On Long-Run Covariance Matrix Estimation with the Truncated Flat Kernel

Chang-Ching Lin[a], Shinichi Sakata[b*]

[a] *Institute of Economics, Academia Sinica, 128, Sec.2 Academia Rd., Nangang, Taipei 115, Taiwan, Republic of China*

[b] *Department of Economics, University of British Columbia, 997-1873 East Mall, Vancouver, BC V6T 1Z1, Canada*

March 31, 2008

Despite its large sample efficiency, the truncated flat (TF) kernel estimator of long-run covariance matrices is seldom used, because it lacks the guaranteed positive semidefiniteness and sometimes performs poorly in small samples, compared to other familiar kernel estimators. This paper proposes simple modifications to the TF estimator to enforce the positive definiteness without sacrificing the large sample efficiency and make the estimator more reliable in small samples through better utilization of the bias-variance tradeoff. We study the large sample properties of the modified TF estimators and verify their improved small-sample performances by Monte Carlo simulations.

[*]Tel: +1-604-822-5360; fax: 1+604-822-5915.

Email addresses: lincc@econ.sinica.edu.tw (C. Lin), ssakata@interchange.ubc.ca (S. Sakata).

1 Introduction

In assessing the accuracy of an extremum estimator or making a statistical inference on unknown parameter values based on an extremum estimator using time series data, it is often necessary to estimate a long-run covariance matrix. The long-run covariance matrix is typically estimated by a kernel estimator that is the weighted average of estimated autocovariance matrices with weights determined by a kernel function and the bandwidth for it.

When a series is known to have an autocovariance function truncated at or before lag m, one can simply estimate each of the autocovariances of the series up to lag m and take a suitable linear combination of the estimated autocovariances to consistently estimate the long-run covariance matrix. This approach is proposed by Hansen (1982). For the case in which the autocovariance function is not truncated, White and Domowitz (1984) show that the above-mentioned method, letting the truncation point m gradually grow to infinity as the sample size approaches infinity, consistently estimates the long-run covariance matrix. Such estimator is a kernel estimator that employs the truncated flat (TF) kernel. We call it the TF estimator in this paper.

A drawback of the TF method is that it sometimes delivers a non-positive semidefinite estimate. A way to avoid non-positive semidefinite estimates is to suitably weight the estimated autocovariances, as the commonly used estimators such as Newey and West's (1987) Bartlett (BT) kernel estimator and Andrews's (1991) Quadratic Spectral (QS) kernel estimator do. As demonstrated in Gallant and White (1988), there are many kernels that generate consistent estimators of long-run covariance matrices. Hansen (1992), de Jong and Davidson (2000), and Jansson (2003) show general conditions sufficient for the kernel estimators to be consistent.

An interesting fact pointed out in the literature on spectral density estimation (e.g., Priestley (1981) and Hannan (1970)) is that the tradeoff between the asymptotic bias and asymptotic variance in choice of the bandwidth sequence does not hold for the TF estimator in the usual way. The asymptotic bias is negligible relative to the asymptotic variance unless the growth rate of the bandwidth becomes very low. This means that use of a slowly growing bandwidth can make the variance of the TF estimator converge fast, still keeping the bias negligible. On the other hand, the other familiar kernel estimators including the BT and QS estimators have upper bounds for the convergence rate of the MSE determined through the tradeoff

between the asymptotic bias and asymptotic variance. It follows that the TF estimator is asymptotically efficient relative to the other familiar kernel estimators in typical scenarios.

Nevertheless, the small-sample behavior of the TF estimator is not necessarily consistent with the large-sample theory. Andrews (1991) conducts Monte Carlo simulations to assess the performances of the TF, QS, BT, and a few other kernel estimators and finds that the TF estimator performs substantially better or worse than the other estimators, depending on the experiment setup.

In this paper, we consider minor modifications to the TF estimator to always deliver a positive semidefinite (p.s.d.) estimate and resolve the puzzling discrepancy between the large sample efficiency and small sample behavior of the TF estimator. The first contribution of the paper is to propose a simple way to modify the TF estimator to enforce the positive semidefiniteness of the estimate, without sacrificing the large sample efficiency of the estimator. Our method pushes the non-p.s.d. TF estimate back to the space of symmetric p.s.d. matrices in a particular way. The resulting estimator, the adjusted TF (ATF) estimator, is guaranteed to have a smaller mean square error (MSE) than the TF estimator. The ATF estimator enjoys the same large sample efficiency in typical scenarios as the TF estimator does, because the probability of the adjustment converges to zero.

The second contribution of the paper is concerned with the puzzling small sample behavior of the TF estimator mentioned above. In the TF (and ATF) estimation, a change in the bandwidth affects the estimate only when crossing an integer value. Compared with the other familiar estimators that are continuously related to their bandwidths, this feature of the TF estimator severely limits the opportunity to balance the bias and variance of the estimator to attain a smaller MSE. To eliminate this limitation, we propose allowing the TF estimator to include the estimated autocovariance matrix at the last lag with a fractional weight. Because the resulting estimator, which we call the TFF estimator, has the same problem of possible non-positive semidefiniteness as the TF estimator does, we further propose its adjusted version, the ATFF estimator. The TFF and ATFF estimators again enjoy the same large sample efficiency as the TF estimator.

Our Monte Carlo simulations verify that the MSE of ATF estimator is only slightly smaller than the TF estimator in most cases, though the difference of the two estimators is more pronounced when the TF estimator is non-p.s.d. with a high probability. The simulations also demonstrate that the relationship between the ATFF and QS estimators in small samples is in line with the large sample theory, unlike that between the TF and QS estimators. The MSE of the ATFF estimator, being often substantially smaller

than that of the ATF estimator, is smaller than or comparable to that of the QS estimator in all of our experiments.

While this paper focuses on consistent estimation of long-run covariance matrices using the kernel method, some other ways to estimate long-run covariance matrices have been considered in the literature. A possible approach is a series approximation approach that fits a vector autoregression (VAR) model to the series and computes the long-run autocovariance matrix implied by the fitted VAR model, described by den Haan and Levin (1997). This method consistently estimates long-run covariance matrices if the lag order of the VAR model gradually grows to infinity when the sample size approaches infinity. It is also possible to combine the series approximation approach and the kernel approach. Fitting a VAR model to the series is "expected" to yield a less persistent residual series. Hence, applying the kernel method after this "prewhitening" may enjoy the advantage of both approaches. Andrews and Monahan (1992) demonstrate that this hybrid approach is highly effective. den Haan and Levin (2000), however, raise some concerns about the performance of the hybrid approach.

Unlike the approaches described above, Kiefer and Vogelsang (2000) and Bunzel, Kiefer, and Vogelsang (2001) consider a way to obtain an asymptotically pivotal test statistic without consistently estimating the asymptotic covariance matrix of the estimator. Further, Kiefer and Vogelsang (2002) show that the approach taken in Kiefer and Vogelsang (2000) and Bunzel, Kiefer, and Vogelsang (2001) is equivalent to use of the BT estimator with the bandwidth set equal to the sample size. While these works are very important, they are beyond the scope of this paper, because they do not pursue consistent estimation of long-run autocovariance matrices.

The rest of the paper is organized as follows. We first propose a way to adjust a long-run covariance matrix estimator for positive semidefiniteness and discuss the basic properties of the proposed adjustment (Section 2). We then describe a method to compute the adjusted estimator (Section 3). Next, we apply the proposed adjustment to the TF estimator to yield a p.s.d. estimator that has a smaller MSE than the TF estimator and shares the same asymptotic MSE (AMSE) as the TF estimator (Section 4). We further propose the TFF estimator that incorporates the autocovariance matrix at the last lag with a fractional weight and adjust it for positive semidefiniteness to obtain the ATFF estimator (Section 5). To assess the performances of the proposed estimators relative to those of the QS and BT estimators in small samples, we conduct Monte Carlo simulations (Sections 6). We finally discuss the behavior of the TF, ATF, TFF,

4

and ATFF estimators with data-based bandwidths (Section 7) and examine the finite sample behavior of the ATF and ATFF estimators with data-based bandwidths by Monte Carlo simulations (Section 8). We collect the mathematical proofs of all theorems, propositions, and lemmas in the Appendix.

Throughout this paper, limits are taken along the sequence of sample sizes (denoted T) growing to infinity, unless otherwise indicated. If sequences of real numbers $\{a_T\}_{T \in \mathbb{N}}$ and $\{b_T\}_{T \in \mathbb{N}}$ satisfy that $a_T = O(b_T)$ and $b_T = O(a_T)$, then we write $a_T \sim b_T$. For each topological space A, $\mathscr{B}(A)$ denotes the Borel σ-field on A. For the Euclidean spaces, write $\mathscr{B}^p \equiv \mathscr{B}(\mathbb{R}^p)$ for simplicity.

2 Estimators Adjusted for Positive Semidefiniteness

We consider the situation described by the next assumption.

ASSUMPTION 1: $(\Omega, \mathscr{F}, P)$ is a probability space, and Θ a nonempty subset of $\mathbb{R}^p$ $(p \in \mathbb{N})$. The sequence $\{Z_t\}_{t \in \mathbb{N}}$ consists of measurable functions from $(\Omega \times \Theta, \mathscr{F} \otimes \mathscr{B}(\Theta))$ to $(\mathbb{R}^v, \mathscr{B}^v)$ $(v \in \mathbb{N})$ such that for each $\theta \in \Theta$ and each $t \in \mathbb{N}$, $\mathrm{E}[Z_t(\cdot, \theta)' Z_t(\cdot, \theta)] < \infty$. Also, $\{\hat{\theta}_T : \Omega \to \Theta\}_{T \in \mathbb{N}}$ is a sequence of $p \times 1$ random vectors, and $\{Z_t^* \equiv Z_t(\cdot, \theta^*)\}_{t \in \mathbb{N}}$ is a zero-mean covariance stationary process.

Our goal is to accurately estimate

$$S_T \equiv \mathrm{var}\left[T^{-1/2} \sum_{t=1}^{T} Z_t^*\right] = \Gamma_T(0) + \sum_{\tau=1}^{T-1} (\Gamma_T(\tau) + \Gamma_T'(\tau)),$$

where $T \in \mathbb{N}$ is the sample size, and each $\tau \in \{1, 2, \ldots, T-1\}$

$$\Gamma_T(\tau) \equiv \frac{T-\tau}{T} \mathrm{cov}[Z_{\tau+1}^*, Z_1^*].$$

$$\boxed{\text{Table 1 around here}}$$

Let k be an even function from $\mathbb{R}$ to $\mathbb{R}$ that is continuous at the origin and discontinuous at most at a finite number of points. Table 1 lists a few such kernels often used in the literature. Suppose that θ^* is known. Then a kernel estimator of S_T using the kernel k and a bandwidth $m_T \in (0, \infty)$ is

$$\tilde{S}_T^k \equiv k(0)\tilde{\Gamma}_T(0) + \sum_{\tau=1}^{T-1} k\left(\frac{\tau}{m_T}\right)\left(\tilde{\Gamma}_T(\tau) + \tilde{\Gamma}_T'(\tau)\right), \quad T \in \mathbb{N},$$

where

$$\tilde{\Gamma}_T(\tau) \equiv \frac{1}{T} \sum_{t=\tau+1}^{T} Z_t^* Z_{t-\tau}^{*\prime}, \quad \tau \in \{1, 2, \ldots, T-1\}, T \in \mathbb{N}. \tag{1}$$

5

When θ^* is unknown, as is the case in typical applications, we would replace the unknown θ^* with its estimator $\hat{\theta}_T$ to obtain a feasible estimator of S_T:

$$\hat{S}_T^k \equiv k(0)\hat{\Gamma}_T(0) + \sum_{\tau=1}^{T-1} k\left(\frac{\tau}{m_T}\right)\left(\hat{\Gamma}_T(\tau) + \hat{\Gamma}_T'(\tau)\right), \quad T \in \mathbb{N},$$

where

$$\hat{\Gamma}_T(\tau) \equiv \frac{1}{T} \sum_{t=\tau+1}^{T} \hat{Z}_{T,t}\hat{Z}_{T,t-\tau}', \quad \tau \in \{1, 2, \ldots, T-1\}, T \in \mathbb{N}, \tag{2}$$

and

$$\hat{Z}_{T,t}(\omega) \equiv Z_t(\omega, \hat{\theta}_T(\omega)), \quad \omega \in \Omega, t \in \{1, 2, \ldots, T\}, T \in \mathbb{N}.$$

Because S_T is a covariance matrix, it is p.s.d. A kernel estimator, on the other hand, may deliver a non-p.s.d. estimate in general. This means that the estimate may lead to a negative estimate of the variance of a statistic. This problem can be avoided by choosing certain kernels such as the BT and QS kernels. The QS kernel yields the efficient estimator among those using such kernels.

We here consider a different way to ensure the positive semidefiniteness of the estimate. Instead of limiting our choice of kernels, our approach pushes an estimate back to the space of symmetric p.s.d. matrices, whenever it is not p.s.d. On $\mathbb{R}^{a_1 \times a_2}$, where $(a_1, a_2) \in \mathbb{N}^2$, define a real valued function $\|\cdot\|_W : \mathbb{R}^{a_1 \times a_2} \to \mathbb{R}$ by

$$\|A\|_W \equiv (\text{vec}(A)' W \text{vec}(A))^{1/2}, \quad A \in \mathbb{R}^{a_1 \times a_2},$$

where W is a $(a_1 a_2) \times (a_1 a_2)$ symmetric p.s.d. matrix, and $\text{vec}(A)$ is the column vector made by stacking the columns of A vertically from the first column to the last. If W is the identity matrix, $\|\cdot\|_W$ becomes the Frobenius norm, which will be denoted $\|\cdot\|$ for simplicity.

DEFINITION 1: Suppose that Assumption 1 holds. Let P_v be the set of all $v \times v$, symmetric p.s.d. matrices. Given an estimator $\{\hat{S}_T : \Omega \to \mathbb{R}^{v \times v}\}_{T \in \mathbb{N}}$ (of $\{S_T\}_{T \in \mathbb{N}}$) and a sequence of $v^2 \times v^2$ symmetric p.s.d. random matrices $\{W_T : \Omega \to \mathbb{R}^{v^2 \times v^2}\}_{t \in \mathbb{N}}$, the sequence of $v \times v$ random matrices $\{\hat{S}_T^A : \Omega \to \mathbb{R}^{v \times v}\}_{T \in \mathbb{N}}$ satisfying that for each $T \in \mathbb{N}$,

$$\|\hat{S}_T - \hat{S}_T^A\|_{W_T} = \inf_{s \in \mathsf{P}_v} \|\hat{S}_T - s\|_{W_T}, \tag{3}$$

provided that it exists, is called *the estimator that adjusts $\{\hat{S}_T\}_{T \in \mathbb{N}}$ for positive semidefiniteness* or simply *the adjusted estimator (with weighting matrix $\{W_T\}$).*

6

The existence of the adjusted estimators can be established by using Brown and Purves (1973, Corollary 1, pp. 904–905).

THEOREM 2.1: *Suppose that Assumption 1 holds. Then for each estimator $\{\hat{S}_T : \Omega \to \mathbb{R}^{v \times v}\}_{T \in \mathbb{N}}$ and each sequence of symmetric p.s.d. random matrices $\{W_T : \Omega \to \mathbb{R}^{v^2 \times v^2}\}_{T \in \mathbb{N}}$, the estimator that adjusts $\{\hat{S}_T\}$ for positive semidefiniteness with the weighting matrix $\{W_T\}$ exists.*

Whenever $\hat{S}_T \in \mathsf{P}_v$, it apparently holds that $\|\hat{S}_T^A - \hat{S}_T\|_{W_T} = 0$. Moreover:

THEOREM 2.2: *Suppose that Assumption 1 holds. Let $\{\hat{S}_T : \Omega \to \mathbb{R}^{v \times v}\}_{T \in \mathbb{N}}$ be an estimator and $\{W_T : \Omega \to \mathbb{R}^{v^2 \times v^2}\}_{T \in \mathbb{N}}$ a sequence of symmetric p.s.d. random matrices. Then the adjusted estimator $\{\hat{S}_T^A : \Omega \to \mathbb{R}^{v \times v}\}_{T \in \mathbb{N}}$ with the weighting matrix $\{W_T\}$ satisfies:*

(a) $\|\hat{S}_T^A - S_T\| = \|\hat{S}_T - S_T\|_{W_T}$, *whenever* $\|\hat{S}_T^A - \hat{S}_T\|_{W_T} = 0$, $T \in \mathbb{N}$.

(b) $\|\hat{S}_T^A - S_T\|_{W_T} \le \|\hat{S}_T - S_T\|_{W_T}$, $T \in \mathbb{N}$.

Because Theorem 2.2(b) means that the adjustment moves the estimator towards S_T in terms of the norm $\| \cdot \|_{W_T}$, the performance of the adjusted estimator cannot be worse than the original estimator. Here are a few implications of this fact.

COROLLARY 2.3: *Suppose that Assumption 1 holds. Let $\{\hat{S}_T : \Omega \to \mathbb{R}^{v \times v}\}_{T \in \mathbb{N}}$ be an estimator and $\{W_T : \Omega \to \mathbb{R}^{v^2 \times v^2}\}_{T \in \mathbb{N}}$ a sequence of $v^2 \times v^2$ symmetric p.s.d. random matrices. Then:*

(a) *For each $T \in \mathbb{N}$,*

$$\mathrm{E}\big[\|\hat{S}_T^A - S_T\|_{W_T}^2\big] \le \mathrm{E}\big[\|\hat{S}_T - S_T\|_{W_T}^2\big].$$

(b) *If $\{\hat{S}_T\}$ is consistent for $\{S_T\}$ (i.e., $\{\|\hat{S}_T - S_T\|\}_{T \in \mathbb{N}}$ converges in probability-P to zero), and $W_T = \mathrm{O}_P(1)$, then $\|\hat{S}_T^A - S_T\|_{W_T}$ converges in probability-P to zero. If in addition $\{W_T\}$ converges in probability-P to a nonsingular matrix W, then $\{\hat{S}_T^A\}$ is consistent for $\{S_T\}$.*

(c) *If $\{\hat{S}_T\}$ is consistent for $\{S_T\}$, and $\{S_T\}$ is asymptotically uniformly positive definite (p.d.), then $\|\hat{S}_T^A - \hat{S}_T\|_{W_T} = \mathrm{o}_P(b_T)$ for each sequence of positive real numbers $\{b_T\}_{T \in \mathbb{N}}$. If in addition $\{W_T\}$ converges in probability-P to a nonsingular matrix W, then $\hat{S}_T^A - \hat{S}_T = \mathrm{o}_P(b_T)$ for each sequence of positive real numbers $\{b_T\}$.*

7

Given Corollary 2.3(c), it is natural to expect that the advantage of the adjusted estimator over the original estimator described in Corollary 2.3(a) becomes negligible in large samples, if the original estimator is a consistent estimator. To make a meaningful statement on this point, we need to suitably magnify the MSE of each of the estimators, because otherwise, the MSEs would converge to zero as $T \to \infty$ in a typical setup. Given an estimator $\hat{S}_T$ of S_T, a $v^2 \times v^2$ symmetric p.s.d. random matrices W_T, and a positive real constant a_T (magnification factor), write

$$\text{MSE}(a_T, \, \hat{S}_T, \, W_T) \equiv a_T \text{E}\big[\|\hat{S}_T - S_T\|_{W_T}^2\big]. \tag{4}$$

Using this scaled MSE, we now state the asymptotic equivalence of the adjusted and original estimators in terms of the MSE.

THEOREM 2.4: *Suppose that Assumption 1 holds. Let $\{\hat{S}_T : \Omega \to \mathbb{R}^{v \times v}\}_{T \in \mathbb{N}}$ be an estimator consistent for $\{S_T \in \mathbb{R}^{v \times v}\}_{T \in \mathbb{N}}$ that is a sequence of symmetric matrices that are asymptotically uniformly p.d. Also, let $\{W_T : \Omega \to \mathbb{R}^{v^2 \times v^2}\}_{T \in \mathbb{N}}$ be a sequence of $v^2 \times v^2$ symmetric p.s.d. random matrices. Suppose that for some sequence of positive real numbers $\{a_T\}_{T \in \mathbb{N}}$, $\{a_T \|\hat{S}_T - S_T\|_{W_T}^2\}_{T \in \mathbb{N}}$ is uniformly integrable. Then*

$$\text{MSE}(a_T, \, \hat{S}_T, \, W_T) - \text{MSE}(a_T, \, \hat{S}_T^A, \, W_T) \to 0.$$

Remark.   The uniform integrability of $\{a_T \|\hat{S}_T - S_T\|_{W_T}^2\}_{T \in \mathbb{N}}$ implies the uniform integrability of $\{a_T \|\hat{S}_T^A - S_T\|_{W_T}^2\}_{T \in \mathbb{N}}$. It follows that both $\text{MSE}(a_T, \, \hat{S}_T, \, W_T)$ and $\text{MSE}(a_T, \, \hat{S}_T^A, \, W_T)$ are finite under the conditions imposed in Theorem 2.4.

When the parameter θ^* is unknown, the effect of the parameter estimation must be taken into account in studying the behavior of a kernel estimator. Because the moments of the parameter estimator sometimes do not exist, the MSE may not be adequate for measuring the performance of the kernel estimators with parameter estimation. Following Andrews (1991), we bypass this potential problem by using the truncated MSE instead. The truncated MSE of an estimator $\hat{S}_T$ of S_T scaled by a_T and truncated at $h \in (0, \infty)$ is defined by

$$\text{MSE}_h(a_T, \hat{S}_T, W_T) \equiv \text{E}\Big[\min\big\{a_T \|\hat{S}_T - S_T\|_{W_T}^2, \, h\big\}\Big].$$

Because for each $h \in (0, \infty)$, the function $x \mapsto \min\{x, h\} : [0, \infty) \to \mathbb{R}$ is a nondecreasing function, the relationship between the adjusted and original estimators stated in Corollary 2.3(a) carries over even if we

8

replace the MSEs with the truncated MSEs.

THEOREM 2.5: *Suppose that Assumption 1 holds. Let $\{\hat{S}_T : \Omega \to \mathbb{R}^{v \times v}\}_{T \in \mathbb{N}}$ be an estimator, $\{a_T\}_{T \in \mathbb{N}}$ a sequence of positive real numbers, and $\{W_T : \Omega \to \mathbb{R}^{v^2 \times v^2}\}_{T \in \mathbb{N}}$ a sequence of $v^2 \times v^2$ symmetric p.s.d. random matrices. Then for each T, each $a_T \in (0, \infty)$, and each $h \in (0, \infty)$,*

$$\mathrm{MSE}_h(a_T, \hat{S}_T^A, W_T) \leq \mathrm{MSE}_h(a_T, \hat{S}_T, W_T).$$

When the difference between two estimators converges in probability to zero fast enough, the two estimators share the same asymptotic truncated MSE.

LEMMA 2.6: *Suppose that Assumption 1 holds. Let $\{\hat{S}_{1,T} : \Omega \to \mathbb{R}^{v \times v}\}_{T \in \mathbb{N}}$ and $\{\hat{S}_{2,T} : \Omega \to \mathbb{R}^{v \times v}\}_{T \in \mathbb{N}}$ be estimators, $\{W_T : \Omega \to \mathbb{R}^{v^2 \times v^2}\}_{T \in \mathbb{N}}$ a sequence of $v^2 \times v^2$ symmetric p.s.d. random matrices, and $\{a_T\}_{T \in \mathbb{N}}$ a sequence of positive real numbers. If $a_T^{1/2}(\hat{S}_{1,T} - S_T) = \mathrm{O}_P(1)$, and $a_T^{1/2}\|\hat{S}_{1,T} - \hat{S}_{2,T}\|_{W_T} \to 0$ in probability-P, then for each $h \in (0, \infty)$ for which $\{\mathrm{MSE}_h(a_T, \hat{S}_{1,T}, W_T)\}_{T \in \mathbb{N}}$ converges to a (finite) real number, it holds that*

$$\lim_{T \to \infty} \mathrm{MSE}_h(a_T, \hat{S}_{2,T}, W_T) = \lim_{T \to \infty} \mathrm{MSE}_h(a_T, \hat{S}_{1,T}, W_T).$$

A consistent estimator and its adjustment are negligibly different, as Corollary 2.3(c) states, if $\{S_T\}_{T \in \mathbb{N}}$ is asymptotically uniformly p.d. It follows that the asymptotic truncated MSEs of the original and adjusted estimators are the same in such situations.

THEOREM 2.7: *Suppose that Assumption 1 holds. Let $\{\hat{S}_T : \Omega \to \mathbb{R}^{v \times v}\}_{T \in \mathbb{N}}$ be an estimator consistent for $\{S_T\}_{T \in \mathbb{N}}$, $\{a_T\}_{T \in \mathbb{N}}$ a sequence of positive real numbers, and $\{W_T : \Omega \to \mathbb{R}^{v^2 \times v^2}\}$ a sequence of $v^2 \times v^2$ symmetric p.s.d. random matrices. If $a_T^{1/2}(\hat{S}_T - S_T) = \mathrm{O}_P(1)$, $\lim_{h \to \infty} \lim_{T \to \infty} \mathrm{MSE}_h(a_T, \hat{S}_T, W_T)$ exists and finite, and $\{S_T\}_{T \in \mathbb{N}}$ is uniformly p.d., then*

$$\lim_{h \to \infty} \lim_{T \to \infty} \mathrm{MSE}_h(a_T, \hat{S}_T^A, W_T) = \lim_{h \to \infty} \lim_{T \to \infty} \mathrm{MSE}_h(a_T, \hat{S}_T, W_T).$$

3   Computation Algorithm for Adjustment for Positive Definiteness

The minimization problem (3) does not have a closed-form solution. Because (3) is a convex programming problem with a smooth convex function, one might think that a gradient search algorithm could be employed to find the solution of (3). A challenge in such approach is that our choice set is P_v, the set of all symmetric p.s.d. matrices. Though Pinheiro and Bates (1996) list a few ways to parameterize P_v, the objective function becomes non-convex in each of the parameterizations. Also, the number of parameters in this approach is large. In addition, the solution of our problem is always on the boundary of the choice set. These features of the problem make the gradient search combined with Pinheiro and Bates's (1996) parameterizations slow and unreliable.

We here take a different approach. Let vech denote the vectorization-half operator, namely, vech transforms each matrix to the vector made by vertically stacking the portions of its columns below the principal diagonal from the first column to the last. Then the set of all $v \times v$ symmetric matrices and $\mathbb{R}^{v(v+1)/2}$ are related to each other in one-to-one manner through vech. It follows that the minimization problem (3) is equivalent to the minimization of $\|\hat{S}_T - \text{vech}^{-1}(x)\|_{W_T}$ with respect to x over $\mathbb{R}^{v(v+1)/2}$ subject to the constraint that $\text{vech}^{-1}(x)$ is p.s.d. Once the problem is solved, the adjusted estimator is given by transforming the solution by vech.

Now, decompose W_T as $W_T = V_T V_T'$, where V_T is a $v \times v$ matrix (e.g., the Cholesky decomposition). Then

$$\|\hat{S}_T - \text{vech}^{-1}(x)\|_{W_T} = (V_T'\text{vec}(\hat{S}_T) - V_T'\text{vec}(\text{vech}^{-1}(x)))' (V_T'\text{vec}(\hat{S}_T) - V_T'\text{vec}(\text{vech}^{-1}(x))).$$

It is straightforward to show that for each $\tau \in \mathbb{R}$, $\|\hat{S}_T - \text{vech}^{-1}(x)\|_{W_T} \leq \tau$ if and only if

$$\begin{pmatrix} \tau I_{v(v+1)/n} & (V_T'\text{vec}(\hat{S}_T) - V_T'\text{vec}(\text{vech}^{-1}(x))) \\ (V_T'\text{vec}(\hat{S}_T) - V_T'\text{vec}(\text{vech}^{-1}(x)))' & \tau \end{pmatrix}$$

is p.s.d. It follows that (3) is equivalent to choosing (x, τ) from $\mathbb{R}^{v(v+1)/2} \times \mathbb{R}$ to minimize τ under the

constant that

$$
\begin{pmatrix}
\text{vech}^{-1}(x) & 0_{v \times v(v+1)/2} & 0_{v \times 1} \\
0_{v(v+1)/2 \times v} & \tau I_{v(v+1)/2} & (V_T' \text{vec}(\hat{S}_T) - V_T' \text{vec}(\text{vech}^{-1}(x))) \\
0_{1 \times v} & (V_T' \text{vec}(\hat{S}_T) - V_T' \text{vec}(\text{vech}^{-1}(x)))' & \tau
\end{pmatrix}
\text{ is p.s.d.} \quad (5)
$$

This problem is a semidefinite programming problem, because the objective function in this minimization problem is linear in (x, τ), and the matrix in the constraint (5) is symmetric and linear in (x, τ) (see, e.g., Vandenberghe and Boyd (1996) for the semidefinite programming in general).

The semidefinite programming has been actively studied in the recent numerical optimization literature, and fast solvers have been developed. In the Monte Carlo simulations in this paper, we employ SeDuMi (Sturm 1999) among them.

4 Truncated Flat Kernel Estimator Adjusted for Positive Semidefiniteness

For each $\tau \in \mathbb{Z}$, write $\Gamma(\tau) \equiv \text{cov}[Z_0^*, Z_\tau^*]$. Also, for arbitrary a_1, a_2, a_3, a_4 in $\{1, \ldots, v\}$ and arbitrary t_1, t_2, t_3, t_4 in $\mathbb{Z}$, let $\kappa_{a_1,a_2,a_3,a_4}(t_1, t_2, t_3, t_4)$ denote the fourth-order cumulant of $(Z_{t_1,a_1}^*, Z_{t_2,a_2}^*, Z_{t_3,a_3}^*, Z_{t_4,a_4}^*)$. Andrews (1991, Proposition 1) shows the asymptotic bias and asymptotic variance of kernel estimators without estimation of θ^*, imposing the following memory conditions on $\{Z_t^*\}_{t \in \mathbb{Z}}$.

ASSUMPTION 2: $\sum_{\tau=-\infty}^{\infty} \|\Gamma(\tau)\| < \infty$ and

$$
\sum_{\tau_1=-\infty}^{\infty} \sum_{\tau_2=-\infty}^{\infty} \sum_{\tau_3=-\infty}^{\infty} |\kappa_{a,b,c,d}(0, \tau_1, \tau_2, \tau_3)| < \infty.
$$

Andrews (1991, pages 827 and 853) also demonstrates that a wide range of kernel estimators satisfy the uniform integrability condition imposed in Theorem 2.4 with a suitably chosen sequence of scaling factors, if:

ASSUMPTION 3: $\{Z_t^*\}$ is eighth-order stationary with

$$
\sum_{\tau_1=-\infty}^{\infty} \cdots \sum_{\tau_7=-\infty}^{\infty} \kappa_{a_1,\ldots a_8}(0, \tau_1, \cdots, \tau_7) < \infty,
$$

11

where for arbitrary $a_1, \ldots a_8$ in $\{1, \ldots, v\}$ and arbitrary $t_1, \ldots, t_8$ in $\mathbb{Z}$, $\kappa_{a_1, \cdots, a_8}(t_1, \ldots, t_8)$ denotes the eight-order cumulant of $(Z^*_{t_1, a_1}, \ldots, Z^*_{t_8, a_8})$.

Write

$$S^{(q)} \equiv \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} |\tau|^q \Gamma(\tau)$$

for each $q \in [0, \infty)$, and $S \equiv S^{(0)}$. When $S^{(q)}$ converges for some $q_0 \in (0, \infty)$,

$$S^{(q_1)} = \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} |\tau|^{-(q_0-q_1)} |\tau|^{q_0} \Gamma(\tau)$$

also converges for each $q_1 \in [0, q_0]$ (Rudin 1976, Theorem 3.42, pp. 70–71). In most applications, it is reasonable to make the following assumption, though not all of our results impose this assumption.

ASSUMPTION 4: *The matrix S is p.d.*

Given Assumptions 1–4, we can assess the asymptotic MSEs of the TF estimator and its adjusted version by applying Andrews (1991, Proposition 1(c)) along with Corollary 2.3(a) and Theorem 2.4 of this paper. Let $\{\hat{S}^{TF}_T(m_T)\}_{T\in\mathbb{N}}$ and $\{\hat{S}^{TF,A}_T(m_T)\}_{T\in\mathbb{N}}$ denote the TF and ATF estimators with bandwidth $\{m_T\}_{T\in\mathbb{N}}$, respectively. Also, let $K_{v,v}$ denote the $v^2 \times v^2$ commutation matrix, i.e., $K_{v,v} \equiv \sum_{i=1}^{v} \sum_{j=1}^{v} e_i e'_j \otimes e_j e'_i$, where e_i is the ith elementary $p \times 1$ vector, and $\otimes$ is the Kronecker product operator.

PROPOSITION 4.1: *Suppose that Assumptions 1 and 2 hold and that $\{m_T\}_{T\in\mathbb{N}}$ is a sequence of positive real numbers such that $m_T^{2q+1}/T \to \gamma \in (0, \infty)$ for some $q \in (0, \infty)$ for which the series $S^{(q)}$ converges. Also, let W be a $v^2 \times v^2$ symmetric p.s.d. matrix. Then we have:*

(a)

$$\lim_{T\to\infty} \text{MSE}(T/m_T, \tilde{S}^{TF,A}_T(m_T), W) \leq \lim_{T\to\infty} \text{MSE}(T/m_T, \tilde{S}^{TF}_T(m_T), W) \tag{6}$$

$$= 8\pi^2 \text{tr}(W(I + K_{v,v})S \otimes S). \tag{7}$$

(b) *If in addition Assumptions 3 and 4 hold, (6) holds with equality.*

Proposition 4.1 means that the convergence rates of both the ATF and TF estimators can be made as fast as $T^{-q/(2q+1)}$, provide that the bandwidth is suitably chosen, and $S^{(q)}$ converges. In particular, when $S^{(q)}$ converges for some $q > 2$, employing a bandwidth $m_T \sim T^{1/(2q+1)}$ makes the TF estimators converge to

12

S_T faster in terms of the MSE than the QT and BT estimator, whose convergence rates never exceed $T^{-1/3}$ and $T^{-2/5}$, respectively.

We now add a few assumptions related to the effect of the parameter estimation on the long-run covariance matrix estimation.

ASSUMPTION 5:   (a) $T^{1/2}(\hat{\theta} - \theta^*) = O_p(1)$.

(b) *There exists a uniformly L_2-bounded sequence of random variables $\{\eta_{1,t}\}_{t \in \mathbb{N}}$ such that for each $t \in \mathbb{N}$,*

$|Z_t^*| \leq \eta_{1,t}$ *and*

$$\sup_{\theta \in \Theta} \left\| \frac{\partial}{\partial \theta} Z_t(\cdot, \theta) \right\| \leq \eta_{1,t}.$$

ASSUMPTION 6:   (a) *The sequence*

$$\left\{ \zeta_t \equiv \left(Z_t^{*\prime}, \mathrm{vec}\left(\frac{\partial}{\partial \theta'} Z_t(\cdot, \theta^*) - \mathrm{E}\left[\frac{\partial}{\partial \theta'} Z_t(\cdot, \theta^*) \right] \right)' \right)' \right\}_{t \in \mathbb{Z}}$$

is a zero-mean, fourth-order stationary sequence of random vectors such that Assumption 2 holds with Z_t replaced by ζ_t.

(b) *There exists a uniformly L_2-bounded sequence of random variables $\{\eta_{2,t}\}_{t \in \mathbb{N}}$ such that for each $t \in \mathbb{N}$*

$$\sup_{\theta \in \Theta} \left\| \frac{\partial^2}{\partial \theta \partial \theta'} Z_{t,a}(\cdot, \theta) \right\| \leq \eta_{2,t}, \quad a = 1, \ldots, v.$$

We hereafter focus on the case where the weighting matrix is convergent in probability.

ASSUMPTION 7: *$\{W_T\}_{T \in \mathbb{N}}$ is a sequence of $v^2 \times v^2$ symmetric p.s.d. random matrices that converges in probability-P to a constant $v^2 \times v^2$ matrix W.*

Under Assumptions 1 and 4, the difference between any estimator consistent for S and the estimator that adjusts it for positive definiteness converges in probability to zero at an arbitrary fast rate by Corollary 2.3(c). The ATF estimator therefore inherits the large sample properties of the TF kernel estimator.

THEOREM 4.2: *Let $\{m_T\}_{T \in \mathbb{N}}$ be a sequence of positive real numbers growing to infinity.*

(a) *If Assumptions 1, 2, 5, and 7 hold, and $m_T^2/T \to 0$, then $\|\hat{S}_T^{TF,A}(m_T) - S_T\|_{W_T} \to 0$ in probability-P. If in addition W is p.d., then $\{\hat{S}_T^{TF,A}(m_T)\}$ is consistent for $\{S_T\}$.*

(b) *If Assumptions 1, 4, 5, 6, 7, hold, and $m_T^{2q+1}/T \to \gamma \in (0, \infty)$ for some $q \in (0, \infty)$ for which $S^{(q)}$ converges, then $(T/m_T)^{1/2}\|\hat{S}_T^{TF,A}(m_T) - S_T\|_{W_T} = O_P(1)$ and $(T/m_T)^{1/2}\|\hat{S}_T^{TF,A}(m_T) - \hat{S}_T^{TF}(m_T)\|_{W_T} \to 0$ in probability-P.*

(c) *If, in addition to the conditions of part (b), W is p.d., then $(T/m_T)^{1/2}(\hat{S}_T^{TF,A}(m_T) - S_T) = O_P(1)$ and $(T/m_T)^{1/2}(\hat{S}_T^{TF,A}(m_T) - \hat{S}_T^{TF}(m_T)) \to 0$ in probability-P.*

(d) *Under the conditions of part (b) plus Assumption 3,*

$$\lim_{h \to \infty} \lim_{T \to \infty} \text{MSE}_h(T/m_T, \hat{S}_T^{TF,A}(m_T), W_T) = \lim_{h \to \infty} \lim_{T \to \infty} \text{MSE}_h(T/m_T, \hat{S}_T^{TF}(m_T), W_T) \tag{8}$$

$$= \lim_{T \to \infty} \text{MSE}(T/m_T, \tilde{S}_T^{TF}(m_T), W) \tag{9}$$

$$= 8\pi^2 \text{tr}(W(I + K_{v,v})S \otimes S) \tag{10}$$

5 Flat Kernel Estimator That Fractionally Incorporates the Autocovariance Matrix at the Last Lag

In the TF estimation, all bandwidths between two adjacent nonnegative integers give the same estimator. Suppose that we have two adjacent integer bandwidths that yield good performances of the TF estimator. Given the familiar argument that the bandwidth should be chosen to balance the bias and variance of the estimator, one might desire to consider an estimator "between" the two estimators picked by the two integer bandwidths. A natural way to create a smooth transition path from an integer bandwidth to the next is to linearly interpolate the TF estimator between each pair of adjacent integer bandwidths. Given the TF estimators $\tilde{S}_T^{TF}$ and $\hat{S}_T^{TF}$ discussed earlier, we now introduce their interpolated versions called the TF estimators that fractionally incorporates the estimated autocovariance matrix at the last lag, which are abbreviated as the TFF estimators hereafter:

$$\tilde{S}_T^{TFF}(m) \equiv (\lfloor m \rfloor + 1 - m)\tilde{S}_T^{TF}(m) + (m - \lfloor m \rfloor)\tilde{S}_T^{TF}(m + 1), \quad m \in [0, \infty), \ T \in \mathbb{N},$$

$$\hat{S}_T^{TFF}(m) \equiv (\lfloor m \rfloor + 1 - m)\hat{S}_T^{TF}(m) + (m - \lfloor m \rfloor)\hat{S}_T^{TF}(m + 1), \quad m \in [0, \infty), \ T \in \mathbb{N},$$

where $\lfloor \cdot \rfloor : \mathbb{R} \to \mathbb{R}$ is the floor function, which returns the greatest integer not exceeding the value of the argument, and we employ the rule that $\tilde{S}^{TF}(0) = \tilde{\Gamma}_T(0)$ and $\hat{S}^{TF}(0) = \hat{\Gamma}_T(0)$. Each version of the TFF

estimators coincides with the corresponding version of the TF estimator if the bandwidth m is an integer. In general, provided that the bandwidth m is less than $T - 1$, each version of the TFF estimator with a bandwidth m, compared to the corresponding version of the TF estimator with the same bandwidth, brings in the fraction $(m - \lfloor m \rfloor)$ of the autocovariance matrix estimator at lag $\lfloor m \rfloor + 1$, namely,

$$\tilde{S}_T^{TFF}(m) = \tilde{S}_T^{TF}(m) + (m - \lfloor m \rfloor)(\tilde{\Gamma}_T(\lfloor m \rfloor + 1) + \tilde{\Gamma}_T(\lfloor m \rfloor + 1)'), \quad m \in [0, T-1), \ T \in \mathbb{N},$$

$$\hat{S}_T^{TFF}(m) = \hat{S}_T^{TF}(m) + (m - \lfloor m \rfloor)(\hat{\Gamma}_T(\lfloor m \rfloor + 1) + \hat{\Gamma}_T(\lfloor m \rfloor + 1)'), \quad m \in [0, T-1), \ T \in \mathbb{N}.$$

These equalities justify the name given to $\tilde{S}_T^{TFF}(m)$ and $\hat{S}_T^{TFF}(m)$.

The behavior of the last autocovariance matrix estimator fractionally incorporated in the TFF estimator, when no parameter is estimated, is described in the next lemma.

LEMMA 5.1: *Suppose that Assumptions 1 and 2 hold. Then:*

(a) $\sup_{\tau \in \{0,1,\ldots,T-1\}} \mathrm{E}[\,\|\tilde{\Gamma}_T(\tau) - \Gamma_T(\tau)\|^2\,] = O(T^{-1})$.

(b) *Let $\{m_T \in (0, T-1)\}_{T \in \mathbb{N}}$ be a sequence that grows to ∞. If $S^{(q)}$ converges for some $q \in (0, \infty)$, then*
$$m_T^q \mathrm{E}[\tilde{\Gamma}_T(\lfloor m_T + 1 \rfloor)] \to 0.$$

(c) *Let $\{m_T \in (0, T-1)\}_{T \in \mathbb{N}}$ be a sequence that grows to ∞. If $m_T \to \infty$ and $m_T^{2q+1}/T \to \gamma \in (0, \infty)$ for some $q \in (0, \infty)$ for which the series $S^{(q)}$ converges, then $(T/m_T)\mathrm{E}[\|\tilde{\Gamma}_T(\lfloor m_T + 1 \rfloor)\|^2] = o(1)$.*

From Lemma 5.1(c), one might conjecture that the autocovariance estimator at the last lag is asymptotically negligible in the TFF estimation. It is indeed the case as the next proposition states.

PROPOSITION 5.2: *Suppose that Assumptions 1 and 2 hold and that $\{m_T \in (0, T-1)\}_{T \in \mathbb{N}}$ is a sequence satisfying that $m_T^{2q+1}/T \to \gamma \in (0, \infty)$ for some $q \in (0, \infty)$ for which the series $S^{(q)}$ converges. Also, let W be a $v^2 \times v^2$ symmetric p.s.d. matrix. Then*

$$\lim_{T \to \infty} \mathrm{MSE}(T/m_T, \tilde{S}_T^{TFF}(m_T), W) = \lim_{T \to \infty} \mathrm{MSE}(T/m_T, \tilde{S}_T^{TF}(m_T), W) \tag{11}$$

$$= 8\pi^2 \mathrm{tr}(W(I + K_{v,v})S \otimes S). \tag{12}$$

The TFF estimator may give an estimate that is not p.s.d., being a convex combination of the TF estimators that have the same problem. Thus, the estimator that adjusts it for positive semidefiniteness is

useful. As is the case with the TF estimation, the adjustment improves the MSE of the TFF estimator in small samples (Corollary 2.3(a)). The next proposition states that the adjusted TFF estimator performs at least as well as the TFF estimator in large samples.

PROPOSITION 5.3: *Suppose that Assumptions 1 and 2 hold and that $\{m_T \in (0, T-1)\}_{T \in \mathbb{N}}$ is a sequence satisfying that $m_T^{2q+1}/T \to \gamma \in (0, \infty)$ for some $q \in (0, \infty)$ for which the series $S^{(q)}$ converges. Also, let W be a $v^2 \times v^2$ symmetric p.s.d. matrix. Then we have:*

(a)

$$\lim_{T \to \infty} \text{MSE}(T/m_T, \tilde{S}_T^{TFF,A}(m_T), W) \leq \lim_{T \to \infty} \text{MSE}(T/m_T, \tilde{S}_T^{TFF}(m_T), W) \tag{13}$$

$$= \lim_{T \to \infty} \text{MSE}(T/m_T, \tilde{S}_T^{TF}(m_T), W) = 8\pi^2 \text{tr}(W(I + K_{v,v})S \otimes S). \tag{14}$$

(b) *If in addition Assumptions 3 and 4 hold, (13) holds with equality.*

Many interesting applications involve unknown parameters θ^*. The TFF estimator with parameter estimation is asymptotically equivalent to the TFF estimator without parameter estimation as described in the next theorem.

THEOREM 5.4: *Let $\{m_T \in (0, T-1)\}_{T \in \mathbb{N}}$ be a sequence growing to infinity.*

(a) *If Assumptions 1, 2, and 5 hold, and $m_T^2/T \to 0$, then $\{\hat{S}_T^{TFF}(m_T)\}_{T \in \mathbb{N}}$ is consistent for $\{S_T\}_{T \in \mathbb{N}}$.*

(b) *If Assumptions 1, 5, and 6 hold, and $m_T^{2q+1}/T \to \gamma \in (0, \infty)$ for some $q \in (0, \infty)$ for which $S^{(q)}$ converges, then $(T/m_T)^{1/2}(\hat{S}_T^{TFF}(m_T) - S_T) = O_P(1)$ and $(T/m_T)^{1/2}(\hat{S}_T^{TFF}(m_T) - \tilde{S}_T^{TFF}(m_T)) \to 0$ in probability-P.*

(c) *Under the conditions of part (b) plus Assumption 3,*

$$\lim_{h \to \infty} \lim_{T \to \infty} \text{MSE}_h(T/m_T, \hat{S}_T^{TFF}(m_T), W_T) = \lim_{T \to \infty} \text{MSE}(T/m_T, \tilde{S}_T^{TFF}(m_T), W) \tag{15}$$

$$= \lim_{T \to \infty} \text{MSE}(T/m_T, \tilde{S}_T^{TF}(m_T), W) \tag{16}$$

$$= 8\pi^2 \text{tr}(W(I + K_{v,v})S \otimes S) \tag{17}$$

When θ^* is estimated, the relationship between the TFF estimator and the adjusted TFF (ATFF) estimator is parallel to that between the TF estimator and the ATF estimator.

16

THEOREM 5.5: *Let $\{m_T \in (0, T-1)\}_{T \in \mathbb{N}}$ be a sequence growing to infinity.*

(a) *If Assumptions 1, 2, 5, and 7 hold, and $m_T^2/T \to 0$, then $\|\hat{S}_T^{TFF,A}(m_T) - S_T\|_{W_T} \to 0$ in probability-P. If in addition W is p.d., then $\{\hat{S}_T^{TFF,A}(m_T)\}$ is consistent for $\{S_T\}$.*

(b) *If Assumptions 1, 4, 5, 6, and 7, hold, and $m_T^{2q+1}/T \to \gamma \in (0, \infty)$ for some $q \in (0, \infty)$ for which $S^{(q)}$ converges, then $(T/m_T)^{1/2}\|\hat{S}_T^{TFF,A}(m_T) - S_T\|_{W_T} = O_P(1)$ and $(T/m_T)^{1/2}\|\hat{S}_T^{TFF,A}(m_T) - \hat{S}_T^{TFF}(m_T)\|_{W_T} \to 0$ in probability-P.*

(c) *If, in addition to the conditions of part (b), W is p.d., then $(T/m_T)^{1/2}(\hat{S}_T^{TFF,A}(m_T) - S_T) = O_P(1)$ and $(T/m_T)^{1/2}(\hat{S}_T^{TFF,A}(m_T) - \hat{S}_T^{TFF}(m_T)) \to 0$ in probability-P.*

(d) *Under the conditions of part (b) plus Assumption 3,*

$$\lim_{h \to \infty} \lim_{T \to \infty} \mathrm{MSE}_h(T/m_T, \hat{S}_T^{TFF,A}(m_T), W_T) = \lim_{h \to \infty} \lim_{T \to \infty} \mathrm{MSE}_h(T/m_T, \hat{S}_T^{TFF}(m_T), W_T) \qquad (18)$$

$$= \lim_{T \to \infty} \mathrm{MSE}(T/m_T, \tilde{S}_T^{TFF}(m_T), W) \qquad (19)$$

$$= 8\pi^2 \mathrm{tr}(W(I + K_{v,v})S \otimes S) \qquad (20)$$

6 Finite-Sample Performance of the ATF and ATFF Estimators

In this section, we conduct Monte Carlo simulations to examine the small-sample performance of the proposed estimators in comparison with the familiar QS and BT estimators, borrowing the experiment setups from Andrews (1991). In each of the experiments, $\{(y_t, x_t')'\}_{t=1}^T$ is a stationary process, where y_t is a random variable, and x_t is a $v \times 1$ random vector. The coefficients θ^* in the population regression of y_t on x_t are parameters of interest. In this setup, we examine the MSE of each of the covariance matrix estimators and the size of the t-test of an exclusion restriction in the OLS regression, using each of the covariance matrix estimators. Thus, we have that

$$Z_t(\cdot, \theta^*) = x_t u_t, \quad t \in \mathbb{N},$$

where

$$u_t = y_t - x_t'\theta^*, \quad t \in \mathbb{N}.$$

The regressor vector x_t consists of a constant equal to one and four random variables x_{t1}, x_{t2}, x_{t3}, and x_{t4}, i.e., $x_t = [1, x_{t2}, x_{t3}, x_{t4}, x_{t5}]'$. The regressors x_t and the disturbance u_t are independent, and θ^* is set equal

to zero.

The experiments are split into four groups: the AR(1)-HOMO, AR(1)-HET1, AR(1)-HET2, and MA(1)-HOMO experiments. In the AR(1)-HOMO experiments, the disturbance u_t is a stationary Gaussian AR(1) process with mean zero and variance one. To generate the four nonconstant regressors, we first generate four independent sequences (that are also independent from the disturbance sequence) in the same way as we generate the disturbance sequence; then normalize them to obtain $\{x_t\}_{t=1}^{T}$ such that $\sum_{t=1}^{T} x_t x_t'/T = I$. The MA(1)-HOMO experiments are the same as the AR(1)-HOMO experiments, except that the disturbance term and the regressors (prior to the normalization) are Gaussian stationary MA(1) processes with mean zero and variance one. In the AR(1)-HET1 and AR(1)-HET2 experiments, the disturbance process is given by

$$
u_t =
\begin{cases}
\tilde{u}_t x_{t2} & \text{in AR(1)-HET1,} \\[2mm]
\frac{1}{2}\tilde{u}_t \,\big|\, \sum_{i=2}^{5} x_{ti} \,\big| & \text{in AR(1)-HET2,}
\end{cases}
$$

where x_t and $\tilde{u}_t$ $(t = 1, \ldots, T)$ are the regressors and errors in the corresponding AR(1)-HOMO experiment. In particular, as pointed out by Andrews (1991), the errors in the AR(1)-HET1 and AR(1)-HET2 are the AR processes with AR parameter ρ^2, where ρ is the slope of the AR(1) process that generates $\{\tilde{u}_t\}$. The number of Monte Carlo replications is 25,000. In each replication, $500 + T$ observations are generated and the last T observations are used. See Table 2 for the summary of the experiment setups.

$$\boxed{\text{Table 2 around here.}}$$

We first compare the performance of the ATF estimator against that of the TF estimator to assess the effect of the adjustment for positive semidefiniteness on the performance. While Corollary 2.3 claims that the MSE of the ATF estimator never exceeds that of the TF estimator, Theorem 4.2(d) suggests that the efficiency gain from the adjustment is asymptotically negligible. We seek to check if the negligibility of the efficiency gain by the adjustment carries over in small samples.

Table 3 displays the efficiency of the ATF estimator relative to that of TF estimator in the AR(1)-HOMO and MA(1)-HOMO experiments with sample size $T = 128$ and bandwidths $m \in \{1, 3, 5, 7\}$. We define the efficiency of an estimator relative to another to be the ratio of the MSEs of the estimators calculated in the form of (4). Following Andrews (1991, p. 836), we employ the weighting matrix

$$
W_T = \left(\left(T^{-1}\sum_{t=1}^{T} x_t x_t' \right)^{-1} \otimes \left(T^{-1}\sum_{t=1}^{T} x_t x_t' \right)^{-1} \right) \tilde{W} \left(\left(T^{-1}\sum_{t=1}^{T} x_t x_t' \right)^{-1} \otimes \left(T^{-1}\sum_{t=1}^{T} x_t x_t' \right)^{-1} \right),
$$

where $\tilde{W}$ is a $v^2 \times v^2$ diagonal matrix that has two for its $((i-1)v+i)$th diagonal elements $(i = 1, 2, \ldots, v)$ and one for all other diagonal elements.

<div align="center">

Table 3 around here.

</div>

The efficiency of the ATF estimator relative to the TF estimator is 1.00 in the vast majority of the experiments in Table 3. This reflects the fact that the probability that that the TF estimator is not p.s.d. is close to zero in many cases. Nevertheless, the adjustment in the ATF estimator sometimes reduces the MSE by a few per cents, if TF estimator is non-p.s.d. with a higher probability. Thus, the efficiency gain of the adjustment is not totally ignorable in small samples though it is often negligibly small. This tendency is also verified with different sample sizes and in AR(1)-HET1 and AR(1)-HET2 experiments, though we do not include the tables for the experiments in this paper.

We next compare the performances of the QS, BT, ATF, and ATFF estimators, letting each of the estimator use its fixed optimum bandwidth in each experiment. We here mean by the fixed optimum bandwidth of a kernel estimator the nonstochastic bandwidth that minimizes the (finite sample) MSE of the estimator, which we numerically find by using the grid search method through the Monte Carlo experiments. Table 4 displays the efficiency of the BT, ATF, and ATFF estimators relative to the QS estimator with sample sizes 64, 128, and 256.

<div align="center">

Table 4 around here.

</div>

The relationship between the ATF and QS estimator is similar to that between the TF and QS estimator reported in Andrews (1991). The ATF outperforms the QS estimator clearly in some cases, and the complete opposite happens in some other cases. On the other hand, the behavior of the ATFF estimator is quite different. The ATFF estimator never has an MSE larger than the ATF estimator and sometimes brings in substantial improvement over the TF estimator, in particular, when the TF estimator poorly performs relatively to the QS estimator. As a result, the MSE of the ATFF estimator is smaller than or about the same as that of the QS estimator in all experiments. Not surprisingly, the fixed optimum bandwidth for the ATFF estimator is close to the midpoint between a pair of adjacent integers when the ATFF estimator outperforms the ATF estimator by a large margin.

The large sample theory indicates that the efficiency of the ATF and ATFF estimators relative to the QS and BT estimators becomes higher as the sample size increases. Table 4 indeed confirms that the relative

<div align="center">

19

</div>

efficiency of the ATFF increases, though slowly, as the sample size grows. On the other hand, the relative efficiency of the ATF estimator shows more complicated moves. That is, the relative efficiency of the ATF may decrease when the sample size increases. To understand why this can happen, it is useful to view the ATF estimator as a restricted version of the ATFF estimator that can only use an integer bandwidth in the ATFF estimation. Suppose that the fixed optimum bandwidth for the ATFF estimator is close to an integer with the initial sample size. Then the ATF and ATFF estimators perform equally well with the initial sample size. When the sample size increases, however, the optimum bandwidth for the ATFF may be close to the midpoint between a pair of adjacent integers. The restriction imposed on the ATF estimator now becomes a severe penalty. Thus, the efficiency of the ATF estimator relative to the QS estimator can decrease, while the relative efficiency of the ATFF increases.

7 TF Estimation with Data-Based Bandwidth

The optimum bandwidth is unknown in practice. We need a way to choose a bandwidth based on data. For consistency of the TF, ATF, TFF, and ATFF estimators with data-based bandwidths, a data-based bandwidth $\hat{m}_T$ only need to satisfy the following assumption.

ASSUMPTION 8: *The sequence $\{m_T \in (0, T-1)\}_{T \in \mathbb{N}}$ satisfies that $m_T \to \infty$ and $m_T^2/T \to 0$. Also, the sequence of random variables $\{\hat{m}_T : \Omega \to (0, T-1)\}_{T \in \mathbb{N}}$ satisfies that $|\log(\hat{m}_T/m_T)| = \mathrm{O}_P(1)$.*

Note that Assumption 8 imposes the same conditions on $\{m_T\}$ as the consistency results for the ATF, TFF, and ATFF estimators in Theorems 4.2(a), 5.4(a), and 5.5(a).

To establish results on the rate of convergence and asymptotic truncated MSE, we impose stronger conditions on the bandwidth.

ASSUMPTION 9: *The sequence $\{m_T \in (0, T-1)\}_{T \in \mathbb{N}}$ satisfies that $m_T \to \infty$ and $m_T^{2q+1}/T \to \gamma \in (0, \infty)$ for some $q \in (0, \infty)$ for which $S^{(q)}$ absolutely converges. Also, the sequence of random variables $\{\hat{m}_T : \Omega \to (0, T-1)\}_{T \in \mathbb{N}}$ satisfies that for some $\{d_T \in (0, \infty)\}_{T \in \mathbb{N}}$ such that $d_T^{-1} m_T^{1/2} \to 0$, $d_T |\hat{m}_T - m_T|/m_T = \mathrm{O}_P(1)$ in probability-P.*

The conditions imposed on $\{m_T\}$ in Assumption 9 are the same as those imposed in Theorems 4.2(b)–(d), 5.4(b)(c), and 5.5(b)–(d).

Remark. In Andrews (1991) and Newey and West (1994), though they do not consider the TF estimator, the data-based bandwidth takes a form of $\hat{m}_T = \hat{c}_T T^r$ where r is some positive real number and $\{\hat{c}_T : \Omega \to (0, \infty)\}_{T \in \mathbb{N}}$ is an estimator of some constant $c \in (0, \infty)$. With such $\hat{m}_T$, the condition $|\log(\hat{m}_T/m_T)| = \mathrm{O}_P(1)$ in Assumption 8 coincides with Assumption E of Andrews (1991), because $\log(\hat{m}_T/m_T) = \log(\hat{c}_T/c_T)$. Also, we have that $d_T|\hat{m}_T - m_T|/m_T = d_T(\hat{c}_T - c) = \mathrm{O}_P(1)$ for a suitably chosen $\{d_T \in (0, \infty)\}_{T \in \mathbb{N}}$, as required in Assumption 9. In order for $\{d_T^{-1} m_T^{1/2}\}_{T \in \mathbb{N}}$ to converge to zero, q in Assumption 9 needs to be sufficiently large. If $d_t = T^{1/2}$, as is the case in the data-based bandwidth of Andrews (1991), it must hold that $q > 1/2$.

We are now ready to state a few results on the large sample behavior of the TF, ATF, TFF, and ATFF estimator with data-based bandwidths.

THEOREM 7.1: (a) *Suppose that Assumptions 1, 2, 5, and 8 hold. Then the estimators* $\{\hat{S}_T^{TF}(\hat{m}_T)\}_{T \in \mathbb{N}}$ *and* $\{\hat{S}_T^{TFF}(\hat{m}_T)\}_{T \in \mathbb{N}}$ *are consistent for* $\{S_T\}_{T \in \mathbb{N}}$, *and it holds that*

$$\|\hat{S}_T^{TF}(\hat{m}_T) - S_T\|_{W_T} \to 0 \ \textit{in probability-P,} \tag{21}$$

$$\|\hat{S}_T^{TF,A}(\hat{m}_T) - S_T\|_{W_T} \to 0 \ \textit{in probability-P,} \tag{22}$$

$$\|\hat{S}_T^{TFF}(\hat{m}_T) - S_T\|_{W_T} \to 0 \ \textit{in probability-P,} \tag{23}$$

$$\|\hat{S}_T^{TFF,A}(\hat{m}_T) - S_T\|_{W_T} \to 0 \ \textit{in probability-P.} \tag{24}$$

If, in addition, W is p.d., then $\{\hat{S}_T^{TF,A}(\hat{m}_T)\}$ *and* $\{\hat{S}_T^{TFF,A}(\hat{m}_T)\}$ *are also consistent for* $\{S_T\}$.

(b) *If Assumptions 1, 4, 5, 6, 7, and 9 hold. Then we have:*

$$(T/m_T)^{1/2}(\hat{S}_T^{TF}(\hat{m}_T) - S_T) = \mathrm{O}_P(1), \tag{25}$$

$$(T/m_T)^{1/2}(\hat{S}_T^{TFF}(\hat{m}_T) - S_T) = \mathrm{O}_P(1), \tag{26}$$

$$(T/m_T)^{1/2}(\hat{S}_T^{TF}(\hat{m}_T) - \hat{S}_T^{TF}(m_T)) = \mathrm{o}_P(1), \tag{27}$$

$$(T/m_T)^{1/2}(\hat{S}_T^{TFF}(\hat{m}_T) - \hat{S}_T^{TFF}(m_T)) = \mathrm{o}_P(1), \tag{28}$$

$$(T/m_T)^{1/2}\|\hat{S}_T^{TF}(\hat{m}_T) - S_T\|_{W_T} = \mathrm{O}_P(1), \tag{29}$$

$$(T/m_T)^{1/2}\|\hat{S}_T^{TF,A}(\hat{m}_T) - S_T\|_{W_T} = \mathrm{O}_P(1), \tag{30}$$

$$(T/m_T)^{1/2}\|\hat{S}_T^{TFF}(\hat{m}_T) - S_T\|_{W_T} = \mathrm{O}_P(1), \tag{31}$$

$$(T/m_T)^{1/2} \|\hat{S}_T^{TFF,A}(\hat{m}_T) - S_T\|_{W_T} = O_P(1), \tag{32}$$

$$(T/m_T)^{1/2} \|\hat{S}_T^{TF}(\hat{m}_T) - \hat{S}_T^{TF}(m_T)\|_{W_T} = o_P(1), \tag{33}$$

$$(T/m_T)^{1/2} \|\hat{S}_T^{TF,A}(\hat{m}_T) - \hat{S}_T^{TF,A}(m_T)\|_{W_T} = o_P(1), \tag{34}$$

$$(T/m_T)^{1/2} \|\hat{S}_T^{TFF}(\hat{m}_T) - \hat{S}_T^{TFF}(m_T)\|_{W_T} = o_P(1), \tag{35}$$

$$(T/m_T)^{1/2} \|\hat{S}_T^{TFF,A}(\hat{m}_T) - \hat{S}_T^{TFF,A}(m_T)\|_{W_T} = o_P(1). \tag{36}$$

(c) *If, in addition to the conditions of part (c), W is p.d., then*

$$(T/m_T)^{1/2} (\hat{S}_T^{TF,A}(\hat{m}_T) - S_T) = O_P(1), \tag{37}$$

$$(T/m_T)^{1/2} (\hat{S}_T^{TFF,A}(\hat{m}_T) - S_T) = O_P(1), \tag{38}$$

$$(T/m_T)^{1/2} (\hat{S}_T^{TF,A}(\hat{m}_T) - \hat{S}_T^{TF,A}(m_T)) = o_P(1), \tag{39}$$

$$(T/m_T)^{1/2} (\hat{S}_T^{TFF,A}(\hat{m}_T) - \hat{S}_T^{TFF,A}(m_T)) = o_P(1). \tag{40}$$

(d) *Under the conditions of part (b) plus Assumption 3, we have that*

$$\lim_{h \to \infty} \lim_{T \to \infty} \mathrm{MSE}_h(T/m_T, \hat{S}_T^{TF}(\hat{m}_T), W_T) = \lim_{h \to \infty} \lim_{T \to \infty} \mathrm{MSE}_h(T/m_T, \hat{S}_T^{TF}(m_T), W_T) \tag{41}$$

$$= \lim_{h \to \infty} \lim_{T \to \infty} \mathrm{MSE}_h(T/m_T, \hat{S}_T^{TF,A}(\hat{m}_T), W_T) = \lim_{h \to \infty} \lim_{T \to \infty} \mathrm{MSE}_h(T/m_T, \hat{S}_T^{TF,A}(m_T), W_T) \tag{42}$$

$$= \lim_{h \to \infty} \lim_{T \to \infty} \mathrm{MSE}_h(T/m_T, \hat{S}_T^{TFF}(\hat{m}_T), W_T) = \lim_{h \to \infty} \lim_{T \to \infty} \mathrm{MSE}_h(T/m_T, \hat{S}_T^{TFF}(m_T), W_T) \tag{43}$$

$$= \lim_{h \to \infty} \lim_{T \to \infty} \mathrm{MSE}_h(T/m_T, \hat{S}_T^{TFF,A}(\hat{m}_T), W_T) = \lim_{h \to \infty} \lim_{T \to \infty} \mathrm{MSE}_h(T/m_T, \hat{S}_T^{TFF,A}(m_T), W_T) \tag{44}$$

$$= 8\pi^2 \mathrm{tr}(W(I + K_{v,v})S \otimes S) \tag{45}$$

The results presented in Theorem 7.1 indicates that the more slowly the bandwidth grows, the faster the MSE shrinks in the TF, ATF, TFF, and ATFF estimation, provided that $\{\Gamma(\tau)\}_{\tau \in \mathbb{N}}$ converges to zero fast enough. The complete flat shape of the TF kernel at the origin makes the convergence rate of the bias so fast that the bias is asymptotically negligible relative to the variance in the TF estimation, virtually regardless of the growth rate of the bandwidth. This means that given a sequence of bandwidths in TF estimation, we can always find another sequence of bandwidths with a slower growth rate that makes faster the convergence rate of the TF estimator. The rate results in Theorem 7.1 reflect this fact.

Andrews (1991) and Newey and West (1994) propose ways to choose bandwidths based on data in kernel estimation. Their approach is based on the tradeoff between the asymptotic bias and asymptotic variance of typical kernel estimators: the more slowly the bandwidth grows, the more slowly the asymptotic bias shrinks and the faster the variance shrinks, loosely speaking. Their approach sets the growth rate of the bandwidth in such a way that the convergence rates of the squared bias and the variance are equated, so that the MSE of the estimator reaches the fastest possible convergence rate. It then chooses the proportional constant for the bandwidth by minimizing the suitably scaled asymptotic MSE.

The approach of Andrews (1991) and Newey and West (1994) is inapplicable in the TF estimation, given the absence of the trade-off between the asymptotic bias and asymptotic variance of the TF estimator. Nevertheless, it is possible to choose a bandwidth sequence that makes the TF estimator asymptotically more efficient than the QS estimator. Let m_T^{QS} and $\tilde{m}_T$ denote the "oracle" and data-based bandwidths of Andrews (1991), respectively (for the precise mathematical formulas of m_T^{QS} and $\tilde{m}_T$, see equations (5.1), (6.1), and (6.8) in Andrews (1991)). If we set $\hat{m}_T = a\tilde{m}_T$ for some $a \in (0, 1/2]$, then we have by Theorem 7.1(d) that

$$\lim_{h \to \infty} \lim_{T \to \infty} \mathrm{MSE}_h(T/m_T^{QS}, \hat{S}_T^{TF}(\hat{m}_T), W_T) = \lim_{h \to \infty} \lim_{T \to \infty} a\mathrm{MSE}_h(T/(am_T^{QS}), \hat{S}_T^{TF}(\hat{m}_T), W_T)$$

$$= 8a\pi^2 \mathrm{tr}(W(I + K_{v,v})S \otimes S) \leq 4\pi^2 \mathrm{tr}(W(I + K_{v,v})S \otimes S).$$

Because the right-hand side of this equality is equal to the asymptotic variance of the QS estimator with bandwidth $\tilde{m}_T$, which is no greater than the asymptotic MSE of the QS estimator, the TF estimator with bandwidth $\hat{m}_T$ is asymptotically more efficient than the QS estimator with bandwidth $\tilde{m}_T$. We can, of course, apply the same bandwidth $\hat{m}_T$ in the ATF and ATFF estimation to attain the same asymptotic MSE.

A practical question is what value we should use for a. Though the asymptotic MSE of the TF estimator with the bandwidth sequence $\{\hat{m}_T\}_{T \in \mathbb{N}}$ can be made arbitrarily small by setting a sufficiently small value to a, too small a value for a would result in a large magnitude of bias in the TF estimation in small samples, because there is a tradeoff between the bias and the variance in finite samples. In our Monte Carlo simulations in the next section, we use $a = 1/2$ for the ATF estimator and $a = 1/3$ for the ATFF estimator, though these choices are arguably ad hoc. We use a larger value for a in the ATF estimation than in the ATFF estimation, because the ATF estimator effectively rounds down the data-based bandwidth $\hat{m}_T$, due to the equality $\hat{S}^{ATF}(\hat{m}_T) = \hat{S}^{ATF}(\lfloor \hat{m}_T \rfloor)$.

8 Finite-Sample Performance of the ATF and ATFF Estimators with Data-Based Bandwidths

In this section, we examine the performances of the ATF and ATFF estimators in comparison with those of the QS and BT estimators, using data-based bandwidths. In the QS and BT estimation, we use the bandwidth selection method of Andrews (1991). For the ATF and ATFF estimators, we use the bandwidth described in the previous section. The experiment setups are the same as in Section 6.

Table 5 reports the efficiency of the BT, ATF, and ATFF estimators relative to the QS estimator. The relationship among the estimators are analogous to that in Table 4 of the experiments with fixed optimum bandwidths, though the ATF and ATFF estimators are slightly more efficient relative to the QS estimator with the data-based bandwidth. The MSE of the ATFF estimator is smaller than or at least comparable to that of the QS estimator in all of our experiments, while the efficiency of the ATF estimator relative to the QS estimator varies from an experiment to another.

Table 5 around here.

Table 6 shows the sizes in the ten- and five-per cent level t-tests of the exclusion of x_{t2} using each of the covariance matrix estimators. The tests using the BT estimator consistently results in the largest size distortion, while the ATF ATFF, and QS estimators tend to have sizes close to each other. Among the ATF, ATFF, and QS estimators, the ATF and QS estimators often, but not always, lead to the smallest and largest size distortion, respectively.

Table 6 around here.

In summary, the relationship between the ATFF and QS estimators is quite consistent with what the large sample theory suggests, unlike the relationship between between the TF and QS estimators. The ATFF estimator seems to be (weakly) more efficient than the QS estimator in practical sample sizes. In terms of the size distortion in hypothesis testing, the ATFF estimator also performs slightly better or equally well, compared to the QS estimator.

9 Concluding Remarks

With the two modifications to the TF estimator proposed in this paper, the TF estimation delivers an estimate guaranteed to be positive semidefinite, enjoys the same large sample efficiency as the original TF estimator does, and shows small-sample performance better than or comparable to that of the QS estimator in terms of the MSE and the size distortion. In particular, the modifications make the relationship between the modified TF estimator and the QS estimator in small samples consistent with what the large sample theory suggests.

The method for adjustment of a matrix estimate for positive definiteness may be necessary and useful in applications other than long-run covariance matrix estimation. We use a general framework in Section 2 to keep our large sample results applicable in such applications. The adjustment of positive semidefiniteness may be in particular effective, when an available estimator does not necessarily deliver a positive semidefinite estimate and the sample size is small.

In this paper, we mainly focus on the accuracy in estimation of the long-run covariance matrix. On the other hand, Sun, Phillips, and Jin (2008) study the size and power properties of statistical inferences using long-run covariance matrix estimators, aiming at the bandwidth selection optimal for statistical inference. Their results seem to suggest that bias reduction in long-run covariance estimation can improve the size and power properties of a test. The ATFF estimator looks promising in this view, given the negligibility of its large-sample bias discussed in Section 7 of this paper, though the analytical framework used in Sun, Phillips, and Jin (2008) do not directly cover the TF estimator. Investigation of this topic is, however, beyond the scope of this paper and left for the future research.

Appendix

For each symmetric p.s.d. matrix A, $A^{1/2}$ denote a p.s.d. matrix such that $A^{1/2}A^{1/2} = A$. Also, for each $(a, b) \in \mathbb{R}^2$, $a \vee b$ and $a \wedge b$ denote the smaller and larger between a and b, respectively.

Proof of Theorem 2.1. Let T be an arbitrary natural number. Define $f : \mathbb{R}^{v \times v} \times \mathsf{P}_{v^2} \times \mathsf{P}_v \to \mathbb{R}$ by

$$f(\hat{s}, w, s) = \|\hat{s} - s\|_w, \quad (\hat{s}, w, s) \in \mathbb{R}^{v \times v} \times \mathsf{P}_{v^2} \times \mathsf{P}_v.$$

If the adjusted estimator exists, it picks s from P_v to minimize $f(\hat{S}_T, W_T, s) = \|\hat{S}_T, s\|_{W_T}$. Note that each finite dimensional normed linear space is complete and separable, and so are $\mathbb{R}^{v \times v} \times \mathsf{P}_{v^2} \times \mathsf{P}_v$ (the domain of f) and P_v (the projection of the domain of f onto the space of the third argument) endowed with the Euclidean metric. Also, $f(\hat{s}, w, \cdot) : \mathsf{P}_v \to \mathbb{R}$ is continuous for each $(\hat{s}, w) \in \mathbb{R}^{v \times v} \times \mathsf{P}_{v^2}$. It follows by Brown and Purves (1973, Corollary 1, pp. 904–905) that the adjusted estimator $\hat{S}_T^A$ exists, if for each $(\hat{s}, w) \in \mathbb{R}^{v \times v} \times \mathsf{P}_{v^2}$, $f(\hat{s}, w, \cdot) : \mathsf{P}_v \to \mathbb{R}$ attains its minimum on P_v.

Pick $\hat{s} \in \mathbb{R}^{v \times v}$ and $w \in \mathsf{P}_{v^2}$ arbitrarily. To establish the desired result, it suffices to show that there exists $\hat{s}^A \in \mathsf{P}_v$ such that $f(\hat{s}, w, \hat{s}^A) = r \equiv \inf_{s \in \mathsf{P}_v} f(\hat{s}, w, s)$. Because w is a symmetric p.s.d. matrix, there exists a full column rank matrix A such that $w = AA'$. With this A, we have that for each $s \in \mathbb{R}^{v \times v}$,

$$f(\hat{s}, w, s) = \|A' \operatorname{vec}(s - \hat{s})\|.$$

Write $\hat{x} \equiv A' \operatorname{vec}(\hat{s})$ and $\mathsf{V} \equiv \{A' \operatorname{vec}(s) : s \in \mathsf{P}_v\}$. Then we have that

$$r = \inf_{s \in \mathsf{P}_v} \|A' \operatorname{vec}(s - \hat{s})\| = \inf_{x \in \mathsf{V}} \|x - \hat{x}\|.$$

If there exists $\hat{x}^A \in \mathsf{V}$ such that $\|\hat{x}^A - \hat{x}\| = r$, there also exists $\hat{s}^A \in \mathsf{P}_v$ such that $f(\hat{s}, w, s) = r$, and the desired result follows.

Let $\mathsf{V}_{2r} \equiv \{x \in \mathsf{V} : \|x - \hat{x}\| \le 2r\}$. Because V_{2r} is a bounded closed subset of a finite-dimensional Euclidean space, it is compact, so that there exists $\hat{x}^A \in \mathsf{V}_{2r}$ such that $\|\hat{x}^A - \hat{x}\| = \inf_{x \in \mathsf{V}_{2r}} \|x - \hat{x}\|$. Further,

$$\inf_{x \in \mathsf{V} \backslash \mathsf{V}_{2r}} \|x - \hat{x}\| \ge 2r > r.$$

It follows that $\inf_{x \in \mathsf{V}_{2r}} \|x - \hat{x}\| = r$. The desired result therefore follows. $\hfill$ *Q.E.D.*

Proof of Theorem 2.2. To prove (a), pick $\omega \in \Omega$ and $T \in \mathbb{N}$ arbitrarily. If $\|\hat{S}_T^A(\omega) - \hat{S}_T(\omega)\|_{W_T(\omega)} = 0$, it holds that $W_T(\omega)\operatorname{vec}(\hat{S}_T^A(\omega) - \hat{S}_T(\omega)) = W_T(\omega)^{1/2} W_T(\omega)^{1/2}\operatorname{vec}(\hat{S}_T^A(\omega) - \hat{S}_T(\omega)) = 0$, because

$$\|\hat{S}_T^A(\omega) - \hat{S}_T(\omega)\|_{W_T} = (W_T(\omega)^{1/2}\operatorname{vec}(\hat{S}_T^A(\omega) - \hat{S}_T(\omega)))'(W_T(\omega)^{1/2}\operatorname{vec}(\hat{S}_T^A(\omega) - \hat{S}_T(\omega))).$$

It follows that if $\|\hat{S}_T^A(\omega) - \hat{S}_T(\omega)\|_{W_T(\omega)} = 0$,

$$\|\hat{S}_T^A(\omega) - S_T\|_{W_T(\omega)}^2 = \|(\hat{S}_T^A(\omega) - \hat{S}_T(\omega)) + (\hat{S}_T(\omega) - S_T)\|_{W_T(\omega)}^2$$
$$= \|\hat{S}_T^A(\omega) - \hat{S}_T(\omega)\|_{W_T}^2 + 2\operatorname{vec}(\hat{S}_T^A(\omega) - \hat{S}_T(\omega))' W_T(\omega) \operatorname{vec}(\hat{S}_T(\omega) - S_T) + \|\hat{S}_T(\omega) - S_T\|_{W_T(\omega)}^2$$
$$= \|\hat{S}_T(\omega) - S_T\|_{W_T(\omega)}^2.$$

For (b), note that the desired inequality holds by (a), whenever $\|\hat{S}_T^A - \hat{S}_T\|_{W_T} = 0$. Fix $T \in \mathbb{N}$ and $\omega \in \Omega$ arbitrarily and suppose that $\hat{S}_T(\omega) \notin \mathsf{P}_v$. Write $\hat{s}^A \equiv \hat{S}_T^A(\omega)$ and $w \equiv \hat{W}_T(\omega)$, and let A, V, $\hat{x}$, and $\hat{x}^A$ be as in the proof of Theorem 2.1. Also, let $\bar{x} \equiv A'\mathrm{vec}(S_T)$, and let B denote the Euclidean closed ball in $\mathbb{R}^{\mathrm{rank}(A)}$ with radius $\|\hat{x}^A - \hat{x}\|$ centered at $\hat{x}$. Because

$$\|\hat{x}^A - \hat{x}\| = \|\hat{s}^A - \hat{s}\|_w = \inf_{s \in \mathsf{P}} \|s - \hat{s}\|_w = \inf_{x \in \mathsf{V}} \|x - \hat{x}\|,$$

it clearly holds that $\mathsf{V} \cap \mathrm{int}\,\mathsf{B} = \emptyset$. Also, V is convex, and B is convex with a nonempty interior, because $\|\hat{s}^A - \hat{s}\|_w > 0$ by hypothesis. By the Eidelheit Separation Theorem (Luenberger 1969, pp. 133–134, Theorem 3), there exists a hyperplane H_1 separating V and B. Because $\hat{x}^A$ belongs to both V and B, H_1 contains $\hat{x}^A$, so that H_1 is the unique tangency plane of the Euclidean closed ball B at $\hat{x}^A$.

Now shift H_1 so that it contains $\bar{x}$ and call the resulting hyperplane H_2. Let $\check{x}$ be the projection of $\hat{x}$ onto H_2. Then $\hat{x}^A$ is on the line segment connecting $\hat{x}$ and $\check{x}$, and $\check{x} - \bar{x}$ is perpendicular to both $\hat{x} - \check{x}$ and $\hat{x}^A - \check{x}$. We thus have that

$$\|\hat{s}_T - S_T\|_w^2 = \|\hat{x} - \bar{x}\|^2 = \|\hat{x} - \check{x}\|^2 + \|\check{x} - \bar{x}\|^2$$
$$\geq \|\hat{x}^A - \check{x}\|^2 + \|\check{x} - \bar{x}\|^2 = \|\hat{x}^A - \bar{x}\|^2 = \|\hat{s}^A - S_T\|_w^2.$$

The desired result therefore follows. $\hfill$ *Q.E.D.*

Proof of Corollary 2.3. Claim (a) immediately follows from Theorem 2.2. To prove (b), let $\lambda_{1,T}$ and $\lambda_{2,T}$ respectively denote the minimum and maximum eigenvalues of W_T. Then we have that

$$\lambda_{1,T}\|\hat{S}_T^A - S_T\|^2 \leq \|\hat{S}_T^A - S_T\|_{W_T}^2 \leq \|\hat{S}_T - S_T\|_{W_T}^2 \leq \lambda_{2,T}\|\hat{S}_T - S_T\|^2,$$

where the second inequality follows from Theorem 2.2. When $W_T = \mathrm{O}_P(1)$, it holds that $\lambda_{2,T} = \mathrm{O}_P(1)$, and the right-hand side of the above inequality converges in probability-P to zero. The first claim of (b) therefore follows. For the second claim of (b), let λ_1 and λ_2 be the minimum and maximum eigenvalues of W. Then $\lambda_1 > 0$, and $\{\lambda_{2,T}/\lambda_{1,T}\}_{T \in \mathbb{N}}$ converges in probability-P to λ_2/λ_1. Because $\|\hat{S}_T^A - S_T\|^2 \leq (\lambda_{2,T}/\lambda_{1,T})\|\hat{S}_T - S_T\|^2$, whenever $\lambda_{1,T} > 0$, the desired result follows.

The first claim of (c) immediately follows from Theorem 2.2(a). For the second claim, note that $\lambda_{1,T}\|\hat{S}_T^A - \hat{S}\| = \mathrm{o}_P(b_T)$, where $\lambda_{1,T}$ converges to $\lambda_1 > 0$. The desired result therefore follows from the first claim.

$\hfill$ *Q.E.D.*

Proof of Theorem 2.4. The sequence $\{a_T\|\hat{S}_T - S_T\|^2_{W_T}\}_{T\in\mathbb{N}}$ converges in probability-P to zero, because $P[\hat{S}_T - S_T = 0] \to 1$ by the consistency of $\{\hat{S}_T\}$ for $\{S_T\}$ and the asymptotic uniform positive definiteness of $\{S_T\}$. Because

$$\text{MSE}(a_T,\ \hat{S}_T,\ W_T) - \text{MSE}(a_T,\ \hat{S}^A_T,\ W_T)$$
$$= \text{E}[a_T\|\hat{S}_T - S_T\|^2_{W_T} - a_T\|\hat{S}^A_T - S_T\|^2_{W_T}],$$

and $\{a_T\|\hat{S}_T - S_T\|^2_{W_T} - a_T\|\hat{S}^A_T - S_T\|^2_{W_T}\}_{T\in\mathbb{N}}$ is uniformly integral under the current assumption, it suffices to show that $\{a_T\|\hat{S}_T - S_T\|^2_{W_T} - a_T\|\hat{S}^A_T - S_T\|^2_{W_T}\}$ converges to zero in probability-P. Let ϵ be an arbitrary positive real number. Then

$$P[|a_T\|\hat{S}_T - S_T\|^2_{W_T} - a_T\|\hat{S}^A_T - S_T\|^2_{W_T}| > \epsilon]$$
$$\leq P[|a_T\|\hat{S}_T - S_T\|^2_{W_T} - a_T\|\hat{S}^A_T - S_T\|^2_{W_T}| \neq 0] \leq P[\hat{S}_T \notin \mathsf{P}] \to 0,$$

where the last inequality follows from the consistency of $\{\hat{S}_T\}$ for $\{S_T\}$ and the asymptotic uniform positive definiteness of $\{S_T\}$ by Theorem 2.2(a). The result therefore follows. *Q.E.D.*

Proof of Theorem 2.5. The result immediately follows from Theorem 2.2(b). *Q.E.D.*

Proof of Lemma 2.6. Note that for each $h \in (0, \infty)$,

$$\min\{a_T\|\hat{S}_{2,T} - S_T\|^2_{W_T},\ h\} - \min\{a_T\|\hat{S}_{1,T} - S_T\|^2_{W_T},\ h\}$$
$$\leq |a_T\|\hat{S}_{2,T} - S_T\|^2_{W_T} - a_T\|\hat{S}_{1,T} - S_T\|^2_{W_T}|$$
$$= \left(a_T^{1/2}\|\hat{S}_{2,T} - S_T\|_{W_T} + a_T^{1/2}\|\hat{S}_{1,T} - S_T\|_{W_T}\right)$$
$$\times \left(a_T^{1/2}\|\hat{S}_{2,T} - S_T\|_{W_T} - a_T^{1/2}\|\hat{S}_{1,T} - S_T\|_{W_T}\right), \quad T \in \mathbb{N}, \tag{A.1}$$

where the inequality holds because for each $h \in (0, \infty)$, $x \mapsto \min\{x_1, h\} : \mathbb{R} \to \mathbb{R}$ is a nondecreasing Lipschitz function with the maximum slope equal to one. The first factor on the right-hand side of (A.1) is $\text{O}_P(1)$, while the second factor converges in probability-P to zero, because

$$\left|\|\hat{S}_{2,T} - S_T\|_{W_T} - \|\hat{S}_{1,T} - S_T\|_{W_T}\right| \leq \|(\hat{S}_{2,T} - S_T) - (\hat{S}_{1,T} - S_T)\|_{W_T}$$
$$= \|\hat{S}_{2,T} - \hat{S}_{1,T}\|_{W_T} = \text{o}_P(a_T^{-1/2}).$$

28

Thus, we have that for each $h \in (0, \infty)$

$$\min\{a_T \|\hat{S}_{2,T} - S_T\|^2_{W_T}, h\} - \min\{a_T \|\hat{S}_{1,T} - S_T\|^2_{W_T}, h\} = o_P(1).$$

Because the left-hand side of this equality is bounded, the mean of it converges to zero by Andrews (1991, Lemma A1). Thus, for each $h \in (0, \infty)$ for which $\{\text{MSE}_h(a_T, \hat{S}_{1,T}, W_T)\}_{T \in \mathbb{N}}$ converges to a real number, $\{\text{MSE}_h(a_T, \hat{S}_{1,T}, W_T)\}_{t \in \mathbb{N}}$ converges to the same number. *Q.E.D.*

Proof of Theorem 2.7. By Corollary 2.3(c), $a_T^{1/2}(\hat{S}_T^A - \hat{S}_T) \to 0$ in probability-P. Applying Lemma 2.6, taking $\{\hat{S}_T\}$ for $\{\hat{S}_{1,T}\}$ and $\{\hat{S}_T^A\}$ for $\{\hat{S}_{2,T}\}$ yields the desired result. *Q.E.D.*

Proof of Proposition 4.1. The inequality in (6) immediately follows from Corollary 2.3(a), and the equality in (7) can be obtained by applying Andrews (1991, Proposition 1(c)). For claim (b), apply Theorem 2.4 with the fact that $\{(T/m_T)(\tilde{S}_T^{TF}(m_T) - S_T)^2\}_{T \in \mathbb{N}}$ is uniformly integrable under the current assumptions (see Andrews (1991, pages 827 and 853)). *Q.E.D.*

Proof of Theorem 4.2. Claim (a) follows from the consistency of $\{\hat{S}_T^{TF}(m_T)\}_{T \in \mathbb{N}}$ for $\{S_T\}_{T \in \mathbb{N}}$ (Andrews 1991, Theorem 1(a)) by Corollary 2.3(b). Claims (b) and (c) can be established by applying Corollary 2.3(c) with Andrews (1991, Theorem 1(b)). For claim (d), (8) follows from Theorem 2.7, while (9) and (10) are given by Andrews (1991, Theorem 1(c)). *Q.E.D.*

Proof of Lemma 5.1. (a) Let a and b be arbitrary natural numbers not exceeding p. Also, let $\tilde{\Gamma}_{T,a,b}(\tau)$ and $\Gamma_{T,a,b}(\tau)$ respectively denote the (a, b)-elements of $\tilde{\Gamma}_T(\tau)$ and $\Gamma_T(\tau)$, $T \in \mathbb{N}$, $\tau \in \{0, 1, \ldots, T-1\}$. For each $T \in \mathbb{N}$ and each $\tau = 0, 1, \ldots, T-1$, define

$$\xi_{T,a,b}(t, \tau) \equiv Z^*_{T,t,a} Z^*_{T,t-\tau,b} - \text{E}[Z^*_{T,t,a} Z^*_{T,t-\tau,b}], \quad t \in \{\tau+1, \ldots, T\}, \tag{A.2}$$

where $Z^*_{T,t,a}$ denote the ath element of $Z^*_{T,t}$. Then we have that for each $T \in \mathbb{N}$ and each $\tau \in \{0, 1, \ldots, T-1\}$,

$$\text{E}[\,(\tilde{\Gamma}_T(\tau) - \Gamma_T(\tau))^2\,] = \text{E}\left[\left(T^{-1} \sum_{t=\tau+1}^{T} \xi_{T,a,b}(t, \tau)\right)^2\right]$$

$$= T^{-2} \text{var}\left[\sum_{t=\tau+1}^{T} \xi_{T,a,b}(t, \tau)\right].$$

To establish the desired result, it suffices to show that for some positive real number $A_{a,b}$ independent of T,

$$\sup_{\tau \in \{0,1,\cdots,T-1\}} \text{var}\left[\sum_{t=\tau+1}^{T} \xi_{T,a,b}(t, \tau)\right] \leq A_{a,b} T, \quad T \in \mathbb{N}. \tag{A.3}$$

Under the fourth order stationarity, as Hannan (1970, equation (3.3), p. 209) shows, we have that

$$\text{var}\left[\sum_{t=\tau+1}^{T} \xi_{a,b,T}(t,\tau)\right] = M_{T,\tau,1} + M_{T,\tau,2} + M_{T,\tau,3},$$

where

$$M_{T,\tau,1} = \sum_{s=\tau+1}^{T} \sum_{t=\tau+1}^{T} \gamma_{a,a}(t-s)\,\gamma_{b,b}(t-s)$$

$$M_{T,\tau,2} = \sum_{s=\tau+1}^{T} \sum_{t=\tau+1}^{T} \gamma_{a,b}(t-s+\tau)\,\gamma_{a,b}(t-s-\tau)$$

$$M_{T,\tau,3} = \sum_{s=\tau+1}^{T} \sum_{t=\tau+1}^{T} \kappa_{a,b,a,b}(-\tau, s-t, s-t+\tau).$$

By setting $k = t - s$, we obtain

$$\sup_{\tau\in\{0,1,\cdots,T-1\}} |M_{T,\tau,1}| = \sup_{\tau\in\{0,1,\cdots,T-1\}}\left|\sum_{t=\tau+1}^{T} \sum_{k=\tau+1-t}^{T-t} \gamma_{a,a}(k)\,\gamma_{b,b}(k)\right|$$

$$\leq \sum_{t=1}^{T} \sum_{k=-T+1}^{T-1} |\gamma_{a,a}(k)\,\gamma_{b,b}(k)|$$

$$\leq T \sum_{k=-T+1}^{T-1} |\gamma_{a,a}(k)\,\gamma_{b,b}(k)|$$

$$\leq A_{a,b,1} T,$$

where $A_{a,b,1} \equiv \left(\sum_{k=-\infty}^{\infty} |\gamma_{a,a}(k)|\right)\left(\sum_{l=-\infty}^{\infty} |\gamma_{b,b}(l)|\right)$. Analogously, we have that

$$\sup_{\tau\in\{0,1,\cdots,T-1\}} |M_{T,\tau,2}| = \sup_{\tau\in\{0,1,\cdots,T-1\}}\left|\sum_{t=\tau+1}^{T} \sum_{l=1-t}^{T-t-\tau} \gamma_{a,b}(l+2\tau)\,\gamma_{a,b}(l)\right|$$

$$\leq \sup_{\tau\in\{0,1,\cdots,T-1\}} \sum_{t=1}^{T} \sum_{l=-T+1}^{T-1-2\tau} |\gamma_{a,b}(l+2\tau)\,\gamma_{a,b}(l)|$$

$$= \sup_{\tau\in\{0,1,\cdots,T-1\}} T \sum_{l=-T+1}^{T-1-2\tau} |\gamma_{a,b}(l+2\tau)\,\gamma_{a,b}(l)|$$

$$\leq A_{a,b,2} T,$$

where $A_{a,b,2} \equiv \left(\sum_{i=-\infty}^{\infty} |\gamma_{a,b}(i)|\right)^2$. Further,

$$\sup_{\tau\in\{0,1,\cdots,T-1\}} |M_{T,\tau,3}| = \sup_{\tau\in\{0,1,\cdots,T-1\}}\left|\sum_{t=\tau+1}^{T} \sum_{k=\tau+1-t}^{T-t} \kappa_{a,b,a,b}(-\tau,-k,-k+\tau)\right| \leq A_{a,b,3},$$

30

where $A_{a,b,3} \equiv \sum_{t,k,l=-\infty}^{\infty} |\kappa_{a,b,a,b}(t,k,l)|$. Thus,

$$\sup_{\tau \in \{0,1,\cdots,T-1\}} \text{var}\left[\sum_{t=\tau+1}^{T} \xi_{T,a,b}(t,\tau)\right] \leq (A_{a,b,1} + A_{a,b,2})T + A_{a,b,3} \leq A_{a,b}T \tag{A.4}$$

where $A_{a,b} \equiv A_{a,b,1} + A_{a,b,2} + A_{a,b,3}$. The desired result therefore follows.

(b) The left-hand side of the equality in question can be rewritten as

$$m_T^q \text{E}[\tilde{\Gamma}_T(\lfloor m_T \rfloor + 1)] = m_T^q \Gamma_T(\lfloor m_T \rfloor + 1) = \frac{T - \lfloor m_T \rfloor - 1}{T} m_T^q \Gamma(\lfloor m_T + 1 \rfloor)$$

In this equality, the right-hand side converges to zero, because $\{(T - \lfloor m_T \rfloor - 1)/T\}_{T \in \mathbb{N}}$ is nonnegative and no greater than one, while $m_T^q \Gamma(\lfloor m_T + 1 \rfloor)$ converges to zero, as $S^{(q)}$ converges. The desired result therefore follows.

(c) Because the second moment of a random variable is equal to the sum of the mean and variance of the random variable, we have that

$$(T/m_T)\text{E}[\|\tilde{\Gamma}_T(\lfloor m_T + 1 \rfloor)\|^2]$$

$$= (T/m_T)\|\text{E}[\tilde{\Gamma}_T(\lfloor m_T + 1 \rfloor)]\|^2 + (T/m_T)\text{E}[\|\tilde{\Gamma}_T(\lfloor m_T + 1 \rfloor) - \Gamma_T(\lfloor m_T + 1 \rfloor)\|^2], \quad T \in \mathbb{N}.$$

By applying (b) and (a) to the first and second terms on the right-hand side of this quality, respectively, we see that the first term is $o(T/m_T^{2q+1}) = o(1)$, because $T/m_T^{2q+1} \to \gamma^{-1}$, and the second term is $O(m_T^{-1}) = o(1)$. The result therefore follows. *Q.E.D.*

Proof of Proposition 5.2. Note that

$$\left| \|\tilde{S}_T^{TFF}(m_T) - S_T\|_W^2 - \|\tilde{S}_T^{TF}(m_T) - S_T\|_W^2 \right|$$

$$= \left| \|\tilde{S}_T^{TFF}(m_T) - S_T\|_W - \|\tilde{S}_T^{TF}(m_T) - S_T\|_W \right|$$

$$\times \left(\|\tilde{S}_T^{TFF}(m_T) - S_T\|_W + \|\tilde{S}_T^{TF}(m_T) - S_T\|_W \right), \quad T \in \mathbb{N}.$$

Because $\|\cdot\|_W$ is a pseudo norm, we have that

$$\left| \|\tilde{S}_T^{TFF}(m_T) - S_T\|_W - \|\tilde{S}_T^{TF}(m_T) - S_T\|_W \right| \leq \|(\tilde{S}_T^{TFF}(m_T) - S_T) - (\tilde{S}_T^{TF}(m_T) - S_T)\|_W$$

$$\leq \|\tilde{S}_T^{TFF}(m_T) - \tilde{S}_T^{TF}(m_T)\|_W = (m - \lfloor m_T \rfloor) \left\| \tilde{\Gamma}_T(\lfloor m_T \rfloor + 1) + \tilde{\Gamma}_T(\lfloor m_T \rfloor + 1)' \right\|_W$$

$$\leq \|\tilde{\Gamma}_T(\lfloor m_T \rfloor + 1)\|_W + \|\tilde{\Gamma}_T(\lfloor m_T \rfloor + 1)'\|_W \leq 2\lambda^{1/2} \|\tilde{\Gamma}_T(\lfloor m_T \rfloor + 1)\|, \quad T \in \mathbb{N},$$

where λ is the maximum eigenvalue of W. We also have that

$$\left| \|\tilde{S}_T^{TFF}(m_T) - S_T\|_W + \|\tilde{S}_T^{TF}(m_T) - S_T\|_W \right|$$

$$= \left| \|\tilde{S}_T^{TF}(m_T) + (m - \lfloor m_T \rfloor)\tilde{\Gamma}_T(\lfloor m_T \rfloor + 1) - S_T\|_W + \|\tilde{S}_T^{TF}(m_T) - S_T\|_W \right|$$

$$\leq \|\tilde{S}_T^{TF}(m_T) - S_T\|_W + (m - \lfloor m_T \rfloor)\|\tilde{\Gamma}_T(\lfloor m_T \rfloor + 1)\|_W + \|\tilde{S}_T^{TF}(m_T) - S_T\|_W$$

$$\leq 2\|\tilde{S}_T^{TF}(m_T) - S_T\|_W + \|\tilde{\Gamma}_T(\lfloor m_T \rfloor + 1)\|_W$$

$$\leq 2\|\tilde{S}_T^{TF}(m_T) - S_T\|_W + \lambda^{1/2}\|\tilde{\Gamma}_T(\lfloor m_T \rfloor + 1)\|, \quad T \in \mathbb{N}.$$

Thus,

$$\left| \|\tilde{S}_T^{TFF}(m_T) - S_T\|_W^2 - \|\tilde{S}_T^{TF}(m_T) - S_T\|_W^2 \right|$$

$$\leq 2\lambda^{1/2}\|\tilde{\Gamma}_T(\lfloor m_T \rfloor + 1)\| \left(2\|\tilde{S}_T^{TF}(m_T) - S_T\|_W + \lambda^{1/2}\|\tilde{\Gamma}_T(\lfloor m_T \rfloor + 1)\| \right)$$

By taking the expectation of both sides of this inequality and applying the Cauchy-Schwarz inequality and the Minkowski inequality, we obtain that

$$\mathrm{E}\left[\left| \|\tilde{S}_T^{TFF}(m_T) - S_T\|_W^2 - \|\tilde{S}_T^{TF}(m_T) - S_T\|_W^2 \right| \right]$$

$$\leq 2\lambda^{1/2}\,\mathrm{E}\left[\|\tilde{\Gamma}_T(\lfloor m_T \rfloor + 1)\| \left(\|\tilde{S}_T^{TF}(m_T) - S_T\|_W + \lambda^{1/2}\|\tilde{\Gamma}_T(\lfloor m_T \rfloor + 1)\| \right) \right]$$

$$\leq 2\lambda^{1/2}\,\mathrm{E}\left[\|\tilde{\Gamma}_T(\lfloor m_T \rfloor + 1)\|^2 \right]^{1/2}\mathrm{E}\left[\left(\|\tilde{S}_T^{TF}(m_T) - S_T\|_W + \lambda^{1/2}\|\tilde{\Gamma}_T(\lfloor m_T \rfloor + 1)\| \right)^2 \right]^{1/2}$$

$$\leq 2\lambda^{1/2}\,\mathrm{E}\left[\|\tilde{\Gamma}_T(\lfloor m_T \rfloor + 1)\|^2 \right]^{1/2} \left(\mathrm{E}\left[\|\tilde{S}_T^{TF}(m_T) - S_T\|_W^2 \right]^{1/2} + \lambda^{1/2}\mathrm{E}\left[\|\tilde{\Gamma}_T(\lfloor m_T \rfloor + 1)\|^2 \right]^{1/2} \right),$$

$T \in \mathbb{N}$. It follows that

$$\left| \mathrm{MSE}(T/m_T, \tilde{S}_T^{TFF}(m_T), W) - \mathrm{MSE}(T/m_T, \tilde{S}_T^{TF}(m_T), W) \right|$$

$$= \left| \mathrm{E}\left[(T/m_T)\tilde{S}_T^{TFF}(m_T) - S_T\|_W^2 - (T/m_T)\|\tilde{S}_T^{TF}(m_T) - S_T\|_W^2 \right] \right|$$

$$\leq (T/m_T)\mathrm{E}\left[\left| \|\tilde{S}_T^{TFF}(m_T) - S_T\|_W^2 - \|\tilde{S}_T^{TF}(m_T) - S_T\|_W^2 \right| \right]$$

$$\leq 2\lambda^{1/2}\,\mathrm{E}\left[(T/m_T)\|\tilde{\Gamma}_T(\lfloor m_T \rfloor + 1)\|^2 \right]^{1/2}$$

$$\times \left(\mathrm{E}\left[(T/m_T)\|\tilde{S}_T^{TF}(m_T) - S_T\|_W^2 \right]^{1/2} + \lambda^{1/2}\mathrm{E}\left[(T/m_T)\|\tilde{\Gamma}_T(\lfloor m_T \rfloor + 1)\|^2 \right]^{1/2} \right)$$

$$\leq 2\lambda^{1/2}\,\mathrm{E}\left[(T/m_T)\|\tilde{\Gamma}_T(\lfloor m_T \rfloor + 1)\|^2 \right]^{1/2}$$

$$\times \left(\mathrm{MSE}(T/m_T, \tilde{S}_T^{TF}(m_T), W)^{1/2} + \lambda^{1/2}\mathrm{E}\left[(T/m_T)\|\tilde{\Gamma}_T(\lfloor m_T \rfloor + 1)\|^2 \right]^{1/2} \right), \quad T \in \mathbb{N}.$$

The desired result follows from this inequality, because the right-hand of this inequality converges to zero by Lemma 5.1(c) and Andrews (1991, Proposition 1(c)). *Q.E.D.*

Proof of Proposition 5.3. The inequality in (13) immediately follows from Corollary 2.3(a), and the equalities in (14) have been established in Proposition 5.2. Thus, claim (a) holds. For claim (b), note that $\{(T/m_T)(\tilde{S}_T^{TF}(m_T) - S_T)^2\}_{T \in \mathbb{N}}$ and $\{(T/m_T)(\tilde{S}_T^{TF}(m_T + 1) - S_T)^2\}_{T \in \mathbb{N}}$ are uniformly integrable under the current assumptions (see Andrews (1991, pages 827 and 853)). Because

$$(T/m_T)\|\tilde{S}_T^{TFF}(m_T) - S_T\|_W^2$$

$$\leq (T/m_T)\big\|(m_T - \lfloor m_T \rfloor)(\tilde{S}_T^{TF}(m_T) - S_T) + (\lfloor m_T \rfloor + 1 - m_T)(\tilde{S}_T^{TF}(m_T + 1) - S_T)\big\|_W^2$$

$$\leq (T/m_T)\big(\|(m_T - \lfloor m_T \rfloor)(\tilde{S}_T^{TF}(m_T) - S_T)\|_W + \|(\lfloor m_T \rfloor + 1 - m_T)^2(\tilde{S}_T^{TF}(m_T + 1) - S_T)\|_W\big)^2$$

$$\leq 2(T/m_T)\big(\|(m_T - \lfloor m_T \rfloor)(\tilde{S}_T^{TF}(m_T) - S_T)\|_W^2 + \|(\lfloor m_T \rfloor + 1 - m_T)^2(\tilde{S}_T^{TF}(m_T + 1) - S_T)\|_W^2\big)$$

$$\leq 2(T/m_T)\|\tilde{S}_T^{TF}(m_T) - S_T\|_W^2 + 2(T/m_T)\|\tilde{S}_T^{TF}(m_T + 1) - S_T\|_W^2, \quad T \in \mathbb{N},$$

$\{(T/m_T)(\tilde{S}_T^{TFF}(m_T) - S_T)^2\}_{T \in \mathbb{N}}$ is also uniformly integrable. The desired result therefore follows by Theorem 2.4. *Q.E.D.*

Proof of Theorem 5.4. (a) If $\{m_T^2/T\}_{T \in \mathbb{N}}$ converges to zero, so does $\{(m_T + 1)^2/T\}_{T \in \mathbb{N}}$. It follows by Andrews (1991, Theorem 1(a)) that under the current assumptions, both $\{\hat{S}_T^{TF}(m_T)\}_{T \in \mathbb{N}}$ and $\{\hat{S}_T^{TF}(m_T + 1)\}_{T \in \mathbb{N}}$ are consistent for $\{S_T\}_{T \in \mathbb{N}}$. Because

$$\|\hat{S}_T^{TFF}(m_T) - S_T\| \leq \|\hat{S}_T^{TF}(m_T) - S_T\| + \|\hat{S}_T^{TF}(m_T + 1) - S_T\|, \quad T \in \mathbb{N}, \tag{A.5}$$

it follows that $\{\hat{S}_T^{TF}(m_T)\}_{T \in \mathbb{N}}$ is consistent for $\{S_T\}_{T \in \mathbb{N}}$.

(b) If $\{m_T^{2q+1}/T\}_{T \in \mathbb{N}}$ converges to γ, so does $\{(m_T + 1)^{2q+1}/T\}_{T \in \mathbb{N}}$. It follows by Andrews (1991, Theorem 1(b)) that under the current assumptions, we have that

$$(T/m_T)^{1/2}\|\hat{S}_T^{TF}(m_T) - S_T\| = O_P(1),$$

$$(T/m_T)^{1/2}\|\hat{S}_T^{TF}(m_T + 1) - S_T\| = O_P(1),$$

$$(T/m_T)^{1/2}\|\hat{S}_T^{TF}(m_T) - \tilde{S}_T^{TF}(m_T)\| = o_P(1),$$

and

$$(T/m_T)^{1/2}\|\hat{S}_T^{TF}(m_T + 1) - \tilde{S}_T^{TF}(m_T + 1)\| = o_P(1).$$

33

Thus,

$$(T/m_T)^{1/2}\|\hat{S}_T^{TFF}(m_T) - S_T\|$$

$$\leq (T/m_T)^{1/2}\|\hat{S}_T^{TF}(m_T) - S_T\| + (T/m_T)^{1/2}\|\hat{S}_T^{TF}(m_T + 1) - S_T\| = O_P(1)$$

and

$$(T/m_T)^{1/2}\|\hat{S}_T^{TFF}(m_T) - \tilde{S}_T^{TFF}(m_T)\|$$

$$\leq (T/m_T)^{1/2}\|\hat{S}_T^{TF}(m_T) - \tilde{S}_T^{TF}(m_T)\| + (T/m_T)^{1/2}\|\hat{S}_T^{TF}(m_T + 1) - \tilde{S}_T^{TF}(m_T + 1)\| = o_P(1).$$

(c) By (b) of the current theorem, $(T/m_T)^{1/2}(\hat{S}_T^{TFF}(m_T) - S_T) = O_P(1)$ and

$$(T/m_T)^{1/2}(\hat{S}_T^{TFF}(m_T) - \tilde{S}_T^{TFF}(m_T)) \to 0 \text{ in probability-}P, \tag{A.6}$$

so that

$$(T/m_T)^{1/2}(\tilde{S}_T^{TFF}(m_T) - S_T)$$

$$= (T/m_T)^{1/2}(\hat{S}_T^{TFF}(m_T) - S_T) - (T/m_T)^{1/2}(\hat{S}_T^{TFF}(m_T) - \tilde{S}_T^{TFF}(m_T)) = O_P(1). \tag{A.7}$$

Applying Lemma 2.6 with (A.6) and (A.7) establishes (15), while (16) and (17) are given by Proposition 5.2.

$$Q.E.D.$$

Proof of Theorem 5.5. Claim (a) follows from the consistency of $\{\hat{S}_T^{TFF}(m_T)\}_{T \in \mathbb{N}}$ for $\{S_T\}_{T \in \mathbb{N}}$ stated in Theorem 5.4(a) by Corollary 2.3. Claims (b) and (c) can be established by applying Corollary 2.3(c) with Theorem 5.4(b). For claim (d), (18) follows from Theorem 2.7, while (9) and (10) are given by Theorem 5.4(c).

$$Q.E.D.$$

Proof of Theorem 7.1. (a) Suppose that

$$\hat{S}_T^{TF}(\hat{m}_T) - \hat{S}_T^{TF}(m_T) \to 0 \text{ in probability-}P. \tag{A.8}$$

Then $\{\hat{S}_T^{TF}(\hat{m}_T)\}_{T \in \mathbb{N}}$ is consistent for $\{S_T\}$, because

$$\|\hat{S}_T^{TF}(\hat{m}_T) - S_T\| \leq \|\hat{S}_T^{TF}(\hat{m}_T) - \hat{S}_T^{TF}(m_T)\| + \|\hat{S}_T^{TF}(m_T) - S_T\|, \quad T \in \mathbb{N},$$

where the first term on the right-hand side converges to zero by hypothesis, and the second term converges in probability-P to zero by Andrews (1991, Theorem 1(a)). Also, $\{\hat{S}_T^{TFF}(\hat{m}_T)\}_{T \in \mathbb{N}}$ is consistent for

34

$\{S_T\}$ by Theorem 5.4(a). The convergences of (21) and (23) respectively follow from the consistency of $\{\hat{S}_T^{TF}(\hat{m}_T)\}$ and $\{\hat{S}_T^{TFF}(\hat{m}_T)\}$ by the Slutsky Theorem. Given (21) and (23), applying Corollary 2.3(b) to $\{\hat{S}_T^{TF,A}(\hat{m}_T)\}_{T \in \mathbb{N}}$ and $\{\hat{S}_T^{TFF,A}(\hat{m}_T)\}_{T \in \mathbb{N}}$ establishes the the rest of the claims in (a). Thus, it suffices to show (A.8).

The condition (A.8) is equivalent to that for each $\epsilon \in (0,1]$,

$$P[\|\hat{S}_T^{TF}(\hat{m}_T) - \hat{S}_T^{TF}(m_T)\| \geq \epsilon] < \epsilon \text{ for almost all } T \in \mathbb{N}. \tag{A.9}$$

Pick $\epsilon \in (0,1]$ arbitrarily. Then, by Assumption 8, there exists $\Delta_\epsilon \in (1, \infty)$ such that for each $T \in \mathbb{N}$, $P[\hat{m}_T \notin [(1/\Delta_\epsilon)m_T, \Delta_\epsilon m_T]] < \epsilon/2$. When $\hat{m}_T \in [(1/\Delta_\epsilon)m_T, \Delta_\epsilon m_T]$, we have that

$$\|\hat{S}_T^{TF}(\hat{m}_T) - \hat{S}_T^{TF}(m_T)\| = \left\| \sum_{\tau=\lfloor \hat{m}_T \wedge m_T \rfloor+1}^{\lfloor \hat{m}_T \vee m_T \rfloor} \left(\hat{\Gamma}_T(\tau) + \hat{\Gamma}_T'(\tau) \right) \right\|$$

$$\leq \left\| \sum_{\tau=\lfloor \hat{m}_T \wedge m_T \rfloor+1}^{\lfloor \hat{m}_T \vee m_T \rfloor} \left((\hat{\Gamma}_T(\tau) - \Gamma_T(\tau)) + (\hat{\Gamma}_T(\tau) - \Gamma_T(\tau))' \right) \right\| + \left\| \sum_{\tau=\lfloor \hat{m}_T \wedge m_T \rfloor+1}^{\lfloor \hat{m}_T \vee m_T \rfloor} \left(\Gamma_T(\tau) + \Gamma(\tau)' \right) \right\|$$

$$\leq 2A_{1,T} + 2A_{2,T}, \quad T \in \mathbb{N}, \tag{A.10}$$

where

$$A_{1,T} \equiv \sum_{\tau=\lfloor (1/\Delta_\epsilon)m_T \rfloor+1}^{\lfloor \Delta_\epsilon m_T \rfloor} \|\hat{\Gamma}_T(\tau) - \Gamma_T(\tau))\|, \quad T \in \mathbb{N}$$

and

$$A_{2,T} \equiv \sum_{\tau=\lfloor (1/\Delta_\epsilon)m_T \rfloor+1}^{\lfloor \Delta_\epsilon m_T \rfloor} \|\Gamma_T(\tau)\|, \quad T \in \mathbb{N}.$$

By using the Minkowski inequality and Lemma 5.1(a), we obtain that

$$\mathrm{E}[A_{1,T}^2]^{1/2} \leq \sum_{\tau=\lfloor (1/\Delta_\epsilon)m_T \rfloor+1}^{\lfloor \Delta_\epsilon m_T \rfloor} \mathrm{E}[\|\hat{\Gamma}_T(\tau) - \Gamma_T(\tau))\|^2]^{1/2} = O(m_T/T^{1/2}) = o(1).$$

By the Markov inequality, it follows that $A_{1,T} \to 0$ in probability-P. Also, the absolute convergence of $S^{(0)}$ implies that

$$A_{2,T} \leq \sum_{\tau=\lfloor (1/\Delta_\epsilon)m_T \rfloor+1}^{\lfloor \Delta_\epsilon m_T \rfloor} \|\Gamma(\tau)\| \leq \sum_{\tau=\lfloor (1/\Delta_\epsilon)m_T \rfloor+1}^{\infty} \|\Gamma(\tau)\| = o(1).$$

Thus, $2A_{1,T} + A_{2,T} \to 0$ in probability-P.

We now have that

$$P[\|\hat{S}_T^{TF}(\hat{m}_T) - \hat{S}_T^{TF}(m_T)\| \geq \epsilon] \leq P[\hat{m}_T \notin [(1/\Delta_\epsilon)m_T, \Delta_\epsilon m_T]] + P[2A_{1,T} + 2A_{2,T} \geq \epsilon], \quad T \in \mathbb{N}.$$

35

Because the first term on the right-hand side of this equality is no greater than $\epsilon/2$ for each $T \in \mathbb{N}$, while the second term is smaller than $\epsilon/2$ for almost all $T \in \mathbb{N}$, (A.9) holds. The desired result therefore follows.

(b) Suppose that (27) holds. Then (25) holds, because

$$\|(T/m_T)^{1/2}(\hat{S}_T^{TF}(\hat{m}_T) - S_T)\|$$
$$\leq \|(T/m_T)^{1/2}(\hat{S}_T^{TF}(\hat{m}_T) - \hat{S}_T^{TF}(m_T))\| + \|(T/m_T)^{1/2}(\hat{S}_T^{TF}(m_T) - S_T)\|, \quad T \in \mathbb{N},$$

where the first term on the right-hand side converges to zero by hypothesis, and the second term is $\mathrm{O}_P(1)$ by Andrews (1991, Theorem 1(b)). Also, (26) and (28) can be easily derived from (25) and (27), respectively, by using the definition of the TFF estimator and the triangle inequality. Given (25)–(28), it is straightforward to establish (29)–(36) by using the definition of $\|\cdot\|_{W_T}$ and applying the basic rules about stochastic order of magnitudes in additions and multiplications. Thus, it suffices to show (27) to prove the current claim.

The condition (27) is equivalent to that for each $\epsilon \in (0, 1]$,

$$P[(T/m_T)^{1/2}\|\hat{S}_T^{TF}(\hat{m}_T) - \hat{S}_T^{TF}(m_T)\| \geq \epsilon] < \epsilon \text{ for almost all } T \in \mathbb{N}. \tag{A.11}$$

Pick $\epsilon \in (0, 1]$ arbitrarily. Then, by Assumption 9, there exists $\Delta_\epsilon \in (0, \infty)$ such that for each $T \in \mathbb{N}$, $P[\hat{m}_T \notin [(1 - d_T^{-1}\Delta_\epsilon)m_T, (1 + d_T^{-1}\Delta_\epsilon)m_T]] < \epsilon/2$. Derivation analogous to (A.10) yields that when $\hat{m}_T \in [(1 - d_T^{-1}\Delta_\epsilon)m_T, (1 + d_T^{-1}\Delta_\epsilon)m_T]$,

$$\|(T/m_T)^{1/2}(\hat{S}_T^{TF}(\hat{m}_T) - \hat{S}_T^{TF}(m_T))\| \leq 2A_{3,T} + 2A_{4,T}, \quad T \in \mathbb{N},$$

where

$$A_{3,T} \equiv (T/m_T)^{1/2} \sum_{\tau = \lfloor(1 - d_T^{-1}\Delta_\epsilon)m_T\rfloor + 1}^{\lfloor(1 + d_T^{-1}\Delta_\epsilon)m_T\rfloor} \|\hat{\Gamma}_T(\tau) - \Gamma_T(\tau))\|, \quad T \in \mathbb{N}$$

and

$$A_{4,T} \equiv (T/m_T)^{1/2} \sum_{\tau = \lfloor(1 - d_T^{-1}\Delta_\epsilon)m_T\rfloor + 1}^{\lfloor(1 + d_T^{-1}\Delta_\epsilon)m_T\rfloor} \|\Gamma_T(\tau)\|, \quad T \in \mathbb{N}.$$

By using the Minkowski inequality and Lemma 5.1(a), we obtain that

$$\mathrm{E}[A_{3,T}^2]^{1/2} \leq (T/m_T)^{1/2} \sum_{\tau = \lfloor(1 - d_T^{-1}\Delta_\epsilon)m_T\rfloor + 1}^{\lfloor(1 + d_T^{-1}\Delta_\epsilon)m_T\rfloor} \mathrm{E}[\|\hat{\Gamma}_T(\tau) - \Gamma_T(\tau))\|^2]^{1/2} = O(d_T^{-1}m_T^{1/2}).$$

Because $d_T^{-1}m_T^{1/2} \to 0$ by Assumptions 9, it follows that $\mathrm{E}[A_{3,T}^2]^{1/2} \to 0$. By the Markov inequality, $A_{3,T} \to 0$ in probability-P. Also, we have that

$$A_{4,T} \le (T/m_T)^{1/2} \sum_{\tau=\lfloor (1-d_T^{-1}\Delta_\epsilon)m_T \rfloor+1}^{\lfloor (1+d_T^{-1}\Delta_\epsilon)m_T \rfloor} \|\Gamma_T(\tau)\| \le (T/m_T)^{1/2} \sum_{\tau=\lfloor (1-d_T^{-1}\Delta_\epsilon)m_T \rfloor+1}^{\lfloor (1+d_T^{-1}\Delta_\epsilon)m_T \rfloor} \|\Gamma(\tau)\|$$

$$\le (T/m_T)^{1/2} \sum_{\tau=\lfloor (1-d_T^{-1}\Delta_\epsilon)m_T \rfloor+1}^{\infty} \|\Gamma(\tau)\| \le (T/m_T)^{1/2} \sum_{\tau=\lfloor m_T/2 \rfloor+1}^{\infty} \|\Gamma(\tau)\|,$$

where the last inequality holds for almost all $T \in \mathbb{N}$, as $1 - d_T^{-1}\Delta_\epsilon \ge 1/2$ for almost all $T \in \mathbb{N}$. Write $\gamma_T \equiv (m_T^{2q+1}/T)$ for each $T \in \mathbb{N}$. Then $\{\gamma_T\}_{T \in \mathbb{N}}$ converges to Γ, and

$$(T/m_T)^{1/2} = (T/m_T)^{1/2}\gamma_T^{1/2}\gamma_T^{-1/2} = \gamma_T^{-1/2}m_T^q = 2^q\gamma_T^{-1/2}(m_T/2)^q, \quad T \in \mathbb{N}.$$

It follows that

$$A_{4,T} \le 2^q\gamma_T^{-1/2} \sum_{\tau=\lfloor m_T/2 \rfloor+1}^{\infty} \tau^q\|\Gamma(\tau)\| = o(1),$$

where the last equality follows by the absolute convergence of $S^{(q)}$.

We now have that

$$P[(T/m_T)^{1/2}\|\hat{S}_T^{TF}(\hat{m}_T) - \hat{S}_T^{TF}(m_T)\| \ge \epsilon]$$

$$\le P[\hat{m}_T \notin [(1 - d_T^{-1}\Delta_\epsilon)m_T, (1 + d_T^{-1}\Delta_\epsilon)m_T]] + P[2A_{3,T} + 2A_{4,T} \ge \epsilon], \quad T \in \mathbb{N}.$$

Because the first term on the right-hand side of this equality is no greater than $\epsilon/2$ for each $T \in \mathbb{N}$, and the second term is smaller than $\epsilon/2$ for almost all $T \in \mathbb{N}$, (A.11) holds, and the desired result follows.

(c) The results follow from (30), (32), (34), and (36) by arguments analogous to the proof of Theorem 2.3(b).

(d) The right-hand sides of (41)–(44) are equal to (45) by Theorems 4.2(d) and 5.5(d). In each (41)–(44), the equality of the left-hand side and the right-hand side follows from the corresponding result among (33)–(36) by Lemma 2.6. *Q.E.D.*

37

REFERENCES

ANDREWS, D. W. K. (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59(3), 817–858.

ANDREWS, D. W. K., AND J. C. MONAHAN (1992): "An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator," *Econometrica*, 60(4), 953–966.

BROWN, L. D., AND R. PURVES (1973): "Measurable Selections of Extrema," *The Annals of Statistics*, 1, 902–912.

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(455), 1088–1096.

DE JONG, R. M., AND J. DAVIDSON (2000): "Consistency of Kernel Estimators of Heteroscedastic and Autocorrelated Covariance Matrices," *Econometrica*, 68(2), 407–423.

DEN HAAN, W. J., AND A. T. LEVIN (1997): "A practitioner's guide to robust covariance matrix estimation.," in *Robust Inference*, ed. by G. S. Maddala, and C. Rao, pp. 299–342. North-Holland, Amsterdam.

——— (2000): "Robust Covariance Matrix Estimation with Data-Dependent VAR Prewhitening Order," Unpublished Mimeo.

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

HANNAN, E. J. (1970): *Multiple Time Series*, A Wiley Publications in Applied Statistics. Wiley.

HANSEN, B. E. (1992): "Consistent Covariance Matrix Estimation for Dependent Heterogeneous Processes," *Econometrica*, 60(4), 967–972.

HANSEN, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50(4), 1029–1054.

JANSSON, M. (2003): "Consistent Covariance Matrix Estimation for Linear Processes," *Econometric Theory*, 18(6), 1449–1459.

KIEFER, N. M., AND T. J. VOGELSANG (2000): "Simple Robust Testing of Regression Hypotheses," *Econometrica*, 68(3), 695–714.

——— (2002): "Heteroskedasticity-Autocorrelation Robust Standard Errors Using the Bartlett Kernel without Truncation," *Econometrica*, 70(5), 2093–2095.

LUENBERGER, D. G. (1969): *Optimization by Vector Space Methods*, Series in Decision and Control. Wiley, New York, NY.

NEWEY, W. K., AND K. D. WEST (1987): "A Simple Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55(3), 703–708.

——— (1994): "Automatic Lag Selection in Covariance Matrix Estimation," *Review of Economic Studies*, 61(4), 631–653.

PINHEIRO, J. C., AND D. M. BATES (1996): "Unconstrained Parametrizations for Variance-Covariance Matrices," *Statistics and Computing*, 6(3), 289–296.

PRIESTLEY, M. B. (1981): *Spectral Analysis and Time Series*. Academic Press, New York.

RUDIN, W. (1976): *Principles of Mathematical Analysis*, International Series in Pure and Applied Mathematics. McGraw-Hill, San Francisco, 3rd edn.

STURM, J. F. (1999): "Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones," *Optimation Methods and Sotfware*, 11(1–4), 625–653.

SUN, Y., P. C. B. PHILLIPS, AND S. JIN (2008): "Optimal Bandwidth Selection in Heteroskedasticity-Autocorrelation Robust Testing," *Econometrica*, 76(1), 175–194.

VANDENBERGHE, L., AND S. BOYD (1996): "Semidefinite Programming," *SIAM Review*, 38(1), 49–95.

WHITE, H., AND I. DOMOWITZ (1984): "Nonlinear Regression with Dependent Observations," *Econometrica*, 52(1), 143–161.

Table 1

The kernels often considered in the literature of long-run covariance matrix estimation.

Truncated flat (TF)	$k^{TF}(x) = \begin{cases} 1 & \text{for } |x| \leq 1, \\ 0 & \text{otherwise,} \end{cases}$
Bartlett (BT)	$k^{BT}(x) = \begin{cases} 1 - |x| & \text{for } |x| \leq 1, \\ 0 & \text{otherwise,} \end{cases}$
Parzen	$k^{PR}(x) = \begin{cases} 1 - 6x^2 + 6|x|^3 & \text{for } |x| \leq 1/2, \\ 2(1 - |x|)^3 & \text{for } 1/2 \leq |x| \leq 1, \\ 0 & \text{otherwise,} \end{cases}$
Quadratic Spectral (QS)	$k^{QS}(x) = \frac{25}{12\pi^2 x^2} \left(\frac{\sin(6\pi x/5)}{6\pi x/5} - \cos(6\pi x/5) \right),$
Tukey-Hanning	$k^{TH}(x) = \begin{cases} 1 + \cos(\pi x)/2 & \text{for } |x| \leq 1, \\ 0 & \text{otherwise,} \end{cases}$
Trapezoid	$k^{TR}(x) = \begin{cases} 1 & \text{for } |x| \leq c_1, 0 < c_1 < 1 \\ \frac{(1-|x|)}{1-c_1} & \text{for } c_1 \leq |x| \leq 1, \\ 0 & \text{otherwise.} \end{cases}$
Sharp Original	$k^{BT}(x) = \begin{cases} (1 - |x|)^\rho & \text{for } |x| \leq 1, \ \rho \geq 1 \\ 0 & \text{otherwise,} \end{cases}$

Table 2

The design of Andrews' (1991) experiments

In all experiments, $y_t = x_t' \theta^* + u_t$, where x_t and u_t are generated in the way described below, and θ^* is set equal to zero (i.e. $y_t = u_t$).

AR(1)-HOME/HET1/HET2 experiments

$\quad \tilde{u}_t = \rho \tilde{u}_{t-1} + \eta_t$, where $\eta_t \sim N(0, 1 - \rho^2)$ (so that $\mathrm{var}[\tilde{u}_t] = 1$),

$\quad \tilde{x}_{ti} = \rho \tilde{x}_{ti-1} + \epsilon_{ti}$, $i \in \{2, \ldots 5\}$, where $\epsilon_{ti} \sim N(0, 1 - \rho^2)$ (so that $\mathrm{var}[\tilde{x}_{ti}] = 1$),

$\quad \rho \in \{0, 0.3, 0.5, 0.7, 0.9, 0.95, -0.1, -0.3, -0.5\}$,

$\quad \tilde{x}_t = [\tilde{x}_{t2}, \tilde{x}_{t3}, \tilde{x}_{t4}, \tilde{x}_{t5}]'$,

$\quad \hat{x}_t = \tilde{x}_t - \frac{1}{T} \sum_{t=1}^{T} \tilde{x}_t$,

$\quad x_t = [1, ((\sum_{t=1}^{T} \hat{x}_t \hat{x}_t')^{-1/2} \hat{x}_t)'] = [1, x_{t2}, x_{t3}, x_{t4}, x_{t5}]$.

$$
u_t = \begin{cases} \tilde{u}_t & \text{in the AR(1)-HOMO experiments,} \\ \tilde{u}_t x_{t2} & \text{in the AR(1)-HET1 experiments,} \\ \frac{1}{2} \tilde{u}_t \left| \sum_{i=2}^{5} x_{ti} \right| & \text{in the AR(1)-HET2 experiments,} \end{cases}
$$

MA(1)-HOMO experiments

$\quad u_t = \eta_t + \vartheta \eta_{t-1}$, where $\eta_t \sim N(0, 1/(1 + \vartheta^2))$ (so that $\mathrm{var}[u_t] = 1$),

$\quad \tilde{x}_{ti} = \epsilon_{ti} + \vartheta \epsilon_{ti-1}$, $i = \in \{2, \ldots 5\}$, where $\epsilon_{ti} \sim N(0, 1/(1 + \vartheta^2))$ (so that $\mathrm{var}[\tilde{x}_{ti}] = 1$),

$\quad \vartheta \in \{\pm 0.1, \pm 0.3, \pm 0.5, \pm 0.7, \pm 0.9\}$

$\quad x_t$ is calculated from $\tilde{x}_t$ in the same way as in the AR(1) experiments.

Hypothesis of the test: The exclusion of x_{t2} in the population regression of y_t on x_t.

Table 3

The efficiency of the ATF estimator relative to the TF estimator.

AR(1)-HOMO										
m	$\rho =0$		0.1		0.3		0.5		0.7	
1	1.00	(0.00)	1.00	(0.00)	1.00	(0.00)	1.00	(0.00)	1.00	(0.00)
3	1.00	(0.02)	1.00	(0.02)	1.00	(0.01)	1.00	(0.00)	1.00	(0.00)
5	1.01	(0.19)	1.01	(0.18)	1.01	(0.14)	1.00	(0.09)	1.00	(0.04)
7	1.04	(0.44)	1.04	(0.43)	1.03	(0.38)	1.02	(0.31)	1.01	(0.21)
m	$\rho =0.9$		0.95		-0.1		-0.3		-0.5	
1	1.00	(0.00)	1.00	(0.00)	1.00	(0.00)	1.00	(0.01)	1.00	(0.40)
3	1.00	(0.00)	1.00	(0.00)	1.00	(0.03)	1.00	(0.03)	1.00	(0.11)
5	1.00	(0.00)	1.00	(0.00)	1.01	(0.19)	1.01	(0.19)	1.00	(0.21)
7	1.00	(0.07)	1.00	(0.04)	1.04	(0.44)	1.03	(0.44)	1.02	(0.42)
MA(1)-HOMO										
m	$\vartheta =0.1$		0.3		0.5		0.7		0.9	
1	1.00	(0.00)	1.00	(0.00)	1.00	(0.00)	1.00	(0.00)	1.00	(0.00)
3	1.00	(0.02)	1.00	(0.01)	1.00	(0.01)	1.00	(0.01)	1.00	(0.01)
5	1.01	(0.18)	1.01	(0.15)	1.01	(0.13)	1.01	(0.12)	1.01	(0.12)
7	1.04	(0.43)	1.03	(0.40)	1.03	(0.37)	1.03	(0.36)	1.02	(0.35)
m	$\vartheta =-.1$		-0.3		-0.5		-0.7		-0.9	
1	1.00	(0.00)	1.00	(0.00)	1.00	(0.06)	1.00	(0.26)	1.00	(0.40)
3	1.00	(0.03)	1.00	(0.05)	1.00	(0.18)	1.00	(0.42)	1.00	(0.55)
5	1.01	(0.20)	1.01	(0.23)	1.01	(0.34)	1.01	(0.51)	1.01	(0.61)
7	1.04	(0.45)	1.03	(0.47)	1.02	(0.54)	1.02	(0.63)	1.02	(0.69)

Notes:

(a) The symbol m denotes the bandwidth.

(b) The efficiency is the ratio of the MSE of the TF estimator to that of the ATF estimator.

(c) The numbers in the parentheses are the relative frequencies of non-p.s.d. estimates in the TF estimation.

(d) The sample size is 128.

Table 4

The Efficiency of the BT, ATF, and ATFF estimators relative to the QS estimator using fixed optimum bandwidths.

AR(1)-HOMO

T	Estimator	ρ							
		0	0.3	0.5	0.7	0.9	0.95	-0.3	-0.5
64	BT	1.00	1.00	0.99	0.96	0.97	0.98	1.00	0.99
	ATF	1.00	1.00	0.91	1.05	1.03	1.02	1.00	0.89
	ATFF	1.00	1.00	0.99	1.05	1.03	1.02	1.00	0.99
		(0.04)	(0.00)	(0.44)	(1.00)	(3.00)	(3.12)	(0.00)	(0.48)
128	BT	1.00	1.00	0.97	0.94	0.95	0.96	1.00	0.97
	ATF	1.00	0.99	0.98	1.04	1.04	1.03	0.99	0.97
	ATFF	1.00	1.00	1.02	1.06	1.04	1.03	1.00	1.02
		(0.00)	(0.08)	(0.80)	(1.64)	(4.60)	(6.16)	(0.12)	(0.76)
256	BT	1.00	1.00	0.95	0.92	0.93	0.95	1.00	0.95
	ATF	1.00	0.91	1.08	1.09	1.06	1.05	0.90	1.08
	ATFF	1.00	1.00	1.08	1.09	1.06	1.05	1.00	1.08
		(0.00)	(0.36)	(1.00)	(2.00)	(6.00)	(10.00)	(0.36)	(1.00)

AR(1)-HET1

T	Estimator	ρ							
		0	0.3	0.5	0.7	0.9	0.95	-0.3	-0.5
64	BT	1.00	1.00	1.00	0.99	0.99	0.99	1.00	1.00
	ATF	1.00	1.00	1.00	1.01	1.01	1.01	1.00	1.00
	ATFF	1.00	1.00	1.00	1.01	1.01	1.01	1.00	1.00
		(0.00)	(0.08)	(0.88)	(1.44)	(3.00)	(3.00)	(0.16)	(0.92)
128	BT	1.00	1.00	0.99	0.98	0.98	0.98	1.00	0.99
	ATF	1.00	0.99	1.01	1.02	1.02	1.01	0.99	1.01
	ATFF	1.00	1.00	1.01	1.02	1.02	1.01	1.00	1.01
		(0.00)	(0.48)	(1.00)	(2.00)	(5.00)	(6.00)	(0.52)	(1.00)
256	BT	1.00	1.00	0.99	0.98	0.98	0.98	1.00	0.99
	ATF	1.00	1.01	1.01	1.02	1.02	1.02	1.01	1.01
	ATFF	1.00	1.01	1.01	1.02	1.02	1.02	1.01	1.01
		(0.00)	(1.00)	(1.48)	(3.00)	(7.00)	(10.76)	(1.00)	(1.52)

(continued from the previous page)

AR(1)-HET2

T	Estimator	ρ							
		0	0.3	0.5	0.7	0.9	0.95	-0.3	-0.5
64	BT	1.00	0.99	0.99	1.00	0.98	0.98	0.99	0.99
	ATF	1.00	0.99	0.99	0.96	1.02	1.02	0.99	0.99
	ATFF	1.00	0.99	0.99	1.00	1.02	1.02	0.99	0.99
		(0.00)	(0.00)	(0.00)	(0.52)	(2.00)	(2.24)	(0.00)	(0.00)
128	BT	1.00	0.99	1.00	0.99	0.98	0.97	0.99	1.00
	ATF	1.00	0.99	1.00	0.98	1.02	1.02	0.99	1.00
	ATFF	1.00	0.99	1.00	1.00	1.02	1.02	0.99	1.00
		(0.08)	(0.00)	(0.00)	(0.76)	(3.00)	(5.00)	(0.00)	(0.00)
256	BT	1.00	0.99	1.00	0.99	0.98	0.97	0.99	1.00
	ATF	1.00	0.99	0.99	1.01	1.02	1.02	0.99	0.99
	ATFF	1.00	0.99	1.00	1.01	1.02	1.02	0.99	1.00
		(0.08)	(0.00)	(0.16)	(0.92)	(3.68)	(7.00)	(0.00)	(0.16)

MA(1)-HOMO

T	Estimator	ϑ							
		0.1	0.3	0.5	0.7	0.9	0.99	-0.3	-0.7
64	BT	1.00	1.00	1.00	0.99	0.99	0.99	0.99	0.99
	ATF	1.00	1.00	0.99	0.94	0.91	0.91	0.99	0.94
	ATFF	1.00	1.00	1.00	0.99	0.99	0.99	0.99	0.99
		(0.00)	(0.00)	(0.12)	(0.28)	(0.36)	(0.36)	(0.00)	(0.28)
128	BT	1.00	1.00	0.99	0.99	0.99	0.99	1.00	0.99
	ATF	1.00	1.00	0.89	0.83	0.86	0.86	1.00	0.80
	ATFF	1.00	1.00	0.99	0.99	1.00	1.00	1.00	0.99
		(0.00)	(0.04)	(0.36)	(0.56)	(0.60)	(0.60)	(0.04)	(0.52)
256	BT	1.00	1.00	0.99	0.96	0.95	0.95	1.00	0.96
	ATF	1.00	0.95	0.84	0.93	0.96	0.96	0.95	0.91
	ATFF	1.00	1.00	1.00	1.03	1.04	1.04	1.00	1.02
		(0.00)	(0.24)	(0.60)	(0.76)	(0.76)	(0.80)	(0.24)	(0.72)

Notes:

(a) The efficiency of each estimator is the ratio of the MSE of the QS estimator to that of the estimator.

(b) The numbers in parentheses are the fixed optimum bandwidths for the ATFF estimator found by grid search.

Table 5

The efficiency of the BT, ATF, and ATFF estimators relative to the QS estimator using data-dependent bandwidths.

AR(1)-HOMO

T	Estimator	ρ							
		0	0.3	0.5	0.7	0.9	0.95	-0.3	-0.5
64	BT	0.74	0.83	0.97	0.98	0.95	0.96	0.87	1.03
	ATF	1.02	0.99	1.03	1.05	1.04	1.03	0.99	1.04
	ATFF	1.02	1.02	1.07	1.09	1.04	1.03	1.03	1.10
128	BT	0.72	0.86	1.00	0.98	0.94	0.94	0.88	1.03
	ATF	1.04	0.96	1.07	1.06	1.05	1.04	0.97	1.07
	ATFF	1.02	1.03	1.12	1.12	1.05	1.03	1.04	1.14
256	BT	0.71	0.87	0.93	0.96	0.94	0.89	0.88	0.95
	ATF	1.08	0.94	1.06	1.08	1.06	1.05	0.95	1.06
	ATFF	1.02	1.03	1.16	1.15	1.07	1.03	1.03	1.17

AR(1)-HET1

T	Estimator	ρ							
		0.0	0.3	0.5	0.7	0.9	0.95	-0.3	-0.5
64	BT	0.93	0.95	0.96	0.96	0.97	0.98	0.96	0.97
	ATF	1.01	1.00	1.00	1.01	1.01	1.01	1.00	1.00
	ATFF	1.00	1.00	1.01	1.01	1.01	1.01	1.00	1.01
128	BT	0.96	0.97	0.98	0.96	0.96	0.97	0.98	0.98
	ATF	1.01	1.00	1.01	1.01	1.01	1.01	1.00	1.01
	ATFF	1.00	1.00	1.01	1.01	1.00	1.00	1.00	1.01
256	BT	0.98	0.98	0.97	0.96	0.95	0.95	0.99	0.97
	ATF	1.01	1.00	1.01	1.01	1.01	1.01	1.00	1.01
	ATFF	1.00	1.00	1.01	1.00	0.99	0.99	1.00	1.01

(continued from the previous page)

AR(1)-HET2

T	Estimator	ρ							
		0.0	0.3	0.5	0.7	0.9	0.95	-0.3	-0.5
64	BT	0.99	1.01	1.05	1.04	0.97	0.98	1.02	1.09
	ATF	0.99	0.99	0.98	1.00	1.02	1.02	0.99	0.98
	ATFF	1.00	1.00	1.00	1.02	1.02	1.02	0.99	1.01
128	BT	0.98	1.01	1.08	1.06	0.99	0.97	1.03	1.10
	ATF	0.99	0.98	0.98	0.99	1.01	1.02	0.98	0.98
	ATFF	1.00	0.99	1.00	1.04	1.02	1.01	0.99	1.01
256	BT	0.98	1.01	1.06	1.06	1.05	0.96	1.02	1.06
	ATF	0.99	0.97	0.98	0.99	1.01	1.02	0.97	0.98
	ATFF	1.00	0.99	1.01	1.06	1.03	1.01	0.99	1.02

MA(1)-HOMO

T	Estimator	ϑ							
		0.1	0.3	0.5	0.7	0.9	0.99	-0.3	-0.7
64	BT	0.75	0.82	0.90	0.95	0.96	0.96	0.84	0.98
	ATF	1.02	0.99	0.98	1.00	1.01	1.01	0.98	0.99
	ATFF	1.02	1.02	1.03	1.04	1.05	1.05	1.02	1.05
128	BT	0.73	0.83	0.93	0.97	0.99	0.99	0.85	1.00
	ATF	1.04	0.96	0.98	1.04	1.05	1.05	0.96	1.03
	ATFF	1.02	1.02	1.03	1.06	1.08	1.08	1.02	1.07
256	BT	0.72	0.84	0.92	0.94	0.95	0.95	0.85	0.96
	ATF	1.05	0.91	1.04	1.09	1.09	1.08	0.91	1.09
	ATFF	1.02	1.02	1.04	1.11	1.13	1.14	1.02	1.11

See the notes of Table 4.

Table 6

The size in the t-test of the exclusion of x_{t2}.

| Estimator | ρ | AR(1) | | | | | | MA(1)-HOMO | | |
| | | HOMO | | HET1 | | HET2 | | | | |
		10%	5%	10%	5%	10%	5%	ϑ	10%	5%
QS	0	11.44	6.08	12.92	6.98	12.19	6.79	0.1	11.71	6.23
BT		12.47	6.99	14.02	8.00	13.30	7.61		12.87	7.07
ATF		11.22	5.78	12.59	6.79	11.98	6.57		11.54	6.02
ATFF		11.41	6.09	12.92	7.02	12.20	6.73		11.71	6.20
QS	0.3	13.13	7.24	14.46	8.27	13.98	7.86	0.3	12.65	6.95
BT		14.02	8.12	15.80	9.34	15.03	8.92		13.66	7.93
ATF		13.05	7.22	14.40	8.24	13.85	7.80		12.63	6.96
ATFF		12.98	7.18	14.31	8.21	13.90	7.82		12.53	6.84
QS	0.5	15.49	9.28	17.09	10.65	16.59	10.00	0.5	13.63	7.61
BT		16.97	10.61	18.76	12.02	17.95	11.14		14.89	8.61
ATF		14.91	8.85	16.72	10.29	16.31	9.70		13.02	7.21
ATFF		15.10	8.92	16.88	10.36	16.31	9.82		13.13	7.28
QS	0.7	19.95	13.10	22.18	14.85	21.57	13.95	0.7	13.98	8.09
BT		23.20	15.81	25.14	17.31	23.74	15.82		15.48	9.31
ATF		19.13	12.46	21.65	14.48	21.17	13.49		13.03	7.23
ATFF		19.65	12.83	22.26	14.78	21.38	13.74		13.23	7.44
QS	0.9	34.52	26.65	37.20	28.97	35.63	27.38	0.9	14.04	8.15
BT		38.48	30.36	41.12	32.76	38.57	30.38		15.70	9.51
ATF		33.10	25.24	36.24	28.16	34.82	26.64		13.00	7.28
ATFF		34.03	26.11	37.58	29.20	35.80	27.45		13.20	7.42
QS	0.95	45.90	38.05	45.69	38.14	44.60	36.48	0.95	14.08	8.17
BT		49.11	41.46	49.81	42.20	47.64	39.56		15.73	9.48
ATF		43.83	36.26	44.87	37.16	43.43	35.22		13.01	7.28
ATFF		44.44	36.96	46.02	38.48	44.37	36.18		13.23	7.46
QS	-0.3	12.60	7.14	14.13	8.23	13.27	7.55	-0.3	12.16	6.82
BT		13.68	7.96	15.33	9.24	14.30	8.49		13.25	7.66
ATF		12.49	7.02	14.04	8.26	13.15	7.50		12.14	6.82
ATFF		12.48	6.96	14.06	8.18	13.07	7.51		12.08	6.73
QS	-0.5	14.66	8.78	16.45	10.01	15.62	9.33	-0.5	13.18	7.71
BT		16.39	10.18	18.22	11.44	17.15	10.39		14.81	8.80
ATF		14.29	8.48	16.29	9.75	15.30	9.02		12.33	6.93
ATFF		14.38	8.51	16.32	9.86	15.43	9.14		12.52	7.11