Robust M Tests without Consistent Estimation of Asymptotic Covariance Matrix

Chung-Ming Kuan
Institute of Economics
Academia Sinica

Wei-Ming Lee
Department of Economics
National Chung Cheng University

Final version: December 11, 2005

† Author for correspondence: Chung-Ming Kuan, Institute of Economics, Academia Sinica, Taipei 115, Taiwan; ckuan@econ.sinica.edu.tw.

†† We are indebted to two anonymous referees and Professor F. J. Samaniego (Theory and Methods Editor) whose comments and suggestions lead to a significantly improved version of the paper. We would like to thank Yi-Ting Chen, Yu-Ching Hsu, Yu-Lieh Huang, Joon Park, Hal White, Jin-Huei Yeh, Sam Yoo, and the participants of the 2005 Econometrics Workshop in Peking University and the 2005 KESG meeting in Jeju, Korea, for their comments on early versions of this paper. All remaining errors are ours. The research supports from the National Science Council of the Republic of China (NSC90-2415-H-001-034 for Kuan and NSC93-2415-H-194-010 for Lee) are gratefully acknowledged. W.-M. Lee also thanks the Institute of Economics, Academia Sinica, for supporting his postdoctoral research.

Abstract

We extend the KVB approach of Kiefer, Vogelsang, and Bunzel (2000, *Econometrica*) to constructing robust M tests without consistent estimation of asymptotic covariance matrix. We demonstrate that, when model parameters have to be estimated, the normalizing matrix computed using a full-sample estimator is able to eliminate the nuisance parameters when there is no estimation effect but not otherwise. To circumvent the problem of estimation effect, we propose using recursive estimators to compute the normalizing matrix and show that the resulting M test is asymptotically pivotal. This M test is thus robust not only to heteroskedasticity and serial correlations of unknown form but also to the presence of estimation effect. As examples, we consider robust tests for serial correlations and robust information matrix tests. The former tests extend that of Lobato (2001, *JASA*) and are applicable to model residuals. For testing higher-order moments, we find that the latter tests are also robust when a lower-order moment is mis-specified. Our simulations confirm that the proposed M tests are properly sized and have power advantage when other tests are computed based on inappropriate user-chosen parameters.

JEL classification: C12, C22

Keywords: information matrix test, M test, Newey-West estimator, test for serial correlations, recursive estimator

1 Introduction

Consistent estimation of asymptotic variance-covariance matrix plays a crucial role in large-sample tests. A leading class of consistent variance-covariance matrix estimators that are robust to heteroskedasticity and serial correlations of unknown form is the nonparametric kernel estimator originated from spectral estimation (Parzen, 1957); see also Hannan (1970) and Priestley (1981). This estimator was brought to econometricians' attention by Newey and West (1987) and Gallant (1987), and it was subsequently elaborated by Andrews (1991), Andrews and Monahan (1992), Hansen (1992), Newey and West (1994), de Jong and Davidson (2000), and Jansson (2002); for an early review see den Haan and Levin (1997). While the performance of this estimator may vary with the choices of the kernel function and truncation lag (i.e., the number of autocovariances to be estimated), such choices are somewhat arbitrary in practice. Therefore, the statistical inferences resulted from the kernel-estimator-based tests are unavoidably vulnerable.

Kiefer, Vogelsang, and Bunzel (2000) proposed an alternative approach to building robust tests in linear regressions. Instead of estimating the asymptotic variance-covariance matrix, they employed a random normalizing matrix to eliminate the nuisance parameters (of the asymptotic variance-covariance matrix) in the limit. This approach, hereafter the KVB approach, makes an important contribution to hypothesis testing because it avoids nonparametric estimation and hence its related problems. Jansson (2004) also showed that, in a Gaussian location model, the test based on the KVB approach compares favorably with the kernel-estimator-based tests in terms of the error in rejection probability. Bunzel, Kiefer, and Vogelsang (2001) and Vogelsang (2002) applied this approach to construct robust tests based on the nonlinear least squares (NLS) estimator and the generalized method of moments (GMM) estimator; a test for serial correlations was also developed by Lobato (2001) along the same line. Whether the KVB approach is applicable to more general specification tests has not been studied, however.

In this paper we 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. The specification tests under the quasi-maximum likelihood framework are leading examples. We show that, when model parameters have to be estimated, the normalizing matrix computed based on a full-sample estimator is able to eliminate the nuisance parameters when there is no estimation effect but *not* otherwise. To circumvent the problem of estimation effect, we also propose using recursive estimators to compute the normalizing matrix and show that the resulting M test is asymptotically pivotal such that its limiting distribution is the same as that of Lobato (2001). This M test is thus robust not only in the KVB sense but also to estimation effect. The latter property is important in M testing because, when estimation effect is present, the asymptotic covariance matrix is usually of a complex form and hence difficult to estimate consistently.

As examples, we consider robust tests for serial correlations and robust information matrix tests for skewness and excess kurtosis. The former tests extend that of Lobato (2001) and are applicable to model residuals. In particular, it is shown that the recursive-estimator-based normalizing matrix is needed for testing the residuals of dynamic models. For information matrix tests, it is quite interesting to find that the proposed tests on higher-order moments are also robust when a lower-order moment is mis-specified. As for the finite-sample performance, our simulations demonstrate that the proposed M tests are properly sized and have power advantage when other tests are computed based on inappropriate user-chosen parameters.

This paper proceeds as follows. In Section 2, we introduce the KVB approach in the context of M testing. We propose two robust M tests in Section 3 and illustrate them using two examples in Section 4. Monte Carlo simulation results are reported in Section 5. Section 6 concludes the paper. All proofs are deferred to the Appendix.

2 The KVB Approach

We first present the KVB approach in the context of M testing. For M tests, the hypothesis of interest can be expressed as a vector of q moment conditions:

$$\mathbb{E}[\boldsymbol{f}(\boldsymbol{\eta}_t;\boldsymbol{\theta}_o)] = \boldsymbol{o}, \text{ for some } \boldsymbol{\theta}_o \in \Theta \subset \mathbb{R}^p, \tag{1}$$

where $\boldsymbol{\eta}_t$ are random data vectors, $\boldsymbol{\theta}_o$ is the $p \times 1$ true parameter vector, and $\boldsymbol{f}$ is a $q \times 1$ vector of functions that are continuously differentiable in the neighborhood of $\boldsymbol{\theta}_o$. To examine the power property of M tests, we consider a sequence of local alternatives:

$$\mathbb{E}[\boldsymbol{f}(\boldsymbol{\eta}_t;\boldsymbol{\theta}_o)] = \boldsymbol{\delta}_o/\sqrt{T}, \tag{2}$$

where T is the sample size and $\boldsymbol{\delta}_o$ is a vector of nonzero constants. Clearly, (2) reduces to (1) when $\boldsymbol{\delta}_o = \boldsymbol{o}$. In what follows, we let $[c]$ denote the integer part of the number c, $\Rightarrow$ denote weak convergence (of associated probability measures), $\xrightarrow{D}$ convergence in distribution, $\stackrel{d}{=}$ equality in distribution, $\boldsymbol{W}_q$ a vector of q independent, standard Wiener processes, and $\boldsymbol{B}_q$ the Brownian bridge with $\boldsymbol{B}_q(r) = \boldsymbol{W}_q(r) - r\boldsymbol{W}_q(1)$ for $0 \leq r \leq 1$.

We first consider the simplest case that $\boldsymbol{\theta}_o$ is known. Define

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

where T is the sample size; for $r = 1$, $\boldsymbol{m}_T(\boldsymbol{\theta})$ is the sample average of $\boldsymbol{f}(\boldsymbol{\eta}_t;\boldsymbol{\theta})$. We can then base an M test on $\boldsymbol{m}_T(\boldsymbol{\theta}_o)$, the sample counterpart of (1). Suppose that $T^{1/2}\boldsymbol{m}_T(\boldsymbol{\theta}_o)$ is governed by a central limit theorem such that under the alternative hypothesis (2), $T^{1/2}\boldsymbol{m}_T(\boldsymbol{\theta}_o) \xrightarrow{D} \mathcal{N}(\boldsymbol{\delta}_o, \boldsymbol{\Sigma}_o)$, where $\boldsymbol{\Sigma}_o$ is nonsingular. As long as we can find a consistent estimator $\widehat{\boldsymbol{\Sigma}}_T$ for $\boldsymbol{\Sigma}_o$, the M test under the alternative hypothesis (2) is:

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

where $\boldsymbol{\delta}_o'\boldsymbol{\Sigma}_o^{-1}\boldsymbol{\delta}_o$ is the non-centrality parameter. When the null hypothesis (1) is true, $\boldsymbol{\delta}_o = \boldsymbol{o}$ so that the limiting distribution of this test is central $\chi^2(q)$.

A leading consistent estimator of $\boldsymbol{\Sigma}_o$ when $\boldsymbol{f}(\boldsymbol{\eta}_t;\boldsymbol{\theta}_o)$ are heteroskedastic and serially correlated is the following nonparametric kernel estimator:

$$\widehat{\boldsymbol{\Sigma}}_T = \frac{1}{T}\sum_{i=1}^{T}\sum_{j=1}^{T}\kappa\left(\frac{|i-j|}{\ell(T)}\right)[\boldsymbol{f}(\boldsymbol{\eta}_i;\boldsymbol{\theta}_o) - \boldsymbol{m}_T(\boldsymbol{\theta}_o)][\boldsymbol{f}(\boldsymbol{\eta}_j;\boldsymbol{\theta}_o) - \boldsymbol{m}_T(\boldsymbol{\theta}_o)]',$$

where κ is a kernel function that vanishes when $|i-j| > \ell(T)$, and $\ell(T)$ grows with T at a slower rate and is known as the truncation lag. The performance of the M test based on $\widehat{\boldsymbol{\Sigma}}_T$ varies with the chosen kernel function κ and/or the truncation lag $\ell(T)$.

The main idea underlying the KVB approach is to employ a random normalizing matrix $\boldsymbol{C}_T(\boldsymbol{\theta}_o)$ in place of a consistent estimator of $\boldsymbol{\Sigma}_o$ so as to avoid the problems arising from kernel estimation. Following this approach, a robust M test of (1) is

$$\mathcal{M}_T = T\,\boldsymbol{m}_T(\boldsymbol{\theta}_o)'\boldsymbol{C}_T(\boldsymbol{\theta}_o)^{-1}\boldsymbol{m}_T(\boldsymbol{\theta}_o), \tag{3}$$

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

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

We impose a "high-level" condition that $\{\boldsymbol{f}(\boldsymbol{\eta}_t;\boldsymbol{\theta}_o)\}$ obeys a functional central limit theorem (FCLT); more primitive regularity conditions will be discussed in Section 3.

[A1] (a) Under the local alternatives (2),

$$\sqrt{T}\boldsymbol{m}_{[rT]}(\boldsymbol{\theta}_o) = \frac{1}{\sqrt{T}}\sum_{t=1}^{[rT]}\boldsymbol{f}(\boldsymbol{\eta}_t;\boldsymbol{\theta}_o) \Rightarrow r\boldsymbol{\delta}_o + \boldsymbol{S}\boldsymbol{W}_q(r), \quad 0 \le r \le 1,$$

where $\boldsymbol{S}$ is the nonsingular, matrix square root of $\boldsymbol{\Sigma}_o$ (i.e., $\boldsymbol{\Sigma}_o = \boldsymbol{S}\boldsymbol{S}'$).

By [A1](a), $T^{1/2}\boldsymbol{m}_T(\boldsymbol{\theta}_o) \Rightarrow \boldsymbol{\delta}_o + \boldsymbol{S}\boldsymbol{W}_q(1) \stackrel{d}{=} \mathcal{N}(\boldsymbol{\delta}_o,\, \boldsymbol{\Sigma}_o)$ under local alternatives, and its null limit is $\boldsymbol{S}\boldsymbol{W}_q(1)$. Yet under both the null and local alternative hypotheses,

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

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

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

under the local alternative (2), and $\mathcal{M}_T \xrightarrow{D} \boldsymbol{W}_q(1)'\boldsymbol{P}_q^{-1}\boldsymbol{W}_q(1)$ under the null hypothesis (1). The advantage of the KVB approach is now clear. Although $\boldsymbol{C}_T(\boldsymbol{\theta}_o)$ is not consistent for $\boldsymbol{\Sigma}_o$, it eliminates the nuisance parameter $\boldsymbol{S}$ arising from the limit of $T^{1/2}\boldsymbol{m}_T(\boldsymbol{\theta}_o)$ and yields an asymptotically pivotal test. This limit is the same as that of Lobato (2001) but differs from that of Kiefer et al. (2000) by a scaling factor; the critical values of this distribution for various q can be found in Lobato (2001). Kiefer and Vogelsang (2002a) showed that $\boldsymbol{C}_T(\boldsymbol{\theta}_o)$ is algebraically equivalent to one half of the nonparametric kernel estimator based on the Bartlett kernel without truncation (i.e., $\ell(T) = T$).

3 Robust M Tests

In practice, $\boldsymbol{\theta}_o$ of (1) is unknown and must be estimated. We now introduce two robust M tests based on different estimators of $\boldsymbol{\theta}_o$.

3.1 M Test Based on a Full-Sample Estimator

Consider the robust M test with $\boldsymbol{\theta}_o$ replaced by a full-sample estimator $\hat{\boldsymbol{\theta}}_T$:

$$\widehat{\mathcal{M}}_T = T\,\boldsymbol{m}_T(\hat{\boldsymbol{\theta}}_T)'\widehat{\boldsymbol{C}}_T^{-1}\boldsymbol{m}_T(\hat{\boldsymbol{\theta}}_T), \tag{4}$$

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

$$\boldsymbol{\varphi}_t(\hat{\boldsymbol{\theta}}_T) = \frac{1}{\sqrt{T}}\sum_{i=1}^{t}\big[\boldsymbol{f}(\boldsymbol{\eta}_i;\hat{\boldsymbol{\theta}}_T) - \boldsymbol{m}_T(\hat{\boldsymbol{\theta}}_T)\big] = \sqrt{T}\boldsymbol{m}_t(\hat{\boldsymbol{\theta}}_T) - \frac{t}{T}\sqrt{T}\boldsymbol{m}_T(\hat{\boldsymbol{\theta}}_T). \tag{5}$$

Although $\widehat{\mathcal{M}}_T$ in (4) is analogous to $\mathcal{M}_T$ in (3), their asymptotic properties are not necessarily the same. To see this, note that

$$\sqrt{T}\boldsymbol{m}_{[rT]}(\hat{\boldsymbol{\theta}}_T) = \sqrt{T}\boldsymbol{m}_{[rT]}(\boldsymbol{\theta}_o) + \frac{[rT]}{T}\boldsymbol{F}_{[rT]}(\boldsymbol{\theta}_o)\big[\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_o)\big] + o_{\mathbb{P}}(1), \tag{6}$$

where $\boldsymbol{F}_{[rT]}(\boldsymbol{\theta}_o) = [rT]^{-1}\sum_{t=1}^{[rT]}\nabla_{\boldsymbol{\theta}}\boldsymbol{f}\big(\boldsymbol{\eta}_t;\boldsymbol{\theta}_o\big)$. Here, the second term on the right-hand side above is $O_{\mathbb{P}}(1)$ and characterizes the estimation effect of replacing $\boldsymbol{\theta}_o$ with $\hat{\boldsymbol{\theta}}_T$.

$T^{1/2}\boldsymbol{m}_T(\hat{\boldsymbol{\theta}}_T)$ in $\widehat{\mathcal{M}}_T$ and $T^{1/2}\boldsymbol{m}_T(\boldsymbol{\theta}_o)$ in $\mathcal{M}_T$ are not asymptotically equivalent and have different asymptotic covariance matrices, unless the estimation effect is absent.

To derive the limit of (4), we observe that many well-known econometric estimators may be expressed as:

$$\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_o) = \boldsymbol{Q}_o \left[\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \boldsymbol{q}(\boldsymbol{\eta}_t; \boldsymbol{\theta}_o) \right] + o_{\mathbb{P}}(1), \tag{7}$$

where $\boldsymbol{Q}_o$ is a $p \times p$ nonsingular matrix and $\boldsymbol{q}$ is a vector-valued function in $\mathbb{R}^p$. 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 NLS and GMM estimators can also be expressed in a similar form. The conditions below require $\{\boldsymbol{g}(\boldsymbol{\eta}_t; \boldsymbol{\theta}_o) = [\boldsymbol{f}(\boldsymbol{\eta}_t; \boldsymbol{\theta}_o)', \boldsymbol{q}(\boldsymbol{\eta}_t; \boldsymbol{\theta}_o)']'\}$ to obey a central limit theorem (CLT) and $\{\nabla_{\boldsymbol{\theta}} \boldsymbol{f}(\boldsymbol{\eta}_t; \boldsymbol{\theta}_o)\}$ to obey a law of large numbers (LLN).

[A1] (b) Under the alternative hypothesis (2),

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

where $\boldsymbol{G}$ is nonsingular with the nonsingular diagonal blocks $\boldsymbol{G}_{11}$ $(q \times q)$ and $\boldsymbol{G}_{22}$ $(p \times p)$ and the off-diagonal blocks $\boldsymbol{G}_{12}$ $(q \times p)$ and $\boldsymbol{G}_{21}$ $(p \times q)$, $\boldsymbol{\zeta}(\boldsymbol{\delta}_o) = [\boldsymbol{\delta}_o' \ \boldsymbol{\phi}(\boldsymbol{\delta}_o)']'$ with $\sup_{\boldsymbol{\delta}_o} \|\boldsymbol{\phi}(\boldsymbol{\delta}_o)\| < \infty$, and $\boldsymbol{\phi}(\boldsymbol{\delta}_o) = \boldsymbol{o}$ if $\boldsymbol{\delta}_o = \boldsymbol{o}$.

[A2] $\boldsymbol{F}_{[rT]}(\boldsymbol{\theta}_o) = [rT]^{-1} \sum_{t=1}^{[rT]} \nabla_{\boldsymbol{\theta}} \boldsymbol{f}(\boldsymbol{\eta}_t; \boldsymbol{\theta}_o) \xrightarrow{\mathbb{P}} \boldsymbol{F}_o$, uniformly in $0 < r \leq 1$, where $\boldsymbol{F}_o$ is a $q \times p$ non-stochastic matrix; $\nabla_{\boldsymbol{\theta}} \boldsymbol{F}_{[rT]}(\boldsymbol{\theta}_o)$ is bounded in probability.

We do not explicitly specify the regularity conditions for [A1] and [A2] so as to reduce technicality and excessive notations. Instead, we note that the conditions ensuring a multivariate FCLT, e.g., Corollary 4.2 of Wooldridge and White (1988) or Theorem 7.30 of White (2001), suffice for [A1], and those in Theorem 3.18 of Gallant and White (1988) suffice for [A2]. These conditions allow for general $\boldsymbol{f}$ functions as well as heterogeneous and weakly dependent data. Such conditions are quite standard in the literature; see Davidson (1994) and White (2001) for more thorough discussions. Note that the upper left $(q \times q)$ block of $\boldsymbol{\Gamma} = \boldsymbol{G}\boldsymbol{G}'$ is $\boldsymbol{\Gamma}_{11} = \boldsymbol{G}_{11}\boldsymbol{G}_{11}' + \boldsymbol{G}_{12}\boldsymbol{G}_{12}'$, which is nothing but $\boldsymbol{\Sigma}_o$ in [A1](a). Thus, $\boldsymbol{S}$ is also the matrix square root of $\boldsymbol{\Gamma}_{11}$. Moreover, [A1](b) implies:

$$\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_o) \xrightarrow{D} \mathcal{N}(\boldsymbol{Q}_o \boldsymbol{\phi}(\boldsymbol{\delta}_o), \ \boldsymbol{Q}_o \boldsymbol{\Gamma}_{22} \boldsymbol{Q}_o') \stackrel{d}{=} \boldsymbol{Q}_o \boldsymbol{\phi}(\boldsymbol{\delta}_o) + \boldsymbol{Q}_o \boldsymbol{\Lambda} \boldsymbol{W}_p(1),$$

where $\boldsymbol{\Lambda}$ is the matrix square root of $\boldsymbol{\Gamma}_{22} = \boldsymbol{G}_{21}\boldsymbol{G}_{21}' + \boldsymbol{G}_{22}\boldsymbol{G}_{22}'$, the lower-right $(p \times p)$ block of $\boldsymbol{\Gamma}$. This is a standard CLT result for many econometric estimators. Therefore, apart from the joint convergence of $\boldsymbol{f}(\boldsymbol{\eta}_t; \boldsymbol{\theta}_o)$ and $\boldsymbol{q}(\boldsymbol{\eta}_t; \boldsymbol{\theta}_o)$, the regularity conditions for [A1] are virtually the same as those in the analysis of M tests (Newey, 1985).

From (5) and (6) we see that $T^{1/2}\,\boldsymbol{m}_T(\hat{\boldsymbol{\theta}}_T)$ has the estimation effect, yet owing to "centering" (i.e., the summand of $\boldsymbol{\varphi}$ being $f(\boldsymbol{\eta}_i, \hat{\boldsymbol{\theta}}_T) - \boldsymbol{m}_T(\hat{\boldsymbol{\theta}}_T)$), $\boldsymbol{\varphi}_{[rT]}(\hat{\boldsymbol{\theta}}_T)$ in $\widehat{\boldsymbol{C}}_T$ does not. It is then clear that the limit of $\widehat{\boldsymbol{C}}_T$ can not eliminate the nuisance parameter arising from $T^{1/2}\,\boldsymbol{m}(\hat{\boldsymbol{\theta}}_T)$ in general, as shown in the result below.

Theorem 3.1 *Suppose that* [A1] *and* [A2] *hold.*

(a) *If $\boldsymbol{F}_o \neq \boldsymbol{o}$, we have under the local alternative* (2) *that*

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

where $\boldsymbol{V}$ is the matrix square root of $[\boldsymbol{I}_q \;\; \boldsymbol{F}_o\boldsymbol{Q}_o]\boldsymbol{G}\boldsymbol{G}'[\boldsymbol{I}_q \;\; \boldsymbol{F}_o\boldsymbol{Q}_o]'$, $\boldsymbol{S}$ is the matrix square root of $\boldsymbol{\Gamma}_{11}$, the upper left block of $\boldsymbol{\Gamma} = \boldsymbol{G}\boldsymbol{G}'$, and $\boldsymbol{P}_q = \int_0^1 \boldsymbol{B}_q(r)\boldsymbol{B}_q(r)'\,dr$; in particular, $\widehat{\mathcal{M}}_T \xrightarrow{D} \boldsymbol{W}_q(1)'\boldsymbol{V}'\big[\boldsymbol{S}\boldsymbol{P}_q\boldsymbol{S}'\big]^{-1}\boldsymbol{V}\boldsymbol{W}_q(1)$ under the null hypothesis (1).

(b) *If $\boldsymbol{F}_o = \boldsymbol{o}$, we have $\boldsymbol{V} = \boldsymbol{S}$ and, under local alternatives* (2),

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

in particular, $\widehat{\mathcal{M}}_T \xrightarrow{D} \boldsymbol{W}_q(1)'\boldsymbol{P}_q^{-1}\boldsymbol{W}_q(1)$ under the null hypothesis (1).

Thus, $\widehat{\mathcal{M}}_T$ is asymptotically pivotal when the estimation effect is absent in the limit (i.e., $\boldsymbol{F}_o = \boldsymbol{o}$) but not otherwise. Note that $\boldsymbol{V}$ in Theorem 3.1(a) must have a complex form because it is obtained from $\boldsymbol{G}$ and depends on the asymptotic variances of the two terms on the right-hand side of (6) and their covariance. Hence, a consistent estimator for the asymptotic covariance matrix $\boldsymbol{V}\boldsymbol{V}'$ may not be readily available.

Remark: The nonsingularity of $\boldsymbol{G}$ required by [A1] excludes the cases that $\boldsymbol{f}$ and $\boldsymbol{q}$ (hence $T^{1/2}\boldsymbol{m}_T(\boldsymbol{\theta}_o)$ and $T^{1/2}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_o)$) are linearly dependent in the limit. This may happen when $\hat{\boldsymbol{\theta}}_T$ is obtained from solving $\boldsymbol{m}_T(\boldsymbol{\theta}) = \boldsymbol{o}$. For example, when $\boldsymbol{m}_T(\boldsymbol{\theta})$ is the average of sample score functions and $\hat{\boldsymbol{\theta}}_T$ is the QML estimator so that $q = p$, we have $\boldsymbol{m}_T(\hat{\boldsymbol{\theta}}_T) = \boldsymbol{o}$ and $T^{1/2}\big(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_o\big) = -\boldsymbol{F}_o^{-1}T^{1/2}\boldsymbol{m}_T(\boldsymbol{\theta}_o) + o_{\mathbb{P}}(1)$. In this case, $\boldsymbol{G}$ is singular because by setting $\boldsymbol{Q}_o = -\boldsymbol{F}_o^{-1}$, $\boldsymbol{f} = \boldsymbol{q}$. In fact, when $\boldsymbol{m}_T(\hat{\boldsymbol{\theta}}_T)$ is identically zero, it can not be used to test (1). For testing over-identifying restrictions in the GMM context, let $\boldsymbol{m}_T(\boldsymbol{\theta}_o)$ denote the average of sample moment functions and $\hat{\boldsymbol{\theta}}_T$ denote the corresponding GMM estimator. Then,

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

where $\boldsymbol{H}_T$ is the weighting matrix in GMM estimation. When $\boldsymbol{H}_T$ converges in probability to a nonsingular matrix $\boldsymbol{H}_o$, we have $\boldsymbol{Q}_o = -(\boldsymbol{F}_o'\boldsymbol{H}_o\boldsymbol{F}_o)^{-1}$ and $\boldsymbol{q} = \boldsymbol{F}_o'\boldsymbol{H}_o\boldsymbol{f}$. This also yields a singular $\boldsymbol{G}$.

3.2 M Test Based on Recursive Estimators

In the light of Theorem 3.1(b), an asymptotically pivotal M test would be available if the estimation effect of $T^{1/2}\boldsymbol{m}_T(\hat{\boldsymbol{\theta}}_T)$ can be eliminated. Estimation effect may be removed using the martingale transformation of Khmaladze (1981); see, e.g., Stute et al. (1998) and Bai (2003). Instead of trying to tackle the estimation effect directly, we propose a different normalizing matrix that can eliminate the nuisance parameter $\boldsymbol{V}$, regardless of the value of $\boldsymbol{F}_o$. This is precisely the spirit of the KVB approach.

To construct a proper normalizing matrix, it is desired that $\boldsymbol{\varphi}_{[rT]}$ preserves the estimation effect and converges to a limit with the same nuisance parameter $\boldsymbol{V}$ for every r. The first objective can be easily achieved by evaluating $\boldsymbol{m}_T(\cdot)$ and $\boldsymbol{m}_{[rT]}(\cdot)$ with $r < 1$ at different estimators, so that "centering" does not cancel out the estimation effect in $\boldsymbol{\varphi}_{[rT]}$. Suppose that $\boldsymbol{m}_T(\cdot)$ is evaluated at $\hat{\boldsymbol{\theta}}_T$ as before and that $\boldsymbol{m}_{[rT]}(\cdot)$ is evaluated at another consistent estimator $\check{\boldsymbol{\theta}}$ which, when normalized by a factor $h(T)$, converges in distribution to a Gaussian random vector. To yield the nuisance parameter $\boldsymbol{V}$, note that the Taylor expansion of $T^{1/2}\boldsymbol{m}_{[rT]}(\check{\boldsymbol{\theta}})$ is:

$$\begin{aligned}
&\sqrt{T}\boldsymbol{m}_{[rT]}(\boldsymbol{\theta}_o) + \frac{[rT]}{\sqrt{T}}\frac{1}{\sqrt{h(T)}}\boldsymbol{F}_{[rT]}(\boldsymbol{\theta}_o)\big[\sqrt{h(T)}(\check{\boldsymbol{\theta}} - \boldsymbol{\theta}_o)\big] + o_{\mathbb{P}}(1)\\
&= \frac{\sqrt{[rT]}}{\sqrt{T}}\left\{\frac{1}{\sqrt{[rT]}}\sum_{t=1}^{[rT]}\boldsymbol{f}(\boldsymbol{\eta}_t;\boldsymbol{\theta}_o) + \frac{\sqrt{[rT]}}{\sqrt{h(T)}}\boldsymbol{F}_{[rT]}(\boldsymbol{\theta}_o)\big[\sqrt{h(T)}(\check{\boldsymbol{\theta}} - \boldsymbol{\theta}_o)\big]\right\} + o_{\mathbb{P}}(1).
\end{aligned}$$

When the estimation effect remains, the second term in the curly brackets above depends on the relative orders of $[rT]$ and $h(T)$ in the limit. As such, the terms in the curly brackets can not behave like $T^{1/2}\boldsymbol{m}_T(\hat{\boldsymbol{\theta}}_T)$ unless $h(T) = [rT]$ and $h(T)^{1/2}(\check{\boldsymbol{\theta}} - \boldsymbol{\theta}_o)$ has the same limiting distribution as $T^{1/2}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_o)$. These suggest us to choose $\check{\boldsymbol{\theta}} = \tilde{\boldsymbol{\theta}}_{[rT]}$, the recursive counterpart of $\hat{\boldsymbol{\theta}}_T$, computed from the subsample of first $[rT]$ observations.

In view of the discussion above, we propose computing the normalizing matrix based on the recursive estimators $\tilde{\boldsymbol{\theta}}_t$: $\widetilde{\boldsymbol{C}}_T = T^{-1}\sum_{t=p+1}^{T}\tilde{\boldsymbol{\varphi}}_t\tilde{\boldsymbol{\varphi}}_t'$ 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 resulting M test is thus

$$\widetilde{\mathcal{M}}_T = T\,\boldsymbol{m}_T(\hat{\boldsymbol{\theta}}_T)'\widetilde{\boldsymbol{C}}_T^{-1}\boldsymbol{m}_T(\hat{\boldsymbol{\theta}}_T). \tag{8}$$

In the light of (7), $[rT]^{1/2}(\tilde{\boldsymbol{\theta}}_{[rT]} - \boldsymbol{\theta}_o) = [rT]^{-1/2}\boldsymbol{Q}_o\sum_{t=1}^{[rT]}\boldsymbol{q}(\boldsymbol{\eta}_t;\boldsymbol{\theta}_o) + o_{\mathbb{P}}(1)$. The first-order Taylor expansion about $\boldsymbol{\theta}_o$ then yields

$$\begin{aligned}\sqrt{T}\boldsymbol{m}_{[rT]}\big(\tilde{\boldsymbol{\theta}}_{[rT]}\big) &= \sqrt{T}\boldsymbol{m}_{[rT]}\big(\boldsymbol{\theta}_o\big) + \frac{\sqrt{[rT]}}{\sqrt{T}}\boldsymbol{F}_{[rT]}(\boldsymbol{\theta}_o)\Big[\sqrt{[rT]}\big(\tilde{\boldsymbol{\theta}}_{[rT]} - \boldsymbol{\theta}_o\big)\Big] + o_{\mathbb{P}}(1) \\ &= \sqrt{T}\boldsymbol{m}_{[rT]}\big(\boldsymbol{\theta}_o\big) + \boldsymbol{F}_{[rT]}(\boldsymbol{\theta}_o)\boldsymbol{Q}_o\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{[rT]}\boldsymbol{q}(\boldsymbol{\eta}_t;\boldsymbol{\theta}_o)\right) + o_{\mathbb{P}}(1).\end{aligned}$$

This expression suggests that we may strengthen [A1] to an FCLT condition.

[B1] Under the local alternatives (2),

$$\begin{bmatrix}\frac{1}{\sqrt{T}}\sum_{t=1}^{[rT]}\boldsymbol{f}(\boldsymbol{\eta}_t;\boldsymbol{\theta}_o) \\ \frac{1}{\sqrt{T}}\sum_{t=1}^{[rT]}\boldsymbol{q}(\boldsymbol{\eta}_t;\boldsymbol{\theta}_o)\end{bmatrix} \Rightarrow r\,\boldsymbol{\zeta}(\boldsymbol{\delta}_o) + \boldsymbol{G}\boldsymbol{W}_{q+p}(r), \quad 0 \le r \le 1,$$

where $\boldsymbol{G}$ and $\boldsymbol{\zeta}(\boldsymbol{\delta}_o)$ are as in [A1](b).

Clearly, [B1] implies [A1]; the main difference is that [B1] also requires $\{\boldsymbol{q}(\boldsymbol{\eta}_t;\boldsymbol{\theta}_o)\}$ to obey an FCLT. This is not an excessively stronger requirement. In fact, as far as the regularity conditions are concerned, those imposed in Corollary 4.2 of Wooldridge and White (1988) or Theorem 7.30 of White (2001) are still sufficient for [B1].

Theorem 3.2 *Given* [B1] *and* [A2], *we have under the local alternatives* (2) *that*

$$\widetilde{\mathcal{M}}_T \xrightarrow{D} \big[\boldsymbol{V}^{-1}\big(\boldsymbol{\delta}_o + \boldsymbol{F}_o\boldsymbol{Q}_o\boldsymbol{\phi}(\boldsymbol{\delta}_o)\big) + \boldsymbol{W}_q(1)\big]'\boldsymbol{P}_q^{-1}\big[\boldsymbol{V}^{-1}\big(\boldsymbol{\delta}_o + \boldsymbol{F}_o\boldsymbol{Q}_o\boldsymbol{\phi}(\boldsymbol{\delta}_o)\big) + \boldsymbol{W}_q(1)\big],$$

where $\boldsymbol{V}$ *and* $\boldsymbol{P}_q$ *are defined in Theorem 3.1; in particular,* $\widetilde{\mathcal{M}}_T \xrightarrow{D} \boldsymbol{W}_q(1)'\boldsymbol{P}_q^{-1}\boldsymbol{W}_q(1)$ *under the null hypothesis* (1).

From the proof of Theorem 3.2 we see that $\tilde{\boldsymbol{\varphi}}_t$ converges weakly to $\boldsymbol{V}\boldsymbol{B}_q(r)$. This is why $\widetilde{\boldsymbol{C}}_T^{-1}$ can eliminate the nuisance parameter of $T^{1/2}\boldsymbol{m}_T(\hat{\boldsymbol{\theta}}_T)$. It follows that $\widetilde{\mathcal{M}}_T$

has the same weak limit as $\mathcal{M}_T$. Clearly, when $\boldsymbol{F}_o$ is indeed zero, both $\widehat{\mathcal{M}}_T$ and $\widetilde{\mathcal{M}}_T$ are asymptotically pivotal, but the former is computationally simpler because it does not require recursive estimation. We stress that $\widetilde{\mathcal{M}}_T$ is not only a robust test in the KVB sense but also an alternative to the estimation effect problem in M testing. The latter is an important feature because estimation effect is typical in M tests and usually renders consistent estimation of the asymptotic covariance matrix difficult.

Remarks:

1. Theorem 3.2 is subject to the same restriction as Theorem 3.1 because $\boldsymbol{G}$ is also required to be nonsingular. That is, $\boldsymbol{f}$ and $\boldsymbol{q}$ (hence $T^{1/2}\boldsymbol{m}_{[rT]}(\boldsymbol{\theta}_o)$ and $[rT]^{1/2}(\tilde{\boldsymbol{\theta}}_{[rT]}-\boldsymbol{\theta}_o)$) can not be linearly dependent in the limit. This requirement may be relaxed slightly. Note that Theorems 3.1 and 3.2 remain valid as long as $\boldsymbol{V}$ is nonsingular, for which the nonsingularity of $\boldsymbol{G}$ is sufficient but not necessary. Thus, our results carry over when $\boldsymbol{G}$ is singular with rank k ($q \le k < p+q$) and $\boldsymbol{V}$ is nonsingular. In practice, it may be difficult to verify whether $\boldsymbol{V}$ is nonsingular, however.

2. Our results fail to hold when $\boldsymbol{V}$ is singular with rank γ, where $0 < \gamma < q$. For example, for testing over-identifying restrictions in the GMM context, we have $\text{rank}(\boldsymbol{G}) = q$ but $\text{rank}(\boldsymbol{V}) = q - p > 0$. Then, as shown in the Appendix:

   $$\boldsymbol{W}_q(1)'\boldsymbol{V}'(\boldsymbol{V}\boldsymbol{P}_q\boldsymbol{V}')^{+}\boldsymbol{V}\boldsymbol{W}_q(1) \stackrel{d}{=} \boldsymbol{W}_\gamma(1)'\boldsymbol{P}_\gamma^{-1}\boldsymbol{W}_\gamma(1), \tag{9}$$

   where $\boldsymbol{A}^{+}$ is the Moore-Penrose generalized inverse of $\boldsymbol{A}$. Although $\widetilde{\boldsymbol{C}}_T \Rightarrow \boldsymbol{V}\boldsymbol{P}_q\boldsymbol{V}'$, there is no guarantee that $\widetilde{\boldsymbol{C}}_T^{+} \Rightarrow (\boldsymbol{V}\boldsymbol{P}_q\boldsymbol{V}')^{+}$ because generalized inverse is not a continuous function. To ensure such convergence, $\widetilde{\boldsymbol{C}}_T$ is required to satisfy more conditions; see, e.g., Scott (1997, 188–190). Hence, further modification of the test and/or $\widetilde{\boldsymbol{C}}_T$ may be needed to deliver an asymptotically pivotal test.

3. Similar to Kiefer and Vogelsang (2002a), we can show that $\widehat{\boldsymbol{C}}_T$ is one half of the Bartlett-kernel-based estimator without truncation. But $\widetilde{\boldsymbol{C}}_T$ does not have this property. Thus, in contrast with Kiefer and Vogelsang (2002b), there is no other kernel estimator without truncation that corresponds to $\widetilde{\boldsymbol{C}}_T$. Yet some variants of $\widetilde{\boldsymbol{C}}_T$ also eliminate the nuisance parameter when estimation effect is present, e.g., the "non-centered" normalizing matrix: $\ddot{\boldsymbol{C}}_T = T^{-1}\sum_{t=p+1}^{T}\ddot{\boldsymbol{\varphi}}_t\ddot{\boldsymbol{\varphi}}_t'$ with $\ddot{\boldsymbol{\varphi}}_t = T^{-1/2}\sum_{i=1}^{t}\boldsymbol{f}(\boldsymbol{\eta}_i,\tilde{\boldsymbol{\theta}}_t)$. Under the null, $\ddot{\boldsymbol{C}}_T \Rightarrow \boldsymbol{V}[\int_0^1 \boldsymbol{W}_q(r)\boldsymbol{W}_q(r)'\,\mathrm{d}r]\boldsymbol{V}'$, so that

   $$\ddot{\mathcal{M}}_T = T\,\boldsymbol{m}_T(\hat{\boldsymbol{\theta}}_T)'\ddot{\boldsymbol{C}}_T^{-1}\boldsymbol{m}_T(\hat{\boldsymbol{\theta}}_T) \Rightarrow \boldsymbol{W}_q(1)'\Big[\int_0^1 \boldsymbol{W}_q(r)\boldsymbol{W}_q(r)'\,\mathrm{d}r\Big]^{-1}\boldsymbol{W}_q(1).$$

This test is asymptotically pivotal but turns out to have little asymptotic local power. For the local powers of $\ddot{\mathcal{M}}_T$ and other tests, see next sub-section.

3.3 Asymptotic Local Power

In this section we compare the asymptotic local powers of the $\widetilde{\mathcal{M}}_T$ test with the "centered" normalizing matrix, $\ddot{\mathcal{M}}_T$ test with the "non-centered" normalizing matrix, and standard M test: $\mathcal{M}_T^{\dagger} = T\boldsymbol{m}_T(\hat{\boldsymbol{\theta}}_T)'\widehat{\boldsymbol{V}}_T^{-1'}\widehat{\boldsymbol{V}}_T^{-1}\boldsymbol{m}_T(\hat{\boldsymbol{\theta}}_T)$, where $\widehat{\boldsymbol{V}}_T$ is a consistent estimator of $\boldsymbol{V}$ defined in Theorem 3.1.

Let $\boldsymbol{\lambda} = \boldsymbol{\delta}_o + \boldsymbol{F}_o\boldsymbol{Q}_o\phi(\boldsymbol{\delta}_o)$. By Theorem 3.2, the asymptotic local power of $\widetilde{\mathcal{M}}_T$ is

$$\mathbb{P}\big\{[\boldsymbol{V}^{-1}\boldsymbol{\lambda} + \boldsymbol{W}_q(1)]'\boldsymbol{P}_q^{-1}[\boldsymbol{V}^{-1}\boldsymbol{\lambda} + \boldsymbol{W}_q(1)] > c_\alpha\big\},$$

with c_α the critical value at α level taken from the distribution of $\boldsymbol{W}_q(1)'\boldsymbol{P}_q^{-1}\boldsymbol{W}_q(1)$. It can be verified that under the local alternative (2), $\ddot{\boldsymbol{C}}_T \Rightarrow \boldsymbol{V}\boldsymbol{\Xi}_q^{-1}\boldsymbol{V}'$ with

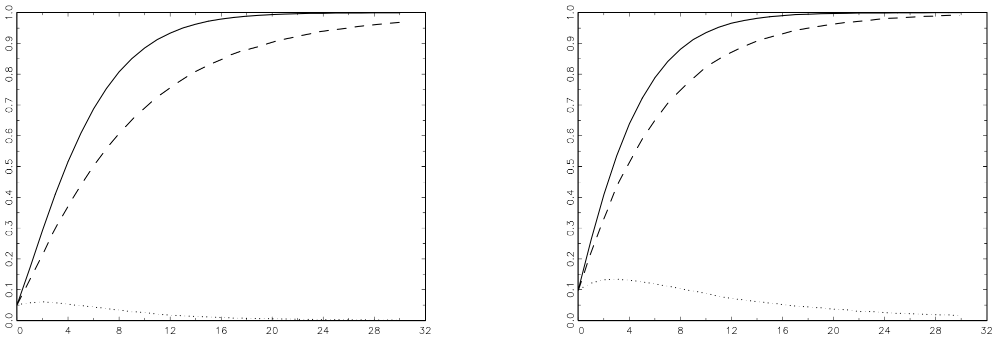
$$\boldsymbol{\Xi}_q = \int_0^1 \big[r\boldsymbol{V}^{-1}\boldsymbol{\lambda} + \boldsymbol{W}_q(r)\big]\big[r\boldsymbol{V}^{-1}\boldsymbol{\lambda} + \boldsymbol{W}_q(r)\big]'\,\mathrm{d}r,$$

which differs from $\boldsymbol{P}_q$ by the term $r\boldsymbol{V}^{-1}\boldsymbol{\lambda}$ in the integral. It follows that the asymptotic local power of $\ddot{\mathcal{M}}_T$ is

$$\mathbb{P}\big\{[\boldsymbol{V}^{-1}\boldsymbol{\lambda} + \boldsymbol{W}_q(1)]'\boldsymbol{\Xi}_q^{-1}[\boldsymbol{V}^{-1}\boldsymbol{\lambda} + \boldsymbol{W}_q(1)] > c_\alpha\big\},$$

with c_α the critical value at α level taken from $\boldsymbol{W}_q(1)'[\int_0^1 \boldsymbol{W}_q(1)\boldsymbol{W}_q(1)'\,\mathrm{d}r]^{-1}\boldsymbol{W}_q(1)$. Further, under the local alternative, $\mathcal{M}_T^{\dagger} \xrightarrow{D} \chi^2(q,\omega)$, a non-central χ^2 distribution with q degrees of freedom and the non-centrality parameter $\omega = \boldsymbol{\lambda}'\boldsymbol{V}^{-1'}\boldsymbol{V}^{-1}\boldsymbol{\lambda}$. The asymptotic local power of $\mathcal{M}_T^{\dagger}$ is then $\mathbb{P}\{\chi^2(q,\omega) > c_\alpha\}$, with c_α the critical value taken from the central χ^2 distribution with q degrees of freedom. Note that the local powers of these tests are essentially due to the same ingredient $\boldsymbol{\lambda}$.

We simulate the asymptotic local powers of $\widetilde{\mathcal{M}}_T$ and $\ddot{\mathcal{M}}_T$ for $q = 1$. In our simulation, the standard Wiener process is approximated by the (normalized) partial sums of $2,000$ pseudo standard normal random variables, and the number of replications is $50,000$. We also compute the local power of $\mathcal{M}_T^{\dagger}$ by comparing $\chi^2(1,\omega)$ with the critical value of $\chi^2(1)$. This is an "ideal" result because the power does not depend on the covariance matrix estimator. The power curves for the nominal sizes 5% and 10% are plotted in Figure 1. Similar to Kiefer, Vogelsang, and Bunzel (2000), the local power of $\mathcal{M}_T^{\dagger}$ dominates that of $\widetilde{\mathcal{M}}_T$, but their difference becomes smaller when the nominal size is 10%. In practice, $\mathcal{M}_T^{\dagger}$ is affected by the covariance matrix estimator and hence need not outperform $\widetilde{\mathcal{M}}_T$.



5% level 10% level

Figure 1: The asymptotic local power of $\mathcal{M}_T^{\dagger}$ (solid), $\widetilde{\mathcal{M}}_T$ (dashed) and $\ddot{\mathcal{M}}_T$ (dotted).

In fact, when estimation effect is present, computing $\mathcal{M}_T^{\dagger}$ may not be easy because a consistent estimator of the covariance matrix may not be readily available.

It is interesting to observe that, while $\widetilde{\mathcal{M}}_T$ and $\ddot{\mathcal{M}}_T$ have similar normalizing matrices and statistics, they have very different local powers. To see why, note that under the local alternative, their limits differ by the extra term $\boldsymbol{V}^{-1}\boldsymbol{\lambda}$ in $\boldsymbol{\Xi}_q$. As this term shows up in both the numerator and denominator of the limit of $\ddot{\mathcal{M}}_T$, their effects cancel out, so that $\ddot{\mathcal{M}}_T$ virtually has no local power. This result demonstrates the importance of "centering" in constructing normalizing matrices and agrees with the conclusion of Hall (2000) on kernel estimators with truncation. We also considered the normalizing matrices: $\bar{\boldsymbol{C}}_T = T^{-1}\sum_{t=[aT]+p+1}^{T} \tilde{\boldsymbol{\varphi}}_t \tilde{\boldsymbol{\varphi}}_t'$ with $0 < a < 1$, but found that the resulting local powers (not reported) are all dominated by $\widetilde{\mathcal{M}}_T$.

4 Examples

We now illustrate the proposed robust M tests using two examples: one contains robust portmanteau tests for serial correlations based on model residuals and the other are robust information matrix tests for conditional asymmetry and excess kurtosis.

4.1 Robust Tests for Serial Correlations

A leading diagnostic test for serial correlations is the Q test of Box and Pierce (1970) and Ljung and Box (1978). Similar to KVB, Lobato (2001) constructed a portmanteau test for serial correlations that does not require consistent estimation of the asymptotic covariance matrix. The Lobato test is applicable to testing raw time series, however.

Consider the nonlinear regression specification: $y_t = h(\boldsymbol{x}_t;\boldsymbol{\theta}) + e_t(\boldsymbol{\theta})$, where h is a measurable 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 $\boldsymbol{\theta}_o$ be such that $\mathbb{E}(y_t|\boldsymbol{x}_t) = h(\boldsymbol{x}_t;\boldsymbol{\theta}_o)$ and assume that $\boldsymbol{\theta}_o$ is the unique solution to $\mathbb{E}[\nabla h(\boldsymbol{x}_t,\boldsymbol{\theta})e_t(\boldsymbol{\theta})] = \boldsymbol{o}$. 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 $\boldsymbol{y}_{t-1,q} = [y_{t-1},\ldots,y_{t-q}]'$ with $q > p$, $\boldsymbol{h}_{t-1,q}(\boldsymbol{\theta})$ and $\boldsymbol{e}_{t-1,q}(\boldsymbol{\theta})$ are similarly defined. Note that $\boldsymbol{\epsilon}_{t-1,q} := \boldsymbol{e}_{t-1,q}(\boldsymbol{\theta}_o)$. Let $\hat{\boldsymbol{\theta}}_T$ denote the NLS estimator for the nonlinear specification above, which is consistent for $\boldsymbol{\theta}_o$ under quite general conditions; see, e.g., Gallant and White (1988). 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.

As the Q test, we are interested in testing

$$\mathbb{E}[\boldsymbol{f}_{t,q}(\boldsymbol{\theta}_o)] = \mathbb{E}(\varepsilon_t\boldsymbol{\epsilon}_{t-1,q}) = \boldsymbol{o}. \tag{10}$$

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 (10) 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 $\boldsymbol{F}_o = \boldsymbol{o}$, where

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

For example, $\boldsymbol{F}_o = \boldsymbol{o}$ when $\{\boldsymbol{x}_t\}$ and $\{\varepsilon_t\}$ are mutually independent. If $h(\boldsymbol{x}_t;\boldsymbol{\theta}_o)$ is a linear function $\boldsymbol{x}_t'\boldsymbol{\theta}_o$, $\boldsymbol{F}_o = \boldsymbol{o}$ when $\{\boldsymbol{x}_t\}$ and $\{\varepsilon_t\}$ are mutually uncorrelated.

For the models that $\boldsymbol{F}_{T_q}(\boldsymbol{\theta}_o) \xrightarrow{\mathbb{P}} \boldsymbol{o}$, the robust M test for 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

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

By Theorem 3.1, $\widehat{\mathcal{M}}_{T_q}$ is asymptotically pivotal with the null limit $\boldsymbol{W}_q(1)'\boldsymbol{P}_q^{-1}\boldsymbol{W}_q(1)$. Note that $\widehat{\mathcal{M}}_{T_q}$ includes the test of Lobato (2001) for raw time series as a special case. To see this, suppose that $h(\boldsymbol{x}_t;\boldsymbol{\theta})$ contains only the constant term. Then, 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. In this case, $\widehat{\mathcal{M}}_{T_q}$ is exactly the test of Lobato (2001) and is asymptotically pivotal because the constant is uncorrelated with all ε_t (so that $\boldsymbol{F}_o = \boldsymbol{o}$).

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, as shown in Theorem 3.1. Nonetheless, we can compute the robust M test (8) 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

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

where $e_i(\tilde{\boldsymbol{\theta}}_t) = y_i - h(\boldsymbol{x}_i;\tilde{\boldsymbol{\theta}}_t)$ is the i-th residual evaluated at the estimator $\tilde{\boldsymbol{\theta}}_t$, 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 thus $\widetilde{\mathcal{M}}_{T_q} = T\,\boldsymbol{m}_{T_q}'(\hat{\boldsymbol{\theta}}_T)\widetilde{\boldsymbol{C}}_{T_q}^{-1}\boldsymbol{m}_{T_q}(\hat{\boldsymbol{\theta}}_T) \xrightarrow{D} \boldsymbol{W}_q(1)'\boldsymbol{P}_q^{-1}\boldsymbol{W}_q(1)$, the weak limit of the test of Lobato (2001).

4.2 Robust Information Matrix Tests

A well known class of specification tests is the information matrix test proposed by White (1982). This test checks model mis-specifications by examining whether the information matrix equality holds; see White (1994) for a thorough discussion.

Consider the quasi-log-likelihood function $\mathcal{L}_T(\boldsymbol{\theta}) = T^{-1}\sum_{t=1}^{T}\ell_t(y_t|\boldsymbol{x}_t;\boldsymbol{\theta})$, with the p-dimensional vector $\boldsymbol{\theta}$. Let $s_t(y_t,\boldsymbol{x}_t;\boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}}\ell_t(y_t|\boldsymbol{x}_t;\boldsymbol{\theta})$, which will be denoted as $s_t(\boldsymbol{\theta})$ for simplicity. In general, the QML estimator $\hat{\boldsymbol{\theta}}_T$ converges to a minimizer of the Kullback-Leibler information criterion, $\boldsymbol{\theta}^*$. When $\ell_t(y_t|\boldsymbol{x}_t;\boldsymbol{\theta})$ is correctly specified in the sense that there exists a $\boldsymbol{\theta}_o$ such that $\ell_t(y_t|\boldsymbol{x}_t;\boldsymbol{\theta}_o)$ is the true conditional density, $\boldsymbol{\theta}^* = \boldsymbol{\theta}_o$, and the information matrix equality below should hold:

$$\mathbb{E}[\nabla_{\boldsymbol{\theta}}\boldsymbol{s}_t(\boldsymbol{\theta}_o) + \boldsymbol{s}_t(\boldsymbol{\theta}_o)\boldsymbol{s}_t(\boldsymbol{\theta}_o)'] = \boldsymbol{o}.$$

An information matrix test is then designed to check if

$$\mathbb{E}[\nabla_{\boldsymbol{\theta}}\boldsymbol{s}_t(\boldsymbol{\theta}^*) + \boldsymbol{s}_t(\boldsymbol{\theta}^*)\boldsymbol{s}_t(\boldsymbol{\theta}^*)'] = \boldsymbol{o};$$

failure of this equality signifies model mis-specification. Different model mis-specifications can be tested by checking different elements of the equality above. The resulting tests are also referred to as the second-order information matrix tests by White (1994).

To test certain pre-specified elements of the information matrix equality, let $\mathbf{\Lambda}$ be a $q \times p^2$ selection matrix and $\boldsymbol{f}(\boldsymbol{\eta}_t;\boldsymbol{\theta}) := \mathbf{\Lambda}\,\mathrm{vec}\big[\nabla_{\boldsymbol{\theta}}\boldsymbol{s}_t(\boldsymbol{\theta}) + \boldsymbol{s}_t(\boldsymbol{\theta})\boldsymbol{s}_t(\boldsymbol{\theta})'\big]$. The hypothesis of interest is $\mathbb{E}[\boldsymbol{f}(\boldsymbol{\eta}_t;\boldsymbol{\theta}^*)] = \boldsymbol{o}$. We can then compute a robust information matrix test ($\widehat{\mathcal{M}}_T$ or $\widetilde{\mathcal{M}}_T$) as discussed in Section 3.

To illustrate, we consider the following simple example:

$$\ell_t(y_t|\boldsymbol{x}_t;\boldsymbol{\theta}) = -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}(y_t - \alpha - \beta x_t)^2,$$

with $\boldsymbol{\theta} = (\alpha, \beta, \sigma^2)'$. Then, $s_t(\boldsymbol{\theta}) = [e_t(\boldsymbol{\theta})/\sigma^2, e_t(\boldsymbol{\theta})x_t/\sigma^2, (e_t(\boldsymbol{\theta})^2/\sigma^2 - 1)/2\sigma^2]'$, where $e_t(\boldsymbol{\theta}) = y_t - \alpha - \beta x_t$. Let $e_t^* = e_t(\boldsymbol{\theta}^*)$ and $\epsilon_t = e_t(\boldsymbol{\theta}_o)$. Straightforward calculation shows that the (3,2) element of the information matrix equality is

$$\mathbb{E}[f(\boldsymbol{\eta}_t;\boldsymbol{\theta}^*)] = \mathbb{E}\left[\frac{x_t}{2(\sigma^*)^3}\left(\frac{(e_t^*)^3}{(\sigma^*)^3} - \frac{3e_t^*}{\sigma^*}\right)\right] = 0.$$

When the maintained assumption is that the conditional mean function is correctly specified (i.e., $\epsilon_t = e_t^*$ has conditional mean zero), testing this equality amounts to testing conditional symmetry ($\mathbb{E}[(e_t^*)^3|\boldsymbol{x}_t] = 0$). It is easy to verify that

$$\nabla_{\boldsymbol{\theta}} f(\boldsymbol{\eta}_t;\boldsymbol{\theta}) = \left[-\frac{3x_t}{2\sigma^4}\left(\frac{e_t^2}{\sigma^2} - 1\right),\ -\frac{3x_t^2}{2\sigma^4}\left(\frac{e_t^2}{\sigma^2} - 1\right),\ -\frac{3x_t}{2\sigma^5}\left(\frac{e_t^3}{\sigma^3} - \frac{2e_t}{\sigma}\right)\right]'.$$

Hence, $\sum_{t=1}^{[Tr]} \nabla_{\boldsymbol{\theta}} f(\boldsymbol{\eta}_t;\boldsymbol{\theta}^*)/[Tr]$ would converge to zero under the null hypothesis of conditional symmetry as well as the additional requirement of conditional homoskedasticity, yet this average does not converge to zero when e_t^* are conditionally heteroskedastic. An information matrix test for conditional symmetry can be calculated as $\widehat{\mathcal{M}}_T$ of (4) when the conditional variance of e_t^* is indeed a constant. We may also compute $\widetilde{\mathcal{M}}_T$ as (8) based on the recursive QML estimators. This test is robust not only to potential serial correlations but also to neglected conditional heteroskedasticity.

Similarly, the (3,3) element of the information matrix equality is

$$\mathbb{E}[f(\boldsymbol{\eta}_t;\boldsymbol{\theta}^*)] = \frac{1}{4(\sigma^*)^4}\,\mathbb{E}\left(\frac{(e_t^*)^4}{(\sigma^*)^4} - 6\frac{(e_t^*)^2}{(\sigma^*)^2} + 3\right) = 0.$$

When the maintained assumptions are: e_t^* has conditional mean zero and is conditionally homoskedastic, testing this equality amounts to testing if the conditional kurtosis of e_t^* is $3(\sigma^*)^4$. It can be seen that

$$\nabla_{\boldsymbol{\theta}} f(\boldsymbol{\eta}_t;\boldsymbol{\theta}) = \frac{1}{\sigma^5}\left[-\frac{e_t^3}{\sigma^3} + \frac{3e_t}{\sigma},\ -\frac{e_t^3 x_t}{\sigma^3} + \frac{3e_t x_t}{\sigma},\ -\frac{1}{\sigma}\left(\frac{e_t^4}{\sigma^4} - \frac{18e_t^2}{4\sigma^2} + \frac{3}{2}\right)\right]'.$$

In addition to the maintained assumptions, the sample averages of the first two terms above converge to zero when e_t^* is also conditionally symmetric. Thus, $\widetilde{\mathcal{M}}_T$ in (8) for excess kurtosis would be robust to both serial correlations and conditional asymmetry.

5 Monte Carlo Simulations

In this section, we first examine the finite sample performance of the $\widehat{\mathcal{M}}_{T_q}$ and $\widetilde{\mathcal{M}}_{T_q}$ tests for serial correlations discussed in Section 4.1. We consider two nominal sizes: 5% and 10%, and two different samples: $T = 100$ and 500. The number of replications is 5000 for size simulations and 1000 for power simulations. As the results for different nominal sizes are qualitatively similar, we report only the results for 5% nominal size.

In size simulations, eight data generating processes (DGPs) are considered: DGP1–DGP4 are a linear regression model with i.i.d., ARCH, GARCH, and bilinear errors; DGP5–DGP8 are an AR(1) model with these four errors, respectively. The details of these DGPs are summarized in Table 1. For comparison, we simulate the Q test of Ljung and Box (1978), the robust test of Wooldridge (1991), the spectral-density-based test of Hong (1996), and the test of Francq, Roy, and Zaköian (2004). For the Q test on residuals, it is well known that the null distribution is $\chi^2(q)$ under DGP1 (a model with an exogenous regressor and i.i.i. errors) and $\chi^2(q-1)$ under DGP5 (an AR(1) model with i.i.d. errors). Yet this distribution may vary under different DGPs; see Davidson and MacKinnon (1993, p. 364) and Haysahi (2000, p. 146) for details. For convenience, the critical values of the Q test are taken from $\chi^2(q)$ for models with an exogenous regressor and from $\chi^2(q-1)$ for AR(1) models in our simulations. Aside from the Q test, the other tests are selected because they all require setting some user-chosen parameters. Note that Wooldridge's test is also robust to the estimation effect problem.

The robust test of Wooldridge (1991) involves several steps: weighted NLS estimation, a step to eliminate the estimation effect, and a VAR regression to remove serial correlations. In our simulation, we set the conditional variance function to one so that only LS estimation is needed. The order G of the VAR regression is set as $G = 1, 3, 5$ and also determined by AIC and BIC, with the maximum lag order being 5. As the results based on AIC and BIC are similar and usually better than those based on different G values, we report only those based on BIC, denote as WL_{BIC}. This test has an asymptotic χ^2 distribution. It should be noted that in order to eliminate the estimation effect, this test in fact loses its consistency. By contrast, our test is robust to the estimation effect and remains consistent.

The test of Hong (1996) involves nonparametric kernel estimation of the spectral density and hence depends on the chosen kernel function and truncation lag. We consider in our simulation the Bartlett, Parzen, Daniell and quadratic-spectral kernels and the truncation lags that are of different rates: (i) $\lfloor 3T^{0.3} \rfloor$, (ii) $\lfloor 9T^{0.3} \rfloor$, and (iii) $\lfloor 12T^{0.3} \rfloor$, where $\lfloor c \rfloor$ stands for the integer that is closest to the real number c. We find that Hong's tests with different kernel functions perform quite similarly and report only those based on the quadratic-spectral kernel. These tests are denoted as H_r, with r the rates (i), (ii), and (iii) above.

Francq et al. (2004) derived the limiting distribution of the Q test under more general conditions. This test depends on consistent estimation of an asymptotic covariance matrix and requires computing its eigenvalues. A consistent estimator may be obtained via a kernel estimator or a VAR method. For the former, we choose the Bartlett kernel and follow Newey and West (1994) to determine the data-dependent bandwidth. In particular, the Newey-West method is implemented by setting the weighting vector to the vector of ones and the parameter $c = 4$ and 12, where the c values were also considered by Newey and West (1994). For simplicity, we report only the result for $c = 12$, denoted as Q^{FRZ}_{NW12}. For the VAR method, the VAR order G is set as $G = 1, 3, 5$ and determined by AIC and BIC; we report only the results based on BIC, denoted as Q^{FRZ}_{BIC}, because they are usually better. The asymptotic critical values of this test are obtained via simulation. Note that these tests are valid only for testing the residuals of ARMA models but not general regression residuals.

The empirical sizes of these tests are summarized in Table 2. As the first four DGPs are based on a linear regression with an exogenous regressor, the empirical sizes of $\widehat{\mathcal{M}}_{T_q}$ and $\widetilde{\mathcal{M}}_{T_q}$ are all close to the nominal size in these cases. For the other four DGPs that involve an AR(1) model, $\widehat{\mathcal{M}}_{T_q}$ is clearly under-sized for all cases, but $\widetilde{\mathcal{M}}_{T_q}$ has better size performance in most cases (though it may be somewhat undersized when T is small). These simulation results are consistent with the analysis of Section 4.1. On the other hand, the Q test and the tests of Hong (1996) are over-sized when there are ARCH, GARCH and bilinear errors, and the size distortions deteriorate when the sample increases from 100 to 500. This is the case because these tests in fact require i.i.d. errors under the null. We also find that the size distortions are more severe when the DGP is a linear regression with an exogenous regressor (DGP2–4). Moreover, the performance of Hong's test varies with the chosen truncation lag. The WL test also depends on the VAR order G, but WL_{BIC} performs well, though it may be undersized when the sample is small. As for the tests of Francq et al. (2004), those based on the VAR method have better size performance than those based on the kernel estimator. In fact, Q_{NW12}^{FRZ} is severely over-sized when the sample is small, and the empirical sizes of Q_{BIC}^{FRZ} are close to the nominal size for $q = 2$ and 3 but usually under-sized for $q = 4$.

In the power simulations, we consider seven DGPs, including a linear regression with AR errors (DGP9) and ARMA processes with innovations being i.i.d. (DGP10 and DGP11), bilinear (DGP12), and ARCH (DGP13–DGP15); see Table 3 for detail. Note that the last three DGPs were also simulated in Francq et al. (2004). The empirical powers of these tests are summarized in Tables 4. It can be seen that $\widetilde{\mathcal{M}}_{T_q}$ dominates $\widehat{\mathcal{M}}_{T_q}$ for all DGPs, except that they perform similarly for DGP9. This may not be surprising because $\widehat{\mathcal{M}}_{T_q}$ is typically under-sized when the DGPs are dynamic models with a lagged dependent variable.

Although we do not report all the results here, we note that the powers of the Q^{FRZ}, H, and WL tests are all sensitive to the user-chosen parameters. The $\widetilde{\mathcal{M}}_{T_q}$ test usually outperforms the tests with inappropriate user-chosen parameters in a small sample ($T = 100$), yet those with BIC-determined parameters, which require more computing, compare favorably with $\widetilde{\mathcal{M}}_{T_q}$. The Q and the kernel-estimator-based Q_{NW12}^{FRZ} tests have higher powers, but these powers may be inflated because of their severe size distortions. These results show that the proposed robust tests do suffer from power loss. Yet they are still practically useful because they are free from user-chosen parameters and are computationally simpler than those rely on model selection criteria.

We also simulate the robust information matrix test for conditional symmetry discussed in Section 4.2. The nominal size is 5%; the samples are $T = 100, 500, 1000$. The number of replications is 5000 for size simulations and 1000 for power simulations. We consider eight DGPs which are a linear regression $y_t = 1 + 1 \cdot x_t + e_t$, where x_t are i.i.d. $\mathcal{N}(1, 1)$ random variables and e_t are generated as: $\mathcal{N}(0, 1)$ (IID-N), $e_t = 0.5e_{t-1} + u_t$ with u_t i.i.d. $\mathcal{N}(0, 1)$ (AR(1)-N), $e_t = \sigma_t u_t$ with $\sigma_t^2 = 1.0 + 0.5e_{t-1}^2$ (ARCH-N), $t(7)$ (IID-t(7)), logistic (IID-Lo), centered $\chi^2(2)$ (IID-$\chi^2(2)$), centered log-normal (IID-LN), and positively skewed t distribution of Jones and Faddy (2003) with the parameters $a = 5$ and $b = 3.5$ (IID-t_{JF}). For these simulations, we do not consider non-robust information matrix test because it is not clear how the asymptotic covariance matrix should be estimated when the estimation effect is present.

The simulation results are reported in Table 5, where the first five DGPs are empirical sizes and the others are empirical powers. The empirical sizes of $\widetilde{\mathcal{M}}_T$ are close to the nominal size 5% for the first two DGPs. When e_t are leptokurtic (ARCH-N, IID-$t(7)$, IID-Lo), $\widetilde{\mathcal{M}}_T$ are over-sized when the sample is small but have proper sizes when the sample gets larger. Note that ARCH errors seem to have a stronger effect on the size performance. By contrast, $\widehat{\mathcal{M}}_T$ are under-sized for all cases and all samples considered. For example, under ARCH-N, the size distortion of $\widetilde{\mathcal{M}}_T$ diminishes when the sample increases, yet the size distortion of $\widehat{\mathcal{M}}_T$ remains even when $T = 1000$. This confirms that $\widehat{\mathcal{M}}_T$ is valid only under conditional homoskedasticity, as discussed in Section 4.2. In power simulations, the errors in the last three DGPs are conditionally homoskedastic, so that both $\widehat{\mathcal{M}}_T$ and $\widetilde{\mathcal{M}}_T$ are valid. Clearly, $\widetilde{\mathcal{M}}_T$ dominates $\widehat{\mathcal{M}}_T$ for all samples.

6 Conclusions

In this paper, the KVB approach is extended to construct robust M tests on moment conditions that involve unknown parameters. We show that the normalizing matrix required by the KVB approach may be computed using a full-sample estimator when there is no estimation effect or using recursive estimators when estimation effect is present. Although recursive estimation is computationally more demanding, it delivers an asymptotically pivotal M test that does not require consistent estimation of the asymptotic covariance matrix and is robust to the presence of estimation effect. The latter feature suggests that the proposed approach may serve as a useful alternative to the estimation effect problem usually encountered in constructing specification tests. As applications, we provide new tests for serial correlations based on models residuals and information matrix tests on higher-order moments that are also robust when a lower-order moment is mis-specified. Many other existing specification tests can be robustified along the line discussed in this paper.

As discussed in the Remark after Theorem 3.2, there are different ways to eliminate the nuisance parameter. Although our choice of the normalizing matrix compares favorably with some alternatives considered in the paper, it does not carry any optimality property. Therefore, the optimal choice of the normalizing matrix and a thorough study on different approaches to constructing robust M tests are important research directions. As far as test implementation is concerned, it would be practically useful if the required normalizing matrix can be computed without recursive estimation. This is another topic for future research.

Appendix

Proof of Theorem 3.1: By the first-order Taylor expansion about $\boldsymbol{\theta}_o$,

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

Using (7), we have from [A1](b) and [A2] that

$$\begin{aligned}
\sqrt{T}\boldsymbol{m}_T(\hat{\boldsymbol{\theta}}_T) &= \sqrt{T}\boldsymbol{m}_T(\boldsymbol{\theta}_o) + \boldsymbol{F}_T(\boldsymbol{\theta}_o)\big[\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_o)\big] + o_{\mathbb{P}}(1) \\
&= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\boldsymbol{f}(\boldsymbol{\eta}_t;\boldsymbol{\theta}_o) + \boldsymbol{F}_T(\boldsymbol{\theta}_o)\boldsymbol{Q}_o\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\boldsymbol{q}(\boldsymbol{\eta}_t;\boldsymbol{\theta}_o)\right) \\
&\Rightarrow \boldsymbol{\delta}_o + \boldsymbol{F}_o\boldsymbol{Q}_o\boldsymbol{\phi}(\boldsymbol{\delta}_o) + [\boldsymbol{I}_q \;\; \boldsymbol{F}_o\boldsymbol{Q}_o]\boldsymbol{G}\boldsymbol{W}_{q+p}(1).
\end{aligned}$$

As $[\boldsymbol{I}_q \;\; \boldsymbol{F}_o\boldsymbol{Q}_o]$ has full row rank q, $[\boldsymbol{I}_q \;\; \boldsymbol{F}_o\boldsymbol{Q}_o]\boldsymbol{G}\boldsymbol{G}'[\boldsymbol{I}_q \;\; \boldsymbol{F}_o\boldsymbol{Q}_o]'$ is nonsingular with the nonsingular, matrix square root $\boldsymbol{V}$. Thus, $[\boldsymbol{I}_q \; \boldsymbol{F}_o\boldsymbol{Q}_o]\boldsymbol{G}\boldsymbol{W}_{q+p}(1)$ has the same distribution as $\boldsymbol{V}\boldsymbol{W}_q(1)$. Owing to "centering," we have from [A1](a) that

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

which does not involve the estimation effect. It follows that $\widehat{\boldsymbol{C}}_T \Rightarrow \boldsymbol{S}\boldsymbol{P}_q\boldsymbol{S}'$, regardless of the value of $\boldsymbol{F}_o$. These results lead to the limit under the local alternative hypothesis (2). The null limit follows by setting $\boldsymbol{\delta}_o = \boldsymbol{o}$. To derive the limits in (b), note that when $\boldsymbol{F}_o = \boldsymbol{o}$, $[\boldsymbol{I}_q \;\; \boldsymbol{F}_o\boldsymbol{Q}_o]\boldsymbol{G}\boldsymbol{G}'[\boldsymbol{I}_q \;\; \boldsymbol{F}_o\boldsymbol{Q}_o]' = \boldsymbol{\Gamma}_{11}$, the upper left $(q \times q)$ block of $\boldsymbol{\Gamma} = \boldsymbol{G}\boldsymbol{G}'$. As $\boldsymbol{S}$ is also the matrix square root of $\boldsymbol{\Gamma}_{11}$, we have $\boldsymbol{V} = \boldsymbol{S}$. The limits under the null and local alternative hypotheses now follow immediately from (a). □

Proof of Theorem 3.2: The first-order Taylor expansion about $\boldsymbol{\theta}_o$ yields

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

Given [A2] and [B1],

$$\sqrt{T}\boldsymbol{m}_{[rT]}\big(\tilde{\boldsymbol{\theta}}_{[rT]}\big) \Rightarrow r\boldsymbol{\delta}_o + r\boldsymbol{F}_o\boldsymbol{Q}_o\boldsymbol{\phi}(\boldsymbol{\delta}_o) + [\boldsymbol{I}_q \;\; \boldsymbol{F}_o\boldsymbol{Q}_o]\boldsymbol{G}\boldsymbol{W}_{q+p}(r).$$

Similar to the preceding proof, we have $[\boldsymbol{I}_q \;\; \boldsymbol{F}_o\boldsymbol{Q}_o]\boldsymbol{G}\boldsymbol{W}_{q+p}(r) \overset{d}{=} \boldsymbol{V}\boldsymbol{W}_q(r)$, where $\boldsymbol{V}$ is, again, the matrix square root of $[\boldsymbol{I}_q \;\; \boldsymbol{F}_o\boldsymbol{Q}_o]\boldsymbol{G}\boldsymbol{G}'[\boldsymbol{I}_q \;\; \boldsymbol{F}_o\boldsymbol{Q}_o]'$. Thus,

$$\sqrt{T}\boldsymbol{m}_{[rT]}\big(\tilde{\boldsymbol{\theta}}_{[rT]}\big) \Rightarrow r\boldsymbol{\delta}_o + r\boldsymbol{F}_o\boldsymbol{Q}_o\boldsymbol{\phi}(\boldsymbol{\delta}_o) + \boldsymbol{V}\boldsymbol{W}_q(r),$$

and $T^{1/2}\boldsymbol{m}_T(\hat{\boldsymbol{\theta}}_T) \Rightarrow \boldsymbol{\delta}_o + \boldsymbol{F}_o\boldsymbol{Q}_o\boldsymbol{\phi}(\boldsymbol{\delta}_o) + \boldsymbol{V}\boldsymbol{W}_q(1)$. It can also be verified that

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

and hence $\widetilde{\boldsymbol{C}}_T \Rightarrow \boldsymbol{V}\boldsymbol{P}_q\boldsymbol{V}'$. This proves the limit under the local alternatives. By setting $\boldsymbol{\delta}_o = \boldsymbol{o}$, the null limit follows. □.

Proof of Equation (9): By the singular value decomposition, we have $\boldsymbol{V} = \boldsymbol{A}\Delta\boldsymbol{B}'$, where $\boldsymbol{A}$ and $\boldsymbol{B}$ are $q \times \gamma$ matrices such that $\boldsymbol{A}'\boldsymbol{A} = \boldsymbol{B}'\boldsymbol{B} = \boldsymbol{I}_\gamma$, and Δ is a $\gamma \times \gamma$ diagonal matrix with positive diagonal elements. As $\boldsymbol{V}\boldsymbol{V}' = \boldsymbol{A}\Delta^2\boldsymbol{A}'$, it can be easily seen that $\boldsymbol{V}\boldsymbol{W}_q(r) \stackrel{d}{=} \boldsymbol{A}\Delta\boldsymbol{W}_\gamma(r)$ and $\boldsymbol{V}\boldsymbol{P}_q\boldsymbol{V}' \stackrel{d}{=} \boldsymbol{A}\Delta\boldsymbol{P}_\gamma\Delta\boldsymbol{A}'$. The Moore-Penrose generalized inverse of $\boldsymbol{A}\Delta\boldsymbol{P}_\gamma\Delta\boldsymbol{A}'$ can be written as

$$(\boldsymbol{A}\Delta\boldsymbol{P}_\gamma\Delta\boldsymbol{A}')^+ = (\Delta\boldsymbol{A}')^+(\boldsymbol{A}\Delta\boldsymbol{P}_\gamma)^+ = (\Delta\boldsymbol{A}')^+\boldsymbol{P}_\gamma^{-1}(\boldsymbol{A}\Delta)^+,$$

because $\operatorname{rank}(\boldsymbol{A}\Delta) = \operatorname{rank}(\boldsymbol{A}\Delta\boldsymbol{P}_\gamma) = \operatorname{rank}(\boldsymbol{P}_\gamma) = \gamma$; see Theorem 5.9 of Scott (1997, page 181). Since $\boldsymbol{A}\Delta$ is of full column rank and $\Delta\boldsymbol{A}'$ is of full row rank, their Moore-Penrose generalized inverses are simply

$$\begin{aligned}(\boldsymbol{A}\Delta)^+ &= (\Delta\boldsymbol{A}'\boldsymbol{A}\Delta)^{-1}\Delta\boldsymbol{A}' = \Delta^{-1}\boldsymbol{A}',\\ (\Delta\boldsymbol{A}')^+ &= \boldsymbol{A}\Delta(\Delta\boldsymbol{A}'\boldsymbol{A}\Delta)^{-1} = \boldsymbol{A}\Delta^{-1}.\end{aligned}$$

Given the result above, we can now show that

$$\begin{aligned}&\boldsymbol{W}_q(1)'\boldsymbol{V}'(\boldsymbol{V}\boldsymbol{P}_q\boldsymbol{V}')^+\boldsymbol{V}\boldsymbol{W}_q(1)\\ &\qquad \stackrel{d}{=} \boldsymbol{W}_\gamma(1)'\Delta\boldsymbol{A}'(\boldsymbol{A}\Delta\boldsymbol{P}_\gamma\Delta\boldsymbol{A}')^+\boldsymbol{A}\Delta\boldsymbol{W}_\gamma(1)\\ &\qquad = \boldsymbol{W}_\gamma(1)'\Delta\boldsymbol{A}'(\Delta\boldsymbol{A}')^+\boldsymbol{P}_\gamma^{-1}(\boldsymbol{A}\Delta)^+\boldsymbol{A}\Delta\boldsymbol{W}_\gamma(1)\\ &\qquad = \boldsymbol{W}_\gamma(1)'\Delta\boldsymbol{A}'\boldsymbol{A}\Delta^{-1}\boldsymbol{P}_\gamma^{-1}\Delta^{-1}\boldsymbol{A}'\boldsymbol{A}\Delta\boldsymbol{W}_\gamma(1)\\ &\qquad = \boldsymbol{W}_\gamma(1)'\boldsymbol{P}_\gamma^{-1}\boldsymbol{W}_\gamma(1). \quad \square\end{aligned}$$

References

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.

Bai, J. (2003). Testing parametric conditional distributions of dynamic models, *Review of Economics and Statistics*, **85**, 531–549.

Box, G. E. P. and D. A. Pierce (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models, *Journal of the American Statistical Association*, **65**, 1509–1526.

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.

Davidson, J. (1994). *Stochastic Limit Theory*, New York: Oxford University Press.

Davidson, R. and J. G. MacKinnon (1993). *Estimation and Inference in Econometrics*, New York: Oxford University Press.

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.

Francq, C., R. Roy, and J.-M. Zakoïan (2004). Diagnostic checking in ARMA models with uncorrelated errors, *Journal of the American Statistical Association*, forthcoming.

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

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

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

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.

Hayashi, F. (2000). *Econometrics*, New Jersey: Princeton University Press.

Hong, Y. (1996). Consistent testing for serial correlation of unknown form, *Econometrica*, **64**, 837–864.

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.

Jones, M. C. and M. J. Faddy (2003). A skew extension of the t-distribution, with applications, *Journal of the Royal Statistical Society: Series B*, **65**, 159–174.

Khmaladze, E. V. (1981). Martingale approach in the theory of goodness-of-fit tests, *Theory of Probability and Its Applications*, **26**, 240–257.

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

Ljung, G. M. and G. E. P. Box (1978). On a measure of lack of fit in time series models, *Biometrika*, **65**, 297–303.

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.

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

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

Scott, J. R. (1997). *Matrix Analysis for Statistics*, New York: Wiley.

Stute, W., S. Thies, and L.-X. Zhu (1998). Model checks for regression: An innovation process approach, *Annals of Statistics*, **26**, 1916–1934.

Tauchen, G. (1985). Diagnostic testing and evaluation of maximum likelihood model, *Journal of Econometrics*, **30**, 415–443.

Vogelsang, T. J. (2002). Testing in GMM models without truncation, Working Paper, Department of Economics and Statistical Science, Cornell University.

White, H. (1982). Maximum likelihood estimation of misspecified models, *Econometrica*, **50**, 1–25.

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.

White, H. (1994). *Estimation, Inference, and Specification Analysis*, Cambridge: Cambridge University Press.

White, H. (2001). *Asymptotic Theory for Econometricians*, revised edition, San Diego: Academic Press.

Wooldridge, J. M. (1991). On the application of robust, regression-based diagnostics to models of conditional means and conditional variances, *Journal of Econometrics*, **47**, 5– 46.

Wooldridge, J. M. and H. White (1988). Some invariance principles and central limit theorems for dependent heterogeneous processes, *Econometric Theory*, **4**, 210–230.

Table 1: The data generating processes for size simulations.

DGP1:	$y_t = 1.0 + 1.0x_t + u_t$
DGP2:	$y_t = 1.0 + 1.0x_t + e_t,\ e_t = \sigma_t u_t,\ \sigma_t^2 = 1.0 + 0.5e_{t-1}^2$
DGP3:	$y_t = 1.0 + 1.0x_t + e_t,\ e_t = \sigma_t u_t,\ \sigma_t^2 = 0.001 + 0.02e_{t-1}^2 + 0.8\sigma_{t-1}^2$
DGP4:	$y_t = 1.0 + 1.0x_t + e_t,\ e_t = u_t + 0.3u_{t-1}e_{t-2}$
DGP5:	$y_t = 1.0 + 0.5y_{t-1} + u_t$
DGP6:	$y_t = 1.0 + 0.5y_{t-1} + e_t,\ e_t = \sigma_t u_t,\ \sigma_t^2 = 1.0 + 0.5e_{t-1}^2$
DGP7:	$y_t = 1.0 + 0.5y_{t-1} + e_t,\ e_t = \sigma_t u_t,\ \sigma_t^2 = 0.001 + 0.02e_{t-1}^2 + 0.8\sigma_{t-1}^2$
DGP8:	$y_t = 1.0 + 0.5y_{t-1} + e_t,\ e_t = u_t + 0.3u_{t-1}e_{t-2}$

Note: $\{x_t\}$ and $\{u_t\}$ are sequences of i.i.d. $\mathcal{N}(0,1)$ random variables and independent to each other.

Table 2: Empirical sizes of the serial correlation tests.

		DGP1		DGP2		DGP3		DGP4		DGP5		DGP6		DGP7		DGP8	
	q	$T=100$	500	100	500	100	500	100	500	100	500	100	500	100	500	100	500
$\widehat{\mathcal{M}}_{T_q}$	1	5.4	4.8	3.9	4.2	4.8	4.8	4.5	4.4	0.3	0.2	0.2	0.2	0.5	0.3	0.2	0.1
	2	5.2	4.9	3.3	4.0	4.9	5.5	3.8	4.4	2.3	1.6	1.7	2.1	1.7	1.7	1.9	1.5
	3	5.2	5.4	3.1	4.4	4.6	5.1	3.6	4.9	2.9	2.3	1.7	2.7	2.5	2.7	2.1	2.7
	4	5.4	6.1	2.5	4.5	4.5	5.1	3.1	4.6	2.9	3.0	1.7	3.3	2.2	3.1	2.3	3.3
$\widetilde{\mathcal{M}}_{T_q}$	1	4.6	4.7	3.6	4.3	4.4	5.0	5.4	5.2	4.6	5.2	3.5	5.6	4.7	5.0	4.8	5.2
	2	4.9	5.2	4.2	4.0	4.8	4.6	5.4	4.7	3.2	4.6	3.0	4.1	3.5	4.5	2.9	4.2
	3	5.0	4.7	3.2	4.2	4.7	5.1	6.0	5.7	3.5	4.2	2.5	3.6	3.4	3.9	3.7	3.1
	4	4.5	5.3	2.8	4.3	4.6	5.4	5.1	6.5	3.5	3.9	2.4	3.1	3.6	3.5	3.8	3.4
Q_q	1	5.1	4.8	16.8	23.7	5.2	5.4	14.8	18.8	-	-	-	-	-	-	-	-
	2	4.8	4.4	16.9	25.3	4.5	6.1	19.1	27.9	4.7	4.8	10.7	15.1	4.6	5.9	5.3	6.2
	3	4.7	4.4	15.0	24.1	4.9	5.9	20.0	30.9	4.9	4.7	10.3	16.1	5.7	5.5	6.2	7.3
	4	5.2	5.0	14.5	22.4	5.2	5.5	20.2	32.1	4.3	4.9	9.5	14.4	4.4	5.5	6.0	7.0
Q_{BIC}^{FRZ}	2	-	-	-	-	-	-	-	-	5.6	5.0	4.7	4.2	4.8	5.3	4.9	5.1
	3	-	-	-	-	-	-	-	-	4.4	4.6	3.3	3.4	4.0	4.8	4.1	4.9
	4	-	-	-	-	-	-	-	-	3.3	5.0	2.6	3.3	3.5	4.5	3.2	4.9
Q_{NW12}^{FRZ}	2	-	-	-	-	-	-	-	-	13.2	6.4	13.2	5.8	14.1	5.2	12.6	6.1
	3	-	-	-	-	-	-	-	-	13.3	5.5	11.3	4.2	13.5	5.8	13.5	5.8
	4	-	-	-	-	-	-	-	-	12.9	5.5	12.1	4.0	12.4	5.0	12.4	5.2
WL_{BIC}	1	3.8	4.9	3.1	3.6	4.4	4.9	5.7	6.2	4.7	5.3	3.9	4.5	4.1	5.3	4.9	5.7
	2	5.2	4.8	3.6	4.5	5.1	5.3	5.2	5.7	3.8	4.6	3.0	3.8	3.9	4.7	3.9	4.8
	3	4.4	4.4	3.5	3.9	4.5	4.8	5.5	6.3	4.3	5.0	2.9	4.3	3.9	4.8	3.3	4.5
	4	4.6	5.1	3.7	4.2	4.5	4.5	5.3	6.2	3.5	4.4	2.3	3.5	3.4	4.9	3.3	5.3
$H_{(i)}$		7.1	7.2	17.0	22.3	7.2	7.5	22.1	32.4	3.4	4.1	5.6	10.9	3.3	4.3	3.8	4.6
$H_{(ii)}$		7.9	7.4	12.4	17.0	8.8	8.0	25.3	21.2	4.8	4.6	5.5	8.4	5.0	5.9	4.9	5.1
$H_{(iii)}$		8.1	7.9	11.9	16.1	9.3	7.6	15.9	20.3	6.3	6.0	5.8	9.0	5.5	6.7	6.0	5.8

Note: The entries are rejection frequencies in percentage; the nominal size is 5%.

Table 3: The data generating processes for power simulations.

DGP9:	$y_t = 1.0 + 1.0x_t + e_t,\ e_t = 0.5e_{t-1} + u_t$
DGP10:	$y_t = 1.0 + 0.5y_{t-1} + u_t + 0.2u_{t-1}$
DGP11:	$y_t = 1.0 + 0.5y_{t-1} + u_t + 0.5u_{t-1}$
DGP12:	$y_t = 1.0 + 0.5y_{t-1} + e_t + 0.5e_{t-1},\ e_t = u_t + 0.3u_{t-1}e_{t-2}$
DGP13:	$y_t = 0.5y_{t-1} + e_t + 0.2e_{t-1},\ e_t = (1.0 + 0.2e_{t-1}^2)^{1/2}u_t$
DGP14:	$y_t = 0.9y_{t-1} + e_t + 0.2e_{t-1},\ e_t = (1.0 + 0.2e_{t-1}^2)^{1/2}u_t$
DGP15:	$y_t = 0.9y_{t-1} + e_t + 0.2e_{t-1},\ e_t = (1.0 + 0.4e_{t-1}^2)^{1/2}u_t$

Note: $\{x_t\}$ and $\{u_t\}$ are sequences of i.i.d. $\mathcal{N}(0,1)$ random variables and independent to each other.

Table 4: Empirical powers of the serial correlation tests.

		DGP9		DGP10		DGP11		DGP12		DGP13		DGP14		DGP15	
	q	$T=100$	500	100	500	100	500	100	500	100	500	100	500	100	500
$\widehat{\mathcal{M}}_{T_q}$	1	76.9	99.9	6.5	29.1	54.6	97.8	46.4	95.9	5.2	23.4	23.5	70.0	18.0	57.5
	2	66.1	99.2	10.2	45.3	59.9	98.5	60.1	98.5	9.3	37.8	16.8	63.7	13.3	51.3
	3	57.0	98.6	8.8	36.8	57.0	98.4	56.0	97.9	6.8	30.3	13.9	56.6	10.9	46.1
	4	50.6	98.4	7.8	35.9	51.5	98.8	46.2	97.0	6.9	29.1	11.1	54.7	8.0	40.7
$\widetilde{\mathcal{M}}_{T_q}$	1	76.3	99.7	17.9	60.6	67.8	99.6	66.4	99.0	18.2	51.0	28.8	74.8	23.1	63.3
	2	64.5	99.3	11.6	47.2	63.6	98.6	59.8	98.6	12.9	41.9	19.6	65.1	14.4	53.7
	3	56.1	98.9	11.8	45.1	57.4	98.8	59.1	98.3	8.5	36.4	16.7	62.7	11.6	48.9
	4	48.6	98.1	10.8	38.3	57.3	98.5	50.6	98.6	9.4	33.1	13.2	55.9	11.9	44.7
Q_q	1	99.5	100.0	-	-	-	-	-	-	-	-	-	-	-	-
	2	99.2	100.0	22.3	76.6	94.5	100.0	93.0	100.0	24.1	76.1	49.6	97.5	51.6	94.6
	3	98.1	100.0	17.0	71.3	86.0	100.0	86.9	100.0	18.4	69.3	38.5	95.4	40.4	91.6
	4	97.9	100.0	14.9	64.7	81.2	100.0	79.6	100.0	18.1	64.7	31.3	92.1	33.1	91.2
Q^{FRZ}_{BIC}	2	-	-	22.7	78.0	91.7	100.0	88.1	100.0	21.0	74.0	26.9	90.8	24.4	80.8
	3	-	-	15.7	69.5	81.2	100.0	73.3	100.0	17.3	64.0	21.5	89.1	17.1	77.3
	4	-	-	13.4	63.3	71.1	100.0	67.1	100.0	11.5	61.2	16.0	84.4	15.8	72.3
Q^{FRZ}_{NW12}	2	-	-	35.5	78.3	96.8	100.0	95.2	100.0	33.7	75.8	50.2	92.6	45.4	82.4
	3	-	-	29.5	69.0	89.0	100.0	85.7	100.0	28.7	64.7	43.6	89.0	38.2	79.6
	4	-	-	24.5	61.4	80.1	100.0	77.4	100.0	23.4	59.8	43.0	88.0	38.2	77.0
WL_{BIC}	1	89.9	100.0	22.2	82.6	86.8	100.0	82.1	100.0	20.3	72.9	30.1	93.3	23.1	83.8
	2	73.0	100.0	16.9	73.9	81.5	100.0	78.8	100.0	16.1	64.8	24.0	89.2	17.7	75.4
	3	59.8	100.0	12.9	63.8	76.3	100.0	76.7	100.0	11.1	57.7	16.5	85.7	15.4	72.4
	4	42.9	100.0	8.4	56.4	68.5	100.0	66.3	100.0	10.5	52.1	12.6	81.4	10.6	65.0
$H_{(\text{i})}$		99.0	100.0	11.0	50.6	78.1	100.0	79.4	100.0	13.6	47.4	28.4	84.4	28.0	81.9
$H_{(\text{ii})}$		95.5	100.0	11.4	30.0	62.4	100.0	60.4	100.0	14.4	36.0	21.1	68.7	21.8	66.0
$H_{(\text{iii})}$		94.6	100.0	12.4	28.9	59.0	100.0	56.4	100.0	12.5	32.0	24.8	64.0	20.7	61.5

Note: The entries are rejection frequencies in percentage; the nominal size is 5%.

Table 5: Empirical sizes and powers of the information matrix tests.

		Sizes					Powers		
	T	IID-N	AR(1)-N	ARCH-N	IID-$t(7)$	IID-Lo	IID-$\chi^2(2)$	IID-LN	IID-t_{JF}
	100	4.1	3.9	2.3	3.9	4.1	33.8	17.3	6.6
$\widehat{\mathcal{M}}_T$	500	4.7	4.6	3.5	3.6	4.2	68.2	28.8	19.3
	1000	4.6	5.3	3.2	4.1	4.3	85.5	41.5	36.0
	100	5.2	4.8	9.1	7.3	6.4	56.8	51.3	11.3
$\widetilde{\mathcal{M}}_T$	500	4.8	5.0	8.1	5.5	5.1	82.5	62.3	27.6
	1000	5.0	5.2	6.3	4.8	5.0	90.9	65.9	40.1

Note: The entries are rejection frequencies in percentage; the nominal size is 5%.