Quasi-Maximum Likelihood Estimation of Volatility with High Frequency Data

Dacheng Xiu

Bendheim Center for Finance

Princeton University

Workshop on Financial Econometrics, Fields Institute

April 24, 2010

Outline

1. Objective
   - Estimate (ex post) daily realized volatility ($\int_0^T \sigma_t^2 dt$).
     - Continuous time model of the tick-by-tick data.
     - Alternative to lower frequency models (GARCH, Heston...)
       Historical changes, Restrictive assumptions, Inconvenient...

2. Dilemma
   - Microstructure noise ruins the RV estimator $\sum(X_{i+1} - X_i)^2$, when using full sample. Empirically,

$$\sum(X_{i+1} - X_i)^2 \longrightarrow \infty, \text{as } \Delta \to 0.$$

   - Suffering information loss if we sample every 5 minutes (99.7%).

1. Objective
   - Estimate (ex post) daily realized volatility ($\int_0^T \sigma_t^2 dt$).
     - Continuous time model of the tick-by-tick data.
     - Alternative to lower frequency models (GARCH, Heston...)
       Historical changes, Restrictive assumptions, Inconvenient...
   - Estimate the variance of the noise.
     - Transaction cost: Roll (1984)
2. Dilemma
   - Microstructure noise ruins the RV estimator $\sum(X_{i+1} - X_i)^2$, when using full sample. Empirically,

$$\sum(X_{i+1} - X_i)^2 \longrightarrow \infty, \text{as } \Delta \to 0.$$

   - Suffering information loss if we sample every 5 minutes (99.7%).

Parametric Model

- Aït-Sahalia, Mykland and Zhang(2005)

1. The logarithm of the transaction price $\tilde{X}$ is observed at $0 = \tau_0 < \tau_1 < ... < \tau_n = T$, such that $\tilde{X}_{\tau_i} = X_{\tau_i} + U_{\tau_i}$.
2. The latent process X_t satisfies

$$dX_t = \sigma dW_t$$

3. The microstructure noise U_{τ_i} is i.i.d. $N(0, a^2)$, $\perp \{W_t\}$, and $EU_{\tau_i}^4 < \infty$.
4. The time intervals $\Delta_i = \tau_i - \tau_{i-1} = \bar{\Delta}$.

Therefore, $Y_i = \tilde{X}_{\tau_i} - \tilde{X}_{\tau_{i-1}}$ has an MA(1) structure.

Likelihood:

$$l(a^2,\sigma^2) = -\log\det(\Omega)/2 - n\log(2\pi)/2 - Y'\Omega^{-1}Y/2 \qquad (1)$$

where

$$\Omega = \begin{pmatrix} \sigma^2\Delta + 2a^2 & -a^2 & 0 & \cdots & 0 \\ -a^2 & \sigma^2\Delta + 2a^2 & -a^2 & \ddots & \vdots \\ 0 & -a^2 & \sigma^2\Delta + 2a^2 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & -a^2 \\ 0 & \cdots & 0 & -a^2 & \sigma^2\Delta + 2a^2 \end{pmatrix}$$

Then the MLE $(\hat{\sigma}^2, \hat{a}^2)$ proves to be consistent:

$$\begin{pmatrix} n^{\frac{1}{4}}(\hat{\sigma}^2 - \sigma_0^2) \\ n^{\frac{1}{2}}(\hat{a}^2 - a_0^2) \end{pmatrix} \xrightarrow{\mathcal{L}} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 8a_0\sigma_0^3 T^{-\frac{1}{2}} & 0 \\ 0 & 2a_0^4 \end{pmatrix}\right)$$

Pros and Cons of MLE

Theory vs Practice

- In Theory:
  - Parametric Estimator
  - Constant Volatlity
- In Practice:
  - pretty good!
    - Simulation Aït-Sahalia and Yu (2009), Hansen, Large and Lunde (2008)
    - Comparison Gatheral and Oomen (2007)
  - but why?

    "The interesting situation is the one with nonconstant volatility...leading to our conjecture that the consistency continues to hold when... The asymptotic normality also seems to hold under nonconstant volatility; however, the asymptotic variance is far more complicated due to autocorrelation in the score function." -Hansen, Large and Lunde (2008)

So, it is natural to ask:

- What is the impact of stochastic volatility on the MLE?
- Will the MLE remain consistent and robust?
- What are the convergence rate and asymptotic distribution of the MLE?
- How does it compare with other alternative nonparametric estimators?
- How to rationalize this approach?

Model Setup and Assumptions

1. $\tilde{X}_{\tau_i} = X_{\tau_i} + U_{\tau_i}$.
2. The latent process X_t satisfies

$$dX_t = \mu_t dt + \sigma_{t-} dW_t$$

   with μ_t locally bounded, σ_t a positive, locally bounded Itô semimartingale. e.g.

$$d\sigma_t^2 = \kappa(\bar{\sigma}^2 - \sigma_t^2)dt + \gamma(\sigma_t^2)dB_t + \beta(\sigma_{t-})J_t dN_t$$

3. The microstructure noise U_{τ_i} is i.i.d., $\perp \{W_t\}$ and $\{\sigma_t\}$ $EU_{\tau_i} = 0$, and $EU_{\tau_i}^4 < \infty$.
4. The time intervals Δ_is satisfy i.i.d. with mean $\bar{\Delta}$ and $\perp \{X_t\}$, $\{\sigma_t\}$ and $\{U_t\}$, so $n_T = O_P(\bar{\Delta}^{-1})$ and $n = T/\bar{\Delta}$.

Review of Prior Works

Biased!

Popular Estimators include but not limited to

- Two/ Multi Scales Realized Volatility
  - Zhang, Mkyland and Aït-Sahalia (2005)
  - Zhang (2006)
- Realized Kernels
  - Barndorff-Nielsen, Hansen, Lunde and Shephard (2008)
- Pre-Averaging Method
  - Jacod, Li, Mykland, Podolskij, and Vetter (2007)
- ...

Intuition

- The parameter of interest $(\frac{1}{T})\int_0^T \sigma_t^2 dt$, happens to be the average of the volatility process.
- No microstructure noise case:

$$l(\omega, \sigma^2) = -\frac{n}{2}\log(\sigma^2\Delta) - \frac{n}{2}\log(2\pi) - \frac{1}{2\sigma^2\Delta}Y'Y$$

  where $Y_i = X_{\tau_i} - X_{\tau_{i-1}} = \int_{\tau_{i-1}}^{\tau_i} \sigma_t dW_t$. Apparently, the MLE is

$$\hat{\sigma}^2 = \frac{1}{T}\sum_{i=1}^{n} Y_i^2 = \frac{1}{T}\sum_{i=1}^{n}(X_{\tau_i} - X_{\tau_{i-1}})^2 \xrightarrow{\mathcal{P}} \frac{1}{T}\int_0^T \sigma_t^2 dt$$

  Here, the RV estimator recurs.
- Volatility will not change too much in a small period?

Intuition

- The parameter of interest $(\frac{1}{T})\int_0^T \sigma_t^2 dt$, happens to be the average of the volatility process.
- No microstructure noise case:

$$l(\omega, \sigma^2) = -\frac{n}{2}\log(\sigma^2\Delta) - \frac{n}{2}\log(2\pi) - \frac{1}{2\sigma^2\Delta}Y'Y$$

where $Y_i = X_{\tau_i} - X_{\tau_{i-1}} = \int_{\tau_{i-1}}^{\tau_i} \sigma_t dW_t$. Apparently, the MLE is

$$\hat{\sigma}^2 = \frac{1}{T}\sum_{i=1}^{n} Y_i^2 = \frac{1}{T}\sum_{i=1}^{n}(X_{\tau_i} - X_{\tau_{i-1}})^2 \xrightarrow{\mathcal{P}} \frac{1}{T}\int_0^T \sigma_t^2 dt$$

Here, the RV estimator recurs.

- Volatility will not change too much in a small period? No!
- Quadratic Representation

Review of the QMLE

Suppose the true data generating distribution is $g(U)$.

- Possibly Misspecified Assumption

$$g(U) \in \{f(\theta, U) : \theta \in \Theta\}, \text{i.e.} \exists\ \theta_0 \in \Theta, \text{s.t.} g(U) = f(\theta_0, U)$$

- $\hat{\theta}$ maximizes the log likelihood:

$$f(\theta, U) = \sum_{i=1}^{n} \log f(\theta, U_i)$$

Review of the QMLE

Suppose the true data generating distribution is $g(U)$.

- Possibly Misspecified Assumption

$$g(U) \in \{f(\theta, U) : \theta \in \Theta\}, \text{i.e.} \exists\ \theta_0 \in \Theta, \text{s.t.} g(U) = f(\theta_0, U)$$

- $\hat{\theta}$ maximizes the log likelihood:

$$f(\theta, U) = \sum_{i=1}^{n} \log f(\theta, U_i)$$

- θ^* minimizes the Kullback - Leibler Information Criterion:

$$I(g : f, \theta) = E(\log[g(U)/f(U, \theta)])$$

Under some conditions, see White (1982)

$$\hat{\theta} \xrightarrow{\mathrm{P}} \theta^*$$

Remark on the QMLE

- Roughly speaking, the rationale behind it is the law of large numbers. Because

$$\frac{1}{n}\sum_{i=1}^{n} \log f(\theta, U_i) \xrightarrow{\mathrm{P}} E(\log f(\theta, U))$$

  Under some conditions, their maximizers should be close to each other.
- Correctly Specified Model: $\theta^* = \theta_0$
- Misspecified Model: Intuitively, θ^* minimizes one's ignorance about the true structure.

Theorem 1: Consistency of the QMLE

Let $Q_n(\omega, \theta)$ and $\bar{Q}_n(\omega, \theta)$ be two random functions such that for each θ in Θ, a compact subset of R^k, they are measurable functions on Ω and, for each $\omega \in \Omega$, continuous functions on Θ. In addition, $\bar{Q}_n(\omega, \theta)$ is almost surely maximized at $\theta_n^*(\omega)$. Further, as $n \to \infty$:

1. Uniform Convergence:

$$\sup_{\theta \in \Theta} \| Q_n(\omega, \theta) - \bar{Q}_n(\omega, \theta) \| \xrightarrow{\mathrm{P}} 0.$$

2. Identifiability: for every $\epsilon > 0$, there exists a constant $\delta_0 > 0$, such that

$$P(\bar{Q}_n(\omega, \theta_n^*) - \max_{\theta \in \Theta : \|\theta - \theta_n^*\| \geq \epsilon} \bar{Q}_n(\omega, \theta) > \delta_0) \to 1.$$

Then any sequence $\hat{\theta}_n$ s.t. $Q_n(\omega, \hat{\theta}_n) \geq \sup_{\theta \in \Theta} Q_n(\omega, \theta) + o_P(1)$ converges in probability to θ_n^*, i.e., $\hat{\theta}_n - \theta_n^* \xrightarrow{\mathrm{P}} 0$.

Theorem 2: Consistency of the Quasi-M-Estimators

Let $\Psi_n(\omega,\theta)$and $\bar{\Psi}_n(\omega,\theta)$ be random vector-valued functions. For each θ in Θ, a compact subset of $\mathbb{R}^k$, they are measurable function on Ω, and for each ω in Ω, continuous functions on Θ. In addition, there exists a sequence of θ_n^*, satisfying $\bar{\Psi}_n(\omega,\theta_n^*) = 0$ almost surely, such that as $n \to \infty$,

1. Uniform Convergence:

$$\sup_{\theta\in\Theta} \|\Psi_n(\omega,\theta) - \bar{\Psi}_n(\omega,\theta)\| \xrightarrow{\mathrm{P}} 0.$$

2. Identifiability: For every $\epsilon > 0$, there exists a constant $\delta_0 > 0$, such that,

$$P(\min_{\theta\in\Theta:\|\theta-\theta_n^*\|\geq\epsilon} \|\bar{\Psi}_n(\omega,\theta)\| > \delta_0) \to 1.$$

Then any sequence of estimators $\hat{\theta}_n$ such that $\Psi_n(\omega,\hat{\theta}_n) = o_P(1)$ converges in probability to θ_n^*, i.e., $\hat{\theta}_n - \theta_n^* \xrightarrow{\mathrm{P}} 0$.

Theorem 3: CLT of the Quasi-M-Estimators

Suppose that the conditions of Consistency Theorem are satisfied. In addition, $\Psi_n(\omega, \theta)$ and $\bar{\Psi}_n(\omega, \theta)$ are continuously differentiable of order 1 on Θ. Also, there exists a sequence of positive definite matrices $\{V_n(\omega)\}$ such that

$$V_n(\omega)(\Psi_n(\omega, \hat{\theta}_n) - \Psi_n(\omega, \theta_n^*)) \xrightarrow{\mathcal{L}} N(0, I_k)$$

If $\nabla \bar{\Psi}_n(\omega, \theta)$ is stochastic equicontinuous, and $|\nabla \Psi_n(\omega, \theta) - \nabla \bar{\Psi}_n(\omega, \theta)| \xrightarrow{\mathrm{P}} 0$, uniformly for all $\theta \in \Theta$, then

$$V_n(\omega) \nabla \bar{\Psi}_n(\omega, \theta_n^*)(\hat{\theta}_n - \theta_n^*) \xrightarrow{\mathcal{L}} N(0, I_k)$$

Including the Market Microstructure Noise

1. Misspecification
   - $\sigma_t = \sigma$
   - $U \sim N(0, a^2)$.
2. The log likelihood function is

$$l(a^2, \sigma^2) = -\log\det(\Omega)/2 - n\log(2\pi)/2 - Y'\Omega^{-1}Y/2$$

$$\Omega = \begin{pmatrix} \sigma^2\Delta + 2a^2 & -a^2 & 0 & \cdots & 0 \\ -a^2 & \sigma^2\Delta + 2a^2 & -a^2 & \ddots & \vdots \\ 0 & -a^2 & \sigma^2\Delta + 2a^2 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & -a^2 \\ 0 & \cdots & 0 & -a^2 & \sigma^2\Delta + 2a^2 \end{pmatrix}$$

where Y_is are observed log returns.

Main Theorem

1. Consistency:
   - $\hat{a}_n^2 - a_0^2 = O_P(n^{-\frac{1}{2}})$
   - $\hat{\sigma}_n^2 - \frac{1}{T}\int_0^T \sigma_t^2 dt = O_P(n^{-\frac{1}{4}})$
2. Central Limit Theorem (stable convergence):

$$\begin{pmatrix} n^{\frac{1}{4}}(\hat{\sigma}^2 - \frac{1}{T}\int_0^T \sigma_t^2 dt) \\ n^{\frac{1}{2}}(\hat{a}^2 - a_0^2) \end{pmatrix} \xrightarrow{\mathcal{L}\mathcal{X}}$$

$$N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{5a_0 \int_0^T \sigma_t^4 dt}{T(\int_0^T \sigma_t^2 dt)^{\frac{1}{2}}} + \frac{3(\int_0^T \sigma_t^2 dt)^{\frac{3}{2}} a_0}{T^2} & 0 \\ 0 & 2a_0^4 + \mathrm{cum}_4[U] \end{pmatrix} \right)$$

Proof

3. Robustness:
   - Non-Gaussian and Serial Dependent Microstructure Noise
   - Jumps in Prices: $\hat{\sigma}^2 \xrightarrow{\mathcal{P}} \frac{1}{T}(\int_0^T \sigma_t^2 dt + \sum_{0\le t\le T}(\Delta X_t)^2)$
   - Random Intervals

QMLE vs RKs: Estimation Methods

- Parametric vs Nonparametric
  Realized Kernels (RKs):

$$K(\tilde{X}_\tau) = \gamma_0(\tilde{X}_\tau) + \sum_{h=1}^{H} k(\frac{h-1}{H})(\gamma_h(\tilde{X}_\tau) + \gamma_{-h}(\tilde{X}_\tau))$$

  where

$$\gamma_h(\tilde{X}_\tau) = \sum_{j=1}^{n}(\tilde{X}_{\tau_j} - \tilde{X}_{\tau_{j-1}})(\tilde{X}_{\tau_{j-h}} - \tilde{X}_{\tau_{j-h-1}})$$

- Bandwidth and Kernel Selection: burden vs flexibility
  1. rule-of-thumb approximation / ad hoc ways
  2. sensitivity

QMLE vs RKs: Asymptotic and Finite Sample

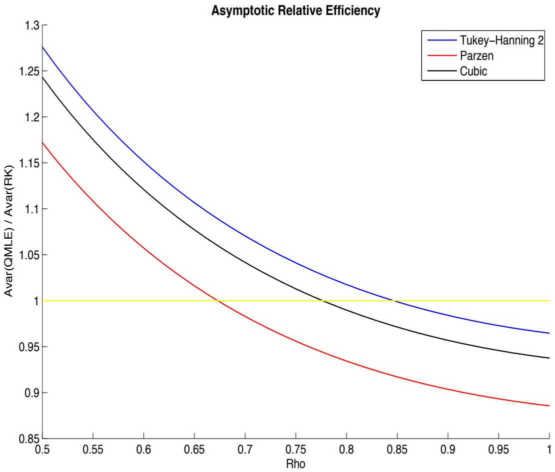
- Asymptotic Behavior
  - Constant Volatility: the QMLE becomes the MLE
  - Stochastic Volatility: depending on heteroskedasticity $\rho = \int_0^T \sigma_u^2 du / \sqrt{T \int_0^T \sigma_u^4 du}$

$$\frac{Avar(\mathrm{RK})}{Avar(\mathrm{QMLE})} = \frac{16\sqrt{\rho k_0 k_1}}{3} \frac{\left(1+\sqrt{1+3k_0k_2/\rho k_1^2}\right)^{-\frac{1}{2}} + \left(1+\sqrt{1+3k_0k_2/\rho k_1^2}\right)^{\frac{1}{2}}}{5\rho^{-\frac{1}{2}} + 3\rho^{\frac{3}{2}}}$$

- Finite Sample Performance
  - Infeasible RK: require out-of-period data
  - Edge Effect

$$\tilde{\gamma}_{\pm h}(\tilde{X}_\tau) = \sum_{j=|H|+1}^{n-H} (\tilde{X}_{\tau_j} - \tilde{X}_{\tau_{j-1}})(\tilde{X}_{\tau_{j-h}} - \tilde{X}_{\tau_{j-h-1}})$$

QMLE vs RKs: Relative Efficiency Plot



QMLE vs RKs: Quadratic Representation and Weighting Matrices

- Quadratic iterative representation of the QMLE:
  The QMLE $(\hat{\sigma}^2, \hat{a}^2)$ satisfies the following equations:

$$\begin{aligned} \hat{\sigma}^2 T &= Y' W_1 Y \\ \hat{a}^2 &= Y' W_2 Y \end{aligned}$$

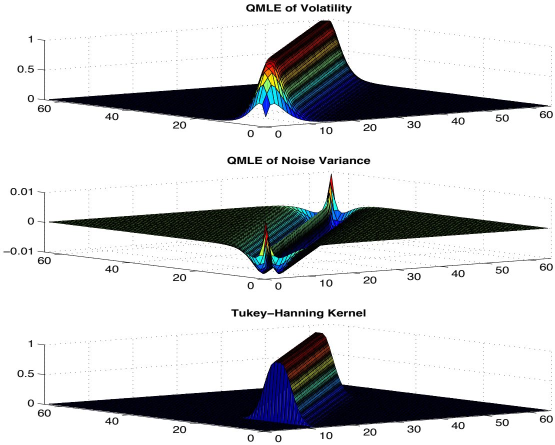
$$\begin{aligned} W_1 &= \frac{n \cdot tr(\Omega^{-2}\Lambda) \cdot \Omega^{-1}\Lambda\Omega^{-1} - n \cdot tr(\Omega^{-2}\Lambda^2) \cdot \Omega^{-2}}{(tr(\Omega^{-2}\Lambda))^2 - tr(\Omega^{-2}) \cdot tr(\Omega^{-2}\Lambda^2)} \\ W_2 &= \frac{tr(\Omega^{-2}\Lambda) \cdot \Omega^{-2} - tr(\Omega^{-2}) \cdot \Omega^{-1}\Lambda\Omega^{-1}}{(tr(\Omega^{-2}\Lambda))^2 - tr(\Omega^{-2}) \cdot tr(\Omega^{-2}\Lambda^2)} \end{aligned}$$

  W_1 and W_2 depend on σ^2 and a^2 only through $\lambda = a^2/\sigma^2 T$.
- The feasible TH_2 kernel estimator can be expressed as

$$K(\tilde{X}_\tau) = Y' W Y$$

$$\begin{aligned} W_{i,i} &= 1_{\{1+H \le i \le n-H\}} \\ W_{i,j} &= k\left(\frac{|i-j|-1}{H}\right) \cdot 1_{\{1 \le |i-j| \le H\}} \cdot 1_{\{1+H \le j \le n-H\}} \end{aligned}$$

QMLE vs RKs: Weighting Matrices Plots



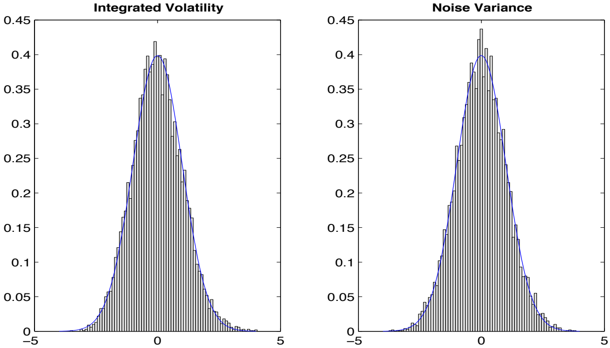
We fix T as 1 day. Within $[0, T]$, the data are simulated using Euler scheme based on stochastic volatility models, for instance the Heston Model with jumps in volatility process.

$$\begin{aligned} dX_t &= \mu dt + \sigma_{t-} dW_t \\ d\sigma_t^2 &= \kappa(\bar{\sigma}^2 - \sigma_t^2)dt + \delta\sigma_{t-} dB_t + \sigma_{t-} J_t^V dN_{2t} \end{aligned}$$

where $E(dW_t \cdot dB_t) = \rho dt$.

- $\mu = 0.03$, $\rho = -0.75$, $\kappa = 5$, $\delta = 0.4$, $a_0 = 0.5\%$.
- $J_t^V = \exp(z)$, where $z \sim N(-5, 1)$, $\lambda = 12$.
- The arrival of transactions $\{\tau_i\}$ follows a Poisson process with mean=1 sec.

Histogram of the Standardized Estimates



Comparisons with RKs: Simulation Design

The implementation of the Tukey-Hanning$_2$ (TH$_2$) kernel

$$k(x) = \sin^2\left(\frac{\pi}{2}(1-x)^2\right)$$

1. RK$_1$ as a benchmark: infeasible bandwidth + out-of-period data: $H = c^* \xi n^{\frac{1}{2}}$ with c^* given by

$$c^* = \sqrt{\rho \frac{k_1}{k_0}\left(1 + \sqrt{1 + \frac{3k_0 k_2}{\rho k_1^2}}\right)}$$

2. RK$_2$: infeasible bandwidth + edge effect
3. RK$_3$: rule-of-thumb bandwidth + out-of-period data

$$\hat{H} = 5.74 \hat{a} \sqrt{n / RV_{10}(\tilde{X})}$$

   where $\hat{a} = \sqrt{RV(\tilde{X})/2n}$ and $RV_{10}(\tilde{X})$ is the RV estimator based on 10-min returns.
4. RK$_4$: rule-of-thumb bandwidth + edge effect

Comparisons with RKs: Results

Table: This table reports the estimates for $100 \cdot (\hat{\sigma}^2 - \frac{1}{T}\int_0^T \sigma_t^2 dt)$.

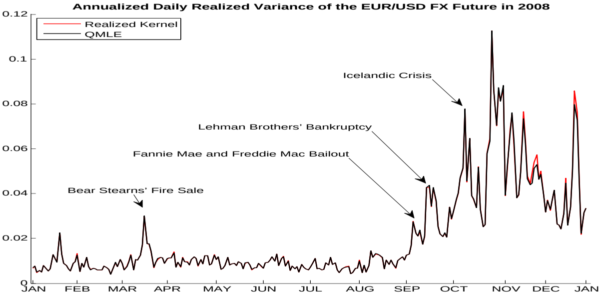
		1 sec	5 sec	30 sec	1 min	3 min
QMLE	Bias	0.0189	0.0480	-0.0341	-0.0452	-0.0838
	RMSE	1.1861	1.8178	2.9272	3.5831	4.9210
RK_1	Bias	0.0178	0.0465	-0.0079	-0.0192	0.0003
	RMSE	1.2329	1.8850	3.0087	3.6851	5.0185
RK_2	Bias	-0.1620	-0.3659	-1.0288	-1.4913	-2.5874
	RMSE	1.2305	1.8827	3.0492	3.7345	5.1255
RK_3	Bias	0.0186	0.0247	-0.0452	0.0009	0.3615
	RMSE	1.8604	2.7556	4.1697	4.8708	6.2360
RK_4	Bias	-0.0641	-0.1568	-0.4934	-0.6316	-0.7731
	RMSE	1.8557	2.7337	4.1066	4.7863	6.0173

Comment:

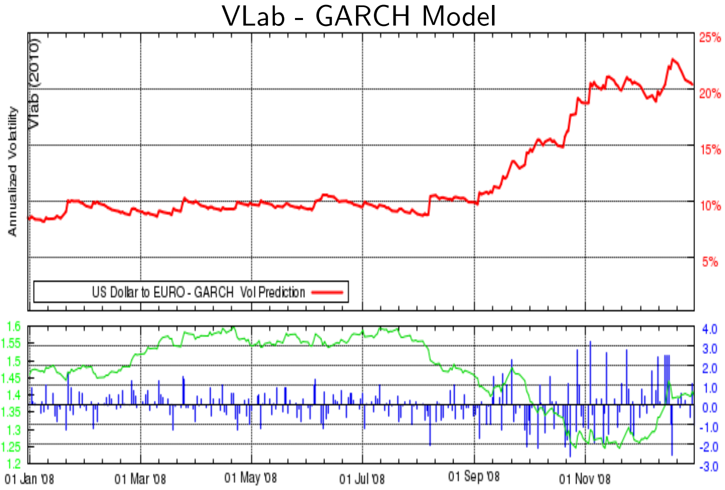
- QMLE dominates.
- The edge effect bias is large, but negligible in large sample.

Empirical Work with the Euro/US Dollar Future Prices

1. Goal: Estimate the Daily Realized Variance $\frac{1}{T}\int_0^T \sigma_t^2 dt$.
2. Data: Euro/Dollar FX Future on CME in 2008.



Empirical Work with the Euro/US Dollar Future Prices



Take Home Message

1. This paper proposes
   - A QMLE, which
     - is Parametric and Free of Tuning-Parameters
     - is Consistent and Rate Efficient
     - has No Edge Effect
     - has a Quadratic Iterative Representation
   - A Framework, which
     - can Deal with Stochastic Parameters using Model Misspecification
     - Links Parametric Approach to Nonparametric Methods
   - An Empirical Study with Euro/Dollar Future, which
     - Uncovers Realized Volatility Trajectory in the FX Market
     - Identifies Abnormal Volatility Movements with News Impact
2. Future work
   - Covariance/Correlation: Aït-Sahalia, Fan and Xiu (2009)

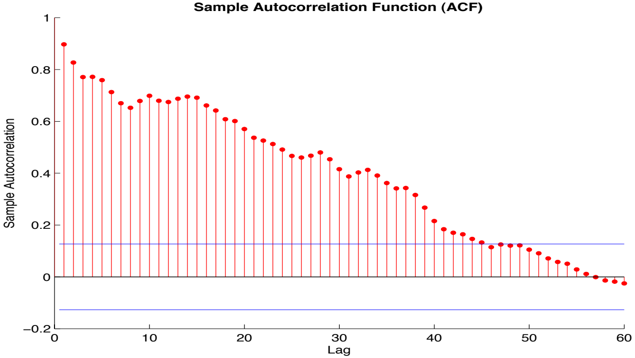
Part II

Thanks!

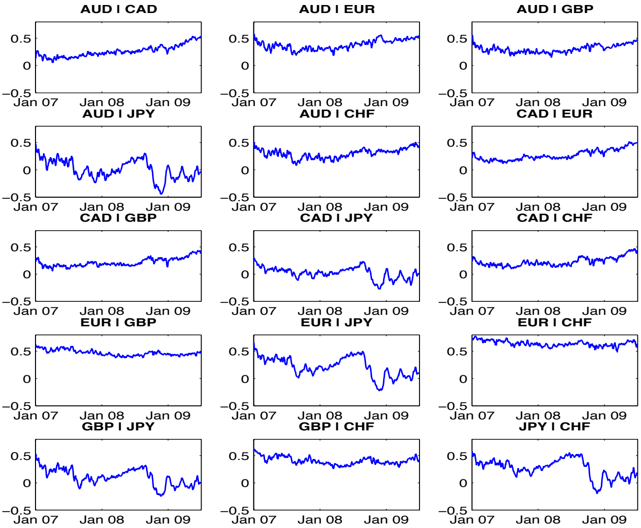
More on Empirical Work

Table: Summary Statistics

Avg No of Obs	Avg Freq	Mean	Std Err	1st Lag	2nd Lag
19550	4.42s	-8.83e-09	6.95e-05	-0.073	0.0091



More on Empirical Work



Step 1: Applying Theorem 2

$$\Psi_n = (\Psi_n^1(\omega,\theta), \Psi_n^2(\omega,\theta))' = \left(-\frac{1}{\sqrt{n}}\frac{\partial l(a^2,\sigma^2)}{\partial \sigma^2}, -\frac{1}{n}\frac{\partial l(a^2,\sigma^2)}{\partial a^2}\right)'$$
$$= \left(\frac{1}{2\sqrt{n}}\left\{\frac{\partial \ln(\det \Omega)}{\partial \sigma^2} + Y'\frac{\partial \Omega^{-1}}{\partial \sigma^2}Y\right\}, \frac{1}{2n}\left\{\frac{\partial \ln(\det \Omega)}{\partial a^2} + Y'\frac{\partial \Omega^{-1}}{\partial a^2}Y\right\}\right)'$$

$$\bar{\Psi}_n = \left(\frac{1}{2\sqrt{n}}\left\{\frac{\partial \ln(\det \Omega)}{\partial \sigma^2} + Tr\left(\frac{\partial \Omega^{-1}}{\partial \sigma^2}\Sigma_0\right)\right\}, \frac{1}{2n}\left\{\frac{\partial \ln(\det \Omega)}{\partial a^2} + Tr\left(\frac{\partial \Omega^{-1}}{\partial a^2}\Sigma_0\right)\right\}\right)'$$

$$\Sigma_0 = \begin{pmatrix} \int_0^{\tau_1}\sigma_t^2 dt + 2a_0^2 & -a_0^2 & 0 & \cdots & 0 \\ -a_0^2 & \int_{\tau_1}^{\tau_2}\sigma_t^2 dt + 2a_0^2 & -a_0^2 & \ddots & \vdots \\ 0 & -a_0^2 & \int_{\tau_2}^{\tau_3}\sigma_t^2 dt + 2a_0^2 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & -a_0^2 \\ 0 & \cdots & 0 & -a_0^2 & \int_{\tau_{n-1}}^{T}\sigma_t^2 dt + 2a_0^2 \end{pmatrix}$$

Step 2: Consistency

The proof can be divided into two steps:

1. Applying Theorem 2:
   - $\hat{\sigma}_n^2 - \sigma_n^{2*} = o_P(1)$
   - $\hat{a}_n^2 - a_n^{2*} = o_P(1)$
2. By direct calculations:
   - $\sigma_n^{2*} - \frac{1}{T}\int_0^T \sigma_t^2 dt = O_P(n^{-\frac{1}{2}})$
   - $a_n^{2*} - a_0^2 = O_P(n^{-1})$
3. Combining the two:
   - $\hat{a}_n^2 - a_0^2 = o_P(1)$
   - $\hat{\sigma}_n^2 - \frac{1}{T}\int_0^T \sigma_t^2 dt = o_P(1)$

Step 3: CLT

$$\begin{pmatrix} n^{\frac{1}{4}}(\Psi_n^1 - \bar{\Psi}_n^1) \\ n^{\frac{1}{2}}(\Psi_n^2 - \bar{\Psi}_n^2) \end{pmatrix} \xrightarrow{\mathcal{L}\mathcal{X}}$$

$$N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{1}{4}\left(\frac{5\int_0^T \sigma_t^4 dt}{16 a \sigma^7 \sqrt{T}} + \frac{a_0^2 \int_0^T \sigma_t^2 dt}{8\sigma^5 a^3 \sqrt{T}} + \frac{a_0^4 \sqrt{T}}{16 a^5 \sigma^3}\right) & 0 \\ 0 & \frac{2a_0^4 + \mathrm{cum}_4[U]}{4a^8} \end{pmatrix} \right)$$

Return