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
Consider the specification: \( y_t = x'_t \beta + \epsilon_t. \)

[A1] For some \( \beta_o, \epsilon_t = y_t - x'_t \beta_o \) such that \( IE(x_t \epsilon_t) = 0 \) and

\[
\frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} x_t \epsilon_t \Rightarrow S_o W_k(r), \quad r \in [0, 1],
\]

\( \Sigma_o = \lim_{T \to \infty} \var(T^{-1/2} \sum_{t=1}^{T} x_t \epsilon_t) \) with \( S_o \) its matrix square root (i.e., \( \Sigma_o = S_o S'_o \)).

[A2] \( M_{[Tr]} := [Tr]^{-1} \sum_{t=1}^{[Tr]} x_t x'_t \xrightarrow{IP} M_o \) uniformly in \( r \in (0, 1] \).

Asymptotic normality of the OLS estimator:

\[
\sqrt{T}(\hat{\beta}_T - \beta_o) = M_T^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} x_t \epsilon_t \xrightarrow{D} M_o^{-1} S_o W_k(1),
\]

which has the distribution \( \mathcal{N}(0, M_o^{-1} \Sigma_o M_o^{-1}) \).
• The null hypothesis is \( R_o \beta_o = r \), where \( R \) (\( q \times k \)) has full row rank.

\[
\sqrt{T}(R\hat{\beta}_T - r) \xrightarrow{D} \mathcal{N}(0, RM_o^{-1}\Sigma_o M_o^{-1}R').
\]

\[
(RM_T^{-1}\hat{\Sigma}_T M_T^{-1}R')^{-1/2} \sqrt{T}(R\hat{\beta}_T - r) \xrightarrow{D} \mathcal{N}(0, I_q).
\]

The Wald test is

\[
\mathcal{W}_T = T(R\hat{\beta}_T - r)'(RM_T^{-1}\hat{\Sigma}_T M_T^{-1}R')^{-1}(R\hat{\beta}_T - r)
\]

\[
\xrightarrow{D} \chi^2(q).
\]

• Consistent estimation of \( \Sigma_o \) is crucial; \( \mathcal{W}_T \) would not have a limiting \( \chi^2 \) distribution if \( \hat{\Sigma}_T \) is not a consistent estimator.

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

• When \( \hat{\Sigma}_T \) is consistent, the resulting tests are said to be robust to heteroskedasticity and serial correlations of unknown form.
Asymptotic Covariance Matrix $\Sigma_o$

A general form:

$$\Sigma_o = \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} \text{IE}(x_t \epsilon_t \epsilon_s x'_s) = \lim_{T \to \infty} \sum_{j=-T+1}^{T-1} \Gamma_T(j),$$

with autocovariances:

$$\Gamma_T(j) = \begin{cases} 
\frac{1}{T} \sum_{t=j+1}^{T} \text{IE}(x_t \epsilon_t \epsilon_{t-j} x'_{t-j}), & j = 0, 1, 2, \ldots, \\
\frac{1}{T} \sum_{t=-j+1}^{T} \text{IE}(x_{t+j} \epsilon_{t+j} \epsilon_t x'_t), & j = -1, -2, \ldots.
\end{cases}$$

- When $x_t \epsilon_t$ are covariance stationary, $\Gamma_T(j) = \Gamma(j)$, and the spectral density of $x_t \epsilon_t$ at frequency $\omega$ is

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

- $\Sigma_o = 2\pi f(0)$ and hence is also known as the long-run variance of $x_t \epsilon_t$. 
Examples of $\Sigma_o$

- When $x_t\epsilon_t$ are serially uncorrelated:

$$
\Sigma_o = \lim_{T \to \infty} \Gamma_T(0) = \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \text{IE}(\epsilon_t^2 x_t x'_t),
$$

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

- When $x_t\epsilon_t$ are serially uncorrelated and $\epsilon_t$ conditionally homoskedastic:

$$
\Sigma_o = \sigma_o^2 \left( \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \text{IE}(x_t x'_t) \right) = \sigma_o^2 M_o,
$$

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

- Estimating $\Sigma_o$ is much more difficult when heteroskedasticity and serial correlations are present and of unknown form.
A Consistent Estimator of $\Sigma_o$

- Recall $\Sigma_o = \lim_{T \to \infty} \sum_{j=-T+1}^{T-1} \Gamma_T(j)$.

- A consistent estimator (White, 1984):

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

with sample autocovariances

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

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:

\[
\hat{\Sigma}_T^\kappa = \sum_{j=-T+1}^{T-1} \kappa \left( \frac{j}{\ell(T)} \right) \hat{\Gamma}_T(j),
\]

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

• \( \kappa \) and \( \ell(T) \) jointly determine the weighting scheme on \( \hat{\Gamma}_T(j) \) 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),

\[
\int_{-\infty}^{\infty} \kappa(x) e^{-ix\omega} \, dx \geq 0, \quad \forall \omega \in \mathbb{R}.
\]
Kernel Functions

- **Bartlett kernel** (Newey and West, 1987):
  \[
  \kappa(x) = \begin{cases} 
  1 - |x|, & |x| \leq 1, \\
  0, & \text{otherwise}; 
  \end{cases}
  \]

- **Parzen kernel** (Gallant, 1987):
  \[
  \kappa(x) = \begin{cases} 
  1 - 6x^2 + 6|x|^3, & |x| \leq 1/2, \\
  2(1 - |x|)^3, & 1/2 \leq |x| \leq 1, \\
  0, & \text{otherwise}; 
  \end{cases}
  \]

- **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:
  \[ \ell^*(T) = 1.1447(c_1 T)^{1/3}, \text{ (Bartlett)}, \]
  \[ \ell^*(T) = 2.6614(c_2 T)^{1/5}, \text{ (Parzen)}, \]
  \[ \ell^*(T) = 1.3221(c_2 T)^{1/5}, \text{ (quadratic spectral)}, \]

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

  \[ \sqrt{T/\ell^*(T)} \left( \hat{\Sigma}_T^k - \Sigma_o \right) = O_{\text{IP}}(1). \]

- As far as MSE is concerned, the quadratic spectral kernel is to be preferred.
  
  \( \Diamond \) \( \hat{\Sigma}_T^B = O_{\text{IP}}(T^{-1/3}) \); \( \hat{\Sigma}_T^P \) nd \( \hat{\Sigma}_T^{QS} \) are \( O_{\text{IP}}(T^{-2/5}) \).

  \( \Diamond \) \( \hat{\Sigma}_T^{QS} \) is more efficient than \( \hat{\Sigma}_T^P \); \( \hat{\Sigma}_T^B \) 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

  - Apply a VAR model to whiten $x_t \hat{e}_t$ 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 $(y_t - x_t'\tilde{\beta}_{t-1})$, 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

$$\hat{W}_T^\dagger = T (R\hat{\beta}_T - r)' \left( RM_T^{-1} \hat{C}_T M_T^{-1} R' \right)^{-1} (R\hat{\beta}_T - r),$$

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

- $\hat{C}_T$ is inconsistent for $\Sigma_o$ but is able to eliminate the nuisance parameters in $\Sigma_o$.

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

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

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

$$
\hat{\varphi}_{[Tr]} = \frac{1}{\sqrt{T}} \sum_{i=1}^{[Tr]} x_i \epsilon_i - \frac{[Tr]}{T} \left( \frac{1}{[Tr]} \sum_{i=1}^{[Tr]} x_i x_i' \right) \sqrt{T} (\beta_T - \beta_o)
$$

$$
= S_o W_k(r) - r M_o M_o^{-1} S_o W_k(1)
$$

$$
= S_o B_k(r).
$$

Hence,

$$
\hat{C}_T \Rightarrow S_o \left( \int_{0}^{1} B_k(r) B_k(r)' \, dr \right) S'_o =: S_o P_k S'_o.
$$
• Let $G_o$ denote the matrix square root of $RM_o^{-1}S_oS'M_o^{-1}R'$. Then,

$$RM_T^{-1} \hat{C}_T M_T^{-1} R' \Rightarrow RM_o^{-1} S_o P_k S'_o M_o^{-1} R' \overset{d}{=} G_o P_q G'_o.$$ 

and $\sqrt{T} R(\hat{\beta}_T - \beta_o) \overset{D}{\rightarrow} RM_o^{-1} S_o W_k(1) \overset{d}{=} G_o W_q(1).$

• $W_T^\dagger$ is thus asymptotically pivotal:

$$W_T^\dagger \Rightarrow W_q(1)' G'_o (G_o P_q G'_o)^{-1} G_o W_q(1) = W_q(1)' P_q^{-1} W_q(1).$$


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

$$t^\dagger = \sqrt{T}(\hat{\beta}_{i,T} - r) \overset{D}{\rightarrow} \frac{W(1)}{\sqrt{\delta_i}} \left[ \int_0^1 B(r)^2 \, dr \right]^{1/2}.$$ 

This distribution is more disperse than the standard normal distribution.
Kernel-Based Normalizing Matrices

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

- Kiefer and Vogelsang (2002b): \( \hat{\Sigma}^\kappa_T \Rightarrow S_o P^\kappa_k S_o^\prime \), with

  \[
  P^\kappa_k = - \int_0^1 \int_0^1 \kappa''(r - s) B_k(r) B_k(s)'} \, dr \, ds;
  \]

  The Wald test based on \( \hat{\Sigma}^\kappa_T \) without truncation can also serve as a KVB’s robust test.

- A test based on \( \hat{\Sigma}^B_T \) without truncation compares favorably with that based on \( \hat{\Sigma}^{QS}_T \) 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: \( \mathbb{E}[f(\eta_t; \theta_o)] = 0 \), where \( \theta_o \) is the \( k \times 1 \) true parameter vector, and \( f \) is a \( q \times 1 \) vector of functions.

\( \theta_o \) Is Known

Define \( m_{[rT]}(\theta) = T^{-1} \sum_{t=1}^{[rT]} f(\eta_t; \theta) \), for \( r \in (0, 1] \).

- An \( M \) test is based on \( m_T(\theta_o) \), the sample counterpart of the null.
- By a CLT, \( T^{1/2} m_T(\theta_o) \xrightarrow{D} \mathcal{N}(0, \Sigma_o) \), and the conventional \( M \) test is:
  \[
  T m_T(\theta_o)' \hat{\Sigma}_T^{-1} m_T(\theta_o) \xrightarrow{D} \chi^2(q),
  \]
  where \( \hat{\Sigma}_T \) is a consistent estimator of \( \Sigma_o \).
- The limiting \( \chi^2 \) distribution hinges on consistent estimation of \( \Sigma_o \).
[B1](a) Under the null, $\sqrt{T} m_{rT}(\theta_o) \Rightarrow S_o W_q(r)$ for $0 \leq r \leq 1$, where $S_o$ is the nonsingular, matrix square root of $\Sigma_o$.

- $C_T(\theta_o) = T^{-1} \sum_{t=1}^{T} \varphi_t(\theta_o) \varphi_t(\theta_o)'$ with

$$\varphi_t(\theta_o) = \frac{1}{\sqrt{T}} \sum_{i=1}^{t} \left[ f(\eta_i; \theta_o) - m_T(\theta_o) \right].$$

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

$$M_T = T m_T(\theta_o)' C_T(\theta_o)^{-1} m_T(\theta_o) \xrightarrow{D} W_q(1)' P_q^{-1} W_q(1).$$

- By [B1](a), $T^{1/2} m_T(\theta_o) \Rightarrow S_o W_q(1)$.

- $\varphi_{[rT]}(\theta_o) \Rightarrow S_o \left[ W_q(r) - r W_q(1) \right] = S_o B_q(r), \quad 0 \leq r \leq 1.$

- $C_T(\theta_o) \Rightarrow S_o P_q S_o'$ with $P_q = \int_{0}^{1} B_q(r) B_q(r)' dr.$
**θ_0 Is Unknown**

- Replacing \( \theta_o \) in \( m_T \) and \( \varphi_t \) with a root-\( T \) consistent estimator \( \hat{\theta}_T \) that satisfies
  \[
  \sqrt{T}(\hat{\theta}_T - \theta_o) = Q_o \left[ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} q(\eta_t; \theta_o) \right] + o_P(1).
  \]

- Kuan and Lee (2006): The \( M \) test is
  \[
  \hat{\mathcal{M}}_T = T m_T(\hat{\theta}_T)' \hat{C}_T^{-1} m_T(\hat{\theta}_T),
  \]
  where \( \hat{C}_T = C_T(\hat{\theta}_T) = T^{-1} \sum_{t=1}^{T} \varphi_t(\hat{\theta}_T) \varphi_t(\hat{\theta}_T)' \) with
  \[
  \varphi_t(\hat{\theta}_T) = \frac{1}{\sqrt{T}} \sum_{i=1}^{t} \left[ f(\eta_i; \hat{\theta}_T) - m_T(\hat{\theta}_T) \right].
  \]

- The limit of \( \hat{\mathcal{M}}_T \) depends on the estimation effect of replacing \( \theta_o \) with \( \hat{\theta}_T \).
[B1](b) Under the null,

\[
\begin{bmatrix}
\frac{1}{\sqrt{T}} \sum_{t=1}^{T} f(\eta_t; \theta_o) \\
\frac{1}{\sqrt{T}} \sum_{t=1}^{T} q(\eta_t; \theta_o)
\end{bmatrix} \Rightarrow G_o W_{q+k}(1),
\]

where \(G_o\) is nonsingular.

[B2] \(F_{[rT]}(\theta_o) = [rT]^{-1} \sum_{t=1}^{[rT]} \nabla_\theta f(\eta_t; \theta_o) \xrightarrow{\text{IP}} F_o\), uniformly in \(r \in (0, 1]\),

where \(F_o\) is a \(q \times k\) non-stochastic matrix; \(\nabla_\theta F_{[rT]}(\theta_o) = O_{\text{IP}}(1)\).

- A Taylor expansion about \(\theta_o\) gives

\[
\sqrt{T} m_{[rT]}(\hat{\theta}_T) = \sqrt{T} m_{[rT]}(\theta_o) + \frac{[rT]}{T} F_{[rT]}(\theta_o) \left[ \sqrt{T}(\hat{\theta}_T - \theta_o) \right] + o_{\text{IP}}(1);
\]

the second term is the estimation effect and converges to \(r F_o Q_o A_o W_k(1)\),

where \(A_o\) is the matrix square root of \(G_{22} G_{22}' + G_{21} G_{21}'\).
• [B1](b) and [B2] imply

\[ \sqrt{T} m_T(\hat{\theta}_T) \Rightarrow [I_q \quad F_o Q_o] G_o W_{q+k}(1) \overset{d}{=} V_o W_q(1), \]

where \( V_o \) is the matrix square root of \([I_q \quad F_o Q_o] G_o G'_o[I_q \quad F_o Q_o]'\). Note \( V_o = S_o \) when \( F_o = 0 \) (i.e., no estimation effect).

• Due to “centering”, the estimation effects in \( \varphi_{[rT]}(\hat{\theta}_T) \) cancel out:

\[ \varphi_{[rT]}(\hat{\theta}_T) = \sqrt{T} m_{[rT]}(\theta_o) + \frac{[rT]}{T} F_{[rT]}(\theta_o) \left[ \sqrt{T}(\hat{\theta}_T - \theta_o) \right] - \frac{[rT]}{T} \sqrt{T} m_T(\theta_o) - \frac{[rT]}{T} F_T(\theta_o) \left[ \sqrt{T}(\hat{\theta}_T - \theta_o) \right] + o_{IP}(1) \]

\[ = \sqrt{T} m_{[rT]}(\theta_o) - \frac{[rT]}{T} \sqrt{T} m_T(\theta_o) + o_{IP}(1). \]

• \( \hat{C}_T = C_T(\theta_o) + o_{IP}(1) \Rightarrow S_o P_q S'_o \), regardless of the estimation effect.
• When estimation effect is present, $\hat{C}_T$ is unable to eliminate $V_o$, and

$$\hat{M}_T \overset{D}{\rightarrow} W_q(1)' V_o' [S_o P_q S_o']^{-1} V_o W_q(1),$$

That is, $\hat{M}_T$ depends on $S_o$ and $V_o$ and is not asymptotically pivotal.

• When there is no estimation effect ($F_o = 0$), $V_o = S_o$, and

$$\hat{M}_T \overset{D}{\rightarrow} W_q(1)' P_q^{-1} W_q(1),$$

which is also the limit of $M_T$.

• Remark: The nonsingularity of $G_o$ required in [B1](b) is crucial for the $M$ tests here. It excludes the cases that the moment functions ($f$) and the estimator (which depends on $q$) are asymptotically correlated, e.g., the over-identifying restrictions in the context of GMM.
M Test under Estimation Effect

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

\[
\tilde{\varphi}_t = \varphi_t(\tilde{\theta}_t, \tilde{\theta}_T) = \frac{1}{\sqrt{T}} \sum_{i=1}^{t} \left[ f(\eta_i, \tilde{\theta}_t) - m_T(\tilde{\theta}_T) \right],
\]

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

- The \( M \) test is

\[
\tilde{\mathcal{M}}_T = T m_T(\hat{\theta}_T)' \tilde{C}_T^{-1} m_T(\hat{\theta}_T) \xrightarrow{D} W_q(1)' P_q^{-1} W_q(1),
\]

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

- \( T^{1/2} m_T(\hat{\theta}_T) \Rightarrow V_o W_q(1) \).

- \( \tilde{\varphi}_{[rT]} \Rightarrow V_o B_q(r) \), and hence \( \tilde{C}_T \Rightarrow V_o P_q V_o' \).

- While HAC estimation of \( V_o \) is practically difficult, \( \tilde{\mathcal{M}}_T \) avoids estimating \( V_o \) and hence is also robust to estimation effect.
Example: Tests of Serial Correlations

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

- $\text{IE}(y_t | x_t) = h(x_t; \theta_o)$ and $\varepsilon_t := e_t(\theta_o) = y_t - h(x_t; \theta_o)$.

- The null hypothesis is

$$\text{IE}[f_{t,q}(\theta_o)] = \text{IE}(\varepsilon_t \varepsilon_{t-1,q}) = 0,$$

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

- Letting $T_q = T - q$, define

$$m_{T_q}(\theta) = \frac{1}{T_q} \sum_{t=q+1}^{T} e_t(\theta) e_{t-1,q}(\theta).$$

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

- $T_q^{1/2} m_{T_q}(\hat{\theta}_T)$ and $T_q^{1/2} m_{T_q}(\theta_o)$ are not asymptotically equivalent unless $F_{T_q}(\theta_o)$ converges to $F_o = 0$. 
Here, \( F_{T_q}(\theta_o) = -T_q^{-1} \sum_{t=q+1}^T [\epsilon_{t-1,q} \nabla_\theta h_t(\theta_o) + \epsilon_t \nabla_\theta h_{t-1,q}(\theta_o)] \).

\( F_o \) would be zero if \( \{x_t\} \) and \( \{\epsilon_t\} \) are mutually independent. When \( h(x_t; \theta_o) = x_t' \theta_o \), \( F_o = 0 \) when \( \{x_t\} \) and \( \{\epsilon_t\} \) are mutually uncorrelated.

\( \diamond \) The \( M \) test based on model residuals is

\[
\hat{M}_{T_q} = T_q m_{T_q}(\hat{\theta}_T)' \hat{C}_{T_q}^{-1} m_{T_q}(\hat{\theta}_T) \xrightarrow{D} W_q(1)' P_q^{-1} W_q(1),
\]

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

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

\( \diamond \) \( \hat{M}_{T_q} \) includes the test of Lobato (2001) for raw time series as a special case.
• $F_o \neq 0$ for the residuals of dynamic models, such as AR models and models with lagged dependent variables.

diamond The $M$ test based on the residuals of dynamic models is

$$
\tilde{M}_{T_q} = T m'_{T_q}(\tilde{\theta}_T)\tilde{C}_{T_q}^{-1} m_{T_q}(\tilde{\theta}_T) \xrightarrow{D} W_q(1)'P_q^{-1}W_q(1),
$$

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

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

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

diamond $\tilde{M}_{T_q}$ 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), $\tilde{M}_T$ (dashed) and $\ddot{M}_T$ (dotted) at 5% level.