Accurate Asymptotic Approximation in the Optimal GMM Framework with Application to Stochastic Volatility Models

Yixiao Sun
UC San Diego

January 21, 2014
Approximations in the Presence of High Variation Scaling

\[ t = \frac{\hat{\beta} - \beta_0}{\sqrt{\hat{V}(\hat{\beta})}} \quad \text{and} \quad \mathbb{W} = (\hat{\beta} - \beta_0)' \hat{V}^{-1}(\hat{\beta})(\hat{\beta} - \beta_0) \]

- Quantile regression: Variance depends on the probability density function which is often estimated nonparametrically.
- Time series regression: Variance depends on the spectral density at the origin, which again is often estimated nonparametrically.
- Other type of regressions or GMM that involve dependent data: spatial data, spatiotemporal data.
- Variance heterogeneity: Clustered data with inter-cluster heterogeneity.
- General nonparametric regression/sieve estimation.
Conventionally, we just ignore it so that

\[ t = \frac{\hat{\beta} - \beta_0}{\sqrt{\hat{V}(\hat{\beta})}} \approx \frac{\hat{\beta} - \beta_0}{\sqrt{V(\hat{\beta})}} \]

\[ W = (\hat{\beta} - \beta_0)' \hat{V}^{-1}(\hat{\beta}) (\hat{\beta} - \beta_0) \approx (\hat{\beta} - \beta_0)' V^{-1}(\hat{\beta}) (\hat{\beta} - \beta_0) \]

This amounts to approximating the distribution of \( \hat{V}(\hat{\beta}) \) by the degenerate distribution concentrating on \( V(\hat{\beta}) \).

The randomness in \( \hat{V}(\hat{\beta}) \) is completely ignored.

This leads to all sorts of size distortion.

This paper develops an asymptotic approximation that captures the random variation in \( \hat{V}(\hat{\beta}) \).
Outline

1. Basic Setting and Problem
   - GMM Estimation
   - Optimal Weighting Matrix
   - Two-step Estimation and Testing

2. Fixed Smoothing Asymptotics and its Representation
   - Big Picture
   - Main Results
   - Representation and Approximation: Series HAR
   - Representation and Approximation: Kernel HAR

3. Implementation, Simulation and Empirical Application
   - Practical Implementation
   - Simulation
   - Empirical Application

4. Conclusions
The Estimation Problem

The moment conditions

$$Ef (v_t, \theta) = 0, \ t = 1, 2, ..., T$$

hold if and only if $\theta = \theta_0$.

- $f (\cdot) \in \mathbb{R}^m$ and $\theta \in \mathbb{R}^d$. $m \geq d \rightarrow$ the model may be overidentified.

- $\{ f (v_t, \theta_0) \}$ may exhibit autocorrelation of unknown forms.

- For the IV estimator: $f (v_t, \theta) = Z_t' (Y_t - X_t \theta) = 0$.

- For the MLE: $f (v_t, \theta) = \partial \log p(v_t, \theta) / \partial \theta$. 
An Example

- Stochastic Volatility Model

\[ r_t = \sigma_t Z_t \]
\[ \log \sigma_t^2 = \omega + \beta (\log \sigma_{t-1}^2 - \omega) + \sigma_u u_t \]

where \( r_t \) is the rate of return and \((Z_t, u_t)\) is iid \( N(0, I_2) \).

- \( \theta = (\omega, \beta, \sigma_u) \)

- Moment conditions:

\[ E |r_t|^\ell = c_\ell E \left( \sigma_t^\ell \right) \text{ for some constant } c_\ell \]
\[ E |r_t r_{t-j}| = 2\pi^{-1} E (\sigma_t \sigma_{t-j}) \text{ and} \]
\[ E r_t^2 r_{t-j}^2 = E (\sigma_t^2 \sigma_{t-j}^2) \]

where the rhs are functions of \( \theta \).
Standard One-step GMM

- Define
  \[ g_T (\theta) = \frac{1}{T} \sum_{t=1}^{T} f(v_t, \theta). \]

- The GMM estimator of \( \theta_0 \) is given by
  \[ \hat{\theta}_{GMM} = \arg \min_{\theta \in \Theta} g_T (\theta) \mathbf{W}_T^{-1} g_T (\theta) \]
  where \( \mathbf{W}_T \) is a positive definite weighting matrix.

- The first-step estimator
  \[ \tilde{\theta}_T = \arg \min_{\theta \in \Theta} g_T (\theta) \mathbf{W}_o^{-1} g_T (\theta) \]
  for \( \mathbf{W}_o = I_m, Z'Z / T, \ldots \)
GMM is of central Importance in Economics and Finance

Large Sample Properties of Generalized Method of Moments Estimators

[PDF] Large sample properties of generalized method of moments estimators
LP Hansen - Econometrica: Journal of the Econometric Society, 1982 - JSTOR
This paper studies estimators that make sample analogues of population orthogonality conditions close to zero. Strong consistency and asymptotic normality of such estimators is established under the assumption that the observable variables are stationary and ...
Outlining the main sections:

1. **Basic Setting and Problem**
   - GMM Estimation
   - Optimal Weighting Matrix
   - Two-step Estimation and Testing

2. **Fixed Smoothing Asymptotics and its Representation**
   - Big Picture
   - Main Results
   - Representation and Approximation: Series HAR
   - Representation and Approximation: Kernel HAR

3. **Implementation, Simulation and Empirical Application**
   - Practical Implementation
   - Simulation
   - Empirical Application

4. **Conclusions**
Two-step GMM: “Optimal” Weighting Matrix

According to Hansen (1982), the optimal weighting matrix $W_T$ is the asymptotic variance matrix of $\sqrt{T}g_T(\theta_0)$.

Lars Hansen was awarded the Nobel Memorial Prize because of this paper.

$$\text{var} \left[ \sqrt{T}g_T(\theta_0) \right] = E \frac{1}{T} \sum_{t=1}^{T} \sum_{t=1}^{T} f(v_t, \theta_0) f'(v_s, \theta_0)$$

Want to estimate the long run variance nonparametrically.

Different names:
- Heteroskedasticity and Autocorrelation Consistent (HAC)
- Heteroskedasticity and Autocorrelation Robust (HAR)
Two-step GMM: “Optimal” Weighting Matrix

- \( \text{var} \left[ \sqrt{T} g_T (\theta_0) \right] = T^{-1} \sum_{t=1}^{T} \text{cov}(u_t, u_s) \) for \( u_t = f(v_t, \theta_0) \)

- In finite samples, not much information is available for \( E u_t u_s = \text{cov}(u_t, u_s) \) when \( |t - s| \) is large.

- It is reasonable to assume that \( E u_t u_s \) is small (relative to the variance of \( u_t \)) when \( |t - s| \) is large.

- Given the above two observations, a good estimator of \( E u_t u_s \) would be

  \[
  Q_h \left( \frac{t}{T}, \frac{s}{T} \right) \hat{u}_t \hat{u}'_s
  \]

  where \( \hat{u}_t \) is an estimated version of \( u_t \) and \( Q_h (\cdot, \cdot) \) satisfies

- \( Q_h (t/T, s/T) \to 0 \) if \( |t - s| / T \to 1 \);

- \( Q_h (t/T, s/T) \to 1 \) if \( |t - s| / T \to 0 \).
Two-step GMM: “Optimal” Weighting Matrix

- Let $\tilde{u}_t = f(v_t, \tilde{\theta}_T)$.
- The optimal Weighting Matrix or the HAR variance estimator

$$W_T (\tilde{\theta}_T) = \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} Q_H \left( \frac{t}{T}, \frac{s}{T} \right) \left( \tilde{u}_t - \frac{1}{T} \sum_{\tau=1}^{T} \tilde{u}_\tau \right)$$

$$\times \left( \tilde{u}_s - \frac{1}{T} \sum_{\tau=1}^{T} \tilde{u}_\tau \right)'$$

for some shrinking function $Q_H (\cdot, \cdot)$.

- The smoothing parameter $h$ controls for the degree of shrinkage.

- General quadratic LRV or HAR variance estimator
Kernel HAR Variance Estimator

- kernel LRV estimator:

\[ Q_h \left( \frac{t}{T}, \frac{s}{T} \right) := k \left( \frac{t - s}{bT} \right) = k_b \left( \frac{t - s}{T} \right) \]

where \( k_b(x) = k(x/b) \).

- \( h = 1/b \) (there is a reason for this parametrization)

- \( Q_h(t, s) \) depends on \((t, s)\) only through \(t - s\).


- Note: as \( b \) increases, the variance of \( W_T(\tilde{\theta}_T) \) increases and the bias of \( W_T(\tilde{\theta}_T) \) decreases (behave like the number of terms in series regression).
Do you HAC?

A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix
WK Newey, KD West - 1986 - nber.org

... Part II presents a simple model of the social costs of monopoly, conceived as the sum of the deadweight loss and the additional loss resulting from ... The significance of this qualification is consider later. 8. p E LW/PC The sign of E is positive in equations (2) and (3). . Page 6. 5. ...
Cited by 10647  Related articles  More
Series HAR Variance Estimator

- Let

\[ Q_h\left( \frac{t}{T}, \frac{s}{T} \right) = \frac{1}{K} \sum_{j=1}^{K} \phi_j \left( \frac{t}{T} \right) \phi_j \left( \frac{s}{T} \right) \quad \text{with} \quad h = K. \]

- Series LRV estimator:

\[ \mathcal{W}_T(\tilde{\theta}_T) = \frac{1}{K} \sum_{j=1}^{K} \mathcal{W}_{Tj}(\tilde{\theta}_T) \]

where \( \mathcal{W}_{Tj}(\tilde{\theta}_T) \) are the “Direct Estimators”:

\[ \mathcal{W}_{Tj}(\tilde{\theta}_T) = \frac{1}{T} \left[ \sum_{t=1}^{T} \phi_j \left( \frac{t}{T} \right) (\tilde{u}_t - \bar{u}) \right] \left[ \sum_{s=1}^{T} \phi_j \left( \frac{s}{T} \right) (\tilde{u}_s - \bar{u}) \right] ' \]

- \( \phi_j \left( t / T \right) \) is a set of orthonormal bases with energy concentrated mostly at the origin and \( \int_0^1 \phi_j(x) \, dx = 0. \)
Kernel HAR v.s. Series HAR

- By Fourier series expansion or Mercer’s Theorem (K-L expansion)

\[ k(t - s) = \sum_{j=1}^{\infty} \lambda_j \tilde{\phi}_j(t) \tilde{\phi}_j(s), \quad \sum_{j=1}^{\infty} \lambda_j = k(0) = 1. \]

- So for \( \phi_j(\cdot) = \tilde{\phi}_j(\cdot / b) \), we have

\[ k\left(\frac{t - s}{bT}\right) = \sum_{j=1}^{\infty} \lambda_j \tilde{\phi}_j\left(\frac{t}{bT}\right) \tilde{\phi}_j\left(\frac{s}{bT}\right) = \sum_{j=1}^{\infty} \lambda_j \phi_j\left(\frac{t}{T}\right) \phi_j\left(\frac{s}{T}\right) \]

- Compare (Hard thresholding vs soft thresholding)

\[ Q_h\left(\frac{t}{T}, \frac{s}{T}\right) = \sum_{j=1}^{K} \frac{1}{K} \phi_j\left(\frac{t}{T}\right) \phi_j\left(\frac{s}{T}\right) \]
Outline

1. Basic Setting and Problem
   - GMM Estimation
   - Optimal Weighting Matrix
   - Two-step Estimation and Testing

2. Fixed Smoothing Asymptotics and its Representation
   - Big Picture
   - Main Results
   - Representation and Approximation: Series HAR
   - Representation and Approximation: Kernel HAR

3. Implementation, Simulation and Empirical Application
   - Practical Implementation
   - Simulation
   - Empirical Application

4. Conclusions
Two-step GMM Estimator

- One-step (First-step) GMM estimator

\[ \tilde{\theta}_T = \arg \min_{\theta \in \Theta} g_T(\theta)' W_o^{-1} g_T(\theta), \]

for \( W_o = I_m, Z'Z / T \).

- Two-step (Second-step) GMM estimator

\[ \hat{\theta}_T = \arg \min_{\theta \in \Theta} g_T(\theta)' W_{T}^{-1}(\tilde{\theta}_T) g_T(\theta). \]

- The difference lies in the weighting matrix.
- Two-step (Second-step) GMM estimator is asymptotically more efficient (in what sense?)
- But there is a cost in estimating the weighting matrix.
Reasons for Using a Two-step GMM Estimator

- Efficiency of the point estimator.
- Power of the associated test.
- To implement GMM criterion-based tests, we often use the two-step estimator.
- Numerical stability.
Two-step GMM Testing

- The null hypothesis $H_0 : R\theta_0 = r \in \mathbb{R}^q$ and alternative hypothesis $H_1 : R\theta_0 \neq r$.
- Focus on linear restrictions but nonlinear ones are allowed.
- The F-test version of the Wald statistic is

$$W_T = \sqrt{T} (\hat{R}T - r)' \left\{ R \left[ G_T(\hat{T})' W_T^{-1}(\hat{T}) G_T(\hat{T}) \right]^{-1} R' \right\}^{-1} \times \sqrt{T} (\hat{R}T - r) / p,$$

where $G_T(\theta) = \frac{\partial g_T(\theta)}{\partial \theta'}$.
- Compare with the asymptotic variance in the case of GLS:

$$G_T(\hat{T})' \quad W_T^{-1}(\hat{T}) \quad G_T(\hat{T})$$

$$X' \quad var(u|X) \quad X$$
Two-step GMM Testing

- Let $\hat{\theta}_{T,R}$ be the restricted second-step GMM estimator:

$$
\hat{\theta}_{T,R} = \arg\min_{\theta \in \Theta} g_T(\theta)' W_T^{-1}(\tilde{\theta}_T) g_T(\theta) \; \text{s.t.} \; R\theta = r,
$$

and define

$$
pD_T := \begin{bmatrix}
T g_T(\hat{\theta}_T)' W_T^{-1}(\tilde{\theta}_T) g_T(\hat{\theta}_T) \\
- T g_T(\hat{\theta}_{T,R})' W_T^{-1}(\tilde{\theta}_T) g_T(\hat{\theta}_{T,R})
\end{bmatrix}
$$

- The score statistic or Lagrange Multiplier (LM) statistic. Let $\Delta_T(\theta) = G_T'(\theta) W_T^{-1}(\tilde{\theta}_T) g_T(\theta)$ and define

$$
pS_T = T [\Delta_T(\hat{\theta}_{T,R})]' [G_T'(\hat{\theta}_{T,R}) W_T^{-1}(\tilde{\theta}_T) G_T(\hat{\theta}_{T,R})]^{-1} \Delta_T(\hat{\theta}_{T,R})
$$
Overidentification Test

- J Statistic:
  \[
  J_T = Tg_T (\hat{\theta}_T)' [W_T^{-1}(\tilde{\theta}_T)]^{-1} g_T (\hat{\theta}_T)
  \]
  \[
  J_T = Tg_T (\hat{\theta}_T)' [W_T^{-1}(\tilde{\theta}_T)]^{-1} g_T (\hat{\theta}_T)
  \]

- Series HAR:
  \[
  J_T \overset{d}{\to} \frac{K}{K - q + 1} F(q, K - q + 1).
  \]

- Kernel HAR:
  \[
  J_T \overset{d}{\to} B_q(1)' \left[ \int_0^1 \int_0^1 Q_h^* (r, s) dB_q(r) dB'_q (s) \right]^{-1} B_q(1).
  \]

- Details are available in Sun and Kim (2013, JoE).
Outline

1 Basic Setting and Problem
   - GMM Estimation
   - Optimal Weighting Matrix
   - Two-step Estimation and Testing

2 Fixed Smoothing Asymptotics and its Representation
   - Big Picture
   - Main Results
   - Representation and Approximation: Series HAR
   - Representation and Approximation: Kernel HAR

3 Implementation, Simulation and Empirical Application
   - Practical Implementation
   - Simulation
   - Empirical Application

4 Conclusions
The fundamental question is how to approximate the distributions of $W_T$, $D_T$ and $S_T$.

This is a difficult problem as $W_T(\tilde{\theta}_T)$ shows up twice.

A first step is to come up with a good approximation to the distribution of $W_T(\tilde{\theta}_T)$.

Under the conventional asymptotics, we approximate the distribution of $W_T(\tilde{\theta}_T)$ by the degenerate distribution that concentrates at its probability limit.

Under the conventional asymptotics, all three statistics $W_T$, $D_T$ and $S_T$ are asymptotically distributed as $\chi^2_p / p$.

Numerous studies have shown that the chi-square approximation is very poor.
The conventional asymptotics is obtained by embedding the finite sample situation in the asymptotic sequence $h \to \infty$ as $T \to \infty$ but $h / T \to 0$.

However, in any given finite sample, $h$ is a given constant.

A more accurate approximation may be obtained by embedding the finite sample in a new asymptotic sequence where $h$ is held fixed.
Towards the New Asymptotics

- For Series HAR variance, $K$ is fixed.

\[
W_T(\tilde{\theta}_T) = \frac{1}{K} \sum_{j=1}^{K} W_{Tj}(\tilde{\theta}_T)
\]

$W_T(\tilde{\theta}_T)$ is an average of a finite number of direct and inconsistent estimators $W_{Tj}(\tilde{\theta}_T)$. The amount of smoothing does not increase with the sample size.

- For kernel HAR variance, $b$ is fixed. In the frequency domain, $W_T(\tilde{\theta}_T)$ is a smoothing over finite number of periodograms.

Vogelsang and coauthors
Towards the New Asymptotics

- The new asymptotics is called fixed-smoothing asymptotics ($h$ is fixed)
- The conventional asymptotics is called increasing-smoothing asymptotics ($h \to \infty$).

<table>
<thead>
<tr>
<th>Fixed-smoothing asymptotics</th>
<th>Increasing-smoothing asymptotics</th>
</tr>
</thead>
<tbody>
<tr>
<td>$h$ is fixed as $T \to \infty$</td>
<td>$h \to \infty$ as $T \to \infty$ but $h/T \to 0$</td>
</tr>
</tbody>
</table>
Towards the New Asymptotics

\[ h \to \infty \text{ as } T \to \infty \text{ but } h/T \to 0 \]

Increasing-smoothing Asymptotics

Fixed-smoothing Asymptotics
Towards the New Asymptotics

- To a great extent and from a broad perspective, the idea is line with many other areas of research in econometrics where more accurate distributional approximations are the focus of interest.
- A common theme for coming up with a new approximation is to embed the finite sample situation in a different limiting thought experiment:
  - Strong IV vs. Weak IV vs. Many (weak) IV asymptotics
  - Unit root/normal approximation vs. local to unity approximation
  - Extreme Quantile vs. Intermediate Quantile Approximation
  - Asymptotics under a drifting sequence vs. Asymptotics under a fixed parameter.
Related Recent Literature


Outline

1. Basic Setting and Problem
   - GMM Estimation
   - Optimal Weighting Matrix
   - Two-step Estimation and Testing

2. Fixed Smoothing Asymptotics and its Representation
   - Big Picture
   - Main Results
   - Representation and Approximation: Series HAR
   - Representation and Approximation: Kernel HAR

3. Implementation, Simulation and Empirical Application
   - Practical Implementation
   - Simulation
   - Empirical Application

4. Conclusions
Fixed-smoothing Asymptotics

Let

\[ Q_{T,h}^* \left( \frac{ t}{T}, \frac{ s}{T} \right) = Q_h \left( \frac{ t}{T}, \frac{ s}{T} \right) - \frac{1}{T} \sum_{ \tau = 1}^{T} Q_h \left( \frac{ \tau}{T}, \frac{ s}{T} \right) \]

\[ - \frac{1}{T} \sum_{ \tau = 1}^{T} Q_h \left( \frac{ t}{T}, \frac{ \tau}{T} \right) + \frac{1}{T} \sum_{ \tau_1 = 1}^{T} \sum_{ \tau_2 = 1}^{T} Q_h \left( \frac{ \tau_1}{T}, \frac{ \tau_2}{T} \right) \]

\[ = Q_h \left( \frac{ t}{T}, \frac{ s}{T} \right) - \bar{Q}_h (\cdot, \frac{ s}{T}) - \bar{Q}_h \left( \frac{ t}{T}, \cdot \right) + \bar{Q} (\cdot, \cdot) \]

be the (doubly) demeaned weighting function.

Then

\[ W_T (\theta) = \frac{1}{T} \sum_{ t=1}^{T} \sum_{ s=1}^{T} Q_{T,h}^* \left( \frac{ t}{T}, \frac{ s}{T} \right) f(v_t, \theta) f(v_s, \theta)'. \]
Fixed-smoothing Asymptotics

Heuristically

\[ W_T(\tilde{\theta}_T) = \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} Q^*_{T,h}(t/T, s/T) f(v_t, \tilde{\theta}_T)f(v_s, \tilde{\theta}_T)' \]

\[ \Rightarrow \int_0^1 \int_0^1 Q^*_h(r, s) \Lambda dB_m(r)dB_m(s)' \Lambda' := W_\infty \]

where

\[ W_\infty = \Lambda \tilde{W}_\infty \Lambda' \] and \[ \tilde{W}_\infty = \int_0^1 \int_0^1 Q^*_h(r, s) dB_m(r)dB_m(s)' \]

and

\[ Q^*_h(r, s) = Q_h(r, s) - \int_0^1 Q_h(\tau_1, s) d\tau_1 \]

\[ - \int_0^1 Q_h(r, \tau_2) d\tau_2 + \int_0^1 \int_0^1 Q_h(\tau_1, \tau_2) d\tau_1 d\tau_2. \]
Fixed-smoothing Asymptotics

Heuristically

\[ \sqrt{T} (\hat{\theta}_T - \theta_0) \overset{d}{=} - \left[ G_T' W_T(\tilde{\theta}_T)^{-1} G_T \right]^{-1} G_T' W_T(\tilde{\theta}_T)^{-1} \sqrt{T} g_T (\theta_0) \]

\[ \overset{d}{\rightarrow} - \left[ G' W_\infty^{-1} G \right]^{-1} G' W_\infty^{-1} \Lambda B_m (1). \]

where

\[ G = G(\theta_0) \text{ and } G(\theta) = E \frac{\partial f(v_t, \theta)}{\partial \theta'}. \]

- \( \hat{\theta}_T \) is not asymptotically normal but rather asymptotically mixed normal.
- Conditional on \( W_\infty^{-1} \), the limiting distribution is normal with the conditional variance depending on \( W_\infty^{-1} \).
- Due to the asymptotic mixed-normality, it is difficult to show the asymptotic pivotality for \( W_T, D_T, \) and \( S_T \).
Theorem

\[ W_T \]
\[ = \left[ R \sqrt{T} \left( \hat{\theta}_T - \theta_0 \right) \right] ' \left\{ R \left[ G_T (\hat{\theta}_T)' W_T^{-1} (\hat{\theta}_T) G_T (\hat{\theta}_T) \right]^{-1} R' \right\}^{-1} \]
\[ \times \left[ R \sqrt{T} \left( \hat{\theta}_T - \theta_0 \right) \right] ' / p \]
\[ \xrightarrow{d} \left[ R \left( G' W_\infty^{-1} G \right)^{-1} G' W_\infty^{-1} \Lambda B_m(1) \right] ' \left[ R \left( G' W_\infty^{-1} G \right)^{-1} R' \right]^{-1} \]
\[ \times \left[ R \left( G' W_\infty^{-1} G \right)^{-1} G' W_\infty^{-1} \Lambda B_m(1) \right] / p \]
\[ \xrightarrow{d} F_\infty \]
Fixed-smoothing Asymptotics: Main Results

\[
C_{pp} = \int_0^1 \int_0^1 Q^*_h(r,s) dB_p(r) dB_p(s)', \quad C_{pq} = \int_0^1 \int_0^1 Q^*_h(r,s) dB_p(r) dB_q(s)
\]

\[
C_{qq} = \int_0^1 \int_0^1 Q^*_h(r,s) dB_q(r) dB_q(s)', \quad D_{pp} = C_{pp} - C_{pq} C_{qq}^{-1} C_{pq}'
\]

Theorem

\[
F_\infty \overset{d}{=} \left[ B_p(1) - C_{pq} C_{qq}^{-1} B_q(1) \right]' D_{pp}^{-1} \left[ B_p(1) - C_{pq} C_{qq}^{-1} B_q(1) \right] / p,
\]

where

- \( p \) is the number of joint hypotheses
- \( q \) is the degree of overidentification \( q = m - d \).
Fixed-smoothing Asymptotics: Main Results

- $F_\infty$ is pivotal (Neat and Elegant).
- When the model is exactly identified, we have $q = 0$. In this case,

$$F_\infty = B_p (1)' C_{pp}^{-1} B_p (1) / p.$$  

This limit is the same as in the one-step GMM framework.

- The limiting distribution $F_\infty$ depends on the degree of over-identification $q$.
- Typically such a dependence shows up only when the number of moment conditions is assumed to grow with the sample size. Here the number of moment conditions is fixed.
- It remains in the limiting distribution because it contains information on the dimension of the random matrix $\tilde{W}_\infty$. 

Yixiao Sun UC San Diego
Accurate Asymptotic Approximation in the Optimal GMM Framework
Fixed-smoothing Asymptotics: Main Results

**Theorem**

*Under some assumptions, for a fixed $h$,*

$$D_T = W_T + o_p(1) \text{ and } S_T = W_T + o_p(1),$$

*as $T \to \infty$.*
Outline

1. Basic Setting and Problem
   - GMM Estimation
   - Optimal Weighting Matrix
   - Two-step Estimation and Testing

2. Fixed Smoothing Asymptotics and its Representation
   - Big Picture
   - Main Results
   - Representation and Approximation: Series HAR
   - Representation and Approximation: Kernel HAR

3. Implementation, Simulation and Empirical Application
   - Practical Implementation
   - Simulation
   - Empirical Application

4. Conclusions
Series Case

- Note that

\[
K \begin{bmatrix}
  C_{pp} & C_{pq} \\
  C'_{pq} & C_{qq}
\end{bmatrix} \overset{d}{=} \int_0^1 \int_0^1 Q_h^* (r, s) \ dB_{p+q} (r) \ dB'_{p+q} (s)
\]

\[
= \sum_{j=1}^K \left[ \int_0^1 \Phi_j (r) \ dB_{p+q} (r) \right] \left[ \int_0^1 \Phi_j (r) \ dB_{p+q} (r) \right]'
\]

where \( \int_0^1 \Phi_j (r) \ dB_{p+q} (r) \overset{d}{=} iidN(0, I_{p+q}) \), we have

\[
K \begin{bmatrix}
  C_{pp} & C_{pq} \\
  C'_{pq} & C_{qq}
\end{bmatrix} \sim \mathcal{W}_{p+q} (K, I_{p+q}),
\]

a standard Wishart distribution.

- \( D_{pp} = C_{pp} - C_{pq} C_{qq}^{-1} C_{pq}' \sim \mathcal{W}_p (K - q, I_p) / K \) and is independent of both \( C_{pq} \) and \( C_{qq} \).
This brings $F_\infty$ close to Hotelling’s $T^2$ distribution, which is the same as a standard $F$ distribution after some multiplicative adjustment.

The only difference is that $B_p (1) - C_{pq} C_{qq}^{-1} B_q (1)$ is not normal and hence $\| B_p (1) - C_{pq} C_{qq}^{-1} B_q (1) \|^2$ does not follow a chi-square distribution.

However, conditional on $C_{pq} C_{qq}^{-1} B_q (1)$,

$\| B_p (1) - C_{pq} C_{qq}^{-1} B_q (1) \|^2$ follows a noncentral $\chi^2_p$ distribution with noncentrality parameter

$$\Delta^2 = \| C_{pq} C_{qq}^{-1} B_q (1) \|^2.$$

It then follows that $F_\infty$ is conditionally distributed as a (mixed) noncentral $F$ distribution.
(Mixed) Noncentral F Approximation

Let
\[ \kappa = \frac{K}{(K - p - q + 1)}, \quad \delta^2 = E\Delta^2 = \frac{pq}{K - q - 1}. \]

**Theorem**

For OS HAR variance estimation, we have
\[ \kappa^{-1}F_\infty \overset{d}{=} \mathcal{F}_{p, K - p - q + 1} \left( \Delta^2 \right), \]

a mixed noncentral F random variable with random noncentrality parameter \( \Delta^2 \).

**Theorem**

\[ P\left( \kappa^{-1}F_\infty < z \right) = F_{p, K - p - q + 1} \left( z, \delta^2 \right) + o \left( K^{-1} \right) \text{ as } K \to \infty. \]
Basic Setting and Problem
Fixed Smoothing Asymptotics and its Representation
Implementation, Simulation and Empirical Application
Conclusions

Big Picture
Main Results
Representation and Approximation: Series HAR
Representation and Approximation: Kernel HAR

Yixiao Sun UC San Diego
Accurate Asymptotic Approximation in the Optimal GMM Framework
Outline

1. Basic Setting and Problem
   - GMM Estimation
   - Optimal Weighting Matrix
   - Two-step Estimation and Testing

2. Fixed Smoothing Asymptotics and its Representation
   - Big Picture
   - Main Results
   - Representation and Approximation: Series HAR
   - Representation and Approximation: Kernel HAR

3. Implementation, Simulation and Empirical Application
   - Practical Implementation
   - Simulation
   - Empirical Application

4. Conclusions
For kernel HAR variance estimation, we have

\[ pF_\infty \overset{d}{=} \chi_p^2 / \eta^2 \]

where \( \chi_p^2 \) is independent of \( \eta^2 \),

\[ \eta^2 \overset{d}{=} \frac{e_p' \left[ I_p + C_{pq}C^{-1}_{qq}C^{-1}_{qq}C'_{pq} \right] e_p}{e_p' \left[ I_p + C_{pq}C^{-1}_{qq}C^{-1}_{qq}C'_{pq} \right] D_{pp}^{-1} \left[ I_p + C_{pq}C^{-1}_{qq}C^{-1}_{qq}C'_{pq} \right] e_p} \]

and \( e_p = (1, 0, ..., 0)' \in \mathbb{R}^p \).
A Major Breakthrough: High-order refinement

- The fixed-smoothing approximation is high order correct under the conventional asymptotics.
- Let $F_\infty^\alpha$ and $\chi_\infty^\alpha$ be the critical values from the fixed-smoothing approximation and chi-square approximation, then as $h \to \infty$ such that $h/T \to 0$,

$$P(W_T < F_\infty^\alpha) = \alpha + o(h^{-1}) + \text{other terms}$$
$$P(W_T < \chi_\infty^\alpha / \rho) = \alpha + O(h^{-1}) + \text{other terms}$$

- Culmination of multiple-year’s research effort.
- Tedious proof (90 pages) in a separate paper.
- Too technical to be presented even in an econometric seminar.
New Asymptotics under the Local Alternatives

- Under $H_1: R\theta_0 = r - \delta_0 / \sqrt{T}$

\[ \mathcal{W}_T(\hat{\theta}_T) \xrightarrow{d} F_\infty(\|\delta_0\|_V^2), \]

where

\[ F_\infty(\|\delta_0\|_V^2) = \left[ B_p(1) - C_{pq} C_{qq}^{-1} B_q(1) + \mathcal{V}^{-1/2} \delta_0 \right]' D_{pp}^{-1} \times \left[ B_p(1) - C_{pq} C_{qq}^{-1} B_q(1) + \mathcal{V}^{-1/2} \delta_0 \right] / p \]

and

\[ \mathcal{V} = R \left[ G' \Omega^{-1} G \right]^{-1} R'. \]
Outline

1 Basic Setting and Problem
   - GMM Estimation
   - Optimal Weighting Matrix
   - Two-step Estimation and Testing

2 Fixed Smoothing Asymptotics and its Representation
   - Big Picture
   - Main Results
   - Representation and Approximation: Series HAR
   - Representation and Approximation: Kernel HAR

3 Implementation, Simulation and Empirical Application
   - Practical Implementation
   - Simulation
   - Empirical Application

4 Conclusions
Basic Framework

- Sample moment conditions: \( g_T(\theta) = \frac{1}{T} \sum_{t=1}^{T} f(v_t, \theta_0) \)
- One-step and two-step GMM estimators

\[
\begin{align*}
\tilde{\theta}_T &= \arg \min_{\theta \in \Theta} g_T(\theta)' g_T(\theta) \quad \text{and} \\
\hat{\theta}_T &= \arg \min_{\theta \in \Theta} g_T(\theta)' W^{-1}_T(\tilde{\theta}_T; h) g_T(\theta)
\end{align*}
\]

where for \( \tilde{u}_t := f(v_t, \tilde{\theta}_T) \)

\[
W_T(\tilde{\theta}_T; h) = \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} Q_h \left( \frac{t}{T}, \frac{s}{T} \right) \left( \tilde{u}_t - \frac{1}{T} \sum_{\tau=1}^{T} \tilde{u}_\tau \right)^2.
\]

If you prefer, the Newey-West estimator can be used.

- The “optimal” smoothing parameter \( h \) can be estimated by \( \tilde{h}_T := h_T(\tilde{\theta}_T) \) using Andrews (1991)' plug-in approach.
Wald, LM and LR

- **Wald Statistic**

\[ W_T = T (R\hat{\theta}_T - r)' \left\{ R \left[ \hat{G}' \tilde{W}_T^{-1} \hat{G} \right]^{-1} R' \right\}^{-1} (R\hat{\theta}_T - r) / p \]

- **GMM Criterion-based LR type statistic**

\[ D_T = \left[ T \hat{g}' \tilde{W}_T^{-1} \hat{g} - Tg_T(\hat{T}_T, R)' \tilde{W}_T^{-1} g_T(\hat{T}_T, R) \right] / p \]

- **Score or LM type of statistic**

\[ S_T = T \left[ \Delta_T(\hat{T}_T, R) \right]' \left[ G'_T(\hat{T}_T, R) \tilde{W}_T^{-1} G_T(\hat{T}_T, R) \right]^{-1} \Delta_T(\hat{T}_T, R) / p \]
Let $e_t := (e_{t,p}', e_{t,d-p}', e_{t,q}')' \sim iidN(0, I_m)$,

$$C_{p,T} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} e_{t,p}, \quad C_{q,T} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} e_{t,q}$$

and

$$C_{pp,T} = \frac{1}{T} \sum_{t=1}^{T} \sum_{\tau=1}^{T} Q_{\tilde{h}_T} \left( \frac{t}{T}, \frac{\tau}{T} \right) \tilde{e}_{t,p} \tilde{e}_{\tau,p}'$$

$$C_{pq,T} = \frac{1}{T} \sum_{t=1}^{T} \sum_{\tau=1}^{T} Q_{\tilde{h}_T} \left( \frac{t}{T}, \frac{\tau}{T} \right) \tilde{e}_{t,p} \tilde{e}_{\tau,q}'$$

$$C_{qq,T} = \frac{1}{T} \sum_{t=1}^{T} \sum_{\tau=1}^{T} Q_{\tilde{h}_T} \left( \frac{t}{T}, \frac{\tau}{T} \right) \tilde{e}_{t,q} \tilde{e}_{\tau,q}'$$

where $\tilde{e}_{i,j} = e_{i,j} - \frac{1}{T} \sum_{s=1}^{T} e_{s,j}$.
Define

\[ D_{pp,T} = C_{pp,T} - C_{pq,T} C_{qq,T}^{-1} C_{pq,T}'. \]

Use \( F_{eT} \) as the reference distribution for \( W_T, D_T, \) and \( S_T \):

\[ F_{eT} = \left[ C_{p,T} - C_{pq,T} C_{qq,T}^{-1} C_{q,T} \right]' D_{pp,T}^{-1} \left[ C_{p,T} - C_{pq,T} C_{qq,T}^{-1} C_{q,T} \right] / p \]

For the t-test, use \( t_{eT} \) as the reference distribution:

\[ t_{eT} = \left[ C_{p,T} - C_{pq,T} C_{qq,T}^{-1} C_{q,T} \right] / \sqrt{D_{pp,T}} \text{ for } p = 1 \]

\( F_{eT} \) and \( t_{eT} \) are finite sample versions of \( F_\infty \) and \( t_\infty \). Easy to simulate.
Reference Distribution in the Case of OS HAR estimation

- Define
  \[ \tilde{\delta}^2_T = \frac{pq}{\tilde{h}_T - q - 1}. \]

- Let \( \mathcal{F}^{1-\alpha}_{p,\tilde{h}_T-p-q+1}(\tilde{\delta}^2_T) \) be the \((1-\alpha)\) quantile of the noncentral \(F\) distribution \( F_{p,\tilde{h}_T-p-q+1}(\tilde{\delta}^2_T) \) with degrees of freedom \((p, \tilde{h}_T - p - q + 1)\) and noncentrality parameter \(\tilde{\delta}^2_T\).

- We can use
  \[ \frac{\tilde{h}_T - p - q + 1}{\tilde{h}_T} \mathcal{F}^{1-\alpha}_{p,\tilde{h}_T-p-q+1}(\tilde{\delta}^2_T) \]

  as the \(\alpha\)-level critical value for \(W_T, D_T, \) and \(S_T\).

- Critical values for t-tests can be similarly approximated.
Outline

1. Basic Setting and Problem
   - GMM Estimation
   - Optimal Weighting Matrix
   - Two-step Estimation and Testing

2. Fixed Smoothing Asymptotics and its Representation
   - Big Picture
   - Main Results
   - Representation and Approximation: Series HAR
   - Representation and Approximation: Kernel HAR

3. Implementation, Simulation and Empirical Application
   - Practical Implementation
   - Simulation
   - Empirical Application

4. Conclusions
DGP

\[ y_t = x_{0,t} \alpha + x_{1,t} \beta_1 + x_{2,t} \beta_2 + x_{3,t} \beta_3 + \epsilon_{y,t} \]

where \( x_{0,t} \equiv 1 \) and \( x_{1,t}, x_{2,t} \) and \( x_{3,t} \) are scalar regressors that are correlated with \( \epsilon_{y,t} \).

\[ \theta = (\alpha, \beta_1, \beta_2, \beta_3)' \]

We have \( m \) instruments \( z_{0,t}, z_{1,t}, \ldots, z_{m-1,t} \) with \( z_{0,t} \equiv 1 \).

The reduced-form equations for \( x_{1,t}, x_{2,t} \) and \( x_{3,t} \) are given by

\[ x_{j,t} = z_{j,t} + \sum_{i=d-1}^{m-1} z_{i,t} + \epsilon_{x_{j,t}} \text{ for } j = 1, 2, 3. \]
We assume that $z_{i,t}$ for $i \geq 1$ follows either an AR(1) process

$$z_{i,t} = \rho z_{i,t-1} + \sqrt{1 - \rho^2} e_{z_{i,t}},$$

or an MA(1) process

$$z_{i,t} = \rho e_{z_{i,t-1}} + \sqrt{1 - \rho^2} e_{z_{i,t}},$$

where

$$e_{z_{i,t}} = \frac{e_{zt}^i + e_{zt}^0}{\sqrt{2}}$$

The DGP for $\varepsilon_t = (\varepsilon_{yt}, \varepsilon_{x_1t}, \varepsilon_{x_2t}, \varepsilon_{x_3t})'$ is the same as that for $(z_{1,t}, \ldots, z_{m-1,t})$ except the dimensionality difference.
A Tale of Four Tests

Null Hypotheses

\[ H_{01} : \beta_1 = 0, \]
\[ H_{02} : \beta_1 = \beta_2 = 0, \]
\[ H_{03} : \beta_1 = \beta_2 = \beta_3 = 0 \]

Four tests with critical values

\[ \frac{\chi_p^{1-\alpha}}{p}, \frac{K}{K - p + 1} F_{p, K-p+1}^{1-\alpha}, \frac{K}{K - p - q + 1} F_{p, K-p-q+1}^{1-\alpha} (\delta^2), \frac{\chi_{\infty}^{1-\alpha}}{p} \]
Simulation Results

Table: Empirical size of the \( \chi^2 \) test, \( F \) test, noncentral \( F \) test and nonstandard \( F_\infty \) test based on the orthonormal series HAR variance estimator for the AR(1) case with \( T = 100 \), number of joint hypotheses \( p \) and number of overidentifying restrictions \( q \)

<table>
<thead>
<tr>
<th>( \rho )</th>
<th>( \chi^2 )</th>
<th>CF</th>
<th>NCF</th>
<th>( F_\infty )</th>
<th>( \chi^2 )</th>
<th>CF</th>
<th>NCF</th>
<th>( F_\infty )</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.8</td>
<td>0.114</td>
<td>0.072</td>
<td>0.072</td>
<td>0.073</td>
<td>0.310</td>
<td>0.109</td>
<td>0.109</td>
<td>0.111</td>
</tr>
<tr>
<td>-0.5</td>
<td>0.081</td>
<td>0.060</td>
<td>0.060</td>
<td>0.059</td>
<td>0.174</td>
<td>0.077</td>
<td>0.077</td>
<td>0.078</td>
</tr>
<tr>
<td>0.0</td>
<td>0.063</td>
<td>0.051</td>
<td>0.051</td>
<td>0.050</td>
<td>0.112</td>
<td>0.060</td>
<td>0.060</td>
<td>0.062</td>
</tr>
<tr>
<td>0.5</td>
<td>0.094</td>
<td>0.063</td>
<td>0.063</td>
<td>0.063</td>
<td>0.222</td>
<td>0.077</td>
<td>0.077</td>
<td>0.078</td>
</tr>
<tr>
<td>0.8</td>
<td>0.134</td>
<td>0.086</td>
<td>0.086</td>
<td>0.088</td>
<td>0.355</td>
<td>0.119</td>
<td>0.119</td>
<td>0.122</td>
</tr>
<tr>
<td>0.9</td>
<td>0.166</td>
<td>0.117</td>
<td>0.117</td>
<td>0.120</td>
<td>0.437</td>
<td>0.181</td>
<td>0.181</td>
<td>0.184</td>
</tr>
</tbody>
</table>
Simulation Results

**Table:** Empirical size of the $\chi^2$ test, $F$ test, noncentral $F$ test and nonstandard $F_\infty$ test based on the orthonormal series HAR variance estimator for the AR(1) case with $T = 100$, number of joint hypotheses $p$ and number of overidentifying restrictions $q$

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>$\chi^2$</th>
<th>$F$</th>
<th>$F_\infty$</th>
<th>$\chi^2$</th>
<th>$F$</th>
<th>$F_\infty$</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.8</td>
<td>0.186</td>
<td>0.129</td>
<td>0.081</td>
<td>0.077</td>
<td>0.457</td>
<td>0.193</td>
</tr>
<tr>
<td>-0.5</td>
<td>0.113</td>
<td>0.086</td>
<td>0.065</td>
<td>0.065</td>
<td>0.247</td>
<td>0.116</td>
</tr>
<tr>
<td>0.0</td>
<td>0.081</td>
<td>0.065</td>
<td>0.053</td>
<td>0.052</td>
<td>0.155</td>
<td>0.085</td>
</tr>
<tr>
<td>0.5</td>
<td>0.128</td>
<td>0.091</td>
<td>0.064</td>
<td>0.063</td>
<td>0.308</td>
<td>0.130</td>
</tr>
<tr>
<td>0.8</td>
<td>0.196</td>
<td>0.140</td>
<td>0.089</td>
<td>0.086</td>
<td>0.489</td>
<td>0.209</td>
</tr>
<tr>
<td>0.9</td>
<td>0.252</td>
<td>0.184</td>
<td>0.126</td>
<td>0.123</td>
<td>0.589</td>
<td>0.301</td>
</tr>
</tbody>
</table>
Simulation Results

**Table:** Empirical size of the $\chi^2$ test, $F$ test, noncentral $F$ test and nonstandard $F_\infty$ test based on the orthonormal series HAR variance estimator for the AR(1) case with $T = 100$, number of joint hypotheses $p$ and number of overidentifying restrictions $q$

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>$\chi^2$</th>
<th>CF</th>
<th>NCF</th>
<th>$F_\infty$</th>
<th>$\chi^2$</th>
<th>CF</th>
<th>NCF</th>
<th>$F_\infty$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$p = 1$, $q = 2$</td>
<td></td>
<td></td>
<td></td>
<td>$p = 3$, $q = 2$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.8</td>
<td>0.260</td>
<td>0.190</td>
<td>0.080</td>
<td>0.079</td>
<td>0.602</td>
<td>0.318</td>
<td>0.100</td>
<td>0.097</td>
</tr>
<tr>
<td>-0.5</td>
<td>0.154</td>
<td>0.117</td>
<td>0.061</td>
<td>0.062</td>
<td>0.351</td>
<td>0.175</td>
<td>0.074</td>
<td>0.073</td>
</tr>
<tr>
<td>0.0</td>
<td>0.104</td>
<td>0.085</td>
<td>0.055</td>
<td>0.055</td>
<td>0.211</td>
<td>0.119</td>
<td>0.065</td>
<td>0.063</td>
</tr>
<tr>
<td>0.5</td>
<td>0.171</td>
<td>0.129</td>
<td>0.065</td>
<td>0.066</td>
<td>0.415</td>
<td>0.200</td>
<td>0.073</td>
<td>0.072</td>
</tr>
<tr>
<td>0.8</td>
<td>0.268</td>
<td>0.199</td>
<td>0.085</td>
<td>0.082</td>
<td>0.623</td>
<td>0.330</td>
<td>0.100</td>
<td>0.097</td>
</tr>
<tr>
<td>0.9</td>
<td>0.332</td>
<td>0.257</td>
<td>0.124</td>
<td>0.121</td>
<td>0.712</td>
<td>0.433</td>
<td>0.160</td>
<td>0.157</td>
</tr>
</tbody>
</table>
Simulation Results

Table: Empirical size of the $\chi^2$ test, $F$ test, noncentral $F$ test and nonstandard $F_\infty$ test based on the kernel HAR variance estimator for the AR(1) case with $T = 100$, number of joint hypotheses $p$ and number of overidentifying restrictions $q$

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>$\chi^2$</th>
<th>$F_\infty[0]$</th>
<th>$F_\infty[q]$</th>
<th>$\chi^2$</th>
<th>$F_\infty[0]$</th>
<th>$F_\infty[q]$</th>
<th>$\chi^2$</th>
<th>$F_\infty[0]$</th>
<th>$F_\infty[q]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p = 1, q = 0$</td>
<td>$p = 2, q = 0$</td>
<td>$p = 3, q = 0$</td>
<td>$p = 1, q = 0$</td>
<td>$p = 2, q = 0$</td>
<td>$p = 3, q = 0$</td>
<td>$p = 1, q = 0$</td>
<td>$p = 2, q = 0$</td>
<td>$p = 3, q = 0$</td>
<td></td>
</tr>
<tr>
<td>-0.8</td>
<td>0.132</td>
<td>0.091</td>
<td>0.091</td>
<td>0.215</td>
<td>0.112</td>
<td>0.112</td>
<td>0.326</td>
<td>0.153</td>
<td>0.153</td>
</tr>
<tr>
<td>-0.5</td>
<td>0.085</td>
<td>0.073</td>
<td>0.073</td>
<td>0.116</td>
<td>0.082</td>
<td>0.082</td>
<td>0.160</td>
<td>0.100</td>
<td>0.100</td>
</tr>
<tr>
<td>0.0</td>
<td>0.063</td>
<td>0.059</td>
<td>0.059</td>
<td>0.071</td>
<td>0.058</td>
<td>0.058</td>
<td>0.093</td>
<td>0.070</td>
<td>0.070</td>
</tr>
<tr>
<td>0.5</td>
<td>0.093</td>
<td>0.074</td>
<td>0.074</td>
<td>0.131</td>
<td>0.076</td>
<td>0.076</td>
<td>0.182</td>
<td>0.096</td>
<td>0.096</td>
</tr>
<tr>
<td>0.8</td>
<td>0.164</td>
<td>0.102</td>
<td>0.102</td>
<td>0.273</td>
<td>0.126</td>
<td>0.126</td>
<td>0.408</td>
<td>0.154</td>
<td>0.154</td>
</tr>
<tr>
<td>0.9</td>
<td>0.216</td>
<td>0.124</td>
<td>0.124</td>
<td>0.382</td>
<td>0.164</td>
<td>0.164</td>
<td>0.559</td>
<td>0.212</td>
<td>0.212</td>
</tr>
</tbody>
</table>
Simulation Results

<table>
<thead>
<tr>
<th>ρ</th>
<th>χ²</th>
<th>F∞ [0]</th>
<th>F∞ [q]</th>
<th>χ²</th>
<th>F∞ [0]</th>
<th>F∞ [q]</th>
<th>χ²</th>
<th>F∞ [0]</th>
<th>F∞ [q]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.8</td>
<td>0.197</td>
<td>0.145</td>
<td>0.110</td>
<td>0.313</td>
<td>0.174</td>
<td>0.127</td>
<td>0.443</td>
<td>0.224</td>
<td>0.164</td>
</tr>
<tr>
<td>-0.5</td>
<td>0.105</td>
<td>0.091</td>
<td>0.076</td>
<td>0.150</td>
<td>0.104</td>
<td>0.088</td>
<td>0.201</td>
<td>0.126</td>
<td>0.104</td>
</tr>
<tr>
<td>0.0</td>
<td>0.073</td>
<td>0.067</td>
<td>0.059</td>
<td>0.096</td>
<td>0.077</td>
<td>0.067</td>
<td>0.121</td>
<td>0.084</td>
<td>0.074</td>
</tr>
<tr>
<td>0.5</td>
<td>0.117</td>
<td>0.095</td>
<td>0.079</td>
<td>0.169</td>
<td>0.111</td>
<td>0.090</td>
<td>0.236</td>
<td>0.132</td>
<td>0.104</td>
</tr>
<tr>
<td>0.8</td>
<td>0.223</td>
<td>0.151</td>
<td>0.110</td>
<td>0.362</td>
<td>0.189</td>
<td>0.132</td>
<td>0.521</td>
<td>0.230</td>
<td>0.160</td>
</tr>
<tr>
<td>0.9</td>
<td>0.315</td>
<td>0.204</td>
<td>0.145</td>
<td>0.521</td>
<td>0.271</td>
<td>0.182</td>
<td>0.704</td>
<td>0.329</td>
<td>0.226</td>
</tr>
</tbody>
</table>
**Simulation Results**

Table: Empirical size of the $\chi^2$ test, the $F_{\infty}(0)$ test and the $F_{\infty}(q)$ test based on the Parzen kernel LRV estimator for the AR(1) case with $T = 100$, number of joint hypotheses $p$ and number of overidentifying restrictions $q$

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>$\chi^2$</th>
<th>$F_{\infty}[0]$</th>
<th>$F_{\infty}[q]$</th>
<th>$\chi^2$</th>
<th>$F_{\infty}[0]$</th>
<th>$F_{\infty}[q]$</th>
<th>$\chi^2$</th>
<th>$F_{\infty}[0]$</th>
<th>$F_{\infty}[q]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p = 1, q = 2$</td>
<td>$p = 2, q = 2$</td>
<td>$p = 3, q = 2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.8</td>
<td>0.257</td>
<td>0.195</td>
<td>0.115</td>
<td>0.406</td>
<td>0.242</td>
<td>0.139</td>
<td>0.555</td>
<td>0.305</td>
<td>0.166</td>
</tr>
<tr>
<td>-0.5</td>
<td>0.128</td>
<td>0.110</td>
<td>0.078</td>
<td>0.188</td>
<td>0.131</td>
<td>0.091</td>
<td>0.248</td>
<td>0.155</td>
<td>0.103</td>
</tr>
<tr>
<td>0.0</td>
<td>0.086</td>
<td>0.079</td>
<td>0.060</td>
<td>0.114</td>
<td>0.091</td>
<td>0.070</td>
<td>0.150</td>
<td>0.108</td>
<td>0.080</td>
</tr>
<tr>
<td>0.5</td>
<td>0.139</td>
<td>0.115</td>
<td>0.081</td>
<td>0.210</td>
<td>0.140</td>
<td>0.088</td>
<td>0.291</td>
<td>0.167</td>
<td>0.105</td>
</tr>
<tr>
<td>0.8</td>
<td>0.282</td>
<td>0.201</td>
<td>0.113</td>
<td>0.456</td>
<td>0.260</td>
<td>0.135</td>
<td>0.613</td>
<td>0.319</td>
<td>0.157</td>
</tr>
<tr>
<td>0.9</td>
<td>0.384</td>
<td>0.269</td>
<td>0.144</td>
<td>0.605</td>
<td>0.365</td>
<td>0.190</td>
<td>0.774</td>
<td>0.438</td>
<td>0.221</td>
</tr>
</tbody>
</table>
Outline

1. Basic Setting and Problem
   - GMM Estimation
   - Optimal Weighting Matrix
   - Two-step Estimation and Testing

2. Fixed Smoothing Asymptotics and its Representation
   - Big Picture
   - Main Results
   - Representation and Approximation: Series HAR
   - Representation and Approximation: Kernel HAR

3. Implementation, Simulation and Empirical Application
   - Practical Implementation
   - Simulation
   - Empirical Application

4. Conclusions
Estimation and Inference in SVM: the Model

- SVM

\[ r_t = \sigma_t z_t \]
\[ \log \sigma_t^2 = \omega + \beta (\log \sigma_{t-1}^2 - \omega) + \sigma_u u_t \]

where \( r_t \) is the rate of return and \( (z_t, u_t) \) is iid \( N(0, I_2) \).

- The parameter vector is \( \theta = (\omega, \beta, \sigma_u) \). We impose the restriction that \( \beta \in (0, 1) \), which is an empirically relevant range.

- The model and the parameter restriction are the same as those considered by Andersen and Sorensen (1996), which gives a detailed discussion on the motivation of the stochastic volatility models and the GMM approach.

- Alternative to ARCH and GARCH.
We employ the GMM to estimate the log-normal stochastic volatility model; MLE is too complicated.

The data are weekly returns of S&P 500 stocks, which are constructed by compounding daily returns with dividends from a CRSP index file.

We consider both value-weighted returns (vwretd) and equal-weighted returns (ewretd).

The weekly returns range from the first week of 2001 to the last week of 2012 with sample size $T = 627$.

We use weekly data in order to minimize problems associated with daily data such as asynchronous trading and bid-ask bounce. This is consistent with Jacquier, Polson and Rossi (1994).
Estimation and Inference in SVM: the Moments

- The GMM approach relies on functions of the time series \( \{r_t\} \) to identify the parameters of the model.

- For the simple log-normal stochastic volatility model, it is easy to obtain the following moment conditions:

\[
E |r_t|^\ell = c_\ell E \left( \sigma_t^{\ell} \right) \quad \text{for } \ell = 1, 2, 3, 4 \text{ with constants } (c_1, c_2, c_3, c_4),
\]

\[
E |r_t r_{t-j}| = 2\pi^{-1} E (\sigma_t \sigma_{t-j}) \quad \text{and } E r_t^2 r_{t-j}^2 = E (\sigma_t^2 \sigma_{t-j}^2) \quad \text{for } j = 1, 2, ...
\]

where

\[
E \left( \sigma_t^{\ell} \right) = \exp \left[ \frac{\omega \ell}{2} + \frac{\ell^2 \sigma_u^2}{8(1 - \beta^2)} \right]
\]

\[
E \left( \sigma_t^{\ell_1} \sigma_{t-j}^{\ell_2} \right) = E \left( \sigma_t^{\ell_1} \right) E \left( \sigma_t^{\ell_2} \right) \exp \left[ \frac{\sigma_u^2 \ell_1 \ell_2 \beta^j}{4(1 - \beta^2)} \right]
\]
Higher order moments can be computed but we choose to focus on a subset of lower order moments.

Andersen and Sorensen (1996) points out that it is generally not optimal to include too many moment conditions when the sample is limited. On the other hand, it is not advisable to include just as many moment conditions as the number of parameters.

When $T = 500$ and $\theta = (-7.36, 0.90, .363)$, which is an empirically relevant parameter vector, Table 1 in Andersen and Sorensen (1996, Table 1) shows that it is MSE-optimal to employ nine moment conditions.
For this reason, we employ two sets of nine moment conditions given in the appendix of Andersen and Sorensen (1996). The baseline set of the nine moment conditions are
\[ Ef_i(r_t, \theta) = 0 \text{ for } i = 1, \ldots, 9 \]
with

\[
\begin{align*}
    f_1(r_t, \theta) &= |r_t|^i - c_\ell E\left(\sigma_t^\ell\right), \ i = 1, \ldots, 4 \\
    f_5(r_t, \theta) &= |r_t r_{t-1}| - 2\pi^{-1} E(\sigma_t \sigma_{t-1}) \\
    f_6(r_t, \theta) &= |r_t r_{t-3}| - 2\pi^{-1} E(\sigma_t \sigma_{t-3}) \\
    f_7(r_t, \theta) &= |r_t r_{t-5}| - 2\pi^{-1} E(\sigma_t \sigma_{t-5}) \\
    f_8(r_t, \theta) &= r_t^2 r_{t-2}^2 - E(\sigma_t^2 \sigma_{t-2}^2) \\
    f_9(r_t, \theta) &= r_t^2 r_{t-4}^2 - E(\sigma_t^2 \sigma_{t-4}^2)
\end{align*}
\]
The alternative set of the nine moment conditions are
\[ E f_i(r_t, \theta) = 0 \] for \( i = 1, \ldots, 9 \) with

\[
\begin{align*}
  f_i(r_t, \theta) &= |r_t|^i - c_\ell E \left( \sigma^\ell_t \right), \quad i = 1, \ldots, 4 \\
  f_5(r_t, \theta) &= |r_t r_{t-2}| - 2\pi^{-1} E (\sigma_t \sigma_{t-2}) \\
  f_6(r_t, \theta) &= |r_t r_{t-4}| - 2\pi^{-1} E (\sigma_t \sigma_{t-4}) \\
  f_7(r_t, \theta) &= |r_t r_{t-6}| - 2\pi^{-1} E (\sigma_t \sigma_{t-6}) \\
  f_8(r_t, \theta) &= r_t^2 r_{t-1}^2 - E (\sigma^2_t \sigma^2_{t-1}) \\
  f_9(r_t, \theta) &= r_t^2 r_{t-3}^2 - E (\sigma^2_t \sigma^2_{t-3})
\end{align*}
\]

In each case \( m = 9 \) and \( d = 3 \), and so the degree of overidentification is \( q = m - d = 6 \).
We focus on constructing 90% and 95% confidence intervals (CI’s) for $\beta$.

Given the high nonlinearity of the moment conditions, we invert the GMM distance statistics to obtain the CI’s, which are invariant to the reparametrization of model parameters.

For example, a 95% confidence interval is the set of $\beta$ values in (0,1) that the $\mathbb{D}_T$ test does not reject at the 5% level.

We search over the grid from 0.01 to 0.99 with increment 0.01 to invert the $\mathbb{D}_T$ test.

As in the simulation study, we employ three different critical values: $\chi_p^{1-\alpha} / p$, $\mathcal{F}_\infty^{1-\alpha} [0]$, and $\mathcal{F}_\infty^{1-\alpha} [q]$, which correspond to three different asymptotic approximations. For the series LRV estimator, $\mathcal{F}_\infty^{1-\alpha} [0]$ is a critical value from the corresponding $F$ approximation.
### Estimation and Inference in SVM

<table>
<thead>
<tr>
<th></th>
<th>Equally-weighted Return</th>
<th>Value-weighted Return</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Baseline</td>
<td>Alternative</td>
</tr>
<tr>
<td>Series $\chi^2$</td>
<td>0.74</td>
<td>0.58</td>
</tr>
<tr>
<td>Series $F_\infty [0]$</td>
<td>0.73</td>
<td>0.57</td>
</tr>
<tr>
<td>Series $F_\infty [q]$</td>
<td>0.70</td>
<td>0.53</td>
</tr>
<tr>
<td>K</td>
<td>(140)</td>
<td>(138)</td>
</tr>
<tr>
<td>Bartlett $\chi^2$</td>
<td>0.56</td>
<td>0.52</td>
</tr>
<tr>
<td>Bartlett $F_\infty [0]$</td>
<td>0.54</td>
<td>0.52</td>
</tr>
<tr>
<td>Bartlett $F_\infty [q]$</td>
<td><strong>0.35</strong></td>
<td>0.42</td>
</tr>
<tr>
<td>b</td>
<td>(.0155)</td>
<td>(.0157)</td>
</tr>
</tbody>
</table>

*Yixiao Sun, UC San Diego*

Accurate Asymptotic Approximation in the Optimal GMM Framework
## Estimation and Inference in SVM

<table>
<thead>
<tr>
<th></th>
<th>Equally-weighted Return</th>
<th>Value-weighted Return</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Baseline</td>
<td>Alternative</td>
</tr>
<tr>
<td><strong>Parzen $\chi^2$</strong></td>
<td>0.72</td>
<td>0.52</td>
</tr>
<tr>
<td><strong>Parzen $F_\infty [0]$</strong></td>
<td>0.73</td>
<td>0.52</td>
</tr>
<tr>
<td><strong>Parzen $F_\infty [q]$</strong></td>
<td>0.69</td>
<td>0.47</td>
</tr>
<tr>
<td><strong>b</strong></td>
<td>(.0136)</td>
<td>(.0139)</td>
</tr>
<tr>
<td><strong>QS $\chi^2$</strong></td>
<td>0.74</td>
<td>0.52</td>
</tr>
<tr>
<td><strong>QS $F_\infty [0]$</strong></td>
<td>0.74</td>
<td>0.52</td>
</tr>
<tr>
<td><strong>QS $F_\infty [q]$</strong></td>
<td>0.71</td>
<td>0.48</td>
</tr>
<tr>
<td><strong>b</strong></td>
<td>(.0068)</td>
<td>(.0069)</td>
</tr>
</tbody>
</table>
There is a noticeable difference between the CI’s based on the $\chi^2$ approximation and the nonstandard $F_\infty[q]$ approximation. This is especially true for the case with baseline moment conditions, the Bartlett kernel, and equally-weighted returns.

Taking into account the randomness of the estimated weighting matrix leads to wider CI’s. The log-volatility may not be as persistent as previously thought.
The paper has developed new, more accurate, and convenient asymptotic approximations for heteroskedasticity and autocorrelation robust inference in a two-step GMM framework.

It is more accurate than the chi-square or normal approximation, especially true when

- the number of joint hypotheses being tested is large
- the degree of over-identification is high
- the underlying processes are persistent.

I hope this line of research can help deliver more reliable statistical inference.
Possible Extensions

- The results of the paper can be extended easily to the spatial setting, spatial-temporal setting or panel data setting. See Kim and Sun (2014) for a paper in the spatial setting.

- The general results of the paper can be extended to a nonparametric sieve GMM framework. See Chen, Liao and Sun (2014) for a recent development on autocorrelation robust sieve inference for time series models.

- Now that we have a new and more accurate asymptotic framework, we can give a more honest assessment of the relative merits of one-step and two-step GMM procedures. Preliminary results are available in Hwang and Sun (2014).
Thank you!