LECTURE ON

Robust Tests with and without Consistent Estimation of Asymptotic Covariance Matrix

CHUNG-MING KUAN

Institute of Economics
Academia Sinica

This version: January 31, 2008

Address for correspondence: Institute of Economics, Academia Sinica, Taipei 115, Taiwan.

E-mail: ckuan@econ.sinica.edu.tw; URL: www.sinica.edu.tw/~ckuan

† I am grateful for many useful suggestions and comments made by Yu-Wei Hsieh.

Contents

1 Introduction

Consistent estimation of asymptotic variance-covariance matrix plays a crucial role in the large-sample tests in the econometrics literature. An estimator that is consistent when there are heteroskedasticity and serial correlations in data is known as the heteroskedasticity and autocorrelation consistent (HAC) covariance matrix estimator, or simply a HAC estimator. A large-sample test that involves a HAC estimator is thus *robust* to the presence of heteroskedasticy and serial correlations of unknown form.

A leading class of HAC estimators is the nonparametric kernel estimator originated from the estimation of spectral density; see, e.g., Parzen (1957), Hannan (1970), and Priestley (1981). The kernel estimator was introduced to the econometrics literature by Newey and West (1987) and Gallant (1987). This class of estimators was subsequently refined by Andrews (1991), Andrews and Monahan (1992), Newey and West (1994), and Phillips, Sun, and Jin (2003, 2006), among others. Robinson (1991), Hansen (1992), de Jong (2000), de Jong and Davidson (2000), and Jansson (2002) also provided various consistency proofs. For an early review of HAC estimation we refer to den Haan and Levin (1997).

A drawback of the kernel HAC estimator is that its performance varies with the choices of the kernel function and its bandwidth. As these choices are somewhat arbitrary in practice, the statistical inferences resulted from the robust tests with the kernel HAC estimator are unavoidably vulnerable. Kiefer, Vogelsang, and Bunzel (2000), henceforth KVB, proposed an asymptotically pivotal test for regression parameters that does not require consistent estimation of asymptotic covariance matrix but relies on a normalizing matrix to eliminate the nuisance parameters; see also Bunzel, Kiefer, and Vogelsang (2001) and Vogelsang (2003). The resulting test is thus robust to heteroskedasticity and serial correlations while circumventing the difficulties in kernel estimation. More generally, Kiefer and Vogelsang (2002a, b) showed that a kernel HAC estimator with the bandwidth equal to sample size can also serve as a normalizing matrix in the KVB approach. This greatly expands the class of KVB's robust tests. Kiefer and Vogelsang (2005) provided a theory that relates the HAC estimation and KVB approach.

The KVB approach can be extended to construct robust model specification tests. Lobato (2001) constructed a robust portmanteau test for serial correlations. For general moment conditions with unknown parameters, Kuan and Lee (2006) proposed two robust

M tests without consistent estimation of asymptotic covariance matrix, which include Lobato's test as a special case. It is worth mentioning that one of Kuan and Lee's tests is valid even when there is estimation effect in the sample moment conditions. This is practically important because HAC estimation is usually very difficult, if not impossible, to implement when estimation effect is present. Lee (2006) also obtained robust M tests with kernel-based normalizing matrices. The robust tests for over-identifying restrictions (OIR) in the context of the generalized method of moments (GMM) were constructed by Lee and Kuan (2006).

It has been documented in the literature that the kernel HAC estimators are downward biased so that the resulting tests are typically over-sized in finite samples. On the other hand, the asymptotic distribution of KVB's robust test provides very good approximation to its finite-sample counterpart. Thus, as far as test size is concerned, KVB's robust test ought to be preferred to the conventional tests with a kernel HAC estimator. In a Gaussian location model, Jansson (2004) found that the error in rejection probability of the KVB test decays at a fast rate $O(T^{-1}\log T)$, whereas such a rate for the conventional tests with a kernel HAC estimator is no better than $O(T^{-1/2})$. It has been found, however, that the KVB tests suffer from power loss in finite samples; such loss may be quite substantial, especially when an inappropriate kernel is used to compute the normalizing matrix and/or the number of restrictions increases.

This note is organized as follows. Section 2 is a review of the basic asymptotic theory for the OLS estimator and the Wald test for linear hypotheses. In Section 3, we present HAC estimation and discuss how a kernel function and its bandwidth can be determined. The KVB approach and robust tests for regression parameters are introduced in Section 4. The KVB approach is extended to construct robust M tests in Section 5. We also consider robust test for serial correlations and robust OIR test.

2 Basic Asymptotic Theory

Consider the linear specification $y_t = \boldsymbol{x}_t'\boldsymbol{\beta} + e_t$, and the OLS estimator $\hat{\boldsymbol{\beta}}_T$ $(k \times 1)$. We shall review some basic asymptotic theory for $\hat{\boldsymbol{\beta}}_T$ and the Wald test of regression parameters. In what follows, we let $[c]$ denote the integer part of c, $\xrightarrow{\mathbb{P}}$ convergence in probability, $\Rightarrow$ weak convergence (of associated probability measures), $\xrightarrow{D}$ convergence in distribution, $\stackrel{d}{=}$ equality in distribution, $\boldsymbol{W}_k$ a vector of k independent, standard Wiener processes, and $\boldsymbol{B}_k$ the Brownian bridge obtained from $\boldsymbol{W}_k$ such that $\boldsymbol{B}_k(r) = \boldsymbol{W}_k(r) - r\boldsymbol{W}_k(1)$, $0 \le r \le 1$. When $k = 1$, we simply write W_1 as W and B_1 as B.

We impose the following "high level" conditions on data.

[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

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

where $\boldsymbol{S}_o$ is the nonsingular, matrix square root of

$$\boldsymbol{\Sigma}_o = \lim_{T\to\infty} \operatorname{var}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \boldsymbol{x}_t\epsilon_t\right),$$

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 $r \in (0,1]$ such that $\boldsymbol{M}_o$ is nonsingular.

By [A1],

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \boldsymbol{x}_t\epsilon_t \xrightarrow{D} \boldsymbol{S}_o\boldsymbol{W}_k(1) \stackrel{d}{=} \mathcal{N}(\mathbf{0},\, \boldsymbol{\Sigma}_o).$$

By [A2], $\boldsymbol{M}_T = T^{-1}\sum_{t=1}^{T} \boldsymbol{x}_t\boldsymbol{x}_t' \xrightarrow{\mathbb{P}} \boldsymbol{M}_o$, and hence

$$\begin{aligned}\sqrt{T}\big(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_o\big) &= \boldsymbol{M}_T^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \boldsymbol{x}_t(y_t - \boldsymbol{x}_t'\boldsymbol{\beta}_o) \\ &\xrightarrow{D} \boldsymbol{M}_o^{-1}\boldsymbol{S}_o\boldsymbol{W}_k(1) \\ &\stackrel{d}{=} \mathcal{N}(\mathbf{0},\, \boldsymbol{M}_o^{-1}\boldsymbol{\Sigma}_o\boldsymbol{M}_o^{-1}).\end{aligned} \tag{1}$$

This is the well known asymptotic normality result for the OLS estimator.

With the results in (1), the limiting distributions of the well known large-sample tests are easily obtained. Consider the null hypothesis $H_0: \boldsymbol{R}\boldsymbol{\beta}_o = \boldsymbol{r}$ with $\boldsymbol{R}$ a $q \times k$ matrix with full row rank. Under the null hypothesis, (1) implies

$$\sqrt{T}\big(\boldsymbol{R}\hat{\boldsymbol{\beta}}_T - \boldsymbol{r}\big) \xrightarrow{D} \mathcal{N}(\mathbf{0},\, \boldsymbol{R}\boldsymbol{M}_o^{-1}\boldsymbol{\Sigma}_o\boldsymbol{M}_o^{-1}\boldsymbol{R}'). \tag{2}$$

Replacing $\boldsymbol{M}_o$ and $\boldsymbol{\Sigma}_o$ with their respective consistent estimators $\boldsymbol{M}_T$ and $\widehat{\boldsymbol{\Sigma}}_T$, we have

$$\big(\boldsymbol{R}\boldsymbol{M}_T^{-1}\widehat{\boldsymbol{\Sigma}}_T\boldsymbol{M}_T^{-1}\boldsymbol{R}\big)^{-1/2}\sqrt{T}\boldsymbol{R}\big(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_o\big) \xrightarrow{D} \mathcal{N}(\mathbf{0},\, \boldsymbol{I}_q).$$

It follows that the Wald test of this hypothesis is

$$\mathcal{W}_T = T\big(\boldsymbol{R}\hat{\boldsymbol{\beta}}_T - \boldsymbol{r}\big)'\big(\boldsymbol{R}\boldsymbol{M}_T^{-1}\widehat{\boldsymbol{\Sigma}}_T\boldsymbol{M}_T^{-1}\boldsymbol{R}'\big)^{-1}\big(\boldsymbol{R}\hat{\boldsymbol{\beta}}_T - \boldsymbol{r}\big) \xrightarrow{D} \chi^2(q). \tag{3}$$

Note that $\mathcal{W}_T$ would not have a limiting χ^2 distribution if $\widehat{\boldsymbol{\Sigma}}_T$ is not a consistent estimator for $\boldsymbol{\Sigma}_o$. For other large-sample tests, such as the LM test and Hausman test, it is also crucial to have a consistent estimator of the asymptotic variance-covariance matrix.

3 HAC Estimation

For consistent estimation of $\boldsymbol{\Sigma}_o$, first note that, by definition,

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

For notation simplicity, we write

$$\boldsymbol{\Sigma}_o = \lim_{T\to\infty} \boldsymbol{\Sigma}_T = \lim_{T\to\infty} \sum_{j=-T+1}^{T-1} \boldsymbol{\Gamma}_T(j), \tag{4}$$

with

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

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

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

where $i = \sqrt{-1}$. In this case, $\boldsymbol{\Sigma}_o$ is $2\pi \times f(0)$ and hence also known as the long-run variance of $\boldsymbol{x}_t\epsilon_t$.

3.1 Kernel HAC Estimators

It is clear that the exact form of $\boldsymbol{\Sigma}_o$ depends on data characteristics. When $\boldsymbol{x}_t\epsilon_t$ are serially uncorrelated, all the autocovariances in (4) vanish, so that

$$\boldsymbol{\Sigma}_o = \lim_{T\to\infty} \boldsymbol{\Gamma}_T(0) = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}(\epsilon_t^2\boldsymbol{x}_t\boldsymbol{x}_t'). \tag{5}$$

This variance-covariance matrix can be consistently estimated by White's heteroskedasticity-consistent estimator:

$$\widehat{\boldsymbol{\Sigma}}_T = \frac{1}{T}\sum_{t=1}^{T} \hat{e}_t^2 \boldsymbol{x}_t \boldsymbol{x}_t',$$

with $\hat{e}_t$ the OLS residuals; see White (1984). The matrix (5) can be further simplified when ϵ_t are conditionally homoskedastic: $\mathbb{E}(\epsilon_t^2 \mid \mathcal{F}^{t-1}) = \sigma_o^2$, where $\mathcal{F}^{t-1}$ denotes the σ-algebra generated by $\{(\boldsymbol{x}_i, \epsilon_i), i \le t-1\}$. In this case, (5) is simplified as

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

which can be consistently estimated by $\widehat{\boldsymbol{\Sigma}}_T = \hat{\sigma}_T^2 \boldsymbol{M}_T$, with $\hat{\sigma}_T^2 = T^{-1}\sum_{t=1}^{T} \hat{e}_t^2$.

When $\boldsymbol{x}_t\epsilon_t$ exhibit serial correlations, it is still possible to estimate (4) consistently. Letting $\ell(T)$ denote a function of T that diverges with T we have

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

as T tends to infinity. It is then natural to estimate $\boldsymbol{\Sigma}_T^{\dagger}$ by its sample counterpart:

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

with the sample autocovariances:

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

The estimator $\widehat{\boldsymbol{\Sigma}}_T^{\dagger}$ would be consistent for $\boldsymbol{\Sigma}_o$ provided that $\ell(T)$ does not grow too fast with T; see the discussion in Section 3.2.

A problem with $\widehat{\boldsymbol{\Sigma}}_T^{\dagger}$ is that it need not be a positive semi-definite matrix and hence can not be a proper variance-covariance matrix. A consistent estimator that is also positive semi-definite is the following estimator of the spectral density:

$$\widehat{\boldsymbol{\Sigma}}_T^{\kappa} = \sum_{j=-T+1}^{T-1} \kappa\Big(\frac{j}{\ell(T)}\Big) \widehat{\boldsymbol{\Gamma}}_T(j), \tag{6}$$

where κ is a proper kernel function and $\ell(T)$ is its bandwidth, which jointly determine the weights assigned to $\widehat{\boldsymbol{\Gamma}}_T(j)$. Typically, κ is required to satisfy: $|\kappa(x)| \le 1$, $\kappa(0) = 1$,

$\kappa(x) = \kappa(-x)$ for all $x \in \mathbb{R}$, $\int |\kappa(x)| \, \mathrm{d}x < \infty$, κ is continuous at 0 and at all but a finite number of other points in $\mathbb{R}$, and

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

Note that the last condition ensures positive semi-definiteness; see Andrews (1991).

Below are some commonly used kernel functions:[1]

(i) Bartlett kernel (Newey and West, 1987):

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

(ii) Parzen kernel (Gallant, 1987):

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

(iii) Quadratic spectral kernel (Andrews, 1991):

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

(iv) Daniel kernel (Ng and Perron, 1996):

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

These kernels are all symmetric about the vertical axis, where the first two kernels have a bounded support $[-1, 1]$, but the other two have unbounded support. These kernel functions with non-negative x are depicted in Figure 1. For the Bartlett and Parzen kernels, the weight assigned to $\widehat{\mathbf{\Gamma}}_T(j)$ decreases with $|j|$ and becomes zero for $|j| \geq \ell(T)$. Hence, $\ell(T)$ in these functions is also known as a truncation lag parameter. For the quadratic spectral and Daniel kernels, $\ell(T)$ does not have this interpretation because the weight decreases to zero at, respectively, $|j| = 1.2\ell(T)$ and $j = \ell(T)$, but then exhibits damped sine waves afterwards. In what follows, the HAC estimator (6) with the Bartlett, Parzen, quadratic spectral and Daniel kernels will be denoted as, respectively, $\widehat{\mathbf{\Sigma}}_T^B$, $\widehat{\mathbf{\Sigma}}_T^P$, $\widehat{\mathbf{\Sigma}}_T^{QS}$ and $\widehat{\mathbf{\Sigma}}_T^D$.

[1] Unlike Andrews (1991), we do not discuss the Tukey-Hanning kernel because it does not result in a positive semi-definite HAC estimator.

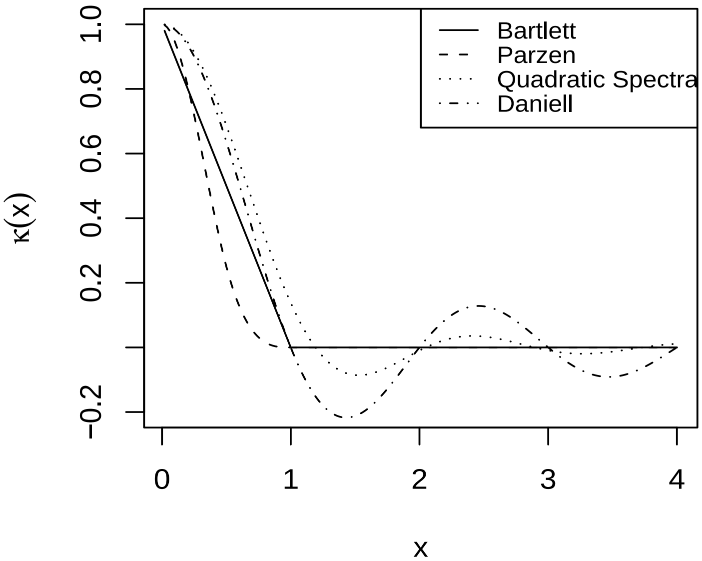

Figure 1: The Bartlett, Parzen, quandratic spectral and Daniel kernels.

Remark: The kernel weighting scheme brings bias to the estimated autocovariances. As $\ell(T)$ diverges with T, $\widehat{\boldsymbol{\Gamma}}_T(j)$ receive weights close to unity asymptotically. Thus, the kernel function entails little asymptotic bias and hence does not affect the consistency of $\widehat{\boldsymbol{\Sigma}}_T^{\kappa}$. Such bias may not be negligible in finite samples, however.

3.2 Choices of Kernel Function and Bandwidth

In theory, the bandwidth $\ell(T)$ must diverge with T at a rate slower than T so as to ensure the consistency of $\widehat{\boldsymbol{\Sigma}}_T^{\kappa}$. It should be clear from (4) that, when $\ell(T)$ does not depend on T, $\widehat{\boldsymbol{\Sigma}}_T^{\kappa}$ can not be a consistent estimator because it does not estimate infinitely many autocovariances in $\boldsymbol{\Sigma}_o$. When $\ell(T)$ grows too fast, the bias resulted from estimating the autocovariances will accumulate too rapidly and can not be averaged out in the limit. In this case, $\widehat{\boldsymbol{\Sigma}}_T^{\kappa}$ also loses consistency.

Newey and West (1987) found that, under certain mixing conditions on data and the Bartlett kernel, $\ell(T)$ is $o(T^{1/4})$. Andrews (1991) improved on their result by showing that for the Bartlett, Parzen, and quadratic spectral kernels, $\ell(T)$ can grow at the rate $o(T^{1/2})$ when data are fourth order stationary; such a rate may be further improved to $o(T)$ when the assumption on data is strengthened, as also shown in de Jong and Davidson (2000) under the condition that the data are near epoch dependent on some mixing sequence.[2] Note that a faster divergence rate of $\ell(T)$ permits estimating more autocovariances as sample size grows.

The growth rate of $\ell(T)$ provides little guidance on how $\ell(T)$ should be chosen in a given sample. Based on an asymptotically truncated mean squared error (MSE) criterion, Andrews (1991) showed that the optimal growth rate $\ell^*(T)$ are

$$\begin{aligned}
\ell^*(T) &= 1.1447(c_1 T)^{1/3}, \quad \text{(Bartlett)} \\
\ell^*(T) &= 2.6614(c_2 T)^{1/5}, \quad \text{(Parzen)} \\
\ell^*(T) &= 1.3221(c_2 T)^{1/5}, \quad \text{(quadractic spectral)}
\end{aligned} \tag{7}$$

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

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

This implies that the rate of convergence of $\widehat{\boldsymbol{\Sigma}}_T^B$ is $T^{-1/3}$ and that of $\widehat{\boldsymbol{\Sigma}}_T^P$ and $\widehat{\boldsymbol{\Sigma}}_T^{QS}$ is $T^{-2/5}$. That is, the commonly used Bartlett kernel in fact has a slower convergence rate. In terms of asymptotic MSE, Andrews (1991) found that the quardratic spectral kernel is 8.6% more efficient asymptotically than the Parzen kernel, while the Bartlett kernel is the least efficient among those considered in Andrews (1991) due to its slower MSE convergence rate. Thus, as far as MSE is concerned, the quadratic spectral kernel is to be preferred, at least asymptotically.

Based on the optimal growth rates of (7), Andrews (1991) suggested an "automatic" method to determine the bandwidth in a given sample. In particular, one first estimates an AR(1) model for each element of $\boldsymbol{x}_t\epsilon_t$ and obtains the AR(1) coefficients $\hat{\rho}_a$ and variance estimates $\hat{\sigma}_a^2$ for $a = 1, \ldots, k$. Then, the unknown numbers (c_1 and c_2) in (7) are computed according to:

$$\begin{aligned}
\hat{c}_1 &= \sum_{a=1}^{k} \frac{4\hat{\rho}_a^2\hat{\sigma}_a^4}{(1-\hat{\rho}_a)^6(1+\hat{\rho}_a)^2} \Bigg/ \sum_{a=1}^{k} \frac{\hat{\sigma}_a^4}{(1-\hat{\rho}_a)^4}, \\
\hat{c}_2 &= \sum_{a=1}^{k} \frac{4\hat{\rho}_a^2\hat{\sigma}_a^4}{(1-\hat{\rho}_a)^8} \Bigg/ \sum_{a=1}^{k} \frac{\hat{\sigma}_a^4}{(1-\hat{\rho}_a)^4}.
\end{aligned}$$

The desired bandwidth is obtained by plugging $\hat{c}_1$ and $\hat{c}_2$ into (7). Other models, such as VAR(1), ARMA(1,1), MA(m) models, may also be employed in the first step; the corresponding formulae for $\hat{c}_1$ and $\hat{c}_2$ are given in Andrews (1991, pp. 835–836).

[2] For the definition of near epoch dependence and related asymptotic theory see Gallant and White (1988).

3.3 Other Improved HAC Estimators

Other than finding a suitable kernel function and a proper bandwidth, the performance of kernel HAC estimators may also be improved by pre-whitening the data, as proposed by Andrews and Monahan (1992), or by computing sample autocovariances based on forecast errors, instead of OLS residuals, as proposed by Kuan and Hsieh (2008).

The whitening method relies on the model residuals from regressing $\hat{\boldsymbol{v}}_t = \boldsymbol{x}_t \hat{e}_t$ on its lagged values. Intuitively, these residuals will be less serially correlated and hence have a smoother spectral density around frequency zero. This usually renders the estimate of spectral density more accurate. Consider a b th order vector AR (VAR) model of $\hat{\boldsymbol{v}}_t$:

$$\hat{\boldsymbol{v}}_t = \sum_{j=1}^{b} \widehat{\boldsymbol{A}}_j \hat{\boldsymbol{v}}_{t-j} + \ddot{\boldsymbol{v}}_t,$$

where $\widehat{\boldsymbol{A}}_j$ are the estimated coefficient matrices, and $\ddot{\boldsymbol{v}}_t$ are the VAR residuals. By computing the sample autocovariances of $\ddot{\boldsymbol{v}}_t$, we can easily construct a kernel HAC estimator $\ddot{\boldsymbol{\Sigma}}_T^{\kappa}$ according to (6). The pre-whitened HAC estimator of $\boldsymbol{\Sigma}_o$ is then recovered by "re-coloring" $\ddot{\boldsymbol{\Sigma}}_T^{\kappa}$, i.e.,

$$\widehat{\ddot{\boldsymbol{\Sigma}}_T^{\kappa}} = \widehat{\boldsymbol{D}} \ddot{\boldsymbol{\Sigma}}_T^{\kappa} \widehat{\boldsymbol{D}}', \qquad \widehat{\boldsymbol{D}} = \Big(\boldsymbol{I}_k - \sum_{j=1}^{b} \widehat{\boldsymbol{A}}_j \Big)^{-1}.$$

While this estimator performs quite well in finite samples, its dependence on another user-chosen parameter, the VAR lag order b, makes the implementation more difficult.

Kuan and Hsieh (2008) found that the kernel HAC estimator may also be improved by replacing $\hat{e}_t$ with one-step-ahead forecast errors: $\tilde{e}_t = y_t - \boldsymbol{x}_t \tilde{\boldsymbol{\beta}}_{t-1}$, where $\tilde{\boldsymbol{\beta}}_{t-1}$ are the recursive OLS estimators based on the subsample of first $t-1$ observation. This estimator is computationally more demanding because of recursive estimation, but it avoids an unnecessary constraint resulted from OLS estimation, namely, $\sum_{t=1}^{T} \boldsymbol{x}_t \hat{e}_t = \boldsymbol{0}$. Kuan and Hsieh (2008) demonstrated that, compared with the kernel HAC estimator with OLS residuals, the forecast-error-based estimator has a smaller bias but a larger MSE, and it usually results in smaller size distortion than do the conventional and pre-whitened HAC estimators. This suggests that bias-reduction may be more important than MSE-minimization in HAC estimation, cf. Andrews (1991).

4 KVB Approach

We have seen that kernel HAC estimation requires the choices of the kernel function and its bandwidth. Such choices are somewhat arbitrary in practice. To circumvent these difficulties, KVB proposed an approach that yields an asymptotically pivotal test without consistent estimation of the asymptotic covariance matrix. The main idea of the KVB approach is to employ a normalizing matrix that is capable of eliminating the nuisance parameters in $\boldsymbol{S}_o$, the matrix square root of $\boldsymbol{\Sigma}_o$. This normalizing matrix is inconsistent for $\boldsymbol{\Sigma}_o$ but is free from the choice of the kernel bandwidth, in contrast with kernel HAC estimators.

4.1 Robust Tests with KVB's Normalizing Matrix

Let $\hat{\boldsymbol{\varphi}}_t$ denote the normalized partial sum of $\boldsymbol{x}_i \hat{e}_i$:

$$\hat{\boldsymbol{\varphi}}_t = \frac{1}{\sqrt{T}} \sum_{i=1}^{t} \boldsymbol{x}_i \hat{e}_i,$$

and define the normalizing matrix $\widehat{\boldsymbol{C}}_T$ as

$$\widehat{\boldsymbol{C}}_T = \frac{1}{T} \sum_{t=1}^{T} \hat{\boldsymbol{\varphi}}_t \hat{\boldsymbol{\varphi}}_t' = \frac{1}{T^2} \sum_{t=1}^{T} \left(\sum_{i=1}^{t} \boldsymbol{x}_i \hat{e}_i \right) \left(\sum_{i=1}^{t} \hat{e}_i \boldsymbol{x}_i' \right). \tag{8}$$

By [A1]–[A2], we have from (1) that

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

Hence,

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

While the kerenl HAC estimator (6) has a nonstochastic limit, $\widehat{\boldsymbol{C}}_T$ in (8) has a random limit depending on $\boldsymbol{S}_o$ and a functional of the Brownian bridge.

By replacing $\widehat{\boldsymbol{\Sigma}}_T$ in the Wald test (3) with $\widehat{\boldsymbol{C}}_T$, we obtain the following statistic:

$$\mathcal{W}_T^{\dagger} = T\big(\boldsymbol{R}\hat{\boldsymbol{\beta}}_T - \boldsymbol{r}\big)' \big(\boldsymbol{R}\boldsymbol{M}_T^{-1} \widehat{\boldsymbol{C}}_T \boldsymbol{M}_T^{-1} \boldsymbol{R}'\big)^{-1} \big(\boldsymbol{R}\hat{\boldsymbol{\beta}}_T - \boldsymbol{r}\big). \tag{9}$$

To derive its null limit, first note that

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

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, $\boldsymbol{R}\boldsymbol{M}_o^{-1}\boldsymbol{S}_o\boldsymbol{W}_k(r) \stackrel{d}{=} \boldsymbol{G}_o\boldsymbol{W}_q(r)$, and hence

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

Note also that (2) can be expressed as

$$\sqrt{T}\boldsymbol{R}\big(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_o\big) \xrightarrow{D} \boldsymbol{G}_o\boldsymbol{W}_q(1).$$

Putting these results together, the statistic (9) is such that

$$\mathcal{W}_T^{\dagger} \Rightarrow \boldsymbol{W}_q(1)'\boldsymbol{G}_o'\big(\boldsymbol{G}_o\boldsymbol{P}_q\boldsymbol{G}_o'\big)^{-1}\boldsymbol{G}_o\boldsymbol{W}_q(1) = \boldsymbol{W}_q(1)'\boldsymbol{P}_q^{-1}\boldsymbol{W}_q(1). \tag{10}$$

Compared with (3), $\mathcal{W}_T^{\dagger}$ does not have a limiting χ^2 distribution, yet it is asymptotically pivotal because the null limit in (10) does not depend on the matrix of nuisance parameters, $\boldsymbol{G}_o$. Although the asymptotic distribution of $\mathcal{W}_T^{\dagger}$ is non-standard, it can be easily simulated. Lobato (2001) reported the critical values of this distribution for various values of q. Note that Kiefer et al. (2000) considered a statistic, analogous to the classical F test: $F_T^{\dagger} = \mathcal{W}_T^{\dagger}/q$, which has the null limit $\boldsymbol{W}_q(1)'\boldsymbol{P}_q^{-1}\boldsymbol{W}_q(1)/q$.

When the null hypothesis is $\beta_i = r$, a robust test analogous to the conventional t test is thus

$$t^{\dagger} = \frac{\sqrt{T}\big(\hat{\beta}_{i,T} - r\big)}{\sqrt{\hat{\delta}_i}} \xrightarrow{D} \frac{W(1)}{\big[\int_0^1 B(r)^2\,\mathrm{d}r\big]^{1/2}}, \tag{11}$$

where $\hat{\delta}_i$ is the i th diagonal element of $\boldsymbol{M}_T^{-1}\widehat{\boldsymbol{C}}_T\boldsymbol{M}_T^{-1}$. Some quantiles of this asymptotic distribution, taken from Kiefer et al. (2000), are summarized in the second row of Table 1. This distribution is symmetric about the vertical axis but is more disperse than the standard normal distribution. Abadir and Paruolo (2002) showed that the distribution of the limit in (11) is the same as that analyzed in Abadir and Paruolo (1997) which contains analytic formulae of its density function and moments.

Remark: An advantage of the robust test with KVB's normalizing matrix is that its asymptotic distribution is usually a very good approximation to its finite-sample counterpart. The empirical size of such test is thus close to the nominal size; that is, the

Table 1: The quantiles of the $t^{\dagger}$ and t tests based on $\widehat{\boldsymbol{\Sigma}}_T^{\kappa}$ without truncation.

prob.	90%	95%	97.5%	99%
$t^{\dagger}$ with $\widehat{\boldsymbol{C}}_T$	3.890	5.374	6.811	8.544
t with $\widehat{\boldsymbol{\Sigma}}_T^{B}$	2.740	3.764	4.771	6.090
t with $\widehat{\boldsymbol{\Sigma}}_T^{P}$	2.840	4.228	5.671	8.112
t with $\widehat{\boldsymbol{\Sigma}}_T^{QS}$	5.188	8.283	12.374	20.380
t with $\widehat{\boldsymbol{\Sigma}}_T^{D}$	4.822	7.711	11.573	19.180

probability of type I error is properly controlled. By contrast, kernel HAC estimators are typically downward biased, so that the resulting tests are over-sized in finite samples. As a result, the tests based on a kernel HAC estimator may incorrectly reject the null more frequently than it should.

4.2 Kernel-Based Normalizing Matrices

Kiefer and Vogelsang (2002a) showed that $2\widehat{\boldsymbol{C}}_T$ is algebraically equivalent to $\widehat{\boldsymbol{\Sigma}}_T^{B}$ without truncation, i.e., $\ell(T) = T$. To see this, note that summation by parts gives

$$\sum_{j=1}^{T} a_j b_j = a_T \sum_{i=1}^{T} b_i - \sum_{j=1}^{T-1} (a_{j+1} - a_j) \left(\sum_{i=1}^{j} b_i \right),$$

and $\sum_{i=1}^{T} \hat{\boldsymbol{v}}_i = \sum_{i=1}^{T} \boldsymbol{x}_i \hat{e}_i = \mathbf{0}$. Then, $\widehat{\boldsymbol{\Sigma}}_T^{B}$ without truncation is

$$\begin{aligned}
\widehat{\boldsymbol{\Sigma}}_T^{B} &= \frac{1}{T} \sum_{i=1}^{T} \hat{\boldsymbol{v}}_i \sum_{j=1}^{T} \Big(1 - \frac{|i-j|}{T} \Big) \hat{\boldsymbol{v}}_j' \\
&= \frac{1}{T} \frac{1}{\sqrt{T}} \sum_{i=1}^{T} \hat{\boldsymbol{v}}_i \sum_{j=1}^{T} (|i-j-1| - |i-j|) \left(\frac{1}{\sqrt{T}} \sum_{h=1}^{j} \hat{\boldsymbol{v}}_h' \right).
\end{aligned}$$

As $(|i-j-1| - |i-j|) = -1$ if $j > j$ and $= 1$ if $i \leq j$,

$$\widehat{\boldsymbol{\Sigma}}_T^{B} = \frac{1}{T} \sum_{j=1}^{T} \left(\frac{1}{\sqrt{T}} \sum_{i=1}^{j} \hat{\boldsymbol{v}}_i - \sum_{i=j+1}^{T} \hat{\boldsymbol{v}}_i \right) \hat{\boldsymbol{\varphi}}_j' = \frac{1}{T} \sum_{j=1}^{T} 2 \hat{\boldsymbol{\varphi}}_j \hat{\boldsymbol{\varphi}}_j' = 2\widehat{\boldsymbol{C}}_T.$$

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 $t^{\dagger}/\sqrt{2}$; some of its critical values, computed from the analytic result of Abadir and Paruolo (1997), are summarized in the third row of Table 1. It can be seen that this distribution is still more disperse than the standard normal distribution.

Kiefer and Vogelsang (2002b) extended the relation between $\widehat{\boldsymbol{C}}_T$ and $\widehat{\boldsymbol{\Sigma}}_T^B$ discussed above and showed that, for a general kernel function with the continuous second order derivative κ'', $\widehat{\boldsymbol{\Sigma}}_T^{\kappa} \Rightarrow \boldsymbol{S}_o \boldsymbol{P}_k^{\kappa} \boldsymbol{S}_o'$, with

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

where the last equality follows from integration by parts. Although the Bartlett kernel is not continuously differentiable everywhere, the above result holds with $\boldsymbol{P}_k^B = 2\int_0^1 \boldsymbol{B}_k(r)\boldsymbol{B}_k(r)' \, \mathrm{d}r$. Consequently, the tests based on various kernel HAC estimators with the bandwidth equal to sample size are also asymptotically pivotal. This result is practically useful because kernel HAC estimators have been included in most econometric packages. For the t tests based on $\widehat{\boldsymbol{\Sigma}}_T^{\kappa}$ with the bandwidth equal to sample, some of their critical values are given in Table 1. See Kiefer and Vogelsang (2002b) and Phillips et al. (2006) for other critical values.

It has been demonstrated in Kiefer and Vogelsang (2002b) that, for $\widehat{\boldsymbol{\Sigma}}_T^{\kappa}$ with the bandwidth equal to sample size, $\widehat{\boldsymbol{\Sigma}}_T^B$ compares favorably with $\widehat{\boldsymbol{\Sigma}}_T^{QS}$ in terms of test power. This is in sharp contrast with the result in HAC estimation, where the latter is preferred to other kernels.

5 Robust M Tests

We now extend the KVB approach to constructing robust M tests. The M test considered by Newey (1985), Tauchen (1985), and White (1987) is a general class of specification tests on moment conditions that are functions of unknown parameters. This class of tests includes most model specification tests in the econometrics literature; the specification tests under the quasi-maximum likelihood framework are leading examples.

5.1 Robust M Tests Based on the KVB Approach

We consider the hypothesis that $\mathbb{E}[\boldsymbol{f}(\boldsymbol{\eta}_t;\boldsymbol{\theta}_o)] = \boldsymbol{0}$ for some $\boldsymbol{\theta}_o \in \Theta \subset \mathbb{R}^k$, where $\boldsymbol{\eta}_t$ are random data vectors, $\boldsymbol{\theta}_o$ is the $k \times 1$ true parameter vector, and $\boldsymbol{f}$ is a $q \times 1$ vector of functions. Let $\hat{\boldsymbol{\theta}}_T$ denote a root-T consistent estimator of $\boldsymbol{\theta}_o$ such that

$$\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_o) = \boldsymbol{Q}_o \left[\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \boldsymbol{q}(\boldsymbol{\eta}_t; \boldsymbol{\theta}_o) \right] + o_{\mathbb{P}}(1), \tag{12}$$

where $\boldsymbol{Q}_o$ is a $k \times k$ nonsingular matrix and $\boldsymbol{q}$ is a vector-valued function in $\mathbb{R}^k$. For example, when $\hat{\boldsymbol{\theta}}_T$ is a quasi-maximum likelihood (QML) estimator, $\boldsymbol{Q}_o$ is the inverse of the limit of the Hessian matrix evaluated at $\boldsymbol{\theta}_o$, and $\boldsymbol{q}(\boldsymbol{\eta}_t; \boldsymbol{\theta}_o)$ is the score function evaluated at $\boldsymbol{\theta}_o$. The nonlinear least dquares (NLS) and GMM estimators can also be expressed in a similar form.

To illustrate, we first consider the case that $\boldsymbol{\theta}_o$ is known. Define

$$\boldsymbol{m}_{[rT]}(\boldsymbol{\theta}) = \frac{1}{T} \sum_{t=1}^{[rT]} \boldsymbol{f}(\boldsymbol{\eta}_t; \boldsymbol{\theta}), \qquad 0 < r \leq 1,$$

where T is the sample size; for $r = 1$, $\boldsymbol{m}_T(\boldsymbol{\theta})$ is the sample average of $\boldsymbol{f}(\boldsymbol{\eta}_t; \boldsymbol{\theta})$. An M test is based on $\boldsymbol{m}_T(\boldsymbol{\theta}_o)$. When $T^{1/2}\boldsymbol{m}_T(\boldsymbol{\theta}_o)$ is governed by a central limit theorem such that under the null, $T^{1/2}\boldsymbol{m}_T(\boldsymbol{\theta}_o) \xrightarrow{D} \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_o)$, where $\boldsymbol{\Sigma}_o$ is nonsingular. When $\boldsymbol{\Sigma}_o$ is consistently estimated by an estimator $\widehat{\boldsymbol{\Sigma}}_T$, the conventional M test is:

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

under the null hypothesis. This is analogous to the result of the Wald test (3).

We now impose the following conditions.

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

[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 with the nonsingular diagonal blocks $\boldsymbol{G}_{11}$ $(q \times q)$ and $\boldsymbol{G}_{22}$ $(k \times k)$ and the off-diagonal blocks $\boldsymbol{G}_{12}$ $(q \times k)$ and $\boldsymbol{G}_{21}$ $(k \times q)$.

[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 $0 < r \leq 1$, where $\boldsymbol{F}_o$ is a $q \times k$ non-stochastic matrix; $\nabla_{\boldsymbol{\theta}} \boldsymbol{F}_{[rT]}(\boldsymbol{\theta}_o)$ is bounded in probability.

Observe that the upper left $(q \times q)$ block of $\boldsymbol{G}_o \boldsymbol{G}_o'$, $\boldsymbol{G}_{11}\boldsymbol{G}_{11}' + \boldsymbol{G}_{12}\boldsymbol{G}_{12}'$, is the asymptotic covariance matrix of $T^{-1/2} \sum_{t=1}^{T} \boldsymbol{f}(\boldsymbol{\eta}_t; \boldsymbol{\theta}_o)$, and hence is also $\boldsymbol{\Sigma}_o$ in [B1](a). The lower right $(k \times k)$ block of $\boldsymbol{G}_o \boldsymbol{G}_o'$, $\boldsymbol{G}_{22}\boldsymbol{G}_{22}' + \boldsymbol{G}_{21}\boldsymbol{G}_{21}'$, is the asymptotic covariance matrix of $T^{-1/2} \sum_{t=1}^{T} \boldsymbol{q}(\boldsymbol{\eta}_t; \boldsymbol{\theta}_o)$. Then, in view of (12), $\hat{\boldsymbol{\theta}}_T$ is asymptotically normally distributed with

$$\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_o) \Rightarrow \boldsymbol{Q}_o \boldsymbol{A}_o \boldsymbol{W}_k(1),$$

where $\boldsymbol{A}_o$ is the matrix square root of $\boldsymbol{G}_{22}\boldsymbol{G}_{22}' + \boldsymbol{G}_{21}\boldsymbol{G}_{21}'$.

Remark: It is important to note that [B1](b) requires $\boldsymbol{G}_o$ to be nonsingular, so that the sample moments can not be correlated with the estimator asymptotically. This excludes the cases in which the estimator is obtained from solving the sample moment conditions. For example, when $\hat{\boldsymbol{\theta}}_T$ is the quasi-maximum likelihood estimator, $\boldsymbol{f}$ in the moment conditions can not be the score function. See also Section 5.3.

Following the KVB approach discussed in Section 4, a robust M test without consistent estimation of the asymptotic covariance matrix is

$$\mathcal{M}_T = T\, \boldsymbol{m}_T(\boldsymbol{\theta}_o)' \boldsymbol{C}_T(\boldsymbol{\theta}_o)^{-1} \boldsymbol{m}_T(\boldsymbol{\theta}_o), \tag{13}$$

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

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

Note that the summands of $\boldsymbol{\varphi}_t(\boldsymbol{\theta}_o)$ are "centered" (i.e., $\boldsymbol{f}(\boldsymbol{\eta}_i; \boldsymbol{\theta}_o) - \boldsymbol{m}_T(\boldsymbol{\theta}_o)$).

By [B1](a), $T^{1/2} \boldsymbol{m}_T(\boldsymbol{\theta}_o) \Rightarrow \boldsymbol{S}_o \boldsymbol{W}_q(1)$, and

$$\boldsymbol{\varphi}_{[rT]}(\boldsymbol{\theta}_o) \Rightarrow \boldsymbol{S}_o \big[\boldsymbol{W}_q(r) - r \boldsymbol{W}_q(1)\big] = \boldsymbol{S}_o \boldsymbol{B}_q(r), \quad 0 \le r \le 1.$$

Hence, $\boldsymbol{C}_T(\boldsymbol{\theta}_o) \Rightarrow \boldsymbol{S}_o \boldsymbol{P}_q \boldsymbol{S}_o'$ with $\boldsymbol{P}_q = \int_0^1 \boldsymbol{B}_q(r) \boldsymbol{B}_q(r)' \, \mathrm{d}r$. It follows that

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

This result is analogous to the result of the robust "Wald" test in (10).

In practice, $\boldsymbol{\theta}_o$ is typically unknown and must be estimated. In the light of (13), it is natural to compute a robust M test as

$$\widehat{\mathcal{M}}_T = T\, \boldsymbol{m}_T(\hat{\boldsymbol{\theta}}_T)' \widehat{\boldsymbol{C}}_T^{-1} \boldsymbol{m}_T(\hat{\boldsymbol{\theta}}_T), \tag{14}$$

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

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

To derive the limit of (14), we note that a Taylor expansion of $T^{1/2}\boldsymbol{m}_{[rT]}(\hat{\boldsymbol{\theta}}_T)$ about $\boldsymbol{\theta}_o$ gives

$$\sqrt{T}\boldsymbol{m}_{[rT]}(\hat{\boldsymbol{\theta}}_T) = \sqrt{T}\boldsymbol{m}_{[rT]}(\boldsymbol{\theta}_o) + \frac{[rT]}{T}\boldsymbol{F}_{[rT]}(\boldsymbol{\theta}_o)\big[\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_o)\big] + o_{\mathbb{P}}(1), \tag{15}$$

where $\boldsymbol{F}_{[rT]}(\boldsymbol{\theta}_o)$ is defined in [B2]. The second term on the right-hand side of (15) characterizes the estimation effect of replacing $\boldsymbol{\theta}_o$ with $\hat{\boldsymbol{\theta}}_T$ in $\boldsymbol{m}_{[rT]}$, and it converges to $r\boldsymbol{F}_o\boldsymbol{Q}_o\boldsymbol{A}_o\boldsymbol{W}_k(1)$. Clearly, the estimation would be absent in the limit if $\boldsymbol{F}_o = \boldsymbol{0}$.

By (15), $T^{1/2}\,\boldsymbol{m}_T(\hat{\boldsymbol{\theta}}_T)$ involves the estimation effect, and [B1](b) and [B2] imply

$$\begin{aligned}\sqrt{T}\boldsymbol{m}_T(\hat{\boldsymbol{\theta}}_T) &= \sqrt{T}\boldsymbol{m}_T(\boldsymbol{\theta}_o) + \boldsymbol{F}_T(\boldsymbol{\theta}_o)\boldsymbol{Q}_o\frac{1}{\sqrt{T}}\sum_{i=1}^{T}\boldsymbol{q}(\boldsymbol{\eta}_t,\boldsymbol{\theta}_o) + o_{\mathbb{P}}(1),\\ &\Rightarrow [\boldsymbol{I}_q \;\; \boldsymbol{F}_o\boldsymbol{Q}_o]\boldsymbol{G}_o\boldsymbol{W}_{q+k}(1)\\ &\overset{d}{=} \boldsymbol{V}_o\boldsymbol{W}_q(1),\end{aligned}$$

where $\boldsymbol{V}_o$ is the matrix square root of $[\boldsymbol{I}_q \;\; \boldsymbol{F}_o\boldsymbol{Q}_o]\boldsymbol{G}_o\boldsymbol{G}_o'[\boldsymbol{I}_q \;\; \boldsymbol{F}_o\boldsymbol{Q}_o]'$. On the other hand,

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

which is free from the estimation effect due to “centering”. It follows that

$$\widehat{\boldsymbol{C}}_T = \boldsymbol{C}_T(\boldsymbol{\theta}_o) + o_{\mathbb{P}}(1) \Rightarrow \boldsymbol{S}_o\boldsymbol{P}_q\boldsymbol{S}_o',$$

as in the case that $\boldsymbol{\theta}_o$ is known. Hence, using $\widehat{\boldsymbol{C}}_T$ as a normalizing matrix is unable to eliminate the nuisance parameters in $\boldsymbol{V}_o$. Indeed,

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

where the matrices of nuisance parameter, $\boldsymbol{V}_o$ and $\boldsymbol{S}_o$, are present in the limit. It is easily seen that, when there is no estimation effect ($\boldsymbol{F}_o = \boldsymbol{0}$) and $\boldsymbol{V}_o = \boldsymbol{S}_o$ so that

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

Thus, $\widehat{\mathcal{M}}_T$ in (14) is not asymptotically pivotal unless the estimation effect is absent.

The discussion above suggests that, to construct a proper normalizing matrix in the presence of estimation effect, $\boldsymbol{\varphi}_{[rT]}$ ought to preserve the estimation effect and converge to a limit with the matrix of nuisance parameters $\boldsymbol{V}_o$. To this end, Kuan and Lee (2006) proposed computing the normalizing matrix based on $\tilde{\boldsymbol{\theta}}_t$, $t = k+1, \ldots, T$, which are the recursive counterparts of $\hat{\boldsymbol{\theta}}_T$ and computed from the sub-sample of first t observations. Specifically,

$$\widetilde{\boldsymbol{C}}_T = T^{-1} \sum_{t=k+1}^{T} \tilde{\boldsymbol{\varphi}}_t \tilde{\boldsymbol{\varphi}}_t'$$

with

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

The M test analogous to (14) is thus

$$\widetilde{\mathcal{M}}_T = T\, \boldsymbol{m}_T(\hat{\boldsymbol{\theta}}_T)' \widetilde{\boldsymbol{C}}_T^{-1} \boldsymbol{m}_T(\hat{\boldsymbol{\theta}}_T). \tag{16}$$

For an extension of $\widetilde{\mathcal{M}}_T$ that admits kernel-based normalizing matrices, see Lee (2006).

Note that the first-order Taylor expansion of $T^{-1/2}\boldsymbol{m}_{[rT]}(\tilde{\boldsymbol{\theta}}_{[rT]})$ about $\boldsymbol{\theta}_o$ yields

$$\sqrt{T}\boldsymbol{m}_{[rT]}(\tilde{\boldsymbol{\theta}}_{[rT]}) = \sqrt{T}\boldsymbol{m}_{[rT]}(\boldsymbol{\theta}_o) + \boldsymbol{F}_{[rT]}(\boldsymbol{\theta}_o)\boldsymbol{Q}_o \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{[rT]} \boldsymbol{q}(\boldsymbol{\eta}_t; \boldsymbol{\theta}_o) \right) + o_{\mathbb{P}}(1),$$

cf. (15). Given [B1] and [B2],

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

and $T^{1/2}\boldsymbol{m}_T(\hat{\boldsymbol{\theta}}_T) \Rightarrow \boldsymbol{V}_o\boldsymbol{W}_q(1)$. Thus,

$$\tilde{\boldsymbol{\varphi}}_{[rT]} = \sqrt{T}\boldsymbol{m}_{[rT]}(\tilde{\boldsymbol{\theta}}_{[rT]}) - \frac{[rT]}{T}\sqrt{T}\boldsymbol{m}_T(\tilde{\boldsymbol{\theta}}_T) \Rightarrow \boldsymbol{V}_o\boldsymbol{B}_q(r),$$

which involves $\boldsymbol{V}_o$, instead of $\boldsymbol{S}_o$. Consequently, $\widetilde{\boldsymbol{C}}_T \Rightarrow \boldsymbol{V}_o\boldsymbol{P}_q\boldsymbol{V}_o'$, and

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

This shows that $\widetilde{\mathcal{M}}_T$ is asymptotically pivotal regardless of the estimation effect and has the same weak limit as $\widehat{\mathcal{M}}_T$ and $\mathcal{M}_T$. It is, however, computationally more demanding to implement $\widetilde{\mathcal{M}}_T$ because recursive estimation is needed.

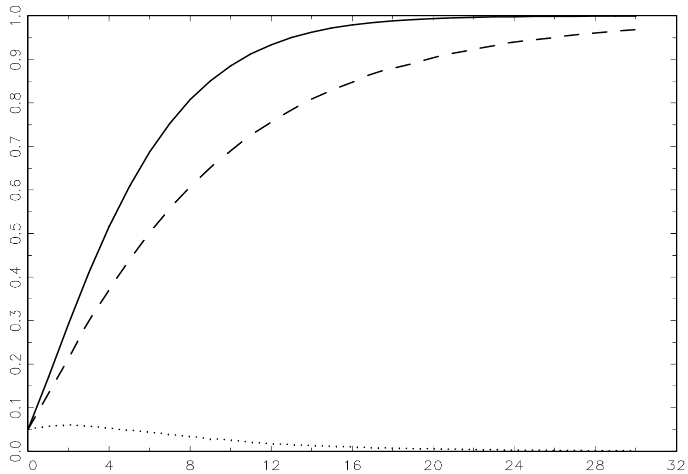

Figure 2: The asymptotic local powers of the standard M test (solid), $\widetilde{\mathcal{M}}_T$ (dashed) and $\ddot{\mathcal{M}}_T$ (dotted) at 5% level.

Remarks:

1. $\widetilde{\mathcal{M}}_T$ is not only a robust M test without consistent estimation of asymptotic covariance matrix, but also an alternative to the estimation effect problem in M testing. This is practically important for M tests because consistent estimation of the asymptotic covariance matrix of $T^{1/2}\boldsymbol{m}_T(\hat{\boldsymbol{\theta}}_T)$, $\boldsymbol{V}_o\boldsymbol{V}_o'$, is typically difficult.

2. It is crucial to employ a "centered" normalizing matrix in the proposed tests. Kuan and Lee (2006) found that the M test with a non-centered normalizing matrix (in which $\tilde{\boldsymbol{\varphi}}_t = T^{-1/2}\sum_{i=1}^{t}\boldsymbol{f}(\boldsymbol{\eta}_i, \tilde{\boldsymbol{\theta}}_t)$) virtually has no power. This is consistent with a conclusion of Hall (2000) on kernel HAC estimators; see also Hall, Inoue, and Peixe (2003). To illustrate, we plot the simulated asymptotic local powers of the standard M test (based on its asymptotic, non-central χ^2 distribution), $\widetilde{\mathcal{M}}_T$ with a centered normalizing matrix, and $\ddot{\mathcal{M}}_T$ with a non-centered normalizing matrix in Figure 2. It can be seen that $\ddot{\mathcal{M}}_T$ has no power, whereas $\widetilde{\mathcal{M}}_T$ suffers from some power loss, compared with the asymptotic power of the standard M test.

5.2 Example: Robust Tests for Serial Correlations

We now present an example of robust M test for serial correlations. Lobato (2001) constructed a portmanteau test that does not require consistent estimation of the asymptotic covariance matrix, yet it is applicable only to raw time series. For a robust information matrix test on high order moments, see Kuan and Lee (2006).

Consider the specification: $y_t = h(\boldsymbol{x}_t;\boldsymbol{\theta}) + e_t(\boldsymbol{\theta})$, where h is a possibly nonlinear function, $\boldsymbol{x}_t$ is a $k \times 1$ vector of observed variables, $\boldsymbol{\theta}$ is a $p \times 1$ vector of unknown parameters, $e_t(\boldsymbol{\theta})$ is a random error. Let $\mathbb{E}(y_t|\boldsymbol{x}_t) = h(\boldsymbol{x}_t;\boldsymbol{\theta}_o)$ and $\boldsymbol{\theta}_o$ is the unique solution to $\mathbb{E}[\nabla h(\boldsymbol{x}_t,\boldsymbol{\theta})e_t(\boldsymbol{\theta})] = \mathbf{0}$. When $h(\boldsymbol{x}_t;\boldsymbol{\theta})$ is evaluated at $\boldsymbol{\theta}_o$, the resulting error is denoted as $\varepsilon_t := e_t(\boldsymbol{\theta}_o)$. For notation simplicity, we write $\boldsymbol{y}_{t-1,q} = [y_{t-1},\dots,y_{t-q}]'$ with $q > p$, $\boldsymbol{h}_{t-1,q}(\boldsymbol{\theta})$ and $\boldsymbol{e}_{t-1,q}(\boldsymbol{\theta})$ are similarly defined. Note that $\boldsymbol{\epsilon}_{t-1,q} := \boldsymbol{e}_{t-1,q}(\boldsymbol{\theta}_o)$. Also let $\hat{\boldsymbol{\theta}}_T$ denote the NLS estimator, which is consistent for $\boldsymbol{\theta}_o$ under quite general conditions. Hence, $e_t(\hat{\boldsymbol{\theta}}_T)$ is the t th model residual evaluated at $\hat{\boldsymbol{\theta}}_T$, and $\boldsymbol{e}_{t-1,q}(\hat{\boldsymbol{\theta}}_T)$ is the vector of q lagged residuals, also evaluated at $\hat{\boldsymbol{\theta}}_T$.

As the well known Q test, we are interested in testing

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

Letting $T_q = T - q$, define

$$\boldsymbol{m}_{T_q}(\boldsymbol{\theta}) = \frac{1}{T_q}\sum_{t=q+1}^{T}\big[y_t - h(\boldsymbol{x}_t;\boldsymbol{\theta})\big]\big[\boldsymbol{y}_{t-1,q} - \boldsymbol{h}_{t-1,q}(\boldsymbol{\theta})\big] = \frac{1}{T_q}\sum_{t=q+1}^{T} e_t(\boldsymbol{\theta})\boldsymbol{e}_{t-1,q}(\boldsymbol{\theta}).$$

We can base an M test of this hypothesis on $\boldsymbol{m}_{T_q}(\hat{\boldsymbol{\theta}}_T) = T_q^{-1}\sum_{t=q+1}^{T} e_t(\hat{\boldsymbol{\theta}}_T)\boldsymbol{e}_{t-1,q}(\hat{\boldsymbol{\theta}}_T)$. We have learned that $T_q^{1/2}\boldsymbol{m}_{T_q}(\hat{\boldsymbol{\theta}}_T)$ 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 $\boldsymbol{F}_o = \mathbf{0}$, where

$$\boldsymbol{F}_{T_q}(\boldsymbol{\theta}_o) = \frac{-1}{T_q}\sum_{t=q+1}^{T}\big[\boldsymbol{\epsilon}_{t-1,q}\nabla_{\boldsymbol{\theta}}h_t(\boldsymbol{\theta}_o) + \varepsilon_t\nabla_{\boldsymbol{\theta}}\boldsymbol{h}_{t-1,q}(\boldsymbol{\theta}_o)\big].$$

Note that $\boldsymbol{F}_o$ would be zero if $\{\boldsymbol{x}_t\}$ and $\{\varepsilon_t\}$ are mutually independent. When $h(\boldsymbol{x}_t;\boldsymbol{\theta}_o)$ is a linear function $\boldsymbol{x}_t'\boldsymbol{\theta}_o$, $\boldsymbol{F}_o = \mathbf{0}$ when $\{\boldsymbol{x}_t\}$ and $\{\varepsilon_t\}$ are mutually uncorrelated.

For the models that $\boldsymbol{F}_{T_q}(\boldsymbol{\theta}_o) \xrightarrow{\mathbb{P}} \mathbf{0}$, the robust M test based on model residuals is

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

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

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

This test is asymptotically pivotal with the null limit $\boldsymbol{W}_q(1)'\boldsymbol{P}_q^{-1}\boldsymbol{W}_q(1)$.

It is readily seen that $\widehat{\mathcal{M}}_{T_q}$ includes the test of Lobato (2001) for raw time series as a special case. To see this, note that Lobato (2001) based his test on sample autocovriances. In the present context, when $h(\boldsymbol{x}_t;\boldsymbol{\theta})$ contains only the constant term, the estimator $\hat{\boldsymbol{\theta}}_T$ is the sample mean of y_t, and $\boldsymbol{m}_{T_q}(\hat{\boldsymbol{\theta}}_T)$ is a vector of sample autocovariances. Hence, $\widehat{\mathcal{M}}_{T_q}$ is exactly the test of Lobato (2001), which is asymptotically pivotal because the constant is uncorrelated with all ε_t (so that $\boldsymbol{F}_o = \boldsymbol{0}$).

The $\widehat{\mathcal{M}}_{T_q}$ test is, however, not valid for testing the residuals of dynamic models, such as AR models and models with lagged dependent variables. Consider now the linear AR(p) specification: $h(\boldsymbol{x}_t;\boldsymbol{\theta}) = \boldsymbol{y}'_{t-1,p}\boldsymbol{\theta}$. As $\nabla_{\boldsymbol{\theta}} h_t(\boldsymbol{x}_t;\boldsymbol{\theta}) = \boldsymbol{y}'_{t-1,p}$ is correlated with $\boldsymbol{\epsilon}_{t-1,q}$, $\boldsymbol{F}_{T_q}(\boldsymbol{\theta}_o)$ does not converge to zero, so that the null limit of $\widehat{\mathcal{M}}_{T_q}$ still contains nuisance parameters. Nonetheless, we can compute the robust M test (16) using the recursive (NLS) estimators $\tilde{\boldsymbol{\theta}}_t$, $t = q+1,\ldots,T$. The required 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

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

where $e_i(\tilde{\boldsymbol{\theta}}_t) = y_i - h(\boldsymbol{x}_i;\tilde{\boldsymbol{\theta}}_t)$ is the i th residual evaluated at $\tilde{\boldsymbol{\theta}}_t$, and $\boldsymbol{e}_{i-1,q}(\tilde{\boldsymbol{\theta}}_t)$ is the vector of q lagged residuals. The robust M test for the residuals of dynamic models is:

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

This test is valid for testing the residuals of both static and dynamic regression models and is a significant generalization of the robust test of Lobato (2001).

5.3 Robust OIR Tests

For the moment conditions $\mathbb{E}[\boldsymbol{f}(\boldsymbol{\eta}_t;\boldsymbol{\theta}_o)] = \boldsymbol{0}$, the parameter $\boldsymbol{\theta}_o$ is said to be over-identified (just-identified) if the number of moment conditions, q, is greater than (equal to) the number of parameters, k. The GMM estimator of $\boldsymbol{\theta}_o$ is

$$\hat{\boldsymbol{\theta}}_T = \operatorname{argmin}_{\boldsymbol{\theta}\in\Theta}\, \boldsymbol{m}_T(\boldsymbol{\theta})'\boldsymbol{H}_T\boldsymbol{m}_T(\boldsymbol{\theta}),$$

where $\boldsymbol{H}_T$ is a symmetric, positive semi-definite weighting matrix that converges to $\boldsymbol{H}_o$.

The normalized GMM estimator can be expressed as

$$\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_o) = -(\boldsymbol{F}'_o\boldsymbol{H}_o\boldsymbol{F}_o)^{-1}\boldsymbol{F}'_o\boldsymbol{H}_o\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\boldsymbol{f}(\boldsymbol{\eta}_t;\boldsymbol{\theta}_o)\right] + o_{\mathbb{P}}(1), \tag{17}$$

and its asymptotic covariance matrix is

$$\boldsymbol{\Omega}_o(\boldsymbol{H}_o) = (\boldsymbol{F}_o'\boldsymbol{H}_o\boldsymbol{F}_o)^{-1}\boldsymbol{F}_o'\boldsymbol{H}_o\boldsymbol{\Sigma}_o\boldsymbol{H}_o\boldsymbol{F}_o(\boldsymbol{F}_o'\boldsymbol{H}_o\boldsymbol{F}_o)^{-1}.$$

In particular, $\boldsymbol{\Omega}_o(\boldsymbol{\Sigma}_o^{-1}) = (\boldsymbol{F}_o'\boldsymbol{\Sigma}_o^{-1}\boldsymbol{F}_o)^{-1}$. It is easily shown that $\boldsymbol{\Omega}_o(\boldsymbol{H}_o) - \boldsymbol{\Omega}_o(\boldsymbol{\Sigma}_o^{-1})$ is a positive semi-definite matrix for any $\boldsymbol{H}_o \neq \boldsymbol{\Sigma}_o^{-1}$. This suggests that the optimal GMM estimator, $\hat{\boldsymbol{\theta}}_T^*$, can be obtained by setting the weighting matrix $\boldsymbol{H}_T$ as $\widehat{\boldsymbol{\Sigma}}_T^{-1}$, a consistent estimator of $\boldsymbol{\Sigma}_o^{-1}$, and minimizing $\boldsymbol{m}_T(\boldsymbol{\theta})'\widehat{\boldsymbol{\Sigma}}_T^{-1}\boldsymbol{m}_T(\boldsymbol{\theta})$. Thus, a preliminary GMM estimator of $\boldsymbol{\theta}_o$ is needed to compute $\widehat{\boldsymbol{\Sigma}}_T$ before doing optimal estimation.

The OIR test for $\mathbb{E}[\boldsymbol{f}(\boldsymbol{\eta}_t;\boldsymbol{\theta}_o)] = \mathbf{0}$ with $q > k$ is an M test based on the GMM estimator. Given the $\boldsymbol{H}_T$-based GMM estimator $\hat{\boldsymbol{\theta}}_T$ and the weighting matrix $\ddot{\boldsymbol{H}}_T$, the conventional OIR test is

$$\mathcal{J}(\hat{\boldsymbol{\theta}}_T, \ddot{\boldsymbol{H}}_T) = T\boldsymbol{m}_T(\hat{\boldsymbol{\theta}}_T)'\ddot{\boldsymbol{H}}_T\boldsymbol{m}_T(\hat{\boldsymbol{\theta}}_T).$$

The OIR test of Hansen (1982) is $\mathcal{J}^* = \mathcal{J}(\hat{\boldsymbol{\theta}}_T^*, \widehat{\boldsymbol{\Sigma}}_T^{-1}) = T\boldsymbol{m}_T(\hat{\boldsymbol{\theta}}_T^*)'\widehat{\boldsymbol{\Sigma}}_T^{-1}\boldsymbol{m}_T(\hat{\boldsymbol{\theta}}_T^*)$. It is not difficult to show that $\mathcal{J}(\hat{\boldsymbol{\theta}}_T, \ddot{\boldsymbol{H}}_T)$ is not asymptotically pivotal, yet $\mathcal{J}^* \xrightarrow{D} \chi^2(q-k)$.

Although the OIR test is an M test, the robust M tests studied in Section 5.1 are not applicable. From (17) we can see that, asymptotically, the GMM estimator is a linear transformation of the moment conditions, and hence $\boldsymbol{G}_o$ in [B1](b) can not be nonsingular. To construct a robust OIR test, let $\boldsymbol{\Lambda}$ denote the matrix square root of $\boldsymbol{H}_o$ such that $\boldsymbol{\Lambda}\boldsymbol{\Lambda}' = \boldsymbol{H}_o$ and

$$\boldsymbol{V}_{\boldsymbol{\Lambda}'\boldsymbol{F}_o} = \boldsymbol{I}_q - \boldsymbol{\Lambda}'\boldsymbol{F}_o(\boldsymbol{F}_o'\boldsymbol{\Lambda}\boldsymbol{\Lambda}'\boldsymbol{F}_o)^{-1}\boldsymbol{F}_o'\boldsymbol{\Lambda}$$

which is symmetric and idempotent with rank $q-p$. Then by (17), we have

$$\begin{aligned}\sqrt{T}\boldsymbol{m}_T(\hat{\boldsymbol{\theta}}_T) &= \left[\boldsymbol{I}_q - \boldsymbol{F}_o(\boldsymbol{F}_o'\boldsymbol{H}_o\boldsymbol{F}_o)^{-1}\boldsymbol{F}_o'\boldsymbol{H}_o\right]\sqrt{T}\boldsymbol{m}_T(\boldsymbol{\theta}_o) + o_{\mathbb{P}}(1) \\ &= \boldsymbol{\Lambda}'^{-1}\boldsymbol{V}_{\boldsymbol{\Lambda}'\boldsymbol{F}_o}\boldsymbol{\Lambda}'\sqrt{T}\boldsymbol{m}_T(\boldsymbol{\theta}_o) + o_{\mathbb{P}}(1) \\ &\xrightarrow{D} \boldsymbol{U}'\boldsymbol{S}_o\boldsymbol{W}_q(1),\end{aligned} \tag{18}$$

where $\boldsymbol{U} := \boldsymbol{\Lambda}\boldsymbol{V}_{\boldsymbol{\Lambda}'\boldsymbol{F}_o}\boldsymbol{\Lambda}^{-1}$ is singular with rank $q-p$.

Lee and Kuan (2006) suggested using the normalizing matrix $\widehat{\boldsymbol{\Gamma}}_T = \widehat{\boldsymbol{U}}_T'\widehat{\boldsymbol{C}}_T\widehat{\boldsymbol{U}}_T$, where $\widehat{\boldsymbol{C}}_T$ is the normalizing matrix defined in Section 5.1 and converges in distribution to $\boldsymbol{S}_o\boldsymbol{P}_q\boldsymbol{S}_o'$, and $\widehat{\boldsymbol{U}}_T = \widehat{\boldsymbol{\Lambda}}_T\widehat{\boldsymbol{V}}_T\widehat{\boldsymbol{\Lambda}}_T^{-1}$ is a consistent estimator of $\boldsymbol{U}$, with $\widehat{\boldsymbol{\Lambda}}_T$ the matrix square root of $\boldsymbol{H}_T$, $\widehat{\boldsymbol{F}}_T = T^{-1}\sum_{t=1}^T \nabla_{\boldsymbol{\theta}}\boldsymbol{f}(\boldsymbol{\eta}_t;\hat{\boldsymbol{\theta}}_T)$, and

$$\widehat{\boldsymbol{V}}_T = \boldsymbol{I}_q - \widehat{\boldsymbol{\Lambda}}_T'\widehat{\boldsymbol{F}}_T[\widehat{\boldsymbol{F}}_T'\boldsymbol{H}_T\widehat{\boldsymbol{F}}_T]^{-1}\widehat{\boldsymbol{F}}_T'\widehat{\boldsymbol{\Lambda}}_T.$$

Then, $\widehat{\boldsymbol{\Gamma}}_T \Rightarrow \big(\boldsymbol{U}'\boldsymbol{S}_o\boldsymbol{P}_q\boldsymbol{S}_o'\boldsymbol{U}\big)$. Note that $\widehat{\boldsymbol{U}}_T$ has rank $q-p$ for all T, and so does $\widehat{\boldsymbol{\Gamma}}_T$. Lee and Kuan (2006) showed that

$$\widehat{\boldsymbol{\Gamma}}_T^{+} \Rightarrow \big(\boldsymbol{U}'\boldsymbol{S}_o\boldsymbol{P}_q\boldsymbol{S}_o'\boldsymbol{U}\big)^{+},$$

where we write $\boldsymbol{A}^{+}$ as the Moore-Penrose generalized inverse of $\boldsymbol{A}$.[3]

A robust OIR test without consistent estimation of the asymptotic covariance matrix can then be computed as

$$\mathcal{J}(\hat{\boldsymbol{\theta}}_T, \widehat{\boldsymbol{\Gamma}}_T^{+}) = T\boldsymbol{m}_T(\hat{\boldsymbol{\theta}}_T)'\widehat{\boldsymbol{\Gamma}}_T^{+}\boldsymbol{m}_T(\hat{\boldsymbol{\theta}}_T).$$

It follows from (18) that

$$\mathcal{J}(\hat{\boldsymbol{\theta}}_T, \widehat{\boldsymbol{\Gamma}}_T^{+}) \xrightarrow{D} \boldsymbol{W}_q(1)'\boldsymbol{S}_o'\boldsymbol{U}\big(\boldsymbol{U}'\boldsymbol{S}_o\boldsymbol{P}_q\boldsymbol{S}_o'\boldsymbol{U}\big)^{+}\boldsymbol{U}'\boldsymbol{S}_o\boldsymbol{W}_q(1),$$

which equals in distribution to $\boldsymbol{W}_{q-p}(1)'\boldsymbol{P}_{q-p}^{-1}\boldsymbol{W}_{q-p}(1)$, as shown in Lee and Kuan (2006). It is worth mentioning that $\mathcal{J}(\hat{\boldsymbol{\theta}}_T, \widehat{\boldsymbol{\Gamma}}_T^{+})$ needs only a consistent GMM estimator, in contrast with Hansen's OIR test that depends on the optimal GMM estimator. Compared with the M test $\widetilde{\mathcal{M}}_T$, $\mathcal{J}(\hat{\boldsymbol{\theta}}_T, \widehat{\boldsymbol{\Gamma}}_T^{+})$ is also robust to the estimation effect, but its normalizing matrix does not require recursive estimation.

[3]This convergence does not follow from the continuous mapping theorem because the generalized inverse is not a continuous function in general. Here, $\widehat{\boldsymbol{\Gamma}}_T$ is constructed such that it has rank $q-k$ for all T, which is also the rank of its limit $\boldsymbol{U}'\boldsymbol{S}_o\boldsymbol{P}_q\boldsymbol{S}_o'\boldsymbol{U}$. The convergence thus carries over under this generalized inverse because there is no rank deficiency in the limit.

References

Abadir, K. M. and P. Paruolo (1997). Two mixed normal densities from cointegration analysis, *Econometrica*, **65**, 671–680.

Abadir, K. M. and P. Paruolo (2002). Simple robust testing of regression hypothesis: A comment *Econometrica*, **70**, 2097–2099.

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation, *Econometrica*, **59**, 817–858.

Andrews, D. W. K. and J. C. Monahan (1992). An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator, *Econometrica*, **60**, 953–966.

Bunzel, H., N. M. Kiefer and T. J. Vogelsang (2001). Simple robust testing of hypotheses in nonlinear models, *Journal of the American Statistical association*, **96**, 1088–1096.

de Jong, R. M. (2000). A strong consistency proof for heteroskedasticity and autocorrelation consistent covariance matrix estimators, *Econometric Theory*, **16**, 262–268.

de Jong, R. M. and J. Davidson (2000). Consistency of kernel estimators of heteroscedastic and autocorrelated covariance matrices, *Econometrica*, **68**, 407–423.

den Haan, W. J. and A. T. Levin (1997). A practitioner's guide to robust covariance matrix estimation, in G. S. Maddala and C. R. Rao (eds.), *Handbook of Statistics*, Vol. 15, 299–342, Amsterdam: Elsevier.

Gallant, A. R. (1987). *Nonlinear Statistical Models*, New York: John Wiley & Sons.

Gallant, A. R. and H. White (1988). *A Unified Theory of Estimation & Inference for Nonlinear Dynamic Models*, New York: Basil Blackwell.

Hall, A. R. (2000). Covariance matrix estimation and the power of the overidentifying restrictions test, *Econometrica*, **68**, 1517–1527.

Hall, A. R., A. Inoue, and F. P. M. Peixe (2003). Covariance matrix estimation and the limiting behavior of the overidentifying restrictions test in the presence of neglected structural instability, *Econometric Theory*, **19**, 962–983.

Hannan, E. J. (1970). *Multiple Time Series*, New York: Wiley.

Hansen, B. E. (1992). Consistent covariance matrix estimation for dependent heterogeneous processes, *Econometrica*, **60**, 967–972.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators, *Econometrica*, **50**, 1029–1054.

Jansson, M. (2002). Consistent covariance matrix estimation for linear processes, *Econometric Theory*, **18**, 1449–1459.

Jansson, M. (2004). The error in rejection probability of simple autocorrelation robust tests, *Econometrica*, **72**, 937–946.

Kiefer, N. M. and T. J. Vogelsang (2002a). Heteroskedasticity-autocorrelation robust standard errors using the Bartlett kernel without truncation, *Econometrica*, **70**, 2093–2095.

Kiefer, N. M. and T. J. Vogelsang (2002b). Heteroskedasticity-autocorrelation robust testing using bandwidth equal to sample size, *Econometric Theory*, **18**, 1350–1366.

Kiefer, N. M. and T. J. Vogelsang (2005). A new asymptotic theory for heteroskedasticity-autocorrelation robust tests, *Econometric Theory*, **21**, 1130–1164

Kiefer, N. M., T. J. Vogelsang and H. Bunzel (2000). Simple robust testing of regression hypothesis, *Econometrica*, **68**, 695–714.

Kuan, C.-M. and Y.-W. Hsieh (2008). Improved HAC covariance matrix estimation based on forecast errors, *Economics Letters*, forthcoming.

Kuan, C.-M. and W.-M. Lee (2006). Robust M tests without consistent estimation of asymptotic covariance matrix, *Journal of the American Statistical Association*, **101**, 1264–1275.

Lee, W.-M. (2006). Robust M tests using kernel-based estimators with bandwidth equal to sample size, Working Paper, Department of Economics, National Chung Cheng University.

Lee, W.-M. and C.-M. Kuan (2006). Testing over-identifying restrictions without consistent estimation of asymptotic covariance matrix, Working Paper, Institute of Economics, Academia Sinica.

Lobato, I. N. (2001). Testing that a dependent process is uncorrelated, *Journal of the American Statistical Association*, **96**, 1066–1076.

Newey, W. K. (1985). Maximum likelihood specification testing and conditional moment tests, *Econometrica*, **53**, 1047–1070.

Newey, W. K. and K. D. West (1987). A simple positive semi-definite heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica*, **55**, 703–708.

Newey, W. K. and K. D. West (1994). Automatic lag selection in covariance matrix estimation, *Review of Economic Studies*, **61**, 631–653.

Ng, S. and P. Perron (1996). The exact error in estimating the spectral density at the origin, *Journal of Time Series Analysis*, **17**, 379–408.

Parzen, E. (1957). On consistent estimates of the spectrum of a time series, *Annals of Mathematical Statistics*, **28**, 329–348.

Phillips, P. C. B., Y. Sun, and S. Jin (2006). Spectral density estimation and robust hypothesis testing using steep origin kernels without truncation, *International Economic Review*, **47**, 837–894.

Priestley, M. B. (1981). *Spectral Analysis and Time Series*, San Diego: Academic Press.

Robinson, P. (1991). Automatic frequncy domain inference on semiparametric and nonparametric models, *Econometrica*, **59**, 1329–1363.

Tauchen, G. (1985). Diagnostic Testing and Evaluation of Maximum Likelihood Model, *Journal of Econometrics*, **30**, 415–443.

Vogelsang, T. J. (2003). Testing in GMM models without truncation, *Advances in Econometrics*, **17**, 199–233.

White, H. (1984). *Asymptotic Theory for Econometricians*, Orlando, FL: Academic Press.

White, H. (1987). Specification testing in Dynamic models, in *Advance in Econometrics — Fifth World Congress*, Vol.1, ed. by T. Bewley, Cambridge: Cambridge University Press.