M Tests with a New Normalization Matrix*

Yi-Ting Chen† Academia Sinica

Zhongjun Qu‡ Boston University

July 1, 2010; this version: September 19, 2012

Abstract

This paper proposes a new family of M tests building on the work of Kuan and Lee (2006) and Kiefer, Vogelsang and Bunzel (2000). The idea is to replace the asymptotic covariance matrix in conventional M tests with an alternative normalization matrix, constructed using moment functions estimated from $(K+1)$ recursive subsamples. The new tests are simple to implement. They automatically account for the effect of parameter estimation and allow for conditional heteroskedasticity and serial correlation of general forms. They converge to central F distributions under the fixed-K asymptotics and to Chi-square distributions if K is allowed to approach infinity. We illustrate their applications using three simulation examples: (1) specification testing for conditional heteroskedastic models, (2) nonnested testing with serially correlated errors, and (3) testing for serial correlation with unknown heteroskedasticity. The results show that the new tests exhibit good size properties with power often comparable to the conventional M tests while being substantially higher than that of Kuan and Lee (2006).

Keywords: Misspecification, Model diagnostics, Hypothesis testing, Self-normalization.

*We thank the editor Esfandiar Maasoumi and an anonymous referee for very helpful comments and suggestions.

†Institute of Economics, Academia Sinica, Taipei 115, Taiwan. Email: ytchen@sinica.edu.tw. Tel: 1-886-2-27822791-622.

‡Department of Economics, Boston University. 270 Bay State Road, Boston, MA, 02215. Email: qu@bu.edu.

1 Introduction

Newey (1985) and Tauchen (1985) developed a unified framework for constructing misspecification tests. The framework encompasses a large family of tests (labeled as M tests) that are quadratic forms consisting of a finite-dimensional moment vector and an asymptotic covariance matrix. Representative members of this family include Newey's (1985) conditional moment test, the Hausman (1978) test, the nonnested hypotheses tests (Cox, 1961, Davidson and MacKinnon, 1981), and the information matrix test (White, 1982). The score tests can also be interpreted as M tests; see White (1984), Chesher and Smith (1997), and Bera and Bilias (2001). M tests are widely applicable. See, for example, Pagan and Vella (1989) for applications to microeconometric models and Berkes et al. (2003), Chen (2008), and Lundergh and Teräsvirta (2002) to GARCH-type models. Newey and McFadden (1994) and White (1994) provided comprehensive reviews of the early literature with more examples.

M-testing requires consistent estimation of the asymptotic covariance matrix. From a practitioner's perspective, this can be cumbersome for two reasons. First, the moment vector depends on unknown parameters. As a result, the covariance matrix is model-specific and estimation-method-specific. Its computation would be simpler if one of the following two conditions held under the null hypothesis: (1) the derivative of the moment vector has mean zero, or (2) the estimator is asymptotically efficient. In the former case, the asymptotic covariance matrix does not depend on the estimation effect. In the latter case, it depends only on the first, but not on the second, order derivatives of the model with respect to unknown parameters due to the information matrix equality. Nonetheless, these derivatives are not necessary easy to compute for complicated time series models, and these cases are exceptions. Second, the desire to allow for heteroskedasticity and/or serial correlation in the moment vector imposes an additional layer of complexity. Indeed, the cost involved in estimating the asymptotic covariance matrix has been considered as an important reason for the infrequent use of the robust M tests in applied work, see Wooldridge (1990) and Cameron and Trivedi (2005, p. 261) for discussions on this issue.

The above difficulty has prompted researchers to construct modified M tests. Wooldridge (1990) is an important contribution along this line. His modification ensures that the moment vector behaves asymptotically as if the parameters were known. The resulting test is robust to heteroskedasticity. However, his approach requires computing the first order derivatives of the moment

vector with respect to the unknown parameters and, more importantly, is limited by the need to know the conditional expectations of these derivatives. Recently, Kuan and Lee (2006, henceforth KL) suggested another approach based on "self-normalization" by building on the insight of Kiefer, Vogelsang and Bunzel (2000). Specifically, they suggested to replace the asymptotic covariance matrix in the test by an alternative normalization matrix, such that the test remains asymptotically pivotal, although its asymptotic distribution will be different. They showed that such a normalization matrix can be constructed by estimating the model recursively using subsamples. Their test is free of the estimation-effect problem, while still being robust to heteroskedasticity and serial correlations of general forms. However, it is asymptotically less powerful than the conventional M test and the power difference can be substantial.

This paper proposes a family of M tests using an alternative normalization matrix. The method involves dividing the full sample into $K+1$ recursive subsamples and constructing a normalization matrix based on them. The resulting tests have the following properties. (1) They do not require direct estimation of the covariance matrix. They are self-normalized. (2) They are straightforward to implement. The main analytical and computational work is in re-estimating the model with $K+1$ subsamples. (3) They automatically account for the effect of parameter estimation. (4) They allow for conditional heteroskedasticity and serial correlation of general forms. (5) They converge to F distributions under the fixed-K asymptotics, therefore the inference is straightforward. (6) They have wide applicability. In particular, the moment functions can be nonlinear and nonsmooth in parameters, therefore permitting models for conditional quantiles.

The number of subsamples is an important tuning parameter. If serial correlation is absent under the null hypothesis (as is typically the case in financial applications), the size of the tests is insensitive to K and the latter can be made large to achieve higher power. We provide values of K under which the maximum asymptotic power loss relative to the conventional M test is bounded by a given small number. On the other hand, if the moment vector is substantially serially correlated and the sample size is relatively small, then a large K can lead to size distortions. In this case, the choice of K will rely on some judgement about the approximate subsample sizes to achieve uncorrelatedness. It is desirable to have a selection rule for K that can adapt to the extent of serial correlation present in the data. However, this turns out to be a challenging task and is beyond the scope of the current paper. In the absence of such a rule, we suggest to construct the test using different K to examine the result sensitivity. In particular, the values of the test statistic can be

plotted against K to provide a full disclosure of the results.

We illustrate the application of the proposed tests using three examples. They are: (1) specification testing for conditional heteroskedastic models, (2) nonnested testing with serially correlated errors, and (3) testing for serial correlation with unknown heteroskedasticity. The results show that the proposed tests have decent sizes, even in relatively small samples, and that their power can be substantially higher than that of KL. The power turns out to be comparable to the conventional M tests in all three cases. Overall, the result suggests that the marriage of "self-normalization" and "fixed-K asymptotics" produces an analytically simple, yet widely applicable, approach to misspecification testing.

Recently, several papers have explored alternative ways of constructing normalizing matrices and delivered test statistics with t or F distributions as limiting distributions. In particular, Foley and Goldman (1999) considered confidence intervals for the mean of a stationary stochastic process. They proposed a series estimator for the variance parameter in the t statistic and showed that this leads to a limiting t distribution. Sun (2011) considered the issue of inference on linear trends with stationary and serially correlated errors. He proposed a series estimator for the long run covariance matrix and proved that the resulting Wald test statistic has an asymptotic F distribution. Sun and Kim (2011) showed that the same idea can be applied to construct tests for over-identifications in a GMM setting. These papers and the current paper share two common features: (1) the asymptotic frameworks adequately account for the uncertainty associated with the normalization matrices, different from the conventional approach which treats such effect as asymptotically negligible; and (2) the resulting test statistics are simple to implement with familiar limiting distributions. Meanwhile, there are two key differences: (1) the current paper exploits the insight that information contained in recursive subsamples can be used to simultaneously eliminate nuisance parameters (i.e., the asymptotic covariance matrix) and account for estimation uncertainty; and (2) the method in this paper can be applied to a wide range of M testing problems while that of Sun and Kim (2011) will in general need to be modified before being applicable to problems other than over-identification.

The remainder of the paper is structured as follows. Section 2 discusses the issue of interest. Section 3 briefly reviews the conventional and KL's M tests. Section 4 introduces the new tests. Section 5 studies their asymptotic properties using alternative asymptotic frameworks. Section 6 studies the local asymptotic power properties. Section 7 illustrates the difference between the new tests and the conventional and KL's M tests using three examples. Section 8 concludes. All proofs

are collected in the Appendix.

The following notation is used. The superscript o indicates the true value of a parameter. For a real valued vector z, $\|z\|$ denotes its Euclidean norm. $[x]$ is the integer part of scalar x. The symbols "$\Rightarrow$", "$\rightarrow^p$" and "$\rightarrow^{a.s.}$" denote weak convergence under Skorohod topology, convergence in probability and convergence almost surely, and $O_p(\cdot)$ and $o_p(\cdot)$ is the usual notation for the orders of stochastic convergence.

2 The issue of interest

Let y_t and x_t be finite dimensional random vectors with y_t being the endogenous variables and x_t the predetermined variables at time t. Let $D_t(.|x_t)$ denote the conditional distribution function of y_t and θ be a finite dimensional parameter vector whose value is θ_o when the model is correctly specified. The M test examines a certain aspect of $D_t(.|x_t)$ using a $p\times 1$ moment vector $m_t(y_t, x_t, \theta)$, such that when the model is correctly specified, $m_t(y_t, x_t, \theta)$ satisfies

$$H_o: \quad \mathbb{E}\left[m_t(y_t, x_t, \theta_o)\right] = 0 \text{ for all } t. \tag{1}$$

The restriction (1) constitutes the null hypothesis in the M testing literature and also in this paper.

Suppose we observe $\{(y_t, x_t) : t = 1, 2, ..., T\}$. Then, the unknown parameter θ_o can be replaced by an estimate $\hat{\theta}_T$ satisfying $\sqrt{T}(\hat{\theta}_T - \theta_o) = O_p(1)$ under H_o. Without loss of generality, we assume $\hat{\theta}_T$ is the solution to

$$T^{-1/2}\sum_{t=1}^{T} s_t(y_t, x_t, \hat{\theta}_T) = o_p\left(1\right). \tag{2}$$

For example, s_t can be the score function of a likelihood or the first order derivative of a GMM criterion function. A key property of s_t is that, under H_o, $\mathbb{E}\left[s_t(y_t, x_t, \theta_o)\right] = 0$ for all t.

The M test is a quadratic form that measures the difference between $T^{-1/2}\sum_{t=1}^{T} m_t(y_t, x_t, \hat{\theta}_T)$ and zero. The usefulness of the M test lies in its flexibility. That is, we can choose m_t to examine a particular type of misspecification without making strong assumption about other aspects of the model. Thus, a useful M test is expected to have the following three features. Firstly, the moment vector m_t needs to be informative about the misspecification. Newey (1985) provides some guidelines on how to choose m_t. The subsequent discussion in this paper is conditional on a pre-specified m_t. Secondly, the test needs to be robust to departures from the distributional assumptions that are not being tested. For example, if the interest is to test the specification of

a conditional mean function, then the test should be made robust to possible heteroskedasticity and/or serial correlation in m_t if these features are deemed relevant. Thirdly, the test should be simple to implement for it to be useful in practice. The goal of the current paper is to propose a modified M test that enjoys these features. We do so by building on the recent work of KL.

To analyze the power of relevant tests, we will consider a set of local alternatives specified by

$$H_{1T}: \quad \begin{array}{l} \mathbb{E}\left[s_t(y_t, x_t, \theta_o)\right] = \delta_s/\sqrt{T} \\ \mathbb{E}\left[m_t(y_t, x_t, \theta_o)\right] = \delta_m/\sqrt{T}, \end{array}$$

where δ_s and δ_m are finite-dimensional vectors. Note that $||\delta_s|| \neq 0$ implies that the model is misspecified, and $||\delta_m|| \neq 0$ implies that the function m_t is informative about the misspecification. To simplify notation, we will write $s_t(y_t, x_t, \theta)$ as $s_t(\theta)$ and $m_t(y_t, x_t, \theta)$ as $m_t(\theta)$ in the remainder of the paper.

3 The conventional and KL's M tests

This section provides a brief review of these two test statistics while focusing on the following two issues: (1) how the effect of parameter estimation is handled and (2) how the robustness to heteroskedasticity or serial correlation is achieved. We stress that the material in this section is not new, and that it is included for the matter of comparison and to motivate the construction of the proposed statistic.

3.1 The conventional M test

The conventional M test statistic is given by

$$\mathcal{M}_{1T} = \left[T^{-1/2}\sum_{t=1}^{T} m_t(\hat{\theta}_T)\right]^{\top} \hat{V}_T^{-1} \left[T^{-1/2}\sum_{t=1}^{T} m_t(\hat{\theta}_T)\right], \tag{3}$$

where $\hat{V}_T$ is a consistent estimate of the limiting covariance of $T^{-1/2}\sum_{t=1}^{T} m_t(\hat{\theta}_T)$. The limiting distribution of the $\mathcal{M}_{1T}$ test can be derived under the following assumptions.

Assumption 1. The observed data are a realization of a stochastic process on a probability space (Ω, F, P); θ_o is an interior point of a compact set $\Theta \subset \mathbb{R}^q$; $m_t(\theta)$ and $s_t(\theta)$ are of dimensions $p \times 1$ and $q \times 1$ respectively, with p and q being finite.

Assumption 2. $\mathbb{E}m_t(\theta)$ and $\mathbb{E}s_t(\theta)$ are differentiable in $\theta \in \Theta$; $T^{-1}\sum_{t=1}^{T}\nabla_{\theta^\top}\mathbb{E}m_t(\theta_o) \to J_{mo}$ and $T^{-1}\sum_{t=1}^{T}\nabla_{\theta^\top}\mathbb{E}s_t(\theta_o) \to J_{so}$ with J_{so} being positive definite[1]. Under H_o and H_{1T},

$$\begin{bmatrix} T^{-1/2}\sum_{t=1}^{T} m_t(\hat{\theta}_T) \\ T^{-1/2}\sum_{t=1}^{T} s_t(\hat{\theta}_T) \end{bmatrix} = \begin{bmatrix} T^{-1/2}\sum_{t=1}^{T} m_t(\theta_o) \\ T^{-1/2}\sum_{t=1}^{T} s_t(\theta_o) \end{bmatrix} + \begin{bmatrix} J_{mo} \\ J_{so} \end{bmatrix} T^{1/2}\left(\hat{\theta}_T - \theta_o\right) + o_p(1).$$

Assumption 3. Under H_o and H_{1T},

$$\begin{bmatrix} T^{-1/2}\sum_{t=1}^{T}\left(m_t(\theta_o) - \mathbb{E}m_t(\theta_o)\right) \\ T^{-1/2}\sum_{t=1}^{T}\left(s_t(\theta_o) - \mathbb{E}s_t(\theta_o)\right) \end{bmatrix} \to^d N(0, \Sigma_o)$$

with Σ_o being positive definite.

The above assumptions imply that the conditions of Theorem 9.5 in White (1994) are satisfied. As in White (1994), we can replace θ_o by a nonstochastic sequence θ_T^* and allow $m_t(\theta)$ to depend on some additional nuisance parameters π_T^* that do not affect s_t. Such generalizations will not change the asymptotic results discussed in this paper. Assumption 2 allows $m_t(\theta)$ and $s_t(\theta)$ to be non-differentiable, permitting the study of least absolute deviation estimation and, more generally, quantile regression of Koenker and Bassett (1978). In the latter case, $T^{-1/2}\sum_{t=1}^{T} s_t(\hat{\theta}_T)$ is given by the directional derivative of the check function. The expansion of $T^{-1/2}\sum_{t=1}^{T} s_t(\hat{\theta}_T)$ can be interpreted as a Bahadur representation for $T^{1/2}(\hat{\theta}_T - \theta_o)$, which holds under quite mild conditions. If $m_t(\theta)$ and $s_t(\theta)$ are continuously differentiable, then Assumption 2 can be replaced by the following set of conditions: $(\hat{\theta}_T - \theta_o) = o_p(1)$ under H_o and H_{1T}, $T^{-1}\sum_{t=1}^{T}\nabla_{\theta^\top} m_t(\theta) \to^p J_{mo}(\theta)$ and $T^{-1}\sum_{t=1}^{T}\nabla_{\theta^\top} s_t(\theta) \to^p J_{so}(\theta)$ uniformly in $\theta \in \Theta$, where $J_{mo}(\theta)$ is non-random and $J_{so}(\theta)$ is non-random and positive definite. The next lemma presents the limiting distributions of $T^{-1/2}\sum_{t=1}^{T} m_t(\hat{\theta}_T)$ and $\mathcal{M}_{1T}$.

Lemma 1 *Let Assumptions 1 to 3 hold, then*

$$T^{-1/2}\sum_{t=1}^{T} m_t(\hat{\theta}_T) = \begin{cases} T^{-1/2}\sum_{t=1}^{T}\xi_t(\theta_o) + o_p(1) & \text{under } H_o, \\ T^{-1/2}\sum_{t=1}^{T}\xi_t(\theta_o) + A_o\delta + o_p(1) & \text{under } H_{1T}, \end{cases}$$

where

$$\xi_t(\theta_o) = \left[m_t(\theta_0) - \mathbb{E}m_t(\theta_0)\right] - J_{mo}J_{so}^{-1}\left[s_t(\theta_o) - \mathbb{E}s_t(\theta_0)\right], \tag{4}$$

[1]Notation: $\nabla_{\theta^\top}\mathbb{E}m_t(\theta_o)$ and $\nabla_{\theta^\top}\mathbb{E}s_t(\theta_o)$ are the partial derivatives of $\mathbb{E}m_t(\theta)$ and $\mathbb{E}s_t(\theta)$ evaluated at θ_o.

$A_o = [I_p, -J_{mo}J_{so}^{-1}]$ and $\delta = (\delta_m^\top, \delta_s^\top)^\top$. Let $V_o = A_o\Sigma_o A_o^\top$ with Σ_o defined in Assumption 3. If, in addition to Assumptions 1-3, $\hat{V}_T \to^p V_o$ under H_o and H_{1T}, then

$$\mathcal{M}_{1T} \to^d \begin{cases} \chi^2(p), & \text{under } H_o, \\ \chi^2(p; \delta^\top A_o^\top V_o^{-1} A_o \delta), & \text{under } H_{1T}. \end{cases}$$

The proof of this lemma is omitted. The result shows that replacing θ_o by an estimate $\hat{\theta}_T$ will in general have a first-order effect on the asymptotic distribution of $T^{-1/2}\sum_{t=1}^T m_t(\hat{\theta}_T)$ unless $J_{mo} = 0$. The limiting covariance matrix V_o will be estimation-method-specific (due to J_{so}) and model-specific (due to both J_{mo} and J_{so}). Note that Wooldridge (1990) proposed a family of modified M tests in which $J_{mo} = 0$ always holds. However, his test is limited by the need to know the analytical expression for $\mathbb{E}(\nabla_{\theta^\top} m_t(\theta_o)|x_t)$.

If $m_t(\theta)$ and $s_t(\theta)$ are differentiable in θ, then J_{mo} and J_{so} can be estimated by $\hat{J}_{mo} = T^{-1}\sum_{t=1}^T \nabla_{\theta^\top} m_t(\hat{\theta}_T)$ and $\hat{J}_{so} = T^{-1}\sum_{t=1}^T \nabla_{\theta^\top} s_t(\hat{\theta}_T)$. Otherwise, the estimator will depend on the model under analysis. The estimation of Σ_o depends on whether $m_t(\theta_o)$ and $s_t(\theta_o)$ are serially correlated. If serial correlation is absent, then Σ_o can be consistently estimated by

$$\hat{\Sigma}_T = T^{-1}\sum_{t=1}^T \left(m_t(\hat{\theta}_T)^\top, s_t(\hat{\theta}_T)^\top\right)^\top \left(m_t(\hat{\theta}_T)^\top, s_t(\hat{\theta}_T)^\top\right)$$

If serial correlation is present, a heteroskedasticity-autocorrelation-consistent (HAC) covariance estimator can be used:

$$\hat{\Sigma}_T = \frac{1}{T}\sum_{t=1}^T\sum_{j=1}^T \kappa\left(\frac{|t-j|}{b(T)}\right)\left(m_t(\hat{\theta}_T)^\top, s_t(\hat{\theta}_T)^\top\right)^\top \left(m_j(\hat{\theta}_T)^\top, s_j(\hat{\theta}_T)^\top\right),$$

where $\kappa(\cdot)$ is a kernel functional and $b(T)$ determines the bandwidth**.** Given such a $\hat{\Sigma}_T$, V_o can be estimated as

$$\hat{V}_T = \hat{A}_T\hat{\Sigma}_T\hat{A}_T^\top \text{ with } \hat{A}_T = [I_n, -\hat{J}_{mo}\hat{J}_{so}^{-1}]. \tag{5}$$

3.2 The modified test of KL

KL use a different normalization matrix in the place of $\hat{V}_T$ in (3). More specifically, let $\hat{\theta}_{[Tr]}$ be the estimate of θ_o from solving (2) but using observations up to $[Tr]$, where $r \in [\varepsilon, 1]$ with ε being an arbitrarily small positive number. Let

$$\psi_{[Tr]} = \frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]}\left(m_t(\hat{\theta}_{[Tr]}) - \frac{1}{T}\sum_{t=1}^T m_t(\hat{\theta}_T)\right). \tag{6}$$

Then, KL construct the following normalization matrix:

$$\hat{S}_T = T^{-1} \sum_{[Tr]=1}^{T} \psi_{[Tr]} \psi_{[Tr]}^{\top},$$

and their modified M test is given by

$$\mathcal{M}_{2T} = \left[T^{-1/2} \sum_{t=1}^{T} m_t(\hat{\theta}_T) \right]^{\top} \hat{S}_T^{-1} \left[T^{-1/2} \sum_{t=1}^{T} m_t(\hat{\theta}_T) \right].$$

KL's modification is inspired by Kiefer et al. (2000). The statistic $\hat{S}_T$ has two properties. First, it has a non-degenerate distribution even asymptotically. Second, it depends on V_o in a particular way (see (7) below). The application of $\hat{S}_T$ yields an asymptotically pivotal test, although the limiting distribution of the test is non-standard. The most important feature of this approach is that it does not require directly estimating the covariance matrix, while maintaining robustness to heteroskedasticity and autocorrelation in m_t of unknown forms. See Bunzel et al. (2001), Lobato (2001) and Shao (2010) for related studies.

KL emphasized that in the current context, it is crucial to re-estimate the model using subsamples. In fact, if $\hat{\theta}_{[Tr]}$ is replaced by $\hat{\theta}_T$, then the resulting test is in general not pivotal. Note that in (6) the subsample statistic is recentered at the full-sample estimate. This is important to ensure that $\hat{S}_T$ does not diverge under the alternative hypothesis. The limiting distribution of $\mathcal{M}_{2T}$ can be established under the following assumptions. They are similar to those used in KL.

Assumption 4. $\mathbb{E}m_t(\theta)$ and $\mathbb{E}s_t(\theta)$ are differentiable in θ; $T^{-1}\sum_{t=1}^{[Tr]} \nabla_{\theta^{\top}} \mathbb{E}m_t(\theta_o) \to r\ J_{mo}$ and $T^{-1}\sum_{t=1}^{[Tr]} \nabla_{\theta^{\top}} \mathbb{E}s_t(\theta_o) \to rJ_{so}$ uniformly in $r \in [0,1]$ with J_{so} being positive definite. Under H_o and H_{1T},

$$\begin{bmatrix} T^{-1/2}\sum_{t=1}^{[Tr]} m_t(\hat{\theta}_{[Tr]}) \\ T^{-1/2}\sum_{t=1}^{[Tr]} s_t(\hat{\theta}_{[Tr]}) \end{bmatrix} = \begin{bmatrix} T^{-1/2}\sum_{t=1}^{[Tr]} m_t(\theta_o) \\ T^{-1/2}\sum_{t=1}^{[Tr]} s_t(\theta_o) \end{bmatrix} + \begin{bmatrix} rJ_{mo} \\ rJ_{so} \end{bmatrix} T^{1/2}\left(\hat{\theta}_{[Tr]} - \theta_o\right) + o_p(1)$$

uniformly in $r \in [\varepsilon, 1]$ with ε being an arbitrarily small positive number.

Assumption 5. Under H_o and H_{1T},

$$\begin{bmatrix} T^{-1/2}\sum_{t=1}^{[Tr]} \left(m_t(\theta_0) - \mathbb{E}m_t(\theta_0)\right) \\ T^{-1/2}\sum_{t=1}^{[Tr]} \left(s_t(\theta_0) - \mathbb{E}s_t(\theta_0)\right) \end{bmatrix} \Rightarrow \Sigma_o^{1/2} W_{p+q}(r),$$

where $W_{p+q}(r)$ is a $p+q$ vector of independent Wiener processes.

Assumptions 4 and 5 strengthen Assumptions 2 and 3 to hold uniformly in $r \in [\varepsilon, 1]$ with ε being some arbitrarily small positive number. In the quantile regression context, the expansion for $T^{-1/2}\sum_{t=1}^{[Tr]} s_t(\hat{\theta}_{[Tr]})$ in Assumption 4 corresponds to a uniform Bahadur representation, sufficient conditions for which can be found in Qu (2008, Lemma 1). In the special case where $m_t(\theta)$ and $s_t(\theta)$ are continuously differentiable, then Assumption 4 can be replaced by the following set of conditions: $(\hat{\theta}_{[Tr]} - \theta_o) = o_p(1)$ uniformly in r under H_o and H_{1T}, $T^{-1}\sum_{t=1}^{[Tr]} \nabla_{\theta^\top} m_t(\theta) \to^p r\ J_{mo}(\theta)$ and $T^{-1}\sum_{t=1}^{[Tr]} \nabla_{\theta^\top} s_t(\theta) \to^p r\ J_{so}(\theta)$ uniformly in $\theta \in \Theta$ and $r \in [0,1]$, where $J_{mo}(\theta)$ is non-random and $J_{so}(\theta)$ is non-random and positive definite. The next lemma gives the limiting distribution of $\hat{S}_T$ and $\mathcal{M}_{2T}$. The proof is omitted.

Lemma 2 *Let Assumptions 1 and 4-5 hold, then under H_o and H_{1T},*

$$\hat{S}_T \Rightarrow V_o^{1/2}\left(\int_0^1 B_p(r)B_p(r)^\top dr\right)(V_o^{1/2})^\top, \tag{7}$$

where $B_p(r)$ is a p-vector of independent Brownian bridges and $V_o = V_o^{1/2}(V_o^{1/2})^\top$. Also,

$$\mathcal{M}_{2T} \Rightarrow \begin{cases} W_p(1)^\top(\int_0^1 B_p(r)B_p(r)^\top dr)^{-1}W_p(1) & \text{under } H_o \\ (V_o^{-1/2}A_o\delta + W_p(1))^\top(\int_0^1 B_p(r)B_p(r)^\top dr)^{-1}(V_o^{-1/2}A_o\delta + W_p(1)) & \text{under } H_{1T} \end{cases}.$$

4 The proposed approach

The subsample statistic $[Tr]^{-1/2}\sum_{t=1}^{[Tr]} m_t(\hat{\theta}_{[Tr]})$ shares the same asymptotic null distribution as its full-sample counterpart $T^{-1/2}\sum_{t=1}^{T} m_t(\hat{\theta}_T)$. This suggests designing an alternative normalization matrix by exploiting the subsample information more efficiently. We propose the following simple procedure.

- **Step 1**. Separate the full sample into $(K+1)$ recursive subsamples, with the first subsample containing the first $b_T = T/(K+1)$ observations and the j^{th} subsample containing observations up to $T_j = jb_T$. Estimate the model using the subsamples to obtain the $Kp \times 1$ vector

$$\Psi_T(K) = \left(\psi_{T_1}^\top, \ldots, \psi_{T_K}^\top\right)^\top, \tag{8}$$

  where ψ_{T_j} is defined as in (6) with $[Tr]$ replaced by T_j.

- **Step 2**. Let $r_j = jb_T/T$ and construct a K by K matrix C with its $ijth$ element being $C_{ij} = r_{\min(i,j)}(1 - r_{\max(i,j)})$. Let G be the upper triangular Cholesky factorization of C^{-1} and obtain

$$\Psi_T^*(K) \equiv (\psi_{T_1}^{*\top}, \ldots, \psi_{T_K}^{*\top})^\top = (G \otimes I_p)\Psi_T(K).$$

- **Step 3**. Construct the following normalization matrix

$$R_T(K) = \frac{1}{K}\sum_{j=1}^{K} \psi_{T_j}^* \psi_{T_j}^{*\top}$$

and the following modified M test

$$\mathcal{M}_{3T} = \left(\frac{K-p+1}{Kp}\right)\left[T^{-1/2}\sum_{t=1}^{T} m_t(\hat{\theta}_T)\right]^\top R_T(K)^{-1}\left[T^{-1/2}\sum_{t=1}^{T} m_t(\hat{\theta}_T)\right]. \tag{9}$$

We first discuss the idea underlying the procedure before presenting a formal analysis of its asymptotic properties. Step 1 entails re-estimating the model and re-computing the moment vector using subsamples. This step closely parallels the procedure of KL, with an important difference in that it involves only K, but not T, recursive subsamples. This can lead to a substantial reduction in the computational cost when the sample size is large or when the model is nonlinear. More importantly, this also opens up a window for constructing tests with better power properties.

The elements of $\Psi_T(K)$, say ψ_{T_j} and ψ_{T_i} with $i \neq j$, are in general correlated and have different variances. Step 2 applies a linear transformation to $\Psi_T(K)$, such that $\psi_{T_j}^*$ and $\psi_{T_i}^*$ become asymptotically independent with the same asymptotic covariance. The normalization matrix $R_T(K)$ is simply the second sample moment of $\{\psi_{T_j}^*\}_{j=1}^K$. As shown later, this matrix has a Wishart distribution with mean equal to V_o if K is fixed as $T \to \infty$ and converges to V_o if K is allowed to increase to infinity with T.

The above procedure is simple to implement. The main work is to re-estimate the model using subsamples. Thus, it is useful for problems where the asymptotic covariance matrix is difficult or cumbersome to estimate but re-estimating the model is not very costly.

5 Asymptotic properties of the $\mathcal{M}_{3T}$ test

We will derive the null limiting distribution of $\mathcal{M}_{3T}$ under two alternative asymptotic frameworks: (i) K fixed as $T \to \infty$, and (ii) $K \to \infty$ but $K/T \to 0$ as $T \to \infty$. Note that if we view $1/(K+1)$

as a bandwidth parameter, then the framework (i) leads to the fixed-b asymptotics along the same lines of Kiefer and Vogelsang (2005) and (ii) the conventional small-b asymptotics.

5.1 Fixed-K asymptotics

The next Lemma establishes the asymptotic properties of $R_T(K)$.

Lemma 3 *Let Assumptions 1, 4 and 5 hold and assume K is fixed as $T \to \infty$, then*

$$\Psi_T^*(K) \to^d N\left(0, I_K \otimes V_o\right) \tag{10}$$

and

$$R_T(K) \to^d W(V_o/K, K) \tag{11}$$

where $W(.,.)$ denotes a Wishart distribution. Meanwhile, $R_T(K)$ is asymptotically independent of $T^{-1/2}\sum_{t=1}^T m_t(\hat{\theta}_T)$.

Note that the asymptotic independence between $\psi_{T_i}^*, \psi_{T_j}^*$ and $T^{-1/2}\sum_{t=1}^T m_t(\hat{\theta}_T)$ is crucial for establishing the asymptotic distribution of the test. We also need the following general result for the distribution of a quadratic form.

Lemma 4 (Anderson, 2003, Theorem 5.2.2) *Let*

$$\mathcal{H}(\omega) = \mathcal{W}(\omega)^\top \left(K^{-1}\sum_{j=1}^K \mathcal{Z}_j\mathcal{Z}_j^\top\right)^{-1}\mathcal{W}(\omega),$$

where $\mathcal{W} \sim N(\omega, I_p)$ for some $\omega \in \mathbb{R}^p$ and $\mathcal{Z}_j \sim N(0, I_p)$ with $j = 1, \ldots, K$. Suppose $\mathcal{W}$ is independent of $\sum_{j=1}^K \mathcal{Z}_j\mathcal{Z}_j^\top$ and $\mathcal{Z}_j$ is independent of $\mathcal{Z}_i$ when $j \neq i$, then

$$\left(\frac{K-p+1}{Kp}\right)\mathcal{H}(\omega) \sim \begin{cases} F(p, K-p+1) & \text{if } \omega = 0, \\ F(p, K-p+1; \omega^\top\omega) & \text{if } \omega \neq 0, \end{cases}$$

where $F(p, K-p+1)$ is the central F distribution with p and $K-p+1$ degrees of freedom, and $F(p, K-p+1; \omega^\top\omega)$ is the non-central F distribution with non-centrality parameter $\omega^\top\omega$.

The next Proposition gives the asymptotic distribution of $\mathcal{M}_{3T}$, which follows from Lemma 3, 4 and the continuous mapping theorem.

Proposition 1 *Let Assumptions 1, 4 and 5 hold and assume K is fixed as $T \to \infty$, then*

$$\mathcal{M}_{3T} \to^d \begin{cases} F(p, K-p+1), & \text{under } H_o, \\ F(p, K-p+1; \delta^\top A_o^\top V_o^{-1} A_o \delta), & \text{under } H_{1T}. \end{cases}$$

The null limiting distribution is a standard F distribution, whose critical values are readily available. They are also monotonically decreasing in K. This is a useful feature because it permits the limiting distribution to provide adequate approximations over a wide range of values for K.

5.2 Large-K asymptotics

It is interesting to study the distribution of the test assuming that both K and T are large. This is not to replace the fixed-K asymptotics, but rather to view the problem from a different angle and to obtain a direct comparison with the conventional M test.

We proceed by assuming $K \to \infty$ but $K/T \to 0$ as $T \to \infty$. Some assumptions need to be strengthened because this alternative framework involves estimation with a vanishing proportion of the sample size. Specifically, the first subsample involves b_T observations with $b_T/T \to 0$.

Assumption 4B. $\mathbb{E}m_t(\theta)$ and $\mathbb{E}s_t(\theta)$ are differentiable in θ; $l^{-1}\sum_{t=1}^{l} \nabla_{\theta^\top} \mathbb{E}m_t(\theta_o) \to J_{mo}$ and $l^{-1}\sum_{t=1}^{l} \nabla_{\theta^\top} \mathbb{E}s_t(\theta_o) \to J_{so}$ uniformly in $l \in [b_T, T]$ with J_{so} being positive definite. Under H_o and H_{1T},

$$\begin{bmatrix} T^{-1/2}\sum_{t=1}^{l} m_t(\hat{\theta}_l) \\ T^{-1/2}\sum_{t=1}^{l} s_t(\hat{\theta}_l) \end{bmatrix} = \begin{bmatrix} T^{-1/2}\sum_{t=1}^{l} m_t(\theta_o) \\ T^{-1/2}\sum_{t=1}^{l} s_t(\theta_o) \end{bmatrix} + lT^{-1/2} \begin{bmatrix} J_{mo}(\theta_o) \\ J_{so}(\theta_o) \end{bmatrix} \left(\hat{\theta}_l - \theta_o\right) + o_p(1)$$

uniformly in $l \in [b_T, T]$.

Assumption 5B. Let Assumption 5 hold. Also, assume $m_t(\theta_0)$ and $s_t(\theta_0)$ are strong mixing with mixing numbers $\alpha_m(j)$ and $\alpha_s(j)$ satisfying $\sum_{j=1}^{\infty} j^2 \alpha_m(j)^{(\nu-1)/\nu} < \infty$ and $\sum_{j=1}^{\infty} j^2 \alpha_s(j)^{(\nu-1)/\nu} < \infty$ for some $\nu > 1$. Also, $sup_{t>1} E\left\|m_t(\theta_o)\right\|^{4\nu} < \infty$ and $sup_{t>1} E\left\|s_t(\theta_o)\right\|^{4\nu} < \infty$.

Similarly as in Assumption 4, in the special case where $m_t(\theta)$ and $s_t(\theta)$ are continuously differentiable, Assumption 4B can replaced by the following requirements: $(\hat{\theta}_l - \theta_o) \to^{a.s.} 0$ under H_o and H_{1T}, $l^{-1}\sum_{t=1}^{l} \nabla_{\theta^\top} m_t(\theta) \to^{a.s.} J_{mo}(\theta)$ and $l^{-1}\sum_{t=1}^{l} \nabla_{\theta^\top} s_t(\theta) \to^{a.s.} J_{so}(\theta)$ uniformly in θ, where $J_{mo}(\theta)$ is non-random and $J_{so}(\theta)$ is non-random and positive definite. Assumption 5B imposes some restrictions on the dependence structure. The same mixing assumption is used in Andrews (1991, Lemma 1) for the analysis of HAC estimators. Assumptions 4 and 5 are more

stringent than Assumptions 4 and 5. However, this does not imply limitations for the $\mathcal{M}_{3T}$, given that in practice the inference will be carried out based on fixed-K asymptotics. The next result gives the limits of $R_T(K)$ and $\mathcal{M}_{3T}$.

Proposition 2 *Let Assumptions 1, 4B and 5B hold. Assume $K \to \infty$ but $K/T \to 0$ as $T \to \infty$.*

1. *Let $Z_j = T^{-1/2} \sum_{t=(j-1)b_T+1}^{jb_T} \xi_t(\theta_o)$ (j=1,...,K+1) and $Z = [Z_1^\top, ..., Z_{K+1}^\top]^\top$. Then*

$$R_T(K) = \sum_{j=1}^{K+1} Z_j Z_j^\top + o_p(1) \to^p V_o \tag{12}$$

   under H_o and H_{1T}.

2. *$\mathcal{M}_{3T}$ test is asymptotically equivalent to the $\mathcal{M}_{1T}$ under both H_o and H_{1T}.*

Because $R_T(K)$ consistently estimates V_o, the $\mathcal{M}_{3T}$ test can potentially achieve the same power as the conventional M test. Thus, in a sense, it provides a bridge between the KL's test and the conventional M test.

The proposition also reveals a connection between $R_T(K)$ and Bartlett's (1950) proposal for estimating the spectral density of a stationary time series. Specifically, to estimate the spectrum, Bartlett suggested splitting the observed sample into $(K+1)$ groups with equal number of observations. The periodogram is then computed for each group, and the estimator for the ordinate associated with a particular frequency is taken to the average of the $(K+1)$ estimators. This estimator, when computed for $\xi_t(\theta_o)$ and evaluated at frequency zero, equals $(1/2\pi)\sum_{j=1}^{K+1} Z_j Z_j^\top$, which is proportional to the leading term in (12). In addition, $\sum_{j=1}^{K+1} Z_j Z_j^\top$ can be equivalently represented as

$$\sum_{s=-b_T+1}^{b_T-1} \left(1 - \frac{|s|}{b_T}\right) \overline{C}_s,$$

where $\overline{C}_s$ is an estimator for the covariance at lag s:

$$\overline{C}_s = \frac{1}{(K+1)(b_T - s)} \sum_{u=0}^{K} \sum_{r=1}^{b_T-s} \xi_{r+ub_T}(\theta_o) \xi_{r+s+ub_T}(\theta_o)^\top \quad (s \geq 0),$$

and $(1 - |s|/b_T)$ with $|s| \leq b_T$ is the Bartlett kernel. The above expression coincides with that of Bartlett (1950, P.5) evaluated at frequency zero.

5.3 Discussions

The fixed-K asymptotics is recommended in practice. This framework allows us to capture the uncertainty associated with a particular choice of K. Note that even in applications where the "large-K" asymptotics is adequate, the "fixed-K" framework is justified because it is more conservative due to the relationship between the F and Chi-square distribution.

Under H_{1T}, the limiting distributions of $\mathcal{M}_{1T}$ and $\mathcal{M}_{3T}$ depend on the model only through $\delta^\top A_o^\top V_o^{-1} A_o \delta$, p (the number of restrictions being tested) and K. Thus, for a given p and significance level α, we can calculate the value of K under which the asymptotic power difference between these two tests is bounded by a given number, say κ. In Table 1, we consider $\alpha = 0.05$, $1 \leq p \leq 10$ and $\kappa = 0.05, 0.02$ and 0.01. These values can provide some guidance as to the choice of K in practice.

Clearly, some caution is needed in using Table 1. The asymptotic results in Propositions 1 and 2 provide good finite sample approximations only if $Cov(Z_j, Z_i) \to 0$ for $i \neq j$. If serial correlation is absent under the null hypothesis, then this requirement is easy to meet and the size of the test is relatively insensitive to the choice of K. In such cases, K can be made large to achieve higher asymptotic power. If m_t or s_t is strongly serially correlated, then a large K may lead to size distortions. In this case, the choice of K will rely on some judgement about the approximate subsample sizes to achieve uncorrelatedness. In practice, it is useful to construct the test using different K to examine the sensitivity of the results.

It is desirable to have a selection rule for K that can adapt to the extent of serial correlation present in the data. Existing results in the literature suggest two possible ways forward. Both of them operate under the large K asymptotics. One is to choose K to minimize a weighted average of type I and type II errors as in Sun, Phillips and Jin (2008). This will require analyzing high order asymptotic properties of the K subsample based estimators in the test statistic. The second is to choose K to minimizes the asymptotic mean squared error for estimating V_o as in Andrews (1991). For this purpose, Proposition 2 will need to be sharpened to have a reminder term of order $o_p(\sqrt{b_T/T})$ rather than $o_p(1)$. Both tasks are quite challenging and beyond the scope of the current paper.

6 Local asymptotic power

We provide some results to illustrate the power differences between $\mathcal{M}_{1T}$, $\mathcal{M}_{2T}$ and $\mathcal{M}_{3T}$ for various $c = V_o^{-1/2} A_o \delta$, p (the number of restrictions being tested) and K. Specifically, we let $p = 1, 5$ and 10, and in each case we vary K between 20 and 80. The parameter c is set to $(d/n^{1/2}, \ldots, d/n^{1/2})^\top$ with d taking values between 0 and 30. For the $\mathcal{M}_{1T}$ and $\mathcal{M}_{3T}$ tests, the power is evaluated using the GAUSS commands for the probability integrals of non-central Chi-square and F distributions. For the $\mathcal{M}_{2T}$ test, since its distribution function does not admit a simple analytical form, we approximate it using simulations. To this end, the Brownian motion $W_p(\cdot)$ is approximated using $T^{-1/2} \sum_{t=1}^{[Tr]} u_t$ with $u_t \sim i.i.d.N(0, I_p)$ and $T = 10,000$. The distribution is based on 50,000 replications.

Figure 1 reports local asymptotic powers at 1%, 5% and 10% level. As predicted by the theory, the power of the $\mathcal{M}_{3T}$ test converges to those of the $\mathcal{M}_{1T}$ test when K increases. The rate at which it approaches the limit depends on $K - p + 1$. For $p = 1$, their powers are already quite close when $K = 20$. For $p = 5$, their maximum power difference is 0.10 when $K = 40$. For $p = 10$, the difference is greater, but is less than 0.10 when $K = 80$. We also observe that, for $p = 1$ ($p = 5, p = 10$), the $\mathcal{M}_{3T}$ test is asymptotically more powerful than the $\mathcal{M}_{2T}$ test when $K \geq 20$ ($K \geq 20, K \geq 40$). These results hold for all three significance levels considered.

7 Applications and simulation comparisons

This section presents three applications of the $\mathcal{M}_{3T}$ test. We consider three values for K: 20, 40 and K^*, with K^* taking values from the first row in Table 1 ($\kappa = 0.05$). All results reported are at a 5% nominal level using 5000 replications.

7.1 Specification testing for conditional heteroskedasticity models

Consider conditional heteroskedasticity models of the general form:

$$y_t = \mu_t(\theta) + h_t(\theta)^{1/2} \varepsilon_t, \tag{13}$$

where y_t is a scalar random variable, $\mu_t(\theta)$ and $h_t(\theta)$ are, respectively, the conditional mean and variance functions of y_t, θ is a finite dimensional parameter, and ε_t is an error term with zero mean and unit variance. The Gaussian quasi-maximum likelihood (QML) estimator can be obtained by

solving the estimating equation (2) with

$$s_t(\theta) = (\nabla_\theta \mu_t(\theta)) h_t^{-1/2}(\theta)\varepsilon_t + \frac{1}{2}(\nabla_\theta h_t(\theta)) h_t^{-1}(\theta)\left(\varepsilon_t^2 - 1\right). \tag{14}$$

One approach to construct diagnostic tests for (13) is to follow Ljung and Box (1978) and McLeod and Li (1983) to consider the first and second moments of ε_t. This leads to a family of M tests considered by Li and Mak (1994), Lundbergh and Teräsvirta (2002), Berkes et al. (2003), Wong and Ling (2005), and Chen (2008). As in their studies, we consider the following moment vector:

$$m_t(\theta) = (\varepsilon_t \lambda_{1,t}^\top, \left(\varepsilon_t^2 - 1\right) \lambda_{2,t}^\top)^\top,$$

where $\lambda_{1,t} = (\varepsilon_{t-1}, ..., \varepsilon_{t-h})^\top$ and $\lambda_{2,t} = \left(\varepsilon_{t-1}^2 - 1, ..., \varepsilon_{t-h}^2 - 1\right)^\top$ with h being a finite constant. The corresponding null hypothesis is given by

$$\mathbb{E}\left[m_t(\theta_o)\right] = 0. \tag{15}$$

Note that the asymptotic covariance matrix in the $\mathcal{M}_{1T}$ test (c.f. Lemma 1) is given by

$$V_o = [I_p, -J_{mo}J_{so}^{-1}]\Sigma_o[I_p, -J_{mo}J_{so}^{-1}]^\top$$

with

$$\begin{aligned} J_{mo} &= -\mathbb{E}\begin{pmatrix} \lambda_{1,t}(\nabla_{\theta^\top}\mu_t(\theta_o)) h_t^{-1/2}(\theta_o) \\ \lambda_{2,t}(\nabla_{\theta^\top} h_t(\theta_o)) h_t^{-1}(\theta_o) \end{pmatrix}, \\ J_{so} &= -\mathbb{E}\left(\nabla_\theta \mu_t(\theta_o)\nabla_{\theta^\top}\mu_t(\theta) h_t^{-1}(\theta_o) + \frac{1}{2}\nabla_\theta h_t(\theta_o)(\nabla_{\theta^\top} h_t(\theta)) h_t^{-2}(\theta_o)\right), \end{aligned}$$

and

$$\Sigma_o = \begin{bmatrix} \mathbb{E}[m_t(\theta_o)m_t(\theta_o)^\top] & \mathbb{E}[m_t(\theta_o)s_t(\theta_o)^\top] \\ \mathbb{E}[s_t(\theta_o)m_t(\theta_o)^\top] & \mathbb{E}[s_t(\theta_o)s_t(\theta_o)^\top] \end{bmatrix}.$$

Consistent estimates for J_{mo}, J_{so} and Σ_o can be obtained by replacing θ_o with the QMLE $\hat{\theta}_T$ and the expectation with the sample average. These may require some nontrivial analytical work depending on the specification of the model. For example, the calculation can be involved if one wants to test models with nonlinear conditional mean and/or variance specifications; see, e.g., Lundbergh and Teräsvirta (2006) for regime-switching conditional mean specifications and Nelson (1991), Glosten et al. (1993), Hansen (1994), and Hentschel (1995) for nonlinear conditional variance specifications.

In contrast, the $\mathcal{M}_{2T}$ and $\mathcal{M}_{3T}$ do not require estimating J_{mo}, J_{so} or Σ_o. To implement the $\mathcal{M}_{3T}$ test, the main computational and analytical task is to solve

$$T_j^{-1/2} \sum_{t=1}^{T_j} s_t(\hat{\theta}_{T_j}) = 0$$

for $(K+1)$ subsamples and construct

$$\psi_{T_j} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T_j} \left(m_t(\hat{\theta}_{T_j}) - \frac{1}{T} \sum_{t=1}^{T} m_t(\hat{\theta}_T) \right).$$

The test can be applied to models with various specifications for (μ_t, h_t) without further modification. Multivariate models do not introduce any complication provided a relatively large sample size is available.

We now conduct some simulations to compare the three tests. We specify GARCH (1,1) models as the null hypothesis and examine the size and power of the $\mathcal{M}_{3T}$ test over $h = 1, 2, 3$ and 4 (note that the number of restrictions being tested is $2h$). To make the simulation empirically relevant, we consider parameter values calibrated to empirical estimates. That is, we first estimate a GARCH (1,1) model using some actual datasets and then use the parameter estimates to generate the data under the null hypothesis. The first dataset is Deutschmark/British Pound daily returns over the period January, 1984 to January, 1992 (1974 observations). We obtain

• **DGP-EXG**: $y_t = -0.006 + u_t$ with $u_t = h_t^{1/2}\varepsilon_t$ and $h_t = 0.011 + 0.806h_{t-1} + 0.153u_{t-1}^2$.

The second is the NASDAQ index daily returns between January, 1990 and January, 2002 (3027 observations). We obtain

• **DGP-IND**: $y_t = 0.086 + u_t$ with $u_t = h_t^{1/2}\varepsilon_t$ and $h_t = 0.023 + 0.875h_{t-1} + 0.116u_{t-1}^2$.

In both cases, $\varepsilon_t \sim i.i.d.N(0,1)$. For size comparisons, we consider $T = 500$ and 1000. The results are reported in Table 2. [2] They show that the $\mathcal{M}_{3T}$ has decent size properties.[3] There is some over rejection when the number of restrictions is large and T is small, but it improves when the sample size is increased.

For power assessment, we consider four alternative models. In the first two models, conditional variances are misspecified, while in the latter two the conditional means are misspecified.

[2] When computing the tests, the likelihood function is maximized using the CML routine for the Gauss environment, under the restriction that the parameters in the variance equation are non-negative.

[3] $\mathcal{M}_{2T}$ is constructed by setting the starting subsample size equal to 20.

• **GARCH(2,1)**: $y_t = -0.006 + u_t$ with $u_t = h_t^{1/2}\varepsilon_t$ and $h_t = 0.011 + 0.3h_{t-1} + 0.3h_{t-2} + 0.3u_{t-1}^2$.

• **GARCH(1,2)**: $y_t = -0.006 + u_t$ with $u_t = h_t^{1/2}\varepsilon_t$ and $h_t = 0.011 + 0.3h_{t-1} + 0.3u_{t-1}^2 + 0.3u_{t-2}^2$.

• **AR(1)-GARCH(1,1)**: $y_t = -0.006 + 0.1y_{t-1} + u_t$ with $u_t = h_t^{1/2}\varepsilon_t$ and $h_t = 0.011 + 0.806h_{t-1} + 0.153u_{t-1}^2$.

• **MA(1)-GARCH(1,1)**: $y_t = -0.006 - 0.1u_{t-1} + u_t$ with $u_t = h_t^{1/2}\varepsilon_t$ and $h_t = 0.011 + 0.806h_{t-1} + 0.153u_{t-1}^2$.

In all cases, $\varepsilon_t \sim i.i.d.N(0,1)$. The results are summarized in Table 3. They show that, across simulations, the maximum power difference between $\mathcal{M}_{3T}$ and $\mathcal{M}_{1T}$ is 0.15 when K=20, 0.08 when K=40 and 0.09 when $K = K^*$. The power of $\mathcal{M}_{3T}$ is uniformly higher than $\mathcal{M}_{2T}$, with the maximum difference being 0.17 when K=20, 0.20 when K=40 and 0.22 when $K = K^*$. This suggests that important power gains are present.

7.2 Nonnested testing with serially correlated errors

This section considers Davidson and MacKinnon (1981) type tests for nonnested models. Suppose there are two specifications for the conditional mean function of y_t:

$$\begin{aligned} H_{1T} &: \mathbb{E}(y_t|x_t) = \mu_t(x_t, \beta_o), \\ H_{2T} &: \mathbb{E}(y_t|z_t) = \eta_t(z_t, \gamma_o), \end{aligned}$$

where x_t and z_t are random vectors that can partially overlap and β_o and γ_o are unknown finite-dimensional parameters. As discussed in Wooldridge (1991), a Davidson-Mackinnon type test for H_{1T} versus H_{2T} can be obtained by letting

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} m_t(\hat{\beta}_T, \hat{\gamma}_T) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(y_t - \mu_t(x_t, \hat{\beta}_T)\right)\left(\mu_t(x_t, \hat{\beta}_T) - \eta_t(z_t, \hat{\gamma}_T)\right). \tag{16}$$

A test for H_{2T} versus H_{1T} can be obtained similarly, by exchanging $\mu_t(x_t, \hat{\beta}_T)$ and $\eta_t(z_t, \hat{\gamma}_T)$. In practice, it is frequently desirable to allow for heteroskedasticity and serial correlation when constructing the test. See, for example, Bernanke, Bohn and Reiss (1988) for testing time series investment models and Elyasiani and Nasseh (1994) for testing Mankiw and Summers' (1986) hypothesis about the U.S. money demand function.

The $\mathcal{M}_{3T}$ test automatically allows for such features. The main computational and analytical work is to re-estimate β_o using subsamples and to construct

$$\psi_{T_j} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T_j} \left(m_t(\hat{\beta}_{T_j}, \hat{\gamma}_T) - \frac{1}{T} \sum_{t=1}^{T} m_t(\hat{\beta}_T, \hat{\gamma}_T) \right).$$

Interestingly, there is no need to re-estimate γ_o. This is because the distribution of the moment vector (16) depends on γ only through γ_o, provided that the two alternative models are not orthogonal. Also, the $\mathcal{M}_{3T}$ test can be applied with different estimators for β_o and γ_o, as long as the same one is used for both the full and the subsamples. No further modification is needed.

We consider a simulation experiment based on the empirical example studied in Choi and Kiefer (2008) and Elyasiani and Nasseh (1994). These two papers revisited the hypothesis of Mankiw and Summers (1986) that consumption (or personal expenditure) rather than income (gross national product [GNP]) is the right scale variable for money demand (for $M1$ or $M2$). The model is

$$y_t = \beta_1 + \beta_2 r_t + \beta_3 r_{t-1} + \beta_4 r_{t-2} + \beta_5 z_t + \beta_6 z_{t-1} + \beta_7 z_{t-2} + \varepsilon_t, \tag{17}$$

where y_t is the difference in log of real money stock $M2$, r_t is the difference in log of the 3-month Treasury Bill rate, z_t is the difference in log of real personal expenditure (for a consumption measure) or real GNP (for an income measure). To account for serial correlation in the errors, Elyasiani and Nasseh (1994) applied the Cochrane–Orcutt procedure, while Choi and Kiefer (2008) applied a HAC estimator with fixed-b asymptotics.

We test the hypothesis that GNP is the right scale variable. Specifically, the regression (17) is first estimated using quarterly observations over 1959.1–2009.7.[4]The estimates, reported in Table 4, are then used to generate simulated samples. To capture the serial correlation in the errors, ε_t is assumed to follow an $AR(1)$ process with the autoregressive coefficient being 0.43 (i.e., the estimate from the null model) and 0.90. To generate a series, we first simulate ε_t and then feed them into (17) to generate y_t. The regressors are fixed at their true values throughout.[5]. The sample size is 203, the same as that used to calibrate the parameter values. Other aspects of the simulation are the same as in the previous section.

The asymptotic covariance matrix in $\mathcal{M}_{1T}$ is estimated using the formula (5). Specifically, let $y_t = x_t'\beta + \varepsilon_t$ and $y_t = w_t'\gamma + \varepsilon_t$ denote the models under the null and alternative hypotheses and

[4]The readers may refer to Choi and Kiefer (2008) for a detailed description about the source of the data set and the construction of the variables.

[5]As an alternative, we tried to model the regressors r_t and z_t as a VAR(1) and simulated them along with ε_t. The results showed no difference.

$\hat{\beta}_T$ and $\hat{\gamma}_T$ the OLS estimates. The relevant quantities are computed as

$$\begin{aligned} m_t(\hat{\beta}_T, \hat{\gamma}_T) &= \left(y_t - x_t'\hat{\beta}_T\right)\left(x_t'\hat{\beta}_T - w_t'\hat{\gamma}_T\right), \\ s_t(\hat{\beta}_T, \hat{\gamma}_T) &= x_t\left(y_t - x_t'\hat{\beta}_T\right), \\ \hat{J}_{mo} &= -\frac{1}{T}\sum_{t=1}^{T} x_t x_t'\hat{\beta}_T + \frac{1}{T}\sum_{t=1}^{T} x_t w_t'\hat{\gamma}_T, \\ \hat{J}_{so} &= -\frac{1}{T}\sum_{t=1}^{T} x_t x_t', \end{aligned}$$

and $\hat{\Sigma}_T$ is computed by applying the Bartlett kernel with the bandwidth determined using Andrews' (1991) method based on an AR(1) specification.

The results are summarized in Table 5. All three tests have decent sizes. The $\mathcal{M}_{3T}$ test has significantly higher power than $\mathcal{M}_{2T}$ in both cases. Its power is comparable to $\mathcal{M}_{1T}$ when $\rho = 0.43$ and higher when $\rho = 0.9$. The latter is because the asymptotic covariance matrix for the $\mathcal{M}_{1T}$ test is estimated without recentering $m_t(\hat{\beta}_T, \hat{\gamma}_T)$ at its sample average, as is typically done in practice. If it was recentered, then its power would be 0.99 and 0.48 respectively. However, its size would be distorted, being 0.08 and 0.15. For the $\mathcal{M}_{3T}$ test, the recentering is done automatically and such an issue does not arise. We also computed size adjusted power for the case $\rho = 0.9$. They are 33.7% ($K = 20$), 35.0% ($K = 40$) and 32.1% ($K = K^*$) for the $\mathcal{M}_{3T}$ test and 28.5% and 23.0% for the $\mathcal{M}_{1T}$ and $\mathcal{M}_{2T}$ test. Therefore the conclusion remains the same.

It is interesting to examine the finite sample properties of $\mathcal{M}_{3T}$ and $\mathcal{M}_{2T}$ when γ_o is estimated recursively because this can reveal whether knowledge about parameter estimation effect is useful for constructing these two tests. We obtain the following results. For $\mathcal{M}_{3T}$, the null rejection frequencies at $K = 20, 40$ and K^* are 1.9%, 2.0% and 1.9% when $\rho = 0.43$, and 1.4%, 2.2% and 1.7% when $\rho = 0.9$. For the $\mathcal{M}_{2T}$ test, the rejection frequencies are 2.1% when $\rho = 0.43$ and 1.0% when $\rho = 0.9$. Under the alternative hypothesis, the respective rejection rates of $\mathcal{M}_{3T}$ are 79.3%, 86.4% and 78.9% when $\rho = 0.43$, and 12.6%, 19.5% and 14.3% when $\rho = 0.9$. The respective values for $\mathcal{M}_{2T}$ the values are 59.6% and 10.0%. Two patterns emerge. First, for both tests, the size and power are less satisfactory when compared with Table 5. This suggests that knowledge about parameter estimation effect can be quite valuable. Second, $\mathcal{M}_{3T}$ is still substantially more powerful than $\mathcal{M}_{2T}$. This provides further evidence for the theoretical results derived above.

7.3 Testing for serial correlation with unknown heteroskedasticity

A leading diagnostic test for serial correlation is the Q test of Box and Pierce (1970) and Ljung and Box (1978). The limiting distributions of these tests depend on whether the model is dynamic and also the unknown heteroskedasticity, and may not have a chi-square limiting distribution under the null hypothesis of no serial correlation. KL provide a careful analysis of such a situation. The goal of this section is to adapt their simulation designs and assess the relative performance of the tests in such a context.

The DGPs are

$$y_t = x_t'\beta_o + u_t,$$

where x_t is a finite dimensional vector that may include lagged values of y_t, u_t is an error term that exhibits heteroskedasticity under the null hypothesis and is serially correlated under the alternative hypothesis. The detailed information on these DGPs are given in Table 6. Let $u_t(\beta) = y_t - x_t'\beta$. The three $\mathcal{M}$ tests are based on

$$m_t(\beta) = (u_t(\beta)u_{t+1}(\beta), ..., u_t(\beta)u_{t+q}(\beta))'.$$

The asymptotic covariance matrix in $\mathcal{M}_{1T}$ is estimated using the formula (5), where (let $\hat{\beta}_T$ be the OLS estimate of β_o using the full sample)

$$\begin{aligned}
m_t(\hat{\beta}_T) &= \left(u_t(\hat{\beta}_T)u_{t+1}(\hat{\beta}_T), ..., u_t(\hat{\beta}_T)u_{t+q}(\hat{\beta}_T)\right)' \\
s_t(\hat{\beta}_T) &= x_t u_t(\hat{\beta}_T), \\
\hat{J}_{mo} &= -\frac{1}{T}\sum_{t=1}^{T-q} u_t(\hat{\beta}_T) \begin{bmatrix} x_{t+1}' \\ ... \\ x_{t+q}' \end{bmatrix}, \\
\hat{J}_{so} &= -\frac{1}{T}\sum_{t=1}^{T} x_t x_t'.
\end{aligned}$$

Because m_t and s_t are serially uncorrelated under the null hypothesis, $\hat{\Sigma}_T$ is computed as

$$\hat{\Sigma}_T = T^{-1}\sum_{t=1}^{T-q}\left(m_t(\hat{\beta}_T)^\top, s_t(\hat{\beta}_T)^\top\right)^\top\left(m_t(\hat{\beta}_T)^\top, s_t(\hat{\beta}_T)^\top\right).$$

Table 7 shows the rejection frequencies of the $\mathcal{M}$ tests under the null hypothesis. The empirical sizes of the $\mathcal{M}_{2T}$ test are taken from Table 2 in KL. The sizes of the $\mathcal{M}_{3T}$ and $\mathcal{M}_{2T}$ are quite close

to the nominal level, however, $\mathcal{M}_{1T}$ displays substantial size distortions for DGPs 5-8 (the AR(1) models). The latter is because the null distribution of the $\mathcal{M}_{1T}$ test is discontinuous at $\rho = 0$ (ρ denotes the coefficient in front of y_{t-1} in Table 6). Indeed, $\hat{\gamma}_{1,T}$ converges to zero at rate $\sqrt{T}$ when $0 < |\rho| < 1$ and at rate T when $\rho = 0$. This feature, first documented by Durbin (1970, p. 419), leads to the vulnerability of the sampling distribution of $\hat{\gamma}_{1,T}$ to sampling errors in $\hat{\rho}$ unless T is sufficiently large or ρ is sufficiently away from zero. In our simulations, the estimated asymptotic matrix in $\mathcal{M}_{1T}$ is often close to being singular and the statistic tends to take very large values. In contrast, $\mathcal{M}_{3T}$ and $\mathcal{M}_{2T}$ tests do not involve direct estimation of the covariance matrix, therefore are more robust to this problem.

The possible size distortion makes the power comparison with $\mathcal{M}_{1T}$ uninformative. Thus, instead of $\mathcal{M}_{1T}$ test, we include the test of Wooldridge (1990) as a benchmark, denoting it by $\mathcal{WL}$. Its rejection frequencies, along with those for the $\mathcal{M}_{2T}$ test are taken from Table 4 in KL. The results, summarized in Table 8, show that $\mathcal{M}_{3T}$ dominates $\mathcal{M}_{2T}$ test and its power is at least comparable to the $\mathcal{WL}$ test.

8 Conclusion

We have proposed a family of modified M tests. We showed that they converge to F distributions under fixed-K asymptotics, and to Chi-square distributions if K is allowed to approach infinity. They are automatically account for the effect of parameter estimation and allow for conditional heteroskedasticity and serial correlation of general forms. We have also showed that the new normalization matrix has a close connection with Bartlett's (1950) estimator for the spectrum of a stationary time series. We conjecture it is possible to generalize the current framework to test specifications in a panel data or spatial regression context.

References

Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis (3rd edition)*, New York: Wiley.

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

Bartlett, M.S. (1950). Periodogram analysis and continuous spectra. *Biometrika*, 37, 1-16.

Bera, A. K. and Y. Bilias (2001). Rao's score, Neyman's $C(\alpha)$, and Silvey's LM tests: An essay on historical developments and some new results, *Journal of Statistical Planning and Inference*, 97, 9–44.

Berkes, I., L. Horváth, and P. Kokoszka (2003). Asymptotics for GARCH squared residual correlations, *Econometric Theory*, 19, 515–540.

Bernanke, B., H. Bohn and P. C. Reiss (1988). Alternative non-nested specification tests of time-series investment models, *Journal of Econometrics,* 37, 293-326.

Box, G. E. P. and D. A. Pierce (1970). Distribution of the autocorrelations in autoregressive moving average time series models, *Journal of 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.

Cameron, A. C. and P. K. Trivedi (2005). *Microeconometrics: Methods and Applications*, Cambridge University Press, New York.

Chen, Y.-T. (2008). A unified approach to standardized-residuals-based correlation tests for GARCH-type models, *Journal of Applied Econometrics*, 23, 111-133.

Chesher, A. and R. Smith (1997). Likelihood ratio specification tests, *Econometrica*, 65, 627–646.

Choi, H. S. and N. M. Kiefer (2008). Robust nonnested testing and the demand for money, *Journal of Business & Economic Statistics* 26*, 9-17.*

Cox, D.R. (1961). Tests of separate families of hypotheses, in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1. Berkeley; University of California Press, 105-123.

Davidson, R. and J. G. MacKinnon (1981). Several tests for model misspecification in the presence of alternative hypotheses, *Econometrica* 49, 781-793.

Durbin, J. (1970). Testing for serial correlation in least squares regression when some of the regressors are lagged dependent variables, *Econometrica*, 38, 410-421.

Elyasiani, E. and A. Nasseh (1994). The appropriate scale variable in the U.S. money demand: an application of nonnested tests of consumption versus income measures, *Journal of Business & Economic Statistics,* 12, 47–55.

Foley, R.D. and D. Goldman (1999). Confidence intervals using orthonormally weighted standardized time series. *ACM Transactions on Modeling and Computer Simulation*, 19, 297-325.

Glosten, L. R., R. Jagannathan, and D. E. Runkle (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks, *Journal of Finance*, 5, 1779–1801.

Hall P. and C.C. Heyde (1980). Martingale limit theory and its application, Academic Press.

Hansen, B.E. (1994). Autoregressive conditional density estimation, *International Economic Review*, 35, 705-730.

Hausman, J.A. (1978). Specification tests in econometrics, *Econometrica*, 46 (6), 1251–1271.

Hentschel, L. (1995). All in the family nesting symmetric and asymmetric GARCH models, *Journal of Financial Economics*, 39, 71–104.

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.

Koenker, R and G. Bassett (1978). Regression quantiles, *Econometrica,* 46, 33–50.

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

Li, W. K. and T. K. Mak (1994). On the squared residual autocorrelations in nonlinear time series with conditional heteroskedasticity, *Journal of Time Series Analysis*, 15, 627–636.

Ljung, G.M. and G.E.P. Box (1978). On a measure of a 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.

Lundbergh S. and T. Teräsvirta (2002). Evaluating GARCH models, *Journal of Econometrics*, 110, 417–435.

Mankiw, N.G. and L. H. Summers (1986). Money demand and the effects of fiscal policies, *Journal of Money, Credit and Banking,* 18, 415–429.

McLeod, A. I. and W. K. Li (1983). Diagnostic checking ARMA time series models using squared-residual autocorrelations, *Journal of Time Series Analysis*, 4, 269–273.

Nelson, D. (1991). Conditional heteroskedasticity in asset returns: a new approach, *Econometrica*, 59, 347–370.

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

Newey, W. K. and D. L. McFadden (1994). Large sample estimation and hypothesis testing. In: R. F. Engle and D. L. McFadden (Eds), *Handbook of Econometrics vol. IV*, Amsterdam: Elsevier.

Pagan, A. and F. Vella (1989). Diagnostic tests for models based on individual data: a survey, *Journal of Applied Econometrics*, 4, S29–S59.

Shao, X.F. (2010). A self-normalized approach to confidence interval construction in time series, *Journal of the Royal Statistical Society, Series, B.* 72, 343-366.

Sun, Y. (2011). Robust trend inference with series variance estimator and testing-optimal smoothing parameter, *Journal of Econometrics*, 164, 345-366.

Sun. Y. and M.S. Kim (2011). Simple and powerful GMM over-identification tests with accurate size, forthcoming in *Journal of Econometrics*.

Sun, Y., P.C.B. Phillips and S. Jin (2008). Optimal bandwidth selection in heteroskedasticity-autocorrelation robust testing, *Econometrica*, 76, 175-194.

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

Qu, Z. (2008). Testing for structural change in regression quantiles, *Journal of Econometrics*, 146, 170-184.

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

——— (1984). Comment on "Tests of specification in econometrics," *Econometric Reviews*, 3, 261–267.

——— (1994). *Estimation, Inference and Specification Analysis*, New York: Cambridge University Press.

Wong, H. and Ling, S. (2005). Mixed portmanteau tests for time series, *Journal of Time Series Analysis* 26, 569-579

Wooldridge, J. M. (1990). A unified approach to robust, regression-based specification tests, *Econometric Theory*, 6, 17–43.

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

Appendix

Proof of Lemma 3. Let $r \in [\varepsilon, 1]$ with ε being an arbitrary small positive number. Using the expansion of $T^{-1/2}\sum_{t=1}^{[Tr]} s_t(\hat{\theta}_{[Tr]})$ in Assumption 4 and the definition of $\hat{\theta}_{[Tr]}$:

$$T^{1/2}(\hat{\theta}_{[Tr]}-\theta_o) = \frac{1}{r} J_{so}^{-1} T^{-1/2} \sum_{t=1}^{[Tr]} s_t(\theta_o) + o_p(1).$$

Combining the above result with the expansion for $T^{-1/2}\sum_{t=1}^{[Tr]} m_t(\hat{\theta}_{[Tr]})$, we have

$$T^{-1/2}\sum_{t=1}^{[Tr]} m_t(\hat{\theta}_{[Tr]}) = T^{-1/2}\sum_{t=1}^{[Tr]} m_t(\theta_o) - J_{mo}J_{so}^{-1}T^{-1/2}\sum_{t=1}^{[Tr]} s_t(\theta_o) + o_p(1), \tag{A.1}$$

where the $o_p(1)$ is uniform in $r \in [\varepsilon, 1]$. This implies

$$\psi_{[Tr]} = T^{-1/2}\sum_{t=1}^{[Tr]} \xi_t(\theta_o) - rT^{-1/2}\sum_{t=1}^{T} \xi_t(\theta_o) + o_p(1) \Rightarrow V_o^{1/2}\left(W_p(r) - rW_p(1)\right).$$

Consequently, $\Psi_T(K) \to^d N(0, C \otimes V_o)$, where the matrix C is defined in Step 2 of the proposed procedure. Therefore,

$$\Psi_T^*(K) \to^d N\left(0, (GCG^\top \otimes V_o)\right) = N\left(0, (I_K \otimes V_o)\right),$$

where we have used $(G \otimes I_p)(C \otimes V_o)(G^\top \otimes I_p) = GCG^\top \otimes V_o$ and the relationship $C = G^{-1}(G^\top)^{-1}$. Thus, (10) holds. (11) holds because of the continuous mapping theorem.

The asymptotic covariance between $\psi_{[Tr]}$ and $T^{-1/2}\sum_{t=1}^{T} m_t(\hat{\theta}_T)$ is given by

$$V_o^{1/2}\mathbb{E}\left[\left(W_p(r) - rW_p(1)\right)W_p(1)\right]\left(V_o^{1/2}\right)^\top = 0,$$

implying the elements of $\Psi_T^*(K)$ are asymptotically uncorrelated with $T^{-1/2}\sum_{t=1}^{T} m_t(\hat{\theta}_T)$. Because they are asymptotically normally distributed, this result also implies asymptotic independence. This completes the proof.

Proof of Proposition 1. Consider (9). Lemma 1 implies

$$T^{-1/2}\sum_{t=1}^{T} m_t(\hat{\theta}_T) \to^d \begin{cases} N(0, V_o) & \text{under } H_o, \\ N(V_o^{-1/2}A_o\delta, V_o) & \text{under } H_{1T}. \end{cases}$$

Lemma 3 states that $R_T(K)$ converges to $W(V_o/K, K)$ with V_o being positive definite. Thus, $R_T(K)^{-1}$ has positive eigenvalues with probability close to 1 for large T and the continuous mapping theorem is applicable. The result then follows by applying Lemma 4 with $\mathcal{W}(\omega)$ and $K^{-1}\sum_{j=1}^{K}\mathcal{Z}_j\mathcal{Z}_j^\top$ replaced by the limits of $T^{-1/2}\sum_{t=1}^{T} m_t(\hat{\theta}_T)$ and $R_T(K)$.

The proof of Proposition 2 requires some intermediate results, given in Lemma A.1-Lemma A.3**.**

Lemma A.1 *Let $(T_1, ..., T_K)$ denote a partition of the sample with $T_j = jb_T$ $(j = 1, 2, ..., K)$, then under H_o, we have*

$$\sup_{1\le j\le K}\left\| T^{-1/2}\sum_{t=1}^{T_j}\{m_t(\hat{\theta}_{T_j}) - \xi_t(\theta_o)\}\right\| = o_p(1),$$

and under H_{1T} we have

$$\sup_{1\le j\le K}\left\|T^{-1/2}\sum_{t=1}^{T_j}\{m_t(\hat\theta_{T_j})-\xi_t(\theta_o)-A_o\delta\}\right\|=o_p(1).$$

Proof of Lemma A.1. The result follows from the same argument as in the proof of Lemma 3.

To obtain some further insight, we now also prove the lemma under the alternative assumptions stated below Assumption 4B. By the mean value theorem,

$$\begin{aligned}
T_j^{-1/2}\sum_{t=1}^{T_j}s_t(\hat\theta_{T_j}) &= T_j^{-1/2}\sum_{t=1}^{T_j}s_t(\theta_o)+T_j^{-1}\sum_{t=1}^{T_j}\nabla_{\theta^\top}s_t(\theta^*_{T_j})T_j^{1/2}(\hat\theta_{T_j}-\theta_o),\\
T^{-1/2}\sum_{t=1}^{T_j}m_t(\hat\theta_{T_j}) &= T^{-1/2}\sum_{t=1}^{T_j}m_t(\theta_o)+T^{-1/2}\sum_{t=1}^{T_j}\nabla_{\theta^\top}m_t(\theta^{**}_{T_j})(\hat\theta_{T_j}-\theta_o),
\end{aligned}\tag{A.2}$$

where $\theta^*_{T_j}$ and $\theta^{**}_{T_j}$ lie between $\hat\theta_{T_j}$ and θ_o. Because $l^{-1}\sum_{t=1}^{l}\nabla_{\theta^\top}s_t(\theta^*_l)$ converges almost surely to a positive definite matrix, it is nonsingular when l is large. Thus

$$T_j^{1/2}(\hat\theta_{T_j}-\theta_o)=-\left(T_j^{-1}\sum_{t=1}^{T_j}\nabla_{\theta^\top}s_t(\theta^*_{T_j})\right)^{-1}T_j^{-1/2}\sum_{t=1}^{T_j}s_t(\theta_o).$$

Substituting into (A.2) and re-arranging terms, we have

$$\begin{aligned}
&\sup_{1\le j\le K}\left\|T^{-1/2}\sum_{t=1}^{T_j}m_t(\hat\theta_{T_j})-T^{-1/2}\sum_{t=1}^{T_j}\xi_t(\theta_o)\right\|\\
\le\ &\sup_{1\le j\le K}\left\|\left(T_j^{-1}\sum_{t=1}^{T_j}\nabla_{\theta^\top}m_t(\theta^{**}_{T_j})\right)\left(T_j^{-1}\sum_{t=1}^{T_j}\nabla_{\theta^\top}s_t(\theta^*_{T_j})\right)^{-1}-J_{mo}J_{so}^{-1}\right\|\\
&\times\sup_{1\le j\le K}\left\|T^{-1/2}\sum_{t=1}^{T_j}s_t(\theta_o)\right\|.
\end{aligned}\tag{A.3}$$

Applying Assumption 4B(i), we have $T_j^{-1}\sum_{t=1}^{T_j}\nabla_{\theta^\top}m_t(\theta^{**}_{T_j})\to^{a.s.}J_{mo}$ and $\left(T_j^{-1}\sum_{t=1}^{T_j}\nabla_{\theta^\top}s_t(\theta^*_{T_j})\right)^{-1}\to^{a.s.}J_{so}^{-1}$. Thus, the first term in (A.3) is $o_{a.s}(1)$. By the functional central limit theorem, $\sup_{1\le j\le K}\left\|T^{-1/2}\sum_{t=1}^{T_j}s_t(\theta_o)\right\|=O_p(1)$. Thus, (A.3) is $o_p(1)$, which establishes the first result in the Lemma. The second result can be proved similarly and the detail is omitted.

Lemma A.2 *Let $(T_1,...,T_K)$ denote a partition of the sample with $T_j=jb_T$ $(j=1,2,...,K)$, then*

$$\sup_{1\le j\le K}\left\|\psi_{T_j}-\left(B_j^\top\otimes I_p\right)Z\right\|=o_p(1),$$

where $B_j^\top$ is a (K+1) row vector with first j elements being $(1-jb_T/T)$ and rest being $-jb_T/T$, and Z is defined in Proposition 2.

Remark 1 *Lemma A.2 implies* $\|\Psi_T(K) - (B^\top \otimes I_p)Z\|_\infty = o_p(1)$, *where* $B^\top$ *is a K by (K+1) matrix with the jth row being* $B_j^\top$ *and* $\|.\|_\infty$ *is the supremum norm.*

Proof of Lemma A.2. We have

$$\psi_{T_j} = T^{-1/2}\sum_{t=1}^{T_j}\left(m_t(\hat{\theta}_{T_j}) - T^{-1}\sum_{t=1}^{T} m_t(\hat{\theta}_T)\right) = T^{-1/2}\sum_{t=1}^{T_j}\xi_t(\theta_o) - \left(\frac{T_j}{T}\right)T^{-1/2}\sum_{t=1}^{T}\xi_t(\theta_o) + o_p(1),$$

where the second equality uses Lemma A.1, and the $o_p(1)$ is uniform in $1 \le j \le K$. The first two terms on the right hand side can be rewritten as

$$\left(1 - \frac{jb_T}{T}\right)\sum_{i=1}^{j}\left(T^{-1/2}\sum_{t=(i-1)b_T+1}^{ib_T}\xi_t(\theta_o)\right) - \frac{jb_T}{T}\sum_{i=j+1}^{K+1}\left(T^{-1/2}\sum_{t=(i-1)b_T+1}^{ib_T}\xi_t(\theta_o)\right) = \left(B_j^\top \otimes I_p\right)Z.$$

Lemma A.3 $\frac{2}{K}\sum_{j=2}^{K+1} Z_j \sum_{s=1}^{j-1}(Z_s^\top) \to^p 0$ *if* $b_T \to \infty$ *but* $b_T/T \to 0$.

Proof of Lemma A.3. For notational simplicity, suppose $p = 1$, let $L_j = Z_j\left(\sum_{s=1}^{j-1} Z_s\right)$ and write $\xi_t(\theta_o)$ as ξ_t. We will establish mean-square convergence, i.e., showing

$$\mathbb{E}\left(\frac{1}{K^2}\sum_i\sum_j L_i L_j\right) = o(1).$$

The proof relies on the following inequality for an α-*mixing* sequence $\{X_t\}$ due to Hall and Heyde (1980,p.278):

$$|\mathbb{E}X_tX_{t-j} - \mathbb{E}X_t\mathbb{E}X_{t-j}| \le 8\left(\mathbb{E}|X_t|^p\right)^{1/p}\left(\mathbb{E}|X_{t-j}|^q\right)^{1/q}\alpha(j)^{1-p^{-1}-q^{-1}},$$

where $p, q > 1$, $p^{-1} + q^{-1} < 1$, $(\mathbb{E}|X_t|^p) < \infty$, $\mathbb{E}|X_{t-j}|^q < \infty$ and $\alpha(j)$ are the mixing numbers.

Applying this inequality, we have, for some $\nu > 1$,

$$|\mathbb{E}L_j| \le b_T T^{-1} 8\left(\mathbb{E}\left|\xi_t\right|^{2\nu}\right)^{1/\nu}\sum_{j=1}^{\infty}\alpha(j)^{(\nu-1)/\nu} = O(b_T T^{-1}).$$

Suppose $h > 0$ and consider

$$\begin{aligned}
& \mathbb{E}\left(L_jL_{j-h}\right) \\
= \;& \mathbb{E}(Z_j\left(\left(Z_{j-1} + ... + Z_{j-h}\right) + \left(Z_{j-h-1}... + Z_1\right)\right)Z_{j-h}\left(Z_{j-h-1} + ... + Z_1\right)) \\
= \;& \mathbb{E}(Z_jZ_{j-h}Z_{j-h}\left(Z_{j-h-1} + ... + Z_1\right))) \qquad \text{(I)} \\
& +\mathbb{E}(Z_j(Z_{j-1} + ... + Z_{j-h+1})Z_{j-h}\left(Z_{j-h-1} + ... + Z_1\right) \qquad \text{(II)} \\
& +\mathbb{E}(Z_j\left(Z_{j-h-1} + ... + Z_1\right)Z_{j-h}\left(Z_{j-h-1} + ... + Z_1\right)). \qquad \text{(III)}
\end{aligned}$$

Term (I) equals to

$$\begin{aligned}
& T^{-2}\mathbb{E}\left(\sum_{t=(j-1)b_T+1}^{jb_T}\xi_t\right)\left(\sum_{s=(j-h-1)b_T+1}^{(j-h)b_T}\xi_s\right)^2\left(\sum_{l=1}^{(j-h-1)b_T}\xi_l\right) \\
= \;& T^{-2}\left(\sum_{t=(j-1)b_T+1}^{jb_T}\sum_{s_1=(j-h-1)b_T+1}^{(j-h)b_T}\sum_{l=1}^{(j-h-1)b_T}\mathbb{E}\left(\xi_t\xi_s^2\xi_l\right)\right) \\
& +2T^{-2}\left(\sum_{t=(j-1)b_T+1}^{jb_T}\sum_{s_1=(j-h-1)b_T+1}^{(j-h)b_T}\sum_{s_2=(j-h-1)b_T+1}^{s_1-1}\sum_{l=1}^{(j-h-1)b_T}\mathbb{E}\left(\xi_t\xi_{s_1}\xi_{s_2}\xi_l\right)\right) \\
= \;& (a) + (b).
\end{aligned}$$

Applying the mixing inequality,

$$\mathbb{E}\left((\xi_t\xi_s^2)\xi_l\right) \leq \left(\mathbb{E}\left|\xi_t\right|^{4\nu}\right)^{1/\nu} \alpha((s-l))^{(\nu-1)/\nu}.$$

Thus, term (a) $=O(b_T^2T^{-2}) = o(1)$. For term (b), note that

$$\begin{aligned}
&\left|\mathbb{E}\left(\xi_t\xi_{s_1}\xi_{s_2}\xi_l\right)\right| \\
\leq\ & 8\left(\mathbb{E}\left|\xi_t\right|^{4\nu}\right)^{1/\nu} \alpha((s_2-s_1)\vee(s_2-l))^{(\nu-1)/\nu} + \left|\mathbb{E}\left(\xi_t\xi_{s_1}\right)\mathbb{E}(\xi_{s_2}\xi_l)\right| \\
\leq\ & 8\left(\mathbb{E}\left|\xi_t\right|^{4\nu}\right)^{1/\nu} \alpha((s_2-s_1)\vee(s_2-l))^{(\nu-1)/\nu} + 8\left(\mathbb{E}\left|\xi_t\right|^{4\nu}\right)^{1/\nu} \alpha(t-s_1)^{(\nu-1)/\nu}\alpha(s_2-l)^{(\nu-1)/\nu}.
\end{aligned}$$

After some tedious algebra,

$$(b) \leq 8T^{-2}b_T^2\left(\mathbb{E}\left|\xi_t\right|^{4\nu}\right)^{1/\nu}\sum_{j=1}^{\infty} j\alpha(j) + O(T^{-2}b_T^2) = O(T^{-2}b_T^2)$$

Term (II) can be analyzed similarly, by noting that

$$(II) = T^{-2}\sum_{t=(j-1)b_T+1}^{jb_T}\sum_{s=(j-h)b_T+1}^{(j-1)b_T}\sum_{u=(j-h-1)b_T+1}^{(j-h)b_T}\sum_{l=1}^{(j-h-1)b_T}\mathbb{E}\left(\xi_t\xi_s\xi_u\xi_l\right)$$

and that

$$\mathbb{E}\left(\xi_t\xi_s\xi_u\xi_l\right) \leq 8\left(\mathbb{E}\left|\xi_t\right|^{4\nu}\right)^{1/\nu}\alpha((s-u)\vee(u-l))^{(\nu-1)/\nu} + \left|\mathbb{E}\left(\xi_t\xi_s\right)\mathbb{E}(\xi_u\xi_l)\right|.$$

Consequently,

$$(II) \leq 8\left(\mathbb{E}\left|\xi_t\right|^{4\nu}\right)^{1/\nu}T^{-2}b_T^2\sum_{j=1}^{\infty} j^2\alpha(j)^{(\nu-1)/\nu} + O(T^{-2}b_T^2) = O(T^{-2}b_T^2)$$

Term (III) equals to

$$\begin{aligned}
&\mathbb{E}(Z_jZ_{j-h}\left(Z_{j-h-1}+...+Z_1\right)^2) \\
=\ & T^{-2}\mathbb{E}\left(\sum_{t=(j-1)b_T+1}^{jb_T}\xi_t\right)\left(\sum_{s=(j-h-1)b_T+1}^{(j-h)b_T}\xi_s\right)\left(\sum_{h=1}^{(j-h-1)b_T}\xi_h\right)^2 \\
=\ & T^{-2}\mathbb{E}\left(\sum_{t=(j-1)b_T+1}^{jb_T}\sum_{s=(j-h-1)b_T+1}^{(j-h)b_T}\sum_{h=1}^{(j-h-1)b_T}\xi_t\xi_s\xi_h^2\right) \\
& +2T^{-2}\mathbb{E}\left(\sum_{t=(j-1)b_T+1}^{jb_T}\sum_{s=(j-h-1)b_T+1}^{(j-h)b_T}\sum_{h=1}^{(j-h-1)b_T}\sum_{l<h}\xi_t\xi_s\xi_h\xi_l\right) \\
=\ & (c)+(d)
\end{aligned}$$

Terms (c) can be analyzed in the same way as term (a), and is of order $O(T^{-1}b_T)$. Term (d) can be analyzed in the same way as term (b). Then, for $b_T \to \infty$,

$$(d) \leq 8\left(E(\xi_t)^{4\nu}\right)^{1/\nu}b_T^2T^{-2}\sum_{j=1}^{\infty} j^2\alpha(j)^{(\nu-1)/\nu} + O(T^{-2}b_T^2) = O(T^{-2}b_T^2).$$

Thus, term $(III) = O(T^{-1}b_T)$. Hence, $\mathbb{E}(L_j L_{j-h}) = O(T^{-1}b_T) \to 0$. Now consider the variance

$$\begin{aligned}\mathbb{E}\left(L_j^2\right) &= \mathbb{E}\sum_{t=(j-1)b_T+1}^{jb_T}\sum_{s=1}^{(j-1)b_T}\xi_t^2\xi_s^2 + 2\mathbb{E}\sum_{t=(j-1)b_T+1}^{jb_T}\sum_{s=1}^{(j-1)b_T}\sum_{l<s}\xi_t^2\xi_s\xi_l \\ &\quad +2\mathbb{E}\sum_{t=(j-1)b_T+1}^{jb_T}\sum_{s=(j-1)b_T+1}^{t-1}\xi_t\xi_s\left(Z_{j-1}+...+Z_1\right)^2 \\ &= O(T^{-1}b_T) = o(1)\end{aligned}$$

The desired result follows immediately upon taking summations.

Proof of Proposition 2. First, note that

$$C^{-1} = (K+1)\begin{bmatrix} 2 & -1 & 0 & \dots & 0 \\ -1 & 2 & -1 & \dots & 0 \\ 0 & -1 & 2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & -1 & 2 \end{bmatrix},$$

which is a K by K matrix with only the diagonal and first off-diagonal elements being nonzero. Let G be the upper triangular Cholesky factorization of C^{-1}, then

$$G = (K+1)^{1/2}\begin{bmatrix} \sqrt{\frac{2}{1}} & -\sqrt{\frac{1}{2}} & 0 & ... & 0 \\ 0 & \sqrt{\frac{3}{2}} & -\sqrt{\frac{2}{3}} & ... & 0 \\ 0 & 0 & \sqrt{\frac{4}{3}} & ... & 0 \\ \vdots & \vdots & \ddots & ... & ... \\ 0 & 0 & 0 & 0 & \sqrt{\frac{K+1}{K}} \end{bmatrix},$$

The diagonal elements are $\sqrt{(i+1)/i}$ and the first upper-diagonal elements are $-\sqrt{i/(i+1)}$ for $i = 1, ..., K$.

By the Remark following Lemma **A.2**, we can write $\Psi_T(K) = (B^\top \otimes I_p) Z + \mathbf{i} * o_p(1)$, where $\mathbf{i}$ is a vector of ones. Thus,

$$\begin{aligned}\Psi_T^* &= (G \otimes I_p)\Psi_T(K) = (G \otimes I_p)\left(\left(B^\top \otimes I_p\right) Z + \mathbf{i} * o_p(1)\right) \\ &= (GB^\top \otimes I_p)Z + \mathbf{i} * o_p(1)\end{aligned}$$

Consider the leading term $(GB^\top \otimes I_p)Z$. Note that

$$GB^\top = (K+1)^{1/2}\left(\begin{bmatrix} \sqrt{\frac{1}{1*2}} & -\sqrt{\frac{1}{1*2}} & 0 & ... & 0 \\ \sqrt{\frac{1}{2*3}} & \sqrt{\frac{1}{2*3}} & -\frac{2}{\sqrt{2*3}} & ... & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sqrt{\frac{1}{K*(K+1)}} & \sqrt{\frac{1}{K*(K+1)}} & \sqrt{\frac{1}{K*(K+1)}} & ... & -\frac{K}{\sqrt{K*(K+1)}} \end{bmatrix}\right).$$

Thus,

$$(GB^\top \otimes I_p)Z = (K+1)^{1/2} \begin{bmatrix} \sqrt{\frac{1}{1*2}} Z_1 - \sqrt{\frac{1}{1*2}} Z_2 \\ \sqrt{\frac{1}{2*3}} Z_1 + \sqrt{\frac{1}{2*3}} Z_2 - \frac{2}{\sqrt{2*3}} Z_3 \\ \vdots \\ \sqrt{\frac{1}{K*(K+1)}} \sum_{j=1}^{K} Z_j - \frac{K}{\sqrt{K*(K+1)}} Z_{K+1} \end{bmatrix}.$$

Thus, we can write $\psi^*_{T_j} = Z^*_{T_j} + \mathcal{E}_{T_j}$ with

$$Z^*_{T_j} = (K+1)^{1/2} \left\{ \sqrt{\frac{1}{j(j+1)}} \sum_{i=1}^{j} Z_i - \frac{j}{\sqrt{j(j+1)}} Z_{j+1} \right\}$$

and

$$\sup_j \left\| \mathcal{E}_{T_j} \right\| = o_p(1). \tag{A.4}$$

We now apply the above results to analyze $R_T(K)$. We have

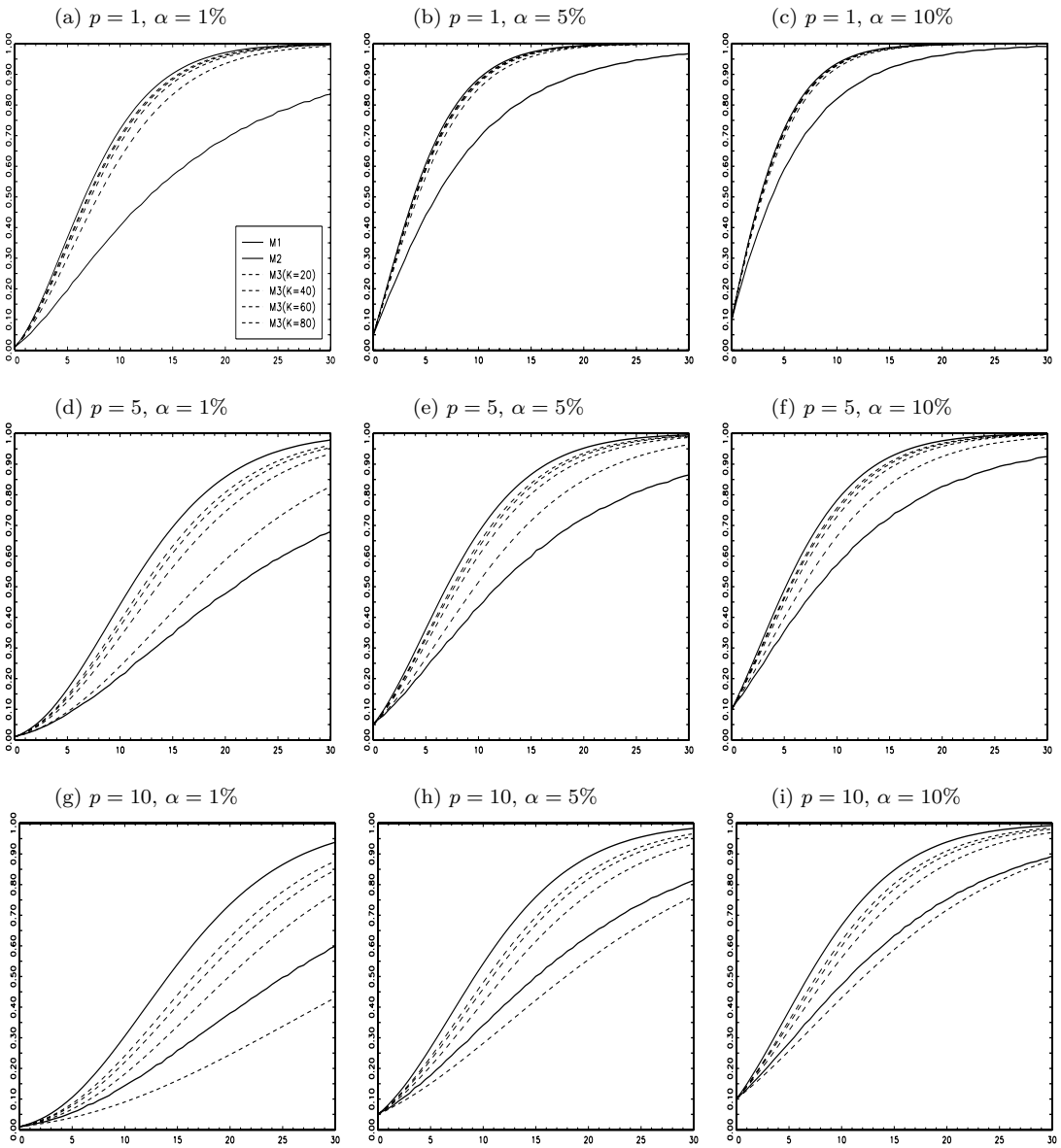
$$\begin{aligned} R_T(K) &= \frac{1}{K} \sum_{j=1}^{K} \psi^*_{T_j} \psi^{*\top}_{T_j} \\ &= \frac{1}{K} \sum_{j=1}^{K} Z^*_{T_j} Z^{*\top}_{T_j} + \frac{1}{K} \sum_{j=1}^{K} \mathcal{E}_{T_j} \mathcal{E}^\top_{T_j} + \frac{1}{K} \sum_{j=1}^{K} Z^*_{T_j} \mathcal{E}^\top_{T_j} + \frac{1}{K} \sum_{j=1}^{K} \mathcal{E}_{T_j} Z^{*\top}_{T_j} \\ &= (a) + (b) + (c) + (d) \end{aligned}$$

We study the four terms separately. Term (a) equals to

$$\sum_{j=1}^{K+1} Z_j Z_j^\top - \frac{1}{K} \sum_{j=2}^{K+1} \sum_{s=1}^{j-1} \left(Z_j Z_s^\top + Z_s Z_j^\top \right).$$

By lemma A.3, $\frac{2}{K} \sum_{j=2}^{K+1} Z_j \sum_{s=1}^{j-1} (Z_s^\top) \to^p 0$. Also, $\sum_{j=1}^{K+1} Z_j Z_j^\top \to^p V_o$. Thus, term (a)=$O_p(1)$. Term $(b) = o_p(1)$ because of (A.4). These two results imply terms (c) and (d) are of order $o_p(1)$ by Holder's inequality. Thus, $R_T(K) \to^p V_o$. This completes the proof.

Figure 1. Local asymptotic powers of the $\mathcal{M}_{1T}$, $\mathcal{M}_{2T}$ and $\mathcal{M}_{3T}$ tests



Note. In each figure, the upper and lower bold curves are, respectively, the power curves of the $\mathcal{M}_{1T}$ (i.e., the conventional) and the $\mathcal{M}_{2T}$ (i.e., KL's) test. The four dash curves from the lowest to the highest are, respectively, the power curves of the $\mathcal{M}_{3T}$ (i.e., the proposed) test with $K = 20, 40, 60$ and 80. All results are at 5% nominal level.

Table 1. Value of K for a given difference in the asymptotic power

	p									
	1	2	3	4	5	6	7	8	9	10
$\kappa = 0.05$	18	30	40	51	60	70	79	88	97	106
$\kappa = 0.02$	43	72	98	122	145	168	191	213	235	256
$\kappa = 0.01$	85	142	193	241	287	332	376	420	463	506

Note. p is the number of restrictions being tested. κ denotes the maximum power difference.

Table 2. Empirical size of the tests for GARCH(1,1) models

Panel (a). DGP-EXG										
	$\mathcal{M}_{3T}(K$=20)		$\mathcal{M}_{3T}(K$=40)		$\mathcal{M}_{3T}(K^*)$		$\mathcal{M}_{2T}$		$\mathcal{M}_{1T}$	
$h\backslash T$	500	1000	500	1000	500	1000	500	1000	500	1000
1	6.0	5.9	5.8	5.8	5.4	5.7	5.5	4.7	5.2	5.4
2	6.6	5.9	7.2	6.3	6.6	6.6	5.8	5.2	5.4	5.3
3	7.2	6.7	8.2	6.7	8.0	7.1	5.7	5.2	5.6	5.5
4	8.2	6.9	8.5	7.4	8.3	7.1	6.5	5.8	5.5	5.5
Panel (b). DGP-IND										
	$\mathcal{M}_{3T}(K$=20)		$\mathcal{M}_{3T}(K$=40)		$\mathcal{M}_{3T}(K^*)$		$\mathcal{M}_{2T}$		$\mathcal{M}_{1T}$	
$h\backslash T$	500	1000	500	1000	500	1000	500	1000	500	1000
1	5.7	5.4	5.4	5.4	5.4	5.6	5.4	6.7	5.6	6.1
2	7.2	6.4	6.6	7.1	7.1	6.6	5.6	6.1	6.3	6.0
3	8.1	7.4	8.1	8.0	7.8	7.3	5.7	6.2	5.8	6.1
4	8.8	8.0	9.6	8.2	8.6	8.5	6.6	6.5	5.5	5.6

Note. $\mathcal{M}_{3T}$: the proposed test. $\mathcal{M}_{2T}$: the test of KL. $\mathcal{M}_{1T}$: the conventional test. T is the sample size and h is the horizon. All results are at 5% nominal level.

Table 3. Empirical power of the tests for GARCH(1,1) models

		GARCH(2,1)		GARCH(1,2)		AR(1)-GARCH(1,1)		MA(1)-GARCH(1,1)	
	$h\backslash T$	500	1000	500	1000	500	1000	500	1000
$\mathcal{M}_{3T}$	1	6.4	12.4	37.3	62.9	35.4	62.9	34.2	63.3
(K=20)	2	12.6	20.1	29.4	53.2	27.2	48.9	28.4	49.0
	3	12.3	17.6	25.4	44.4	23.1	41.0	24.9	40.6
	4	12.6	15.7	23.3	37.4	21.2	34.5	23.8	34.7
$\mathcal{M}_{3T}$	1	5.7	12.7	37.9	65.3	36.3	66.0	38.3	67.2
(K=40)	2	12.7	22.7	31.0	59.0	29.1	55.1	31.4	56.5
	3	12.5	19.3	28.9	52.7	26.0	47.2	27.1	49.5
	4	12.8	20.0	27.0	48.2	24.3	42.2	25.9	44.5
$\mathcal{M}_{3T}$	1	5.6	11.9	37.6	63.9	36.5	65.4	37.3	65.4
(K^*)	2	13.9	23.2	32.2	61.1	29.6	55.3	31.6	57.0
	3	13.1	19.1	31.3	55.7	26.4	49.9	30.1	50.8
	4	12.0	19.2	27.0	49.8	23.6	44.3	26.5	45.1
$\mathcal{M}_{2T}$	1	5.4	10.2	26.5	46.3	25.9	47.3	27.7	49.5
	2	10.5	16.5	21.7	38.9	22.0	38.2	23.4	41.5
	3	9.7	15.8	20.5	36.6	19.1	34.3	21.1	37.4
	4	10.4	15.2	19.8	35.5	19.0	32.5	20.0	34.3
$\mathcal{M}_{1T}$	1	6.7	14.6	47.0	72.4	40.8	71.4	42.1	71.2
	2	13.0	23.3	36.3	67.1	32.4	59.1	32.0	59.9
	3	10.7	19.5	29.0	58.8	25.7	51.6	25.7	52.9
	4	10.0	18.3	24.7	52.4	22.9	46.3	22.2	47.9

Note. $\mathcal{M}_{3T}$: the proposed test. $\mathcal{M}_{2T}$: the test of KL. $\mathcal{M}_{1T}$: the conventional test. T is the sample size and h is the horizon. All results are at 5% nominal level.

Table 4. Parameter values used for nonnested testing

Regression coefficients and R^2		
	GNP	Consumption
Constant	0.0024	-0.0012
r_t	-0.0198	-0.0192
r_{t-1}	-0.0112	-0.0102
r_{t-2}	-0.0082	-0.0086
z_t	0.3209	0.4039
z_{t-1}	0.2058	0.4095
z_{t-2}	0.1059	0.1937
R^2	0.2954	0.4042
Residual process		
$\varepsilon_t = \rho\varepsilon_{t-1}+v_t$, with $\rho = 0.43, \sigma_v$=0.0072		

Table 5. Empirical rejection frequencies for nonnested testing

Panel (a). Empirical size					
	$\mathcal{M}_{3T}(K$=20)	$\mathcal{M}_{3T}(K$=40)	$\mathcal{M}_{3T}(K^*)$	$\mathcal{M}_{2T}$	$\mathcal{M}_{1T}$
ρ=0.43	5.6	6.1	5.5	4.7	6.2
ρ=0.90	6.5	8.1	6.1	5.9	4.9
Panel (b). Empirical power					
	$\mathcal{M}_{3T}(K$=20)	$\mathcal{M}_{3T}(K$=40)	$\mathcal{M}_{3T}(K^*)$	$\mathcal{M}_{2T}$	$\mathcal{M}_{1T}$
ρ=0.43	96.3	98.3	96.1	79.5	98.1
ρ=0.90	35.1	44.7	35.0	25.6	28.1

Note. $\mathcal{M}_{3T}$: the proposed test. $\mathcal{M}_{2T}$: the test of KL. $\mathcal{M}_{1T}$: the conventional test. All results are at 5% nominal level.

Table 6. The Data-Generating Processes for autocorrelation tests

DGPs for the null hypothesis	
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:	$y_t = 1.0 + 0.5y_{t-1} + e_t, e_t = u_t + 0.3u_{t-1}e_{t-2}$
DGPs for the alternative hypothesis	
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 iid N(0,1) and are independent of each other.

Table 7. Empirical sizes of the tests for serial correlation

		DGP1		DGP2		DGP3		DGP4		DGP5		DGP6		DGP7		DGP8	
Tests	$q\backslash T$	100	500	100	500	100	500	100	500	100	500	100	500	100	500	100	500
$\mathcal{M}_{3T}$	1	3.7	3.9	4.0	4.0	4.6	4.7	5.3	5.2	4.1	5.2	3.8	4.8	4.1	3.6	3.6	4.8
(K=20)	2	4.5	4.7	4.3	4.5	5.2	4.9	5.4	5.1	3.2	4.0	3.0	3.5	2.8	3.3	2.7	3.5
	3	4.9	5.1	4.5	4.6	5.4	5.0	5.3	5.5	3.1	3.6	3.2	3.3	3.1	3.5	3.0	3.2
	4	5.1	5.0	4.4	4.6	5.8	5.1	5.5	6.3	3.6	3.3	3.5	3.7	3.3	3.3	3.7	3.4
$\mathcal{M}_{3T}$	1	4.0	4.7	4.6	4.1	4.7	4.6	5.7	5.3	3.4	5.3	3.4	5.2	4.0	5.1	3.9	4.7
(K=40)	2	4.8	4.8	4.8	5.0	5.0	4.5	6.4	5.9	2.6	4.1	2.2	3.3	3.6	3.7	3.0	3.9
	3	4.7	4.5	4.3	4.5	4.6	4.5	6.2	6.1	2.9	3.5	2.6	3.7	4.2	4.1	3.5	3.6
	4	4.7	5.4	4.1	4.6	5.2	5.1	6.3	6.4	3.5	3.8	2.8	3.2	4.4	3.9	3.3	3.6
$\mathcal{M}_{3T}$	1	4.0	4.4	4.1	4.1	4.3	4.8	4.8	5.0	4.5	5.1	3.4	5.0	3.9	4.1	4.5	5.2
(K^*)	2	4.4	4.5	4.1	4.5	4.4	4.8	5.9	4.8	4.3	5.0	4.5	4.8	4.4	4.7	4.8	5.3
	3	4.7	4.7	4.3	4.4	5.4	4.9	5.8	5.6	3.2	3.4	3.2	3.7	4.1	3.6	3.5	3.5
	4	4.7	5.5	3.5	4.4	5.2	5.3	5.4	6.3	3.3	3.6	3.0	3.9	4.3	4.3	3.3	3.3
$\mathcal{M}_{2T}$	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
$\mathcal{M}_{1T}$	1	5.1	5.3	5.5	4.6	5.2	5.0	7.3	6.5	8.1	5.6	10.1	6.4	7.5	5.5	8.4	5.4
	2	4.6	5.4	5.2	3.9	5.2	4.8	6.8	6.1	15.6	5.9	19.5	7.0	14.7	6.6	16.4	6.7
	3	4.8	5.0	4.3	4.1	4.5	4.7	6.5	6.6	31.4	9.5	33.6	10.8	29.3	9.9	31.6	9.4
	4	4.8	5.2	4.3	4.3	4.0	4.2	5.8	6.1	49.1	18.2	49.0	20.6	48.3	19.1	49.7	18.6

Note. $\mathcal{M}_{3T}$: the proposed test. $\mathcal{M}_{2T}$: the test of KL. $\mathcal{M}_{1T}$: the conventional test. T is the sample size and q is the number of restrictions being tested. All results are at 5% nominal level.

Table 8. Empirical power of the tests for serial correlation

		DGP9		DGP10		DGP11		DGP12		DGP13		DGP14		DGP15	
Tests	$q\backslash T$	100	500	100	500	100	500	100	500	100	500	100	500	100	500
$\mathcal{M}_{3T}$	1	96.9	100	21.6	77.8	88.0	100	83.8	100	22.1	68.1	35.7	92.6	30.4	82.7
(K=20)	2	88.9	100	15.3	66.0	83.9	100	82.6	100	14.6	55.5	24.7	85.3	20.4	72.5
	3	78.2	100	13.1	56.5	77.0	100	74.2	100	11.8	45.7	19.8	79.3	16.1	63.3
	4	66.2	100	11.6	48.9	68.8	100	65.5	100	9.98	39.8	16.3	71.8	13.2	55.5
$\mathcal{M}_{3T}$	1	98.8	100	24.1	80.1	90.0	100	86.5	100	22.2	72.0	38.6	94.0	34.4	84.3
(K=40)	2	95.8	100	18.0	70.6	89.7	100	87.3	100	15.2	61.0	28.0	89.4	23.6	76.8
	3	90.1	100	14.6	62.0	83.4	100	81.5	100	11.6	52.3	22.3	83.8	19.0	69.6
	4	82.1	100	11.8	55.3	78.5	100	75.0	100	10.0	46.0	18.0	78.6	14.9	64.2
$\mathcal{M}_{3T}$	1	96.6	100	23.8	76.2	87.9	100	82.6	100	21.3	69.1	35.7	92.6	29.8	81.5
(K^*)	2	93.4	100	17.6	67.5	89.0	100	87.3	100	15.8	59.4	26.6	87.2	22.2	74.7
	3	89.2	100	14.5	61.1	83.9	100	80.9	100	12.7	53.2	22.4	83.4	19.3	71.2
	4	87.5	100	12.0	56.8	78.9	100	75.5	100	10.8	47.3	17.5	80.3	14.5	67.1
$\mathcal{M}_{2T}$	1	76.3	99.7	17.1	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	8.50	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	9.40	33.1	13.2	55.9	11.9	44.7
$\mathcal{WL}$	1	89.9	100	22.2	82.6	86.8	100	82.1	100	20.3	72.9	30.1	93.3	23.1	83.8
	2	73.0	100	16.9	73.9	81.5	100	78.8	100	16.1	64.8	24.0	89.2	17.1	75.4
	3	59.8	100	12.9	63.8	76.3	100	76.7	100	11.1	57.7	16.5	85.7	15.4	72.4
	4	42.9	100	8.4	56.4	68.5	100	66.5	100	10.5	52.1	12.6	81.4	10.6	65.0

Note. $\mathcal{M}_{3T}$: the proposed test. $\mathcal{M}_{2T}$: the test of KL. $\mathcal{M}_{1T}$: the conventional test. $\mathcal{WL}$: Wooldridge's test. T is the sample size and q is the number of restrictions being tested. All results are at 5% nominal level.