LECTURE ON HAC COVARIANCE MATRIX ESTIMATION AND THE KVB APPROACH

CHUNG-MING KUAN

Institute of Economics

Academia Sinica

October 20, 2006

ckuan@econ.sinica.edu.tw

www.sinica.edu.tw/~ckuan

Outline

- Preliminary
- HAC (heteroskedasticity and autocorrelation consistent) covariance matrix estimation
  - ◇ Kernel HAC estimators
  - ◇ Choices of the kernel function and bandwidth
- KVB Approach: Constructing asymptotically pivotal tests without consistent estimation of asymptotic covariance matrix
  - ◇ Tests of parameters
  - ◇ M tests for general moment conditions
  - ◇ Over-identifying restrictions tests

Preliminary

- Consider the specification: $y_t = \boldsymbol{x}_t'\boldsymbol{\beta} + e_t$.

[A1] For some $\boldsymbol{\beta}_o$, $\epsilon_t = y_t - \boldsymbol{x}_t'\boldsymbol{\beta}_o$ such that $\mathbb{E}(\boldsymbol{x}_t\epsilon_t) = \mathbf{0}$ and

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]} \boldsymbol{x}_t\epsilon_t \Rightarrow \boldsymbol{S}_o\boldsymbol{W}_k(r), \quad r \in [0,1],$$

$\boldsymbol{\Sigma}_o = \lim_{T\to\infty} \operatorname{var}\big(T^{-1/2}\sum_{t=1}^{T}\boldsymbol{x}_t\epsilon_t\big)$ with $\boldsymbol{S}_o$ its matrix square root (i.e., $\boldsymbol{\Sigma}_o = \boldsymbol{S}_o\boldsymbol{S}_o'$).

[A2] $\boldsymbol{M}_{[Tr]} := [Tr]^{-1}\sum_{t=1}^{[Tr]}\boldsymbol{x}_t\boldsymbol{x}_t' \xrightarrow{\mathbb{P}} \boldsymbol{M}_o$ uniformly in $r \in (0,1]$.

- Asymptotic normality of the OLS estimator:

$$\sqrt{T}(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_o) = M_T^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\boldsymbol{x}_t\epsilon_t \xrightarrow{D} \boldsymbol{M}_o^{-1}\boldsymbol{S}_o\boldsymbol{W}_k(1),$$

which has the distribution $\mathcal{N}(\mathbf{0},\, \boldsymbol{M}_o^{-1}\boldsymbol{\Sigma}_o\boldsymbol{M}_o^{-1})$.

- The null hypothesis is $\boldsymbol{R}\boldsymbol{\beta}_o = \boldsymbol{r}$, where $\boldsymbol{R}$ $(q \times k)$ has full row rank.
  - ◇ $\sqrt{T}\big(\boldsymbol{R}\hat{\boldsymbol{\beta}}_T - \boldsymbol{r}\big) \xrightarrow{D} \mathcal{N}(\mathbf{0},\, \boldsymbol{R}\boldsymbol{M}_o^{-1}\boldsymbol{\Sigma}_o\boldsymbol{M}_o^{-1}\boldsymbol{R}')$.
  - ◇ $\big(\boldsymbol{R}\boldsymbol{M}_T^{-1}\widehat{\boldsymbol{\Sigma}}_T\boldsymbol{M}_T^{-1}\boldsymbol{R}'\big)^{-1/2}\sqrt{T}\big(\boldsymbol{R}\hat{\boldsymbol{\beta}}_T - \boldsymbol{r}\big) \xrightarrow{D} \mathcal{N}(\mathbf{0},\, \boldsymbol{I}_q)$.
  - ◇ The Wald test is

$$\mathcal{W}_T = T\big(\boldsymbol{R}\hat{\boldsymbol{\beta}}_T - \boldsymbol{r}\big)'\big(\boldsymbol{R}\boldsymbol{M}_T^{-1}\widehat{\boldsymbol{\Sigma}}_T\boldsymbol{M}_T^{-1}\boldsymbol{R}'\big)^{-1}\big(\boldsymbol{R}\hat{\boldsymbol{\beta}}_T - \boldsymbol{r}\big) \xrightarrow{D} \chi^2(q).$$

- Consistent estimation of $\boldsymbol{\Sigma}_o$ is crucial; $\mathcal{W}_T$ would not have a limiting χ^2 distribution if $\widehat{\boldsymbol{\Sigma}}_T$ is not a consistent estimator.

- Othe large sample tests (e.g., the LM test) also depend on consistent estimation of the asymptotic covariance matrix.

- When $\widehat{\boldsymbol{\Sigma}}_T$ is consistent, the resulting tests are said to be robust to heteroskedasticity and serial correlations of unknown form.

Asymptotic Covariance Matrix $\boldsymbol{\Sigma}_o$

A general form:

$$\boldsymbol{\Sigma}_o = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} \mathbb{E}(\boldsymbol{x}_t \epsilon_t \epsilon_s \boldsymbol{x}_s') = \lim_{T\to\infty} \sum_{j=-T+1}^{T-1} \boldsymbol{\Gamma}_T(j),$$

with autocovariances:

$$\boldsymbol{\Gamma}_T(j) = \begin{cases} \frac{1}{T}\sum_{t=j+1}^{T} \mathbb{E}(\boldsymbol{x}_t \epsilon_t \epsilon_{t-j} \boldsymbol{x}_{t-j}'), & j = 0, 1, 2, \ldots, \\ \frac{1}{T}\sum_{t=-j+1}^{T} \mathbb{E}(\boldsymbol{x}_{t+j} \epsilon_{t+j} \epsilon_t \boldsymbol{x}_t'), & j = -1, -2, \ldots. \end{cases}$$

- When $\boldsymbol{x}_t \epsilon_t$ are covariance stationary, $\boldsymbol{\Gamma}_T(j) = \boldsymbol{\Gamma}(j)$, and the spectral density of $\boldsymbol{x}_t \epsilon_t$ at frequency ω is

$$f(\omega) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \boldsymbol{\Gamma}(j) e^{-ij\omega}.$$

- $\boldsymbol{\Sigma}_o = 2\pi\, f(0)$ and hence is also known as the long-run variance of $\boldsymbol{x}_t \epsilon_t$.

Examples of Σ_o

- When $\boldsymbol{x}_t\epsilon_t$ are serially uncorrelated:

$$\boldsymbol{\Sigma}_o = \lim_{T\to\infty} \boldsymbol{\Gamma}_T(0) = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}(\epsilon_t^2 \boldsymbol{x}_t \boldsymbol{x}_t'),$$

which can be consistently estimated by White's heteroskedasticity-consistent estimator: $\widehat{\boldsymbol{\Sigma}}_T = T^{-1}\sum_{t=1}^{T} \hat{e}_t^2 \boldsymbol{x}_t \boldsymbol{x}_t'$.

- When $\boldsymbol{x}_t\epsilon_t$ are serially uncorrelated and ϵ_t conditionally homoskedastic:

$$\boldsymbol{\Sigma}_o = \sigma_o^2 \left(\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}(\boldsymbol{x}_t \boldsymbol{x}_t') \right) = \sigma_o^2 \boldsymbol{M}_o,$$

which can be consistently estimated by $\widehat{\boldsymbol{\Sigma}}_T = \hat{\sigma}_T^2 \boldsymbol{M}_T$.

- Estimating $\boldsymbol{\Sigma}_o$ is much more difficult when heteroskedasticity and serial correlations are present and of unknown form.

A Consistent Estimator of $\boldsymbol{\Sigma}_o$

- Recall $\boldsymbol{\Sigma}_o = \lim_{T\to\infty} \sum_{j=-T+1}^{T-1} \boldsymbol{\Gamma}_T(j)$.
- A consistent estimator (White, 1984):

$$\widehat{\boldsymbol{\Sigma}}_T^{\dagger} = \sum_{j=-\ell(T)}^{\ell(T)} \widehat{\boldsymbol{\Gamma}}_T(j),$$

with sample autocovariances

$$\widehat{\boldsymbol{\Gamma}}_T(j) = \begin{cases} \frac{1}{T}\sum_{t=j+1}^{T} \boldsymbol{x}_t \hat{e}_t \hat{e}_{t-j} \boldsymbol{x}_{t-j}', & j = 0,1,2,\ldots, \\ \frac{1}{T}\sum_{t=-j+1}^{T} \boldsymbol{x}_{t+j} \hat{e}_{t+j} \hat{e}_t \boldsymbol{x}_t', & j = -1,-2,\ldots. \end{cases}$$

where $\ell(T)$ grows with T but at a slower rate.

- Drawbacks:
  - ◇ This estimator is not guaranteed to be a positive semi-definite matrix.
  - ◇ Must determine $\ell(T)$ in a given sample.

Kernel HAC Estimators

- Newey and West (1987) and Gallant (1987): A consistent estimator that is also positive semi-definite is:

$$\widehat{\boldsymbol{\Sigma}}_T^{\kappa} = \sum_{j=-T+1}^{T-1} \kappa\Big(\frac{j}{\ell(T)}\Big)\widehat{\boldsymbol{\Gamma}}_T(j),$$

where κ is a kernel function and $\ell(T)$ is its bandwidth.

- κ and $\ell(T)$ jointly determine the weighting scheme on $\widehat{\boldsymbol{\Gamma}}_T(j)$ and must be selected by users.
  - ◇ For a given j, $\kappa(j/\ell(T)) \approx 1$ as $T \to \infty$, so as to ensure consistency.
  - ◇ For a given T, $\kappa(j/\ell(T))$ should be small when j is large, so as to ensure positive semi-definiteness. More formally (Andrews, 1991),

$$\int_{-\infty}^{\infty} \kappa(x) e^{-ix\omega}\, \mathrm{d}x \geq 0, \quad \forall \omega \in \mathbb{R}.$$

Kernel Functions

- Bartlett kernel (Newey and West, 1987):

$$\kappa(x) = \begin{cases} 1 - |x|, & |x| \le 1, \\ 0, & \text{otherwise;} \end{cases}$$

- Parzen kernel (Gallant, 1987):

$$\kappa(x) = \begin{cases} 1 - 6x^2 + 6|x|^3, & |x| \le 1/2, \\ 2(1 - |x|)^3, & 1/2 \le |x| \le 1, \\ 0, & \text{otherwise;} \end{cases}$$

- Quadratic spectral kernel (Andrews, 1991):

$$\kappa(x) = \frac{25}{12\pi^2 x^2} \left(\frac{\sin(6\pi x/5)}{6\pi x/5} - \cos(6\pi x/5) \right);$$

- Daniel kernel (Ng and Perron, 1996):

$$\kappa(x) = \frac{\sin(\pi x)}{\pi x}.$$

Bandwidths

Andrews (1991):

- By minimizing MSE, the optimal bandwidth growth rates are:

$$\ell^*(T) = 1.1447(c_1 T)^{1/3}, \quad \text{(Bartlett)},$$

$$\ell^*(T) = 2.6614(c_2 T)^{1/5}, \quad \text{(Parzen)},$$

$$\ell^*(T) = 1.3221(c_2 T)^{1/5}, \quad \text{(quadractic spectral)},$$

  c_1 and c_2 are unknown numbers depending on the spectral density, and

$$\sqrt{T/\ell^*(T)}\,(\hat{\boldsymbol{\Sigma}}_T^{\kappa} - \boldsymbol{\Sigma}_o) = O_{\mathbb{P}}(1).$$

- As far as MSE is concerned, the quadratic spectral kernel is to be preferred.
  - ◇ $\widehat{\boldsymbol{\Sigma}}_T^{B} = O_{\mathbb{P}}(T^{-1/3})$; $\widehat{\boldsymbol{\Sigma}}_T^{P}$ nd $\widehat{\boldsymbol{\Sigma}}_T^{QS}$ are $O_{\mathbb{P}}(T^{-2/5})$.
  - ◇ $\widehat{\boldsymbol{\Sigma}}_T^{QS}$ is more efficient than $\widehat{\boldsymbol{\Sigma}}_T^{P}$; $\widehat{\boldsymbol{\Sigma}}_T^{B}$ is the least efficient.

Problems with the Kernel Estimators

- The performance of the kernel HAC estimator varies with the choices of the kernel and its bandwidth.
  - ◇ The kernel weighting scheme yields negative bias, and such bias could be substantial in finite samples.
  - ◇ The tests based on the HAC estimators usually over-reject the null.
- The choices of kernel and bandwidth are somewhat arbitrary in practice, and hence the statistical inferences are vulnerable.
  - ◇ The HAC estimator with the quadratic spectral kernel need not have better performance in finite samples.
  - ◇ Andrews (1991) suggested a "plug-in" method to estimate the optimal growth rates $\ell^*(T)$, but this method requires estimation of a user-selected model to determine c_1 and c_2.

Other Improved HAC Estimators

- Andrews and Monahan (1992): Pre-whitened estimator.
  - ◇ Apply a VAR model to whiten $\boldsymbol{x}_t \hat{e}_t$ and estimate the covariance matrix based on its residuals.
  - ◇ The choices of the model for pre-whitening and VAR lag order are, again, arbitrary.
- Kuan and Hsieh (2006): Computing sample autocovariances based on forecast errors $(y_t - \boldsymbol{x}_t' \tilde{\boldsymbol{\beta}}_{t-1})$, instead of the OLS residuals.
  - ◇ It does not require another user-chosen parameter.
  - ◇ It yields a smaller bias (but a larger MSE); the resulting tests have more accurate test size without sacrificing test power.
  - ◇ Bias reduction seems more important for improving HAC estimators.

KVB Approach

Kiefer, Vogelsang, and Bunzel (2000): A Wald-type test is

$$\mathcal{W}_T^{\dagger} = T(\boldsymbol{R}\hat{\boldsymbol{\beta}}_T - \boldsymbol{r})'\left(\boldsymbol{R}\boldsymbol{M}_T^{-1}\widehat{\boldsymbol{C}}_T\boldsymbol{M}_T^{-1}\boldsymbol{R}'\right)^{-1}(\boldsymbol{R}\hat{\boldsymbol{\beta}}_T - \boldsymbol{r}),$$

where a normalizing matrix $\widehat{\boldsymbol{C}}_T$ is used in place of $\widehat{\boldsymbol{\Sigma}}_T^{\kappa}$.

- $\widehat{\boldsymbol{C}}_T$ is inconsistent for $\boldsymbol{\Sigma}_o$ but is able to eliminate the nuisance parameters in $\boldsymbol{\Sigma}_o$.
- Advantages:
  - ◇ Do not have to choose a kernel bandwidth.
  - ◇ The resulting test remains pivotal asymptotically.
  - ◇ The limiting distribution of the test approximates the finite-sample distribution very well (i.e., little size distortion).

KVB's Normalizing Matrix

- Let $\hat{\boldsymbol{\varphi}}_t = T^{-1/2} \sum_{i=1}^{t} \boldsymbol{x}_i \hat{e}_i$. The normalizing matrix $\widehat{\boldsymbol{C}}_T$ is

$$\widehat{\boldsymbol{C}}_T = \frac{1}{T} \sum_{t=1}^{T} \hat{\boldsymbol{\varphi}}_t \hat{\boldsymbol{\varphi}}_t' = \frac{1}{T^2} \sum_{t=1}^{T} \left(\sum_{i=1}^{t} \boldsymbol{x}_i \hat{e}_i \right) \left(\sum_{i=1}^{t} \hat{e}_i \boldsymbol{x}_i' \right).$$

- The limit of $\hat{\boldsymbol{\varphi}}_{[Tr]}$:

$$\begin{aligned}
\hat{\boldsymbol{\varphi}}_{[Tr]} &= \frac{1}{\sqrt{T}} \sum_{i=1}^{[Tr]} \boldsymbol{x}_i \epsilon_i - \frac{[Tr]}{T} \left(\frac{1}{[Tr]} \sum_{i=1}^{[Tr]} \boldsymbol{x}_i \boldsymbol{x}_i' \right) \sqrt{T} (\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_o) \\
&\Rightarrow \boldsymbol{S}_o \boldsymbol{W}_k(r) - r \boldsymbol{M}_o \boldsymbol{M}_o^{-1} \boldsymbol{S}_o \boldsymbol{W}_k(1) \\
&= \boldsymbol{S}_o \boldsymbol{B}_k(r).
\end{aligned}$$

Hence,

$$\widehat{\boldsymbol{C}}_T \Rightarrow \boldsymbol{S}_o \left(\int_0^1 \boldsymbol{B}_k(r) \boldsymbol{B}_k(r)' \, \mathrm{d}r \right) \boldsymbol{S}_o' =: \boldsymbol{S}_o \boldsymbol{P}_k \boldsymbol{S}_o'.$$

- Let $\boldsymbol{G}_o$ denote the matrix square root of $\boldsymbol{R}\boldsymbol{M}_o^{-1}\boldsymbol{S}_o\boldsymbol{S}_o'\boldsymbol{M}_o^{-1}\boldsymbol{R}'$. Then,

$$\boldsymbol{R}\boldsymbol{M}_T^{-1}\widehat{\boldsymbol{C}}_T\boldsymbol{M}_T^{-1}\boldsymbol{R}' \Rightarrow \boldsymbol{R}\boldsymbol{M}_o^{-1}\boldsymbol{S}_o\boldsymbol{P}_k\boldsymbol{S}_o'\boldsymbol{M}_o^{-1}\boldsymbol{R}' \stackrel{d}{=} \boldsymbol{G}_o\boldsymbol{P}_q\boldsymbol{G}_o'.$$

and $\sqrt{T}\boldsymbol{R}(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_o) \xrightarrow{D} \boldsymbol{R}\boldsymbol{M}_o^{-1}\boldsymbol{S}_o\boldsymbol{W}_k(1) \stackrel{d}{=} \boldsymbol{G}_o\boldsymbol{W}_q(1)$.

- $\mathcal{W}_T^\dagger$ is thus asymptotically pivotal:

$$\mathcal{W}_T^\dagger \Rightarrow \boldsymbol{W}_q(1)'\boldsymbol{G}_o'\big(\boldsymbol{G}_o\boldsymbol{P}_q\boldsymbol{G}_o'\big)^{-1}\boldsymbol{G}_o\boldsymbol{W}_q(1) = \boldsymbol{W}_q(1)'\boldsymbol{P}_q^{-1}\boldsymbol{W}_q(1).$$

Lobato (2001) reported some quantiles, cf. Kiefer et al. (2000).

- For the null of $\beta_i = r$, a t-type test is

$$t^\dagger = \frac{\sqrt{T}\big(\hat{\beta}_{i,T} - r\big)}{\sqrt{\hat{\delta}_i}} \xrightarrow{D} \frac{W(1)}{\big[\int_0^1 B(r)^2\,\mathsf{d}r\big]^{1/2}}.$$

This distribution is more disperse than the standard normal distribution.

Kernel-Based Normalizing Matrices

- Kiefer and Vogelsang (2002a): $2\widehat{\boldsymbol{C}}_T = \widehat{\boldsymbol{\Sigma}}_T^B$ without truncation, i.e., $\ell(T) = T$. The usual Wald test based on $\widehat{\boldsymbol{\Sigma}}_T^B$ without truncation is thus the same as $\mathcal{W}_T^\dagger/2$. In particular, the t test based on $\widehat{\boldsymbol{\Sigma}}_T^B$ without truncation is also $t^\dagger/\sqrt{2}$.

- Kiefer and Vogelsang (2002b): $\widehat{\boldsymbol{\Sigma}}_T^\kappa \Rightarrow \boldsymbol{S}_o \boldsymbol{P}_k^\kappa \boldsymbol{S}_o'$, with

$$\boldsymbol{P}_k^\kappa = -\int_0^1 \int_0^1 \kappa''(r-s) \boldsymbol{B}_k(r) \boldsymbol{B}_k(s)' \,\mathrm{d}r\,\mathrm{d}s;$$

  The Wald test based on $\widehat{\boldsymbol{\Sigma}}_T^\kappa$ without truncation can also serve as a KVB's robust test.

- A test based on $\widehat{\boldsymbol{\Sigma}}_T^B$ without truncation compares favorably with that based on $\widehat{\boldsymbol{\Sigma}}_T^{QS}$ in terms of test power. Hence, the Bartlett kernel is to be preferred in constructing a KVB test, in contrast with HAC estimation.

M Tests

The null hypothesis: $\mathbb{E}[\boldsymbol{f}(\boldsymbol{\eta}_t;\boldsymbol{\theta}_o)] = \mathbf{0}$, where $\boldsymbol{\theta}_o$ is the $k \times 1$ true parameter vector, and $\boldsymbol{f}$ is a $q \times 1$ vector of functions.

$\boldsymbol{\theta}_o$ Is Known

Define $\boldsymbol{m}_{[rT]}(\boldsymbol{\theta}) = T^{-1}\sum_{t=1}^{[rT]} \boldsymbol{f}(\boldsymbol{\eta}_t;\boldsymbol{\theta})$, for $r \in (0,1]$.

- An M test is based on $\boldsymbol{m}_T(\boldsymbol{\theta}_o)$, the sample counterpart of the null.
- By a CLT, $T^{1/2}\boldsymbol{m}_T(\boldsymbol{\theta}_o) \xrightarrow{D} \mathcal{N}(\mathbf{0},\ \boldsymbol{\Sigma}_o)$, and the conventional M test is:

  $$T\,\boldsymbol{m}_T(\boldsymbol{\theta}_o)'\widehat{\boldsymbol{\Sigma}}_T^{-1}\boldsymbol{m}_T(\boldsymbol{\theta}_o) \xrightarrow{D} \chi^2(q),$$

  when $\widehat{\boldsymbol{\Sigma}}_T$ is a consistent estimator of $\boldsymbol{\Sigma}_o$.
- The limiting χ^2 distribution hinges on consistent estimation of $\boldsymbol{\Sigma}_o$.

[B1](a) Under the null, $\sqrt{T}\boldsymbol{m}_{[rT]}(\boldsymbol{\theta}_o) \Rightarrow \boldsymbol{S}_o \boldsymbol{W}_q(r)$ for $0 \le r \le 1$, where $\boldsymbol{S}_o$ is the nonsingular, matrix square root of $\boldsymbol{\Sigma}_o$.

- $\boldsymbol{C}_T(\boldsymbol{\theta}_o) = T^{-1}\sum_{t=1}^{T} \boldsymbol{\varphi}_t(\boldsymbol{\theta}_o)\boldsymbol{\varphi}_t(\boldsymbol{\theta}_o)'$ with

$$\boldsymbol{\varphi}_t(\boldsymbol{\theta}_o) = \frac{1}{\sqrt{T}} \sum_{i=1}^{t} \big[\boldsymbol{f}(\boldsymbol{\eta}_i; \boldsymbol{\theta}_o) - \boldsymbol{m}_T(\boldsymbol{\theta}_o)\big].$$

- Analogous to KVB's Wald-type test, an M test is

$$\mathcal{M}_T = T\, \boldsymbol{m}_T(\boldsymbol{\theta}_o)' \boldsymbol{C}_T(\boldsymbol{\theta}_o)^{-1} \boldsymbol{m}_T(\boldsymbol{\theta}_o) \xrightarrow{D} \boldsymbol{W}_q(1)' \boldsymbol{P}_q^{-1} \boldsymbol{W}_q(1).$$

  - ◇ By [B1](a), $T^{1/2}\boldsymbol{m}_T(\boldsymbol{\theta}_o) \Rightarrow \boldsymbol{S}_o \boldsymbol{W}_q(1)$.
  - ◇ $\boldsymbol{\varphi}_{[rT]}(\boldsymbol{\theta}_o) \Rightarrow \boldsymbol{S}_o\big[\boldsymbol{W}_q(r) - r\boldsymbol{W}_q(1)\big] = \boldsymbol{S}_o \boldsymbol{B}_q(r)$, $0 \le r \le 1$.
  - ◇ $\boldsymbol{C}_T(\boldsymbol{\theta}_o) \Rightarrow \boldsymbol{S}_o \boldsymbol{P}_q \boldsymbol{S}_o'$ with $\boldsymbol{P}_q = \int_0^1 \boldsymbol{B}_q(r)\boldsymbol{B}_q(r)'\,\mathrm{d}r$.

$\boldsymbol{\theta}_o$ Is Unknown

- Replacing $\boldsymbol{\theta}_o$ in $\boldsymbol{m}_T$ and $\boldsymbol{\varphi}_t$ with a root-T consistent estimator $\hat{\boldsymbol{\theta}}_T$ that satisfies

$$\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_o) = \boldsymbol{Q}_o \left[\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \boldsymbol{q}(\boldsymbol{\eta}_t; \boldsymbol{\theta}_o) \right] + o_{\mathbb{P}}(1).$$

- Kuan and Lee (2006): The M test is

$$\widehat{\mathcal{M}}_T = T\, \boldsymbol{m}_T(\hat{\boldsymbol{\theta}}_T)' \widehat{\boldsymbol{C}}_T^{-1} \boldsymbol{m}_T(\hat{\boldsymbol{\theta}}_T),$$

where $\widehat{\boldsymbol{C}}_T = \boldsymbol{C}_T(\hat{\boldsymbol{\theta}}_T) = T^{-1} \sum_{t=1}^{T} \boldsymbol{\varphi}_t(\hat{\boldsymbol{\theta}}_T) \boldsymbol{\varphi}_t(\hat{\boldsymbol{\theta}}_T)'$ with

$$\boldsymbol{\varphi}_t(\hat{\boldsymbol{\theta}}_T) = \frac{1}{\sqrt{T}} \sum_{i=1}^{t} \big[\boldsymbol{f}(\boldsymbol{\eta}_i; \hat{\boldsymbol{\theta}}_T) - \boldsymbol{m}_T(\hat{\boldsymbol{\theta}}_T) \big].$$

- The limit of $\widehat{\mathcal{M}}_T$ depends on the estimation effect of replacing $\boldsymbol{\theta}_o$ with $\hat{\boldsymbol{\theta}}_T$.

[B1](b) Under the null,

$$\begin{bmatrix} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \boldsymbol{f}(\boldsymbol{\eta}_t; \boldsymbol{\theta}_o) \\ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \boldsymbol{q}(\boldsymbol{\eta}_t; \boldsymbol{\theta}_o) \end{bmatrix} \Rightarrow \boldsymbol{G}_o \boldsymbol{W}_{q+k}(1),$$

where $\boldsymbol{G}_o$ is nonsingular.

[B2] $\boldsymbol{F}_{[rT]}(\boldsymbol{\theta}_o) = [rT]^{-1} \sum_{t=1}^{[rT]} \nabla_{\boldsymbol{\theta}} \boldsymbol{f}\big(\boldsymbol{\eta}_t; \boldsymbol{\theta}_o\big) \xrightarrow{\mathbb{P}} \boldsymbol{F}_o$, uniformly in $r \in (0, 1]$, where $\boldsymbol{F}_o$ is a $q \times k$ non-stochastic matrix; $\nabla_{\boldsymbol{\theta}} \boldsymbol{F}_{[rT]}(\boldsymbol{\theta}_o) = O_{\mathbb{P}}(1)$.

- A Taylor expansion about $\boldsymbol{\theta}_o$ gives

$$\sqrt{T} \boldsymbol{m}_{[rT]}(\hat{\boldsymbol{\theta}}_T) = \sqrt{T} \boldsymbol{m}_{[rT]}(\boldsymbol{\theta}_o) + \frac{[rT]}{T} \boldsymbol{F}_{[rT]}(\boldsymbol{\theta}_o) \big[\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_o)\big] + o_{\mathbb{P}}(1);$$

  the second term is the estimation effect and converges to $r \boldsymbol{F}_o \boldsymbol{Q}_o \boldsymbol{A}_o \boldsymbol{W}_k(1)$, where $\boldsymbol{A}_o$ is the matrix square root of $\boldsymbol{G}_{22} \boldsymbol{G}_{22}' + \boldsymbol{G}_{21} \boldsymbol{G}_{21}'$.

- [B1](b) and [B2] imply

$$\sqrt{T}\boldsymbol{m}_T(\hat{\boldsymbol{\theta}}_T) \Rightarrow [\boldsymbol{I}_q \;\; \boldsymbol{F}_o\boldsymbol{Q}_o]\boldsymbol{G}_o\boldsymbol{W}_{q+k}(1) \stackrel{d}{=} \boldsymbol{V}_o\boldsymbol{W}_q(1),$$

where $\boldsymbol{V}_o$ is the matrix square root of $[\boldsymbol{I}_q \;\; \boldsymbol{F}_o\boldsymbol{Q}_o]\boldsymbol{G}_o\boldsymbol{G}_o'[\boldsymbol{I}_q \;\; \boldsymbol{F}_o\boldsymbol{Q}_o]'$. Note $\boldsymbol{V}_o = \boldsymbol{S}_o$ when $\boldsymbol{F}_o = \boldsymbol{0}$ (i.e., no estimation effect).

- Due to "centering", the estimation effects in $\boldsymbol{\varphi}_{[rT]}(\hat{\boldsymbol{\theta}}_T)$ cancel out:

$$\begin{aligned}
\boldsymbol{\varphi}_{[rT]}(\hat{\boldsymbol{\theta}}_T) &= \sqrt{T}\boldsymbol{m}_{[rT]}(\boldsymbol{\theta}_o) + \frac{[rT]}{T}\boldsymbol{F}_{[rT]}(\boldsymbol{\theta}_o)\big[\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_o)\big] \\
&\quad - \frac{[rT]}{T}\sqrt{T}\boldsymbol{m}_T(\boldsymbol{\theta}_o) - \frac{[rT]}{T}\boldsymbol{F}_T(\boldsymbol{\theta}_o)\big[\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_o)\big] + o_{\mathbb{P}}(1) \\
&= \sqrt{T}\boldsymbol{m}_{[rT}(\boldsymbol{\theta}_o) - \frac{[rT]}{T}\sqrt{T}\boldsymbol{m}_T(\boldsymbol{\theta}_o) + o_{\mathbb{P}}(1).
\end{aligned}$$

- $\widehat{\boldsymbol{C}}_T = \boldsymbol{C}_T(\boldsymbol{\theta}_o) + o_{\mathbb{P}}(1) \Rightarrow \boldsymbol{S}_o\boldsymbol{P}_q\boldsymbol{S}_o'$, regardless of the estimation effect.

- When estimation effect is present, $\widehat{\boldsymbol{C}}_T$ is unable to eliminate $\boldsymbol{V}_o$, and

$$\widehat{\mathcal{M}}_T \xrightarrow{D} \boldsymbol{W}_q(1)'\boldsymbol{V}_o'\big[\boldsymbol{S}_o\boldsymbol{P}_q\boldsymbol{S}_o'\big]^{-1}\boldsymbol{V}_o\boldsymbol{W}_q(1),$$

  That is, $\widehat{\mathcal{M}}_T$ depends on $\boldsymbol{S}_o$ and $\boldsymbol{V}_o$ and is not asymptotically pivotal.

- When there is no estimation effect ($\boldsymbol{F}_o = \mathbf{0}$), $\boldsymbol{V}_o = \boldsymbol{S}_o$, and

$$\widehat{\mathcal{M}}_T \xrightarrow{D} \boldsymbol{W}_q(1)'\boldsymbol{P}_q^{-1}\boldsymbol{W}_q(1),$$

  which is also the limit of $\mathcal{M}_T$.

- Remark: The nonsingularity of $\boldsymbol{G}_o$ required in [B1](b) is crucial for the M tests here. It excludes the cases that the moment functions ($\boldsymbol{f}$) and the estimator (which depends on $\boldsymbol{q}$) are asymptotically correlated, e.g., the over-identifying restrictions in the context of GMM.

M Test under Estimation Effect

- Kuan and Lee (2006): $\widetilde{\boldsymbol{C}}_T = T^{-1} \sum_{t=k+1}^{T} \tilde{\boldsymbol{\varphi}}_t \tilde{\boldsymbol{\varphi}}_t'$ with

$$\tilde{\boldsymbol{\varphi}}_t = \boldsymbol{\varphi}_t(\tilde{\boldsymbol{\theta}}_t, \tilde{\boldsymbol{\theta}}_T) = \frac{1}{\sqrt{T}} \sum_{i=1}^{t} \big[\boldsymbol{f}(\boldsymbol{\eta}_i, \tilde{\boldsymbol{\theta}}_t) - \boldsymbol{m}_T(\tilde{\boldsymbol{\theta}}_T) \big],$$

where $\tilde{\boldsymbol{\theta}}_t$ are the recursive estimators based on first t observations.

- The M test is

$$\widetilde{\mathcal{M}}_T = T\, \boldsymbol{m}_T(\hat{\boldsymbol{\theta}}_T)' \widetilde{\boldsymbol{C}}_T^{-1} \boldsymbol{m}_T(\hat{\boldsymbol{\theta}}_T) \xrightarrow{D} \boldsymbol{W}_q(1)' \boldsymbol{P}_q^{-1} \boldsymbol{W}_q(1),$$

which has the same limit as $\mathcal{M}_T$.

  - ◇ $T^{1/2} \boldsymbol{m}_T(\hat{\boldsymbol{\theta}}_T) \Rightarrow \boldsymbol{V}_o \boldsymbol{W}_q(1)$.
  - ◇ $\tilde{\boldsymbol{\varphi}}_{[rT]} \Rightarrow \boldsymbol{V}_o \boldsymbol{B}_q(r)$, and hence $\widetilde{\boldsymbol{C}}_T \Rightarrow \boldsymbol{V}_o \boldsymbol{P}_q \boldsymbol{V}_o'$.

- While HAC estimation of $\boldsymbol{V}_o$ is practically difficult, $\widetilde{\mathcal{M}}_T$ avoids estimating $\boldsymbol{V}_o$ and hence is also robust to estimation effect.

Example: Tests of Serial Correlations

Specification: $y_t = h(\boldsymbol{x}_t; \boldsymbol{\theta}) + e_t(\boldsymbol{\theta})$ with the NLS estimator $\hat{\boldsymbol{\theta}}_T$.

- $\mathbb{E}(y_t|\boldsymbol{x}_t) = h(\boldsymbol{x}_t; \boldsymbol{\theta}_o)$ and $\varepsilon_t := e_t(\boldsymbol{\theta}_o) = y_t - h(\boldsymbol{x}_t; \boldsymbol{\theta}_o)$.
- The null hypothesis is

$$\mathbb{E}[\boldsymbol{f}_{t,q}(\boldsymbol{\theta}_o)] = \mathbb{E}(\varepsilon_t \boldsymbol{\epsilon}_{t-1,q}) = \mathbf{0},$$

  where $\boldsymbol{\epsilon}_{t-1,q} = [\varepsilon_{t-1}, \ldots, \varepsilon_{t-q}]'$.

- Letting $T_q = T - q$, define

$$\boldsymbol{m}_{T_q}(\boldsymbol{\theta}) = \frac{1}{T_q} \sum_{t=q+1}^{T} e_t(\boldsymbol{\theta}) \boldsymbol{e}_{t-1,q}(\boldsymbol{\theta}).$$

  We can base an M test on $\boldsymbol{m}_{T_q}(\hat{\boldsymbol{\theta}}_T) = T_q^{-1} \sum_{t=q+1}^{T} e_t(\hat{\boldsymbol{\theta}}_T) \boldsymbol{e}_{t-1,q}(\hat{\boldsymbol{\theta}}_T)$.

- $T_q^{1/2} \boldsymbol{m}_{T_q}(\hat{\boldsymbol{\theta}}_T)$ and $T_q^{1/2} \boldsymbol{m}_{T_q}(\boldsymbol{\theta}_o)$ are not asymptotically equivalent unless $\boldsymbol{F}_{T_q}(\boldsymbol{\theta}_o)$ converges to $\boldsymbol{F}_o = \mathbf{0}$.

- Here, $\boldsymbol{F}_{T_q}(\boldsymbol{\theta}_o) = -T_q^{-1} \sum_{t=q+1}^{T} \big[\boldsymbol{\epsilon}_{t-1,q} \nabla_{\boldsymbol{\theta}} h_t(\boldsymbol{\theta}_o) + \varepsilon_t \nabla_{\boldsymbol{\theta}} \boldsymbol{h}_{t-1,q}(\boldsymbol{\theta}_o)\big]$.
- $\boldsymbol{F}_o$ would be zero if $\{\boldsymbol{x}_t\}$ and $\{\varepsilon_t\}$ are mutually independent. When $h(\boldsymbol{x}_t; \boldsymbol{\theta}_o) = \boldsymbol{x}_t' \boldsymbol{\theta}_o$, $\boldsymbol{F}_o = \mathbf{0}$ when $\{\boldsymbol{x}_t\}$ and $\{\varepsilon_t\}$ are mutually uncorrelated.
  - ◇ The M test based on model residuals is

$$\widehat{\mathcal{M}}_{T_q} = T_q \, \boldsymbol{m}_{T_q}(\hat{\boldsymbol{\theta}}_T)' \widehat{\boldsymbol{C}}_{T_q}^{-1} \boldsymbol{m}_{T_q}(\hat{\boldsymbol{\theta}}_T) \xrightarrow{D} \boldsymbol{W}_q(1)' \boldsymbol{P}_q^{-1} \boldsymbol{W}_q(1),$$

    where the normalizing matrix is $\widehat{\boldsymbol{C}}_{T_q} = T_q^{-1} \sum_{t=q+1}^{T} \hat{\boldsymbol{\varphi}}_t \hat{\boldsymbol{\varphi}}_t'$ with

$$\hat{\boldsymbol{\varphi}}_t = \frac{1}{\sqrt{T_q}} \sum_{i=q+1}^{t} \big[e_i(\hat{\boldsymbol{\theta}}_T) \boldsymbol{e}_{i-1,q}(\hat{\boldsymbol{\theta}}_T)\big] - \frac{t-q}{T_q} \frac{1}{\sqrt{T_q}} \sum_{i=q+1}^{T} \big[e_i(\hat{\boldsymbol{\theta}}_T) \boldsymbol{e}_{i-1,q}(\hat{\boldsymbol{\theta}}_T)\big].$$

  - ◇ $\widehat{\mathcal{M}}_{T_q}$ includes the test of Lobato (2001) for raw time series as a special case.

- $\boldsymbol{F}_o \neq \mathbf{0}$ for the residuals of dynamic models, such as AR models and models with lagged dependent variables.

  ◇ The M test based on the residuals of dynamic models is

  $$\widetilde{\mathcal{M}}_{T_q} = T\,\boldsymbol{m}'_{T_q}(\hat{\boldsymbol{\theta}}_T)\widetilde{\boldsymbol{C}}_{T_q}^{-1}\boldsymbol{m}_{T_q}(\hat{\boldsymbol{\theta}}_T) \xrightarrow{D} \boldsymbol{W}_q(1)'\boldsymbol{P}_q^{-1}\boldsymbol{W}_q(1),$$

  where the normalizing matrix is $\widetilde{\boldsymbol{C}}_{T_q} = T_q^{-1}\sum_{t=q+1}^{T}\tilde{\boldsymbol{\varphi}}_t\tilde{\boldsymbol{\varphi}}'_t$ with

  $$\tilde{\boldsymbol{\varphi}}_t = \frac{1}{\sqrt{T_q}}\sum_{i=q+1}^{t}\left[e_i(\tilde{\boldsymbol{\theta}}_t)\boldsymbol{e}_{i-1,q}(\tilde{\boldsymbol{\theta}}_t)\right] - \frac{t-q}{T_q}\frac{1}{\sqrt{T_q}}\sum_{i=q+1}^{T}\left[e_i(\tilde{\boldsymbol{\theta}}_T)\boldsymbol{e}_{i-1,q}(\tilde{\boldsymbol{\theta}}_T)\right],$$

  and $e_i(\tilde{\boldsymbol{\theta}}_t) = y_i - h(\boldsymbol{x}_i;\tilde{\boldsymbol{\theta}}_t)$ is the i th residual evaluated at the recursive NLS estimator $\tilde{\boldsymbol{\theta}}_t$.

  ◇ $\widetilde{\mathcal{M}}_{T_q}$ is a specification test without consistent estimation of the asymptotic covariance matrix.

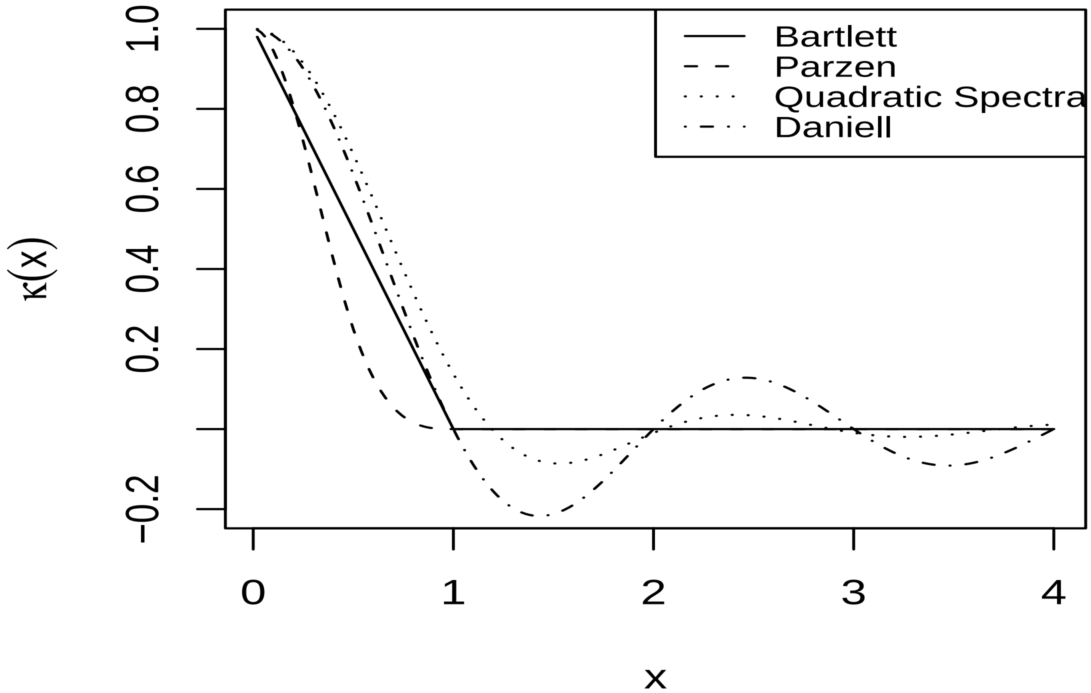

Figure 1: The Bartlett, Parzen, quadratic spectral and Daniel kernels.

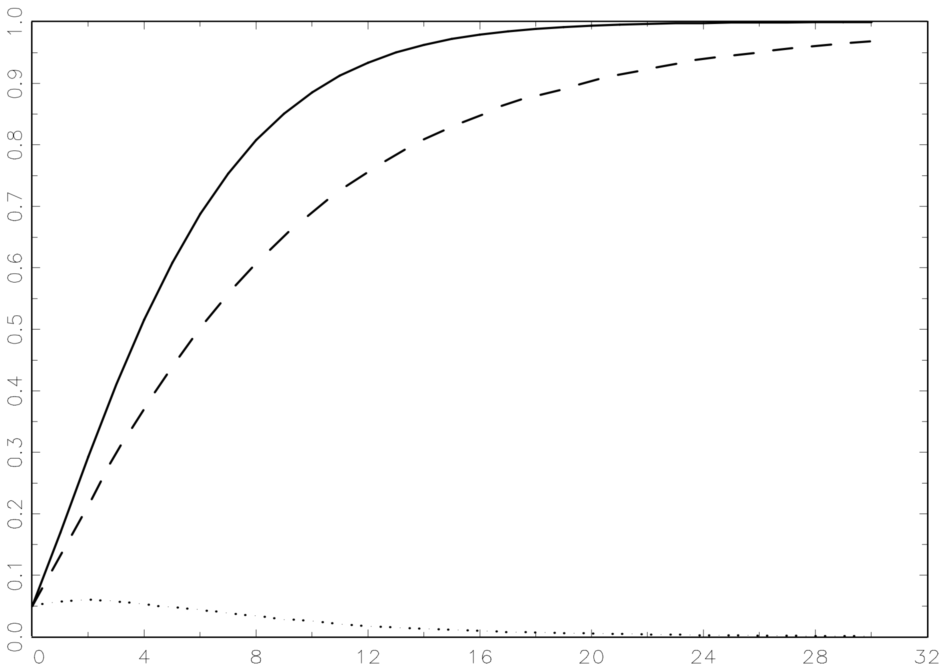

Figure 2: The asymptotic local powers of the standard M test (solid), $\widetilde{\mathcal{M}}_T$ (dashed) and $\ddot{\mathcal{M}}_T$ (dotted) at 5% level.