 0.033 & 0.000 \\
-0.000 & 0.004 & -0.001 & -0.003 & 0.000 & 0.004
\end{pmatrix} \times 0.7469.
\]
 Filtering Argument

- Observed Prices: $X_{T_j}, j = 0, 1, \ldots, n$, with
- $\Delta X_{T_j} = X_{T_j} - X_{T_{j-1}}$ a homogeneous ergodic MC.
- Consider filtered prices
  \[ E(X_{T_{j+h}} | \mathcal{F}_{T_j}) \]

- Easy to compute within the MC framework because
  \[ E(X_{T_{j+h}} | \mathcal{F}_{T_j}) = X_{T_j} + \sum_{i=1}^{h} E(\Delta X_{T_{j+i}} | \mathcal{F}_{T_j}) \]

- Even as $h \to \infty$!
Martingale Representation

- We have
  \[ X_{T_j} = \mu_j + Y_{T_j} + U_{T_j}, \]
  where
  \[ \mu_j = \mu \cdot j, \text{ with } \mu = \mathbb{E}(\Delta X_{T_j}); \]
  \[ Y_{T_j} = \lim_{h \to \infty} \mathbb{E}(X_{T_j+h} - \mu_{j+h}|\mathcal{F}_{T_j}) \]
  is a Martingale; and
  \[ U_{T_j} \] is stationary, ergodic, bounded process.
MC Estimator is RV of Filtered Price

- \[ X_{Tj} = \mu_j + Y_{Tj} + U_{Tj}, \]

- MC estimator is basically the realized variance of \( Y_{Tj} \),

- Consistent with a Gaussian limit distribution under appropriate assumption.

Log-correction

- \[ \text{MC} = \frac{nx' \hat{\Lambda} \hat{\pi} (2\hat{Z} - I)x}{\frac{1}{n} \sum_{j=1}^{n} X_{Tj}^2}. \]
Financial Crisis: Markov Chain Estimates USD/YEN
USD/Yen: September 8, 2008 (Monday)

- Volatility (annualized)
- USD/Yen exchange rate

Fannie Mae/Freddie Mac Conservatorship

Hour (EST)
USD/Yen: September 12, 2008 (Friday)

- **Volatility (annualized)**
- **USD/Yen exchange rate**

**Hour (EST)**

17 18 19 20 21 22 23 00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16

Volatility and USD/Yen exchange rate chart showing fluctuations over time.
USD/Yen: September 16, 2008 (Tuesday)

- Volatility (annualized)
- USD/Yen exchange rate

AIG opens 60% down
AIG credit downgraded
AIG $85 billion bailout
USD/Yen: September 17, 2008 (Wednesday)

Volatility (annualized)

USD/Yen exchange rate

AIG $85 billion bailout.
USD/Yen: September 23, 2008 (Tuesday)

- Volatility (annualized)
- USD/Yen exchange rate

Hour (EST)
Financial Crisis: Markov Chain Estimates

SPY
September 29-30, 2008
High-Frequency Data and Forecasting

- **HF data improves...**
  - understanding of **volatility dynamics** – key for forecasting.
  - understanding of the **driving forces** of volatility.
    For instance: Study of news announcements and their effect on the financial markets.
  - **evaluation** of models/forecast

- **Realized measures...**
  - **good predictors** of future volatility.
  - led to **new and better volatility models**... yield more accurate forecasts.
  - facilitate **better estimation of complex volatility models**.
Realized GARCH Models
joint with
Albert Huang and Howard Shek
ARCH and GARCH Models

- ARCH-type models (Engle, 1982)

\[ \mu_t \equiv E_{t-1}(r_t) \quad h_t \equiv \text{var}_{t-1}(r_t). \]

Studentized returns

\[ z_t = \frac{r_t - \mu_t}{\sqrt{h_t}} \sim (0, 1) \]

- GARCH(1,1) (Bollerslev, 1986)

\[ h_t = \omega + \alpha r_{t-1}^2 + \beta h_{t-1}. \]

\[ \beta \approx 0.95 \text{ and } \alpha \approx 0.05 \text{ in practice.} \]
GARCH Model

- Squared returns, $r_{t-1}^2$, defines the dynamic of the conditional variance
  \[ h_t = \omega + \alpha r_{t-1}^2 + \beta h_{t-1}. \]

- $r_t^2$ can be viewed as a noisy measure of volatility
- Realized measures provide more accurate measurements of volatility.
Engle (2002), and many others

\[ h_t = \omega + \alpha r_{t-1}^2 + \beta h_{t-1} + \gamma x_{t-1}. \]

- \( x_t \) is a realized measure of volatility (e.g. RV)
- Huge improvement in the empirical fit.
- Typically
  - \( \hat{\gamma} \approx 0.5. \)
  - \( \hat{\alpha} \approx 0. \) (ARCH parameter becomes insignificant)
GARCH Model

- Re-write GARCH(1,1) equation:
  \[ h_t = \omega + \pi h_{t-1} + \alpha (r_{t-1}^2 - h_{t-1}) \]

- \( \pi = \alpha + \beta \) measures how persistent is volatility.
- \( \alpha \approx \) “the strength of the signal \( r_{t-1}^2 \)”
- \( \beta \approx 0.95 \) and \( \alpha \approx 0.05 \) in practice.
GARCH is Slow

![Graph showing the comparison between 'New Volatility', 'GARCH', and 'Old Volatility'. The 'New Volatility' line is constant at around 40%, while 'GARCH' and 'Old Volatility' show a slower and more gradual increase over time.](image-url)
GARCH-X with a Realized Measure is Fast
GARCH-X is Incomplete

- Data \((r_t, x_t)\), but model only specifies \(r_t | r_{t-1}, x_{t-1}, \ldots\)
- Simple case

\[
\begin{align*}
  r_t &= \sqrt{h_t} z_t. \\
  h_t &= \omega + \alpha r_{t-1}^2 + \beta h_{t-1} + \gamma x_{t-1}.
\end{align*}
\]

- Need a Model for \(x_t\).
Realized GARCH
Realized GARCH: Simple Case

GARCH-X structure

\[
\begin{align*}
  r_t &= \sqrt{h_t} z_t \\
  h_t &= \omega + \alpha r_{t-1}^2 + \beta h_{t-1} + \gamma x_{t-1}
\end{align*}
\]

Measurement Equation completes the model

\[
x_t = \xi + \varphi h_t + \text{Error}_t.
\]

- \(x_t\) is noisy measurement of \(QV_t\)
- \(QV_t\) is \(h_t + \) volatility shock.
Logarithmic Specification

- Logarithmic specification is preferred

\[ \log h_t = \omega + \beta \log h_{t-1} + \gamma \log x_{t-1}. \]
\[ \log x_t = \xi + \varphi \log h_t + \tau(z_t) + u_t. \]

- Leverage Function:

\[ \tau(z) = \tau_1 z + \tau_2(z^2 - 1) \]

- Captures the joint dependence between
  - return shocks, \( z_t \)
  - volatility shocks, \( \tau(z_t) + u_t. \)
Key Features of Realized GARCH

- **Empirical Features**
  - **Easy** to estimate.
  - Captures return-volatility dependence (**leverage effect**).
  - Properties of multiperiod returns (**skewness and kurtosis**)
  - Outperforms conventional GARCH

- **Theoretical Features** (elegant mathematical structure)
  - Parsimonious
  - Tractable analysis (**quasi maximum likelihood**).
  - Induced **simple ARMA structure** for both $x$ and $h$

- Natural extension of conventional GARCH
Dow Jones Industrial Average stocks and SPY (ETF).
- 2002-01-01 to 2007-12-31 as in-sample data and
- 2008-01-01 to 2008-08-31 as out-of-sample.

For $x$, we use the realized kernel (RK) by BHLS (2008)
- $x_t \approx h_t$ with open-to-close returns
- $x_t < h_t$ (on average) with close-to-close returns
GARCH Equation

\[ h_t = 0.09 + 0.29 h_{t-1} + 0.63 x_{t-1} \]

(0.05) (0.16) (0.18)

Measurement Equation

\[ x_t = -0.05 + 1.01 h_t - 0.02 z_t + 0.06 (z_t^2 - 1) + u_t \]

(0.09) (0.19) (0.02) (0.01)

\[ \tau(z) \]

Standard deviation of \( u_t \): \( \hat{\sigma}_u = 0.51. \)

(0.05)
Linear Model (SPY Close-to-Close)

- **GARCH Equation**
  
  \[ h_t = 0.07 + 0.29 h_{t-1} + 0.87 x_{t-1} \]

- **Measurement Equation**
  
  \[ x_t = +0.00 + 0.74 h_t - 0.07 z_t + 0.03 (z_t^2 - 1) + u_t \]

- Standard deviation of \( u_t \): \( \hat{\sigma}_u = 0.51 \).
Log-Linear Model (SPY Open-to-Close)

- **GARCH Equation**

\[
\log h_t = 0.04 + 0.70 \log h_{t-1} + 0.45 \log x_{t-1} - 0.18 \log x_{t-2}
\]

- **Measurement Equation**

\[
\log x_t = -0.18 + 1.04 \log h_t - 0.07 z_t + 0.07 (z_t^2 - 1) + u_t
\]

- **Persistence Parameter**

\[
\hat{\pi} = \hat{\beta} + (\hat{\gamma}_1 + \hat{\gamma}_2)\hat{\phi} = 0.986
\]
Log-Linear Model (SPY Close-to-Close)

- **GARCH Equation**
  \[
  \log h_t = 0.11 + 0.72 \log h_{t-1} + 0.48 \log x_{t-1} - 0.21 \log x_{t-2}
  \]

- **Measurement Equation**
  \[
  \log x_t = -0.42 + 1.00 \log h_t + -0.11 z_t + 0.04 (z_t^2 - 1) + u_t
  \]

- **Standard deviation of** \( u_t \): \( \hat{\sigma}_u = 0.38 \).

- **Persistence Parameter**
  \[
  \hat{\pi} = \hat{\beta} + (\hat{\gamma}_1 + \hat{\gamma}_2) \hat{\phi} = 0.987
  \]
Estimated News Impact Curve
Skewness and Kurtosis of Cumulative Returns

Figure: Skewness and kurtosis in line with empirical returns
Multi-Period Forecast

- Multi-period ahead predictions with the Realized GARCH model is straightforward.
- When \( p = q = 1 \), we obtain VARMA(1,1) structure

\[
\begin{bmatrix}
\tilde{h}_t \\
\tilde{x}_t
\end{bmatrix} = \begin{bmatrix}
\beta & \gamma \\
\varphi \beta & \varphi \gamma
\end{bmatrix} \begin{bmatrix}
\tilde{h}_{t-1} \\
\tilde{x}_{t-1}
\end{bmatrix} + \begin{bmatrix}
\omega \\
\xi + \varphi \omega
\end{bmatrix} + \begin{bmatrix}
0 \\
\tau(z_t) + u_t
\end{bmatrix},
\]

so we can write, \( Y_t = AY_{t-1} + \mu + \epsilon_t \).
Realized GARCH Volatility during the Global Financial Crisis
Realized GARCH: Global Financial Crisis

Volatility by Realized EGARCH Model

- Freddie Mac: Purchase of the most risky subprime mortgages to be suspended
- Big subprime related writedowns by Citigroup, Merrill Lynch, Morgan Stanley, etc.
- Major drop in stock prices worldwide
- AIG $62bn losses unveiled
- Citigroup
- Liquidity Crunch. Hedge funds suffer big losses
- Lehman AIG
- Bank of America
- Dubai Crisis
- 2 Bear Stearns hedge funds liquidated
- Sharp increase in unemployment Signs of recession
- IndyMac collapse
- Bear Stearns collapse

Figure: Conditional volatility during the global financial crisis with some of the major events.
Extensions

Realized GARCH
Realized EGARCH with Multiple RM (w. Huang)

- Multiple Realized Measures (like MEM by Engle & Gallo)
  
  \[ \log h_t = \omega + \beta \log h_{t-1} + \gamma' \log X_{t-1} + \tau(z_{t-1}) \]
  
  \[ \log X_t = \xi + \varphi \log h_t + \delta(z_t) + U_t. \]

- RK crowds out RV and Range (High minus Low)
- \( \text{\texttt{var}}(U_t) \) yields information about accuracy about RMGs.
Build Realized GARCH for Market Returns, \( r_{0,t} = \mu_0 + \sqrt{h_{0,t}} z_{0,t} \).

Asset \( i \)'s returns, \( r_{i,t} = \mu_i + \sqrt{h_{i,t}} z_{i,t} \), conditional on \( r_{0,t}, x_{0,t} \).

Key in this model:

\[
\rho_t = \text{cov}(z_{0,t}, z_{i,t}|\mathcal{F}_{t-1}).
\]

Measurement equation with Fisher transform

\[
F(\rho_t) = a_{0i} + b_{0i} F(\rho_{t-1}) + c_{0i} F(y_{1,t-1}).
\]

1-factor structure where we can extract

\[
\beta_t = \rho_t \sqrt{h_{i,t}/h_{0,t}}.
\]
Correlation Measurement Equation (CVX)

\[
F(\rho_t) = \frac{\hat{a}_{01}}{0.01} + \frac{\hat{b}_{01}}{0.01} + \frac{\hat{c}_{01}}{0.02} F(\rho_t) + 0.267 F(y_{1t}).
\]

\[
F(y_{1t}) = -\frac{\hat{\psi}_{01}}{0.01} + \frac{\hat{\psi}_{01}}{0.05} F(\rho_t) + \nu_{1t}.
\]

and the covariance structure for the error terms in the three measurement equations

\[
\hat{\Sigma} = \begin{bmatrix}
0.148 & 0.088 & 0.027 \\
0.088 & 0.157 & 0.021 \\
0.027 & 0.021 & 0.029
\end{bmatrix}
\]

\[\text{Persistence} = b + \phi \cdot c = 0.97964\]
Time Series for $\rho_t$ and $\beta_t$ (CVX)
Cross Sectional Beta-Quantiles
Conclusion

- Realized Measures
  - Empirical Issues with High-Frequency Data
  - Realized Variance, Realized Kernel, Markov Chain Estimator

- Volatility Forecasting
- Realized GARCH
  - Easy to estimate, Tractable, can explain empirical “stylized facts”.
  - Multivariate extension very promising.


Jacod, J., 1994. Limit of random measures associated with the increments of a Brownian semimartingalePreprint number 120, Laboratoire de Probabilités, Université Pierre et Marie Curie, Paris.