WAVELET-BASED TESTING FOR SERIAL CORRELATION OF UNKNOWN FORM IN PANEL MODELS

BY YONGMIAO HONG AND CHIHWA KAO

Wavelet analysis is a new mathematical method developed as a unified field of science over the last decade or so. As a spatially adaptive analytic tool, wavelets are useful for capturing serial correlation where the spectrum has peaks or kinks, as can arise from persistent dependence, seasonality, and other kinds of periodicity. This paper proposes a new class of generally applicable wavelet-based tests for serial correlation of unknown form in the estimated residuals of a panel regression model, where error components can be one-way or two-way, individual and time effects can be fixed or random, and regressors may contain lagged dependent variables or deterministic/stochastic trending variables. Our tests are applicable to unbalanced heterogenous panel data. They have a convenient null limit $N(0, 1)$ distribution. No formulation of an alternative model is required, and our tests are consistent against serial correlation of unknown form even in the presence of substantial inhomogeneity in serial correlation across individuals. This is in contrast to existing serial correlation tests for panel models, which ignore inhomogeneity in serial correlation across individuals by assuming a common alternative, and thus have no power against the alternatives where the average of serial correlations among individuals is close to zero. We propose and justify a data-driven method to choose the smoothing parameter—the finest scale in wavelet spectral estimation, making the tests completely operational in practice. The data-driven finest scale automatically converges to zero under the null hypothesis of no serial correlation and diverges to infinity as the sample size increases under the alternative, ensuring the consistency of our tests. Simulation shows that our tests perform well in small and finite samples relative to some existing tests.

KEYWORDS: Error component, hypothesis testing, serial correlation of unknown form, spectral peak, static and dynamic panel models, unbalanced panel data, wavelet.

1. INTRODUCTION

Panel data have been widely used in economics and finance. They often provide insights not available in pure time-series or cross-sectional data (e.g., Baltagi (2002), Granger (1996), Hsiao (2003)). This paper proposes a new class of generally applicable wavelet-based consistent tests for serial correlation of unknown form in the errors of panel models. It is important to test serial correlation for panel models because existence of serial correlation will invalidate conventional tests such as $t$- and $F$-tests which use standard covariance estimators of parameter estimators, and will indicate model misspecification when

---

1We thank the co-editor and three referees for insightful comments that have lead to significant improvement on a previous version. We also thank Badi Baltagi, Pierre Duchesne, Jiti Gao, Jerry Hausman, Cheng Hsiao, Heshem Pesaran, Jim Stock, and seminar participants at MIT-Harvard Econometrics Workshop, 2001 North American Summer Meeting of Econometric Society in Washington, DC, 2001 Far Eastern Meeting of Econometric Society in Kobe, Japan, Fifth ICSA International Conference in Hong Kong, and 10th International Conference on Panel Data Models in Berlin for helpful comments. This research is supported by National Science Foundation via Grant SES-0111769.
regressors include lagged dependent variables. Moreover, the choice of estimation methods may depend on whether there exists serial correlation in the errors of panel models. When the errors are serially correlated, for example, the computation of MLE (e.g., Anderson and Hsiao (1982), Hsiao (2003), Binder, Hsiao, and Pesaran (1999)) and GMM (e.g., Blundell and Bond (1998)) could be complicated, and the feasible GLS estimator will be invalid or have to be modified substantially (e.g., Baltagi and Li (1991)). Some procedures, such as Breusch and Pagan’s (1980) tests for random effects, also assume serial uncorrelatedness in the errors of panel models.


All existing tests for serial correlation in panel models assume a known form of a common alternative, e.g., an AR(1) or MA(1) model. These tests have optimal power when the assumed model is true, but they are not consistent against serial correlation of unknown form. It is useful to test serial correlation of unknown form because prior information about the alternative is usually not available in practice. This is particularly relevant for panel models because there may exist significant inhomogeneity in serial correlation across individuals (e.g., Choi (2002)). By assuming a common alternative model, existing tests ignore inhomogeneity in serial correlation across individuals, and so have little power against the alternatives where the average of serial correlations among individuals is close to zero. Moreover, as Granger and Newbold (1977, p. 92) pointed out, the first few lags of OLS residuals of a misspecified linear dynamic model often appear like a white noise, due to the very nature of OLS. It is therefore important to check serial correlation at higher order lags. Little effort has been made on evaluation of dynamic panel models (Granger (1996)).

Wavelets are a new mathematical tool alternative to the Fourier transform. They can effectively capture nonsmooth features such as singularities and spatial inhomogeneity (e.g., Donoho and Johnstone (1994, 1995a, 1995b), Donoho et al. (1996), Gao (1997), Hong and Lee (2001), Jensen (2000), Neumann (1996), Lee and Hong (2001), Ramsey (1999), Wang (1995)). Many economic and financial time series have a spectrum with peaks and kinks, as can arise from persistent dependence, business cycles, seasonality, and other kinds of periodicity (e.g., Bizer and Durlauf (1992), Granger (1969), Watson (1993)). In this paper we use wavelets to test serial correlation in estimated residuals of panel models. Unlike existing tests, whose constructions are usually
model-dependent, our tests are generally applicable. The panel model can be static or dynamic, and one-way or two-way; both balanced and unbalanced panel data are covered; individual and time effects can be fixed or random; regressors can contain lagged dependent variables or deterministic/stochastic trending variables; and no specific estimation method is required. Our tests have a convenient limit $N(0, 1)$ distribution under the null hypothesis, no matter whether regressors contain lagged dependent variables or deterministic/stochastic trending variables. In contrast to Durbin and Watson’s (1951) test and Box and Pierce’s (1970) portmanteau test, parameter estimation uncertainty has no impact on the limit distribution of our test statistics when applied to dynamic panel models. We do not require an alternative model, and our tests are consistent against serial correlation of unknown form even in the presence of substantial inhomogeneity in serial correlation across individuals. No consistent test for serial correlation of unknown form was available for panel models.

Our asymptotic theory considers a panel model with both large $n$ and $T$, where $n$ is the number of individuals and $T$ is the number of time-series observations. Increasing effort has been devoted to the study of panel models with both large $n$ and $T$, due to the growing use of cross-country data over time to study growth convergence, international R&D spillover and purchasing power parity, and to the growing use of firm- or portfolio-level financial time series. As is well known (e.g., Phillips and Moon (1999), Hahn and Kuersteiner (2002)), asymptotic analysis in panel models is much more involved than in pure time-series analysis, due to the need to handle double indices. As a distinct feature, we treat both $n \to \infty$ and $T \to \infty$ simultaneously, which complements Phillips and Moon (1999) and Hahn and Kuersteiner’s (2002) joint limit theory for panel models. Our general theory does not require that the ratio $n/T$ go to 0 or a constant. We also show that the use of the estimated residuals from a possibly nonstationary panel model rather than the unobservable errors has no impact on the null limit distribution of our test statistics. In addition, we find several interesting features not available in pure time-series analysis. Most remarkably, the limit $N(0, 1)$ distribution of our test statistics is obtained without requiring the smoothing parameter—the finest scale in wavelet estimation to grow with $T$. This not only leads to reasonable asymptotic approximation in finite samples, but also makes it possible to use data-driven methods that deliver a fixed finest scale under the null hypothesis of no serial correlation. This is in sharp contrast to Lee and Hong (2001), who, in testing serial correlation for observed raw time series data, require the finest scale to grow as $T \to \infty$ under the null hypothesis. We further develop a data-driven method to choose the finest scale, making our tests completely operational in practice. The data-driven finest scale converges to 0 under the null and grows to $\infty$ under the alternative, ensuring consistency against serial correlation of unknown form. We also find that a heteroskedasticity-corrected test may be less powerful than a heteroskedasticity-consistent test. This differs
from the well-known estimation result that heteroskedasticity-corrected estimators (e.g., feasible GLS) are more efficient than heteroskedasticity-consistent estimators (e.g., OLS). Our tests work reasonably well in small and finite samples often encountered in economics.

We describe the panel model and hypotheses in Section 2, introduce wavelets and test statistics in Section 3, derive the limit distributions of these tests in Section 4, and establish their consistency in Section 5. Section 6 proposes a data-driven finest scale. Section 7 is a simulation study. Section 8 concludes. All proofs are in the Appendix. Throughout, $\|A\|$ denotes the Euclidean norm $[\text{tr}(A^*A)]^{1/2}$; $A^*$ and $\text{Re}(A)$ the complex conjugate and the real part of $A$; $\mathbb{Z} \equiv \{0, \pm 1, \ldots\}$ and $\mathbb{Z}^+ \equiv \{0, 1, \ldots\}$ the set of integers and the set of nonnegative integers; and $c$ and $C$ generic bounded constants, with $0 < c < C < \infty$. Unless indicated, all limits are taken as both $n, T \to \infty$.

AGAUS Program for implementing our tests is available from the authors upon request. The user only needs to supply estimated residuals.

2. THE FRAMEWORK

Consider a linear panel process

$$Y_{it} = \alpha + X_{it}'\beta + \mu_i + \lambda_t + \varepsilon_{it}$$  \hspace{1cm} (2.1)

$(t = 1, \ldots, T_i; \ i = 1, \ldots, n; \ n, T_i \in \mathbb{Z}^+)$,

where $Y_{it}$ is a scalar, $X_{it}$ is a $p \times 1$ vector of regressors that may contain lagged dependent variables, $\alpha$ is an intercept, $\beta$ is a $p \times 1$ vector of slope parameters, $\mu_i$ is the individual effect, $\lambda_t$ is the time effect, and $\varepsilon_{it}$ is the error such that $\{\varepsilon_{it}\}$ and $\{\varepsilon_{is}\}$ are mutually independent for all $i \neq l$ and all $t, s$. We allow for fixed or random effects. We assume $T_i = c_i T$ for some integer $T$ and $c_i \in [c, C]$, thus permitting unbalanced panel data. Moreover, we allow $Y_{it}, X_{it}, \alpha$, and $\beta$ to depend on $n$ and $T$. (For notational simplicity, we have suppressed such dependence.)

The slope parameter $\beta$ is often of interest. It can be estimated, e.g., by the within estimator

$$\hat{\beta} \equiv \left[ \sum_{i=1}^n \sum_{t=1}^{T_i} \tilde{X}_{it} \tilde{X}_{it}' \right]^{-1} \left[ \sum_{i=1}^n \sum_{t=1}^{T_i} \tilde{X}_{it} \tilde{Y}_{it} \right],$$  \hspace{1cm} (2.2)

where $\tilde{X}_{it} \equiv X_{it} - \bar{X}_i - \bar{X}_t + \bar{X}$, $\bar{X}_i \equiv \frac{1}{T_i} \sum_{t=1}^{T_i} X_{it}$, $\bar{X}_t \equiv \frac{1}{n} \sum_{i=1}^n X_{it}$, and $\bar{X} \equiv n^{-1} \sum_{i=1}^n T_i^{-1} \sum_{t=1}^{T_i} X_{it}$. The variables $\tilde{Y}_{it}, \tilde{Y}_i, \tilde{Y}_t$, and $\bar{Y}$ are defined similarly. For interval estimation and hypothesis testing, one often uses the standard covariance estimator $\hat{\Omega}_{\beta} \equiv \hat{\sigma}_e^2 (\sum_{i=1}^n \sum_{t=1}^{T_i} \tilde{X}_{it} \tilde{X}_{it}')^{-1}$ of $\hat{\beta}$, where $\hat{\sigma}_e^2$ is an estimator for $\sigma_e^2 \equiv \lim_{n \to \infty} n^{-1} \sum_{i=1}^n E(\varepsilon_{it}^2)$. This estimator is valid when
\(\{\varepsilon_{i,t}\}\) in (2.1) is serially uncorrelated, among other things. The existence of serial correlation of any form, however, will generally invalidate the covariance estimator and related inference. In particular, conventional \(t\)- and \(F\)-tests will be misleading. Moreover, when the \(X_{i,t}\) contain lagged dependent variables, serial correlation will further render inconsistent the within estimator \(\hat{\beta}\) for \(\beta\).

We are interested in testing whether the error process \(\{\varepsilon_{i,t}\}\) is serially correlated. The hypotheses of interest are \(H_0: \text{cov}(\varepsilon_{it}, \varepsilon_{it-h}) = 0\) for all \(h \neq 0\) and all \(i\) vs. \(H_A: \text{cov}(\varepsilon_{it}, \varepsilon_{it-h}) \neq 0\) at least for some \(h \neq 0\) and some \(i\). The alternative \(H_A\) allows some (but not all) individual series to be white noises. Prior information about the alternative is usually not available in practice, although there may exist substantial inhomogeneities in serial correlation across \(i\).

To test \(H_0\), we will examine serial correlation in the demeaned estimated residual

\[
(2.3) \quad \hat{v}_{it} \equiv \hat{u}_{it} - \bar{u}_t - \hat{u}_d + \bar{\nu} \quad (t = 1, \ldots, T_i; \ i = 1, \ldots, n),
\]

where \(\hat{u}_{it} \equiv Y_{it} - X_{i,t}'\hat{\beta}, \bar{u}_t \equiv \frac{1}{T_i} \sum_{t=1}^{T_i} \hat{u}_{it}, \hat{u}_d \equiv n^{-1} \sum_{t=1}^{n} \hat{u}_{it}, \bar{\nu} \equiv n^{-1} \sum_{t=1}^{n} T_i^{-1} \times \sum_{t=1}^{T_i} \hat{u}_t,\) and \(\hat{\beta}\) is an estimator consistent for \(\beta\) under \(H_0\). When \(\hat{\beta}\) is the within estimator in (2.2), \(\hat{v}_{it}\) is the well-known within residual. However, we allow using other estimators that are consistent for \(\beta\) under \(H_0\) but not necessarily under \(H_A\).

Suppose \(\hat{\beta} \xrightarrow{p} \beta\), as does the within estimator in (2.2) under \(H_0\); then \(\hat{v}_{it}\) will converge in probability to the true error \(\varepsilon_{it}\). Under \(H_A\), however, \(\hat{\beta}\) may not be consistent for \(\beta\) (as in a dynamic panel model), and \(\hat{v}_{it}\) will converge in probability to the model error

\[
(2.4) \quad v_{it} \equiv \varepsilon_{it} + (\beta - \beta^*)'(X_{it} - E\bar{X}_t - E\bar{X}_d + E\bar{X}),
\]

which contains both the true error \(\varepsilon_{it}\) and the misspecified component, where \(\beta^* \equiv p\lim \hat{\beta}\). This does not invalidate our tests, but it affects the power of the tests in finite samples, because serial correlation in \(\{v_{it}\}\) may differ from serial correlation in \(\{\varepsilon_{it}\}\). However, our tests are still consistent against \(H_A\), because \(H_0\) holds if and only if \(\{v_{it}\}\) is serially uncorrelated: When \(H_0\) holds, \(\{v_{it}\}\) coincides with \(\{\varepsilon_{it}\}\) and so is serially uncorrelated; on the other hand, if \(\{v_{it}\}\) is serially uncorrelated in a linear dynamic panel model, we can view \(\{v_{it}\}\) as the true error for (2.1), and estimation and inference can be implemented in a standard fashion. (We note that in a linear dynamic panel setup, it is possible that \(\{\varepsilon_{it}\}\) is serially correlated but \(\{v_{it}\}\) is serially uncorrelated, due to \(\beta^* \neq \beta\). This occurs when and only when \(\varepsilon_{it}\) contains the misspecified linear component, \((\beta^* - \beta)'(X_{it} - E\bar{X}_t - E\bar{X}_d + E\bar{X})\), plus a white noise. In this case, serial correlation in \(\{\varepsilon_{it}\}\) is solely caused by the misspecified linear component, and it is actually more appropriate to view that (2.1) is a correctly specified linear dynamic panel model, but with \(v_{it}\) as the true error and \(\beta^*\) as the true
model parameter. With such an interpretation, $H_0$ holds when \{\textit{vit}\} is seriously uncorrelated.

Suppose \{\textit{vit}\} has autocovariance function $R_i(h) \equiv E(\textit{vit}\textit{vit}_{i-h})$ and power spectrum

(2.5) \[ f_i(\omega) \equiv (2\pi)^{-1} \sum_{h=-\infty}^{\infty} R_i(h) e^{-ih\omega}, \quad \omega \in [-\pi, \pi], \quad i \equiv \sqrt{-1}. \]

Both $R_i(h)$ and $f_i(\omega)$ contain the same information on serial correlation of \{\textit{vit}\}. One can use $R_i(h)$ or $f_i(\omega)$ to test $H_0$ vs. $H_A$. All existing tests for serial correlation in panel models are based on $R_i(h)$ assuming a common model with some prespecified lags $h$ for all $i$ (e.g., AR(1) and MA(1)). We use $f_i(\omega)$ here, which is a natural tool to test serial correlation of unknown form, because it contains information on serial correlation at all lags. Under $H_0$, $f_i(\omega)$ becomes $f_{i0}(\omega) \equiv (2\pi)^{-1} R_i(0)$ for all $\omega \in [-\pi, \pi]$. Under $H_A$, $f_i(\omega) \neq (2\pi)^{-1} R_i(0)$ at least for some $i$. Thus, a consistent test for $H_0$ vs. $H_A$ can be formed by comparing consistent estimators of $f_i(\omega)$ and $f_{i0}(\omega)$. We will use wavelets to estimate $f_i(\omega)$, which are suitable for time series with spectral peaks and kinks. Of course, other nonparametric methods (e.g., kernel smoothing; see Hong (1996) and Section 7 below) could be used.

3. WAVELET METHOD

3.1. Wavelets

The essence of wavelet analysis is to expand a function as a sum of elementary functions called wavelets centered at a sequence of locations. These wavelets are derived from a single function $\psi(\cdot)$, called the mother wavelet, by translations and dilations. As a spatially adaptive analytic tool, wavelets are powerful in capturing singularities of nonsmooth functions, such as spectral peaks and kinks (e.g., Gao (1997), Neumann (1996)). We first impose a standard condition on $\psi(\cdot)$.

**Assumption 1:** $\psi: \mathbb{R} \to \mathbb{R}$ is an orthonormal wavelet such that $\int_{-\infty}^{\infty} \psi(x) \, dx = 0$, $\int_{-\infty}^{\infty} |\psi(x)| \, dx < \infty$, $\int_{-\infty}^{\infty} \psi(x) \psi(x - k) \, dx = 0$ for all $k \in \mathbb{Z}$, $k \neq 0$, and $\int_{-\infty}^{\infty} \psi^2(x) \, dx = 1$.

The orthonormality of $\psi(\cdot)$ ensures that the doubly infinite sequence $\{\psi_{jk}(\cdot)\}$, where

(3.1) $\psi_{jk}(x) \equiv 2^{j/2} \psi(2^j x - k), \quad j, k \in \mathbb{Z},$

constitutes an orthonormal basis for $L_2(\mathbb{R})$, the space of square-integrable functions on $\mathbb{R}$ (Daubechies (1992)). The integers $j$ and $k$ are called scale and
translation parameters. Intuitively, $j$ localizes analysis in frequency and $k$ localizes analysis in time or space.

Assumption 1 ensures that the Fourier transform of $\psi(\cdot)$,

$$
\hat{\psi}(z) \equiv (2\pi)^{-1/2} \int_{-\infty}^{\infty} \psi(x)e^{-izx} \, dx, \quad z \in \mathbb{R},
$$

exists and is continuous in $z$ almost everywhere. We impose a condition on $\hat{\psi}(\cdot)$.

**Assumption 2:** (a) $|\hat{\psi}(z)| \leq C(1 + |z|)^{-\tau}$ for some $\tau > \frac{3}{2}$; (b) $\hat{\psi}(z) = e^{ik/2}b(z)$ or $\hat{\psi}(z) = -ie^{ik/2}b(z)$, where $b(\cdot)$ is real-valued with $b(0) = 0$.

Many wavelets satisfy this. One example is spline wavelets of positive order $m \in \mathbb{Z}^+$. For odd $m$, this family has $\hat{\psi}(z) = e^{ik/2}b(z)$, where $b(\cdot)$ is real-valued and symmetric. For even $m$, it has $\hat{\psi}(z) = -ie^{ik/2}b(z)$, where $b(\cdot)$ is real-valued and odd (e.g., Hernández and Weiss (1996, (2.16), p. 161)). One member in this family is the first-order spline wavelet, called the Franklin wavelet, whose

$$
\hat{\psi}(z) = e^{iz/2}(2\pi)^{-1/2}\sin^4(z/4)\left[\frac{P_3(z/4 + \pi/4)}{P_3(z/2)P_3(z/4)}\right]^{1/2}, \quad \text{where}
$$

$$
P_3(z) \equiv \frac{2}{3} + \frac{1}{3}\cos(2z).
$$

Another member is the second-order spline wavelet, with

$$
\hat{\psi}(z) = -ie^{iz/2}(2\pi)^{-1/2}\sin^6(z/4)\left[\frac{P_5(z/4 + \pi/4)}{P_5(z/2)P_5(z/4)}\right]^{1/2}, \quad \text{where}
$$

$$
P_5(z) \equiv \frac{1}{30}\cos^2(2z) + \frac{13}{30}\cos(2z) + \frac{8}{15}.
$$

3.2. Wavelet Representation of Spectrum

We now consider wavelet representation of spectral density $f_i(\cdot)$. Given an orthonormal wavelet basis $\{\psi_{jk}(\cdot)\}$ for $L_2(\mathbb{R})$, we define

$$
\Psi_{jk}(\omega) \equiv (2\pi)^{-1/2} \sum_{m=-\infty}^{\infty} \psi_{jk}\left(\frac{\omega}{2\pi} + m\right), \quad \omega \in [-\pi, \pi].
$$

This constitutes an orthonormal wavelet basis for $L_2[-\pi, \pi]$, the space of $2\pi$-periodic functions on $[-\pi, \pi]$. See, e.g., Daubechies (1992, Ch. 9) or Hernández and Weiss (1996, Ch. 4).
One can also compute $\Psi_{jk}(\cdot)$ from its Fourier transform $\hat{\Psi}_{jk}(h) \equiv (2\pi)^{-1/2} \int_{-\pi}^{\pi} \Psi_{jk}(\omega)e^{-ih\omega} d\omega$ via the formula

\begin{equation}
\Psi_{jk}(\omega) = (2\pi)^{-1/2} \sum_{h=-\infty}^{\infty} \hat{\Psi}_{jk}(h)e^{ih\omega}.
\end{equation}

Lee and Hong (2001) show that the spectral density $f_i(\cdot)$ in (2.5) can be decomposed as

\begin{equation}
f_i(\omega) = (2\pi)^{-1}\sigma_i^2 + \sum_{j=0}^{\infty} 2^j \sum_{k=1}^{\infty} \alpha_{ijk} \Psi_{jk}(\omega), \quad \omega \in [-\pi, \pi],
\end{equation}

where the wavelet coefficient $\alpha_{ijk}$ is the orthogonal projection of $f_i(\cdot)$ on the base $\Psi_{jk}(\cdot)$; i.e.,

\begin{equation}
\alpha_{ijk} \equiv \int_{-\pi}^{\pi} f_i(\omega)\Psi_{jk}(\omega) d\omega.
\end{equation}

Unlike the Fourier transforms, $\alpha_{ijk}$ depends on the local behavior of $f_i(\cdot)$, because $\Psi_{jk}(\cdot)$ is effectively 0 outside an interval of width $2^{-j}$ centered at $k/2^j$. Such a spatial adoption feature makes it powerful for capturing nonsmooth features. We can also express $\alpha_{ijk}$ in time domain; i.e.,

\begin{equation}
\alpha_{ijk} = (2\pi)^{-1/2} \sum_{h=-\infty}^{\infty} R_i(h)\hat{\Psi}_{jk}(h),
\end{equation}

3.3. Wavelet Spectral Density Estimator

Define the sample autocovariance function of $\{\hat{u}_t\}$:

\begin{equation}
\hat{R}_i(h) \equiv T_i^{-1} \sum_{t=|h|+1}^{T_i} \hat{u}_t \hat{u}_{t-|h|} \quad (h = 0, \pm 1, \ldots, \pm (T_i - 1)).
\end{equation}

Then a wavelet estimator of the spectral density $f_i(\cdot)$ can be given by

\begin{equation}
\hat{f}_i(\omega) \equiv (2\pi)^{-1}\hat{R}_i(0) + \sum_{j=0}^{\infty} \sum_{k=1}^{2^j} \hat{\alpha}_{ijk} \Psi_{jk}(\omega), \quad \omega \in [-\pi, \pi],
\end{equation}

where the empirical wavelet coefficient

\begin{equation}
\hat{\alpha}_{ijk} \equiv (2\pi)^{-1/2} \sum_{h=-T_i}^{T_i-1} \hat{R}_i(h)\hat{\Psi}_{jk}(h),
\end{equation}
and $J_i \equiv J_i(T_i)$ is the finest scale corresponding to the highest resolution level. Appropriate conditions on $J_i$ will be given. We allow a different $J_i$ for a different $i$. This is useful because the pattern of serial correlation may vary substantially across $i$. We will also propose a data-driven method to choose $J_i$. Note that (3.11) is a linear wavelet estimator. Nonlinear wavelet estimators of Donoho et al. (1996) which are popular in curve estimation could be used. However, under our regularity conditions (see subsequent sections) which are not stronger than standard assumptions in time series panel econometrics, both linear and nonlinear wavelet estimators have the same convergence rate. Nonlinear wavelet estimators have no advantage at least in large samples. Masry (1994, 1997) also finds that the gain from using nonlinear wavelet estimators rather than linear wavelet estimators is marginal for different models. Moreover, the use of nonlinear wavelet estimators would lead to a much more complicated analysis in theory.

3.4. Wavelet-Based Tests

Put $Q(f_1, f_2) \equiv \int_{-\pi}^{\pi} [f_1(\omega) - f_2(\omega)]^2 d\omega$ for any $f_1(\cdot)$ and $f_2(\cdot)$. We use the quadratic form

$$Q(\hat{f}_i, \hat{f}_{i0}) = \sum_{j=0}^{J_i} \sum_{k=1}^{2^j} \hat{\alpha}_{ijk}^2,$$

where $\hat{f}_{i0}(\omega) \equiv (2\pi)^{-1} \hat{R}_i(0)$ and the equality follows by Parseval’s identity. Our first test statistic

$$\hat{W}_i \equiv \left( \sum_{i=1}^{n} 2\pi T_i \sum_{j=0}^{J_i} \sum_{k=1}^{2^j} \hat{\alpha}_{ijk}^2 - \hat{M} \right) / \hat{V}^{1/2},$$

where

$$\hat{M} \equiv \sum_{i=1}^{n} \hat{R}_i^2(0) M_{i0},$$

$$\hat{V} \equiv \sum_{i=1}^{n} \hat{R}_i^4(0) V_{i0},$$

$$M_{i0} \equiv \sum_{h=1}^{T_i-1} (1 - h/T_i) b_{j_i}(h, h),$$

$$V_{i0} \equiv 4 \sum_{h=1}^{T_i} \sum_{m=1}^{T_i} (1 - h/T_i)(1 - m/T_i) b_{j_i}^2(h, m),$$
\[ b_j(h, m) \equiv 2 \text{Re}[a_j(h, m) + a_j(h, -m)], \]
\[ a_j(h, m) \equiv \sum_{j=0}^{J} \sum_{k=1}^{2^j} \hat{\Psi}_{jk}(h) \hat{\Psi}_{jk}^*(m), \]
and \( \hat{\Psi}_{jk}(\cdot) \) is as in (3.6).

Put \( \hat{a}_{ijk} \equiv (2\pi)^{-1/2} \sum_{T^{-1}_i} \hat{\rho}_i(h) \hat{\Psi}_{jk}(h), \) where \( \hat{\rho}_i(h) \equiv \hat{R}_i(h)/\hat{R}_i(0). \) Our second test statistic
\[
(3.15) \quad \hat{W}_2 \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left( 2\pi T_i \sum_{j=0}^{J} \sum_{k=1}^{2^j} \hat{a}_{ijk}^2 - M_{it0} \right) / V_{i0}^{1/2}.
\]

Intuitively, \( \hat{W}_2 \) can be viewed as a heteroskedasticity-corrected test while \( \hat{W}_i \) is a heteroskedasticity-consistent test, where heteroskedasticity arises from different variances \( \sigma_i^2 \) and finest scales \( J_i \). In \( \hat{W}_2 \), these two forms of heteroskedasticity are corrected first for each \( i \). As is shown below, \( \hat{W}_1 \) and \( \hat{W}_2 \) are asymptotically \( N(0, 1) \) under \( \mathbb{H}_0 \), but their power properties generally differ. The heteroskedasticity-robust test \( \hat{W}_1 \) may be more powerful than the heteroskedasticity-consistent test \( \hat{W}_2 \). This differs from the well-known estimation result that correcting heteroskedasticity leads to more efficient estimation (e.g., the feasible GLS is more efficient than OLS).

Both \( \hat{W}_1 \) and \( \hat{W}_2 \) apply to one-way or two-way error component models. For one-way component models, however, one can use \( \hat{v}_{it} \equiv \hat{u}_{it} - \hat{u}_t \) if one knows \( \lambda_t = 0 \), and use \( \hat{v}_{it} \equiv \hat{u}_{it} - \hat{u}_i \) if one knows \( \mu_i = 0 \). The limit distribution of the test statistics is unchanged.

4. ASYMPOTIC DISTRIBUTION

We now impose a set of unified regularity conditions that hold under both \( \mathbb{H}_0 \) and \( \mathbb{H}_A \).

ASSUMPTION 3: \( \sqrt{nT}(\hat{\beta} - \beta^*) = O_p(1) \), where \( \beta^* = \beta \) under \( \mathbb{H}_0 \).

ASSUMPTION 4: (a) For each \( i \), \( \{v_{it}\} \) is covariance-stationary with \( E(v_{it}) = 0 \), \( E(v_{it}^2) = \sigma_i^2 \in [c, C] \), and \( E(v_{it}^4) \in [c, C] \); (b) the individual and time effects, \( \mu_i \) and \( \lambda_t \), can be stochastic (random effects) or deterministic (fixed effects).

ASSUMPTION 5: Put \( \tilde{\Gamma}_{ixv}(h) \equiv T_i^{-1} \sum_{t=h+1}^{T_i} \tilde{X}_{it} \tilde{v}_{it-h} \) if \( h \geq 0 \) and \( \tilde{\Gamma}_{ixv}(h) \equiv \tilde{\Gamma}_{ixv}(-h)' \) if \( h < 0 \), \( \Gamma_{ixv}(h) \equiv p \lim \tilde{\Gamma}_{ixv}(h) \), where \( \tilde{X}_{it} \equiv X_{it} - \tilde{X}_t - \tilde{X}_i + \tilde{X} \)
and \( \tilde{v}_{it} \equiv v_{it} - \tilde{v}_i - \tilde{v}_t + \tilde{v} \). Then (a) \( \sup_{1 \leq i \leq n} T_i^{-1} \sum_{i=1}^{T_i} E \| \tilde{X}_{it} \|^4 \leq C \); (b) \( \sup_{1 \leq i \leq n} \sup_{1 \leq h \leq T_i} E \| \tilde{I}_{iv}(h) - I_{iv}(h) \|^2 \leq CT_i^{-1} \); (c) \( \sum_{h=-\infty}^{\infty} \| I_{iv}(h) \| \leq C \).

We only require that the estimator \( \hat{\beta} \) be consistent for \( \beta \) under \( H_0 \); it need not be asymptotically most efficient under \( H_0 \) and even need not be consistent for \( \beta \) under \( H_A \). Thus, we can use the convenient within estimator \( \hat{\beta} \) in (2.2). Of course, other estimators such as OLS, feasible GLS, MLE, and IV are also allowed.

In Assumption 4, no restrictive assumptions on \( \{ \mu_i \} \) and \( \{ \lambda_i \} \) are imposed. We allow \( \{ \lambda_i \} \) to be serially correlated if \( \lambda_i \) is random, and \( \{ \mu_i \} \) to be spatially correlated if \( \mu_i \) is random. Given Assumption 3, \( \{ v_{it} \} \) coincides with the true error \( \{ \varepsilon_{it} \} \) under \( H_0 \). Here, we allow a certain degree of heterogeneity in panel data under \( H_0 \)—the process \( \{ Y_{it}, X_{it} \} \) need not be stationary for each \( i \), and \( \{ \varepsilon_{it} \} \) may have different variances across \( i \). In particular, we allow some nonstationary processes. One example is the deterministic trend process (e.g., Kao and Emerson (2004))

\[
Y_{it} = \alpha + \gamma_1 t + \gamma_2 t^2 + \cdots + \gamma_p t^p + \mu_i + \lambda_i + \varepsilon_{it}.
\]

This is covered by (2.1) with \( X_{it} \equiv [t/T, \ldots, (t/T)^p]' \) and \( \beta \equiv (T \gamma_1, \ldots, T^p \gamma_p)' \). Another example is the panel cointegration process (e.g., Phillips and Moon (1999), Kao and Chiang (2000)):

\[
Y_{it} = \alpha + \gamma Z_{it} + \mu_i + \lambda_i + \varepsilon_{it},
\]

where \( Z_{it} = Z_{it-1} + \eta_{it}, \{ \eta_{it} \} \) is I(0) for each \( i \), and \( \{ \eta_{it} \} \) may or may not be correlated with \( \{ \varepsilon_{it} \} \). This process is also covered by (2.1) with \( X_{it} \equiv T^{-1} Z_{it} \) and \( \beta \equiv T \gamma \).

However, Assumption 4 and (2.4) generally imply that both \( \{ \varepsilon_{it} \} \) and \( \{ X_{it} \} \) are covariance-stationary when \( \beta^* \neq \beta \) under \( H_A \). This more restrictive condition is needed under \( H_A \) when we investigate the asymptotic power property of the proposed tests.

**THEOREM 1:** Suppose Assumptions 1–5 hold and \( \max_{1 \leq i \leq n} (2^{2i})/(n^2 + T) \to 0 \) as \( n \to \infty \) and \( T \to \infty \). If \( \{ \varepsilon_{it} \} \) in (2.1) is i.i.d. for each \( i \), then \( \hat{W}_1 \overset{d}{\to} N(0, 1) \) and \( \hat{W}_2 \overset{d}{\to} N(0, 1) \).

The large sample properties of the proposed tests are due to the \( n \) independent copies of serially uncorrelated series \( \{ \varepsilon_{it} \} \) and the existence of a \( \sqrt{nT} \) consistent estimator for \( \beta \) under \( H_0 \). In fact, the assumption of a linear panel model is not necessary to obtain the large sample properties. One could extend Assumptions 4 and 5 to cover some nonlinear panel models, with more tedious algebra in the proof. It is also possible to relax the condition on the \( \sqrt{nT} \)-convergence rate of \( \hat{\beta} \) for \( \beta \) at a cost of strengthening conditions on the finest scales \( \{ J_i \} \).
Most remarkably, we permit but do not require $J_i \to \infty$ for any $i$; all $J_i$ can be fixed as $n, T \to \infty$ under $H_0$. This is in sharp contrast to Lee and Hong (2001), who require $J \to \infty$ as $T \to \infty$ to achieve asymptotic normality in testing serial correlation for observed raw time-series data. The reason that all $J_i$ can be fixed is that the additional smoothing provided by $n$ ensures asymptotic normality of $\hat{W}_1$ and $\hat{W}_2$. Intuitively, $\hat{W}_1$ and $\hat{W}_2$ are sums of approximately independent random variables $(2\pi T_i Q(\hat{f}_i, \hat{f}_{i0}))_{i=1}^n$. By the central limit theorem, they will converge to a normal distribution as $n \to \infty$. This occurs no matter whether $J_i \to \infty$. In the time-series or cross-sectional literature it is often found that the normal approximation is inadequate for the finite sample distributions of quadratic forms involving smoothed nonparametric estimation. This is because the asymptotic normality of these quadratic forms requires the smoothing parameter to grow or vanish at a suitable rate as the sample size grows and the convergence rate of test statistics delicately depends on the smoothing parameter. The fact that the asymptotic normality of $\hat{W}_1$ and $\hat{W}_2$ does not depend on whether $J_i \to \infty$ suggests that asymptotic approximation may work well in the panel context. Indeed, our simulation shows that $\hat{W}_1$ and $\hat{W}_2$ perform well in finite samples even when $J_i = 0$ for all $i$. Most importantly, the fact that $J_i$ may be fixed for all $i$ allows use of data-driven methods that may deliver a fixed finest scale under $H_0$. Sensible data-driven methods have this feature because the optimal finest scale $J_0 = 0$ under $H_0$. We will propose a plug-in method to select a finest scale, which automatically converges to 0 under $H_0$ and grows to $\infty$ under $H_A$, thus ensuring consistency against serial correlation of unknown form. Such a data-driven method could not be used for Lee and Hong’s (2001) test.

Although we require that both $n$ and $T$ grow to $\infty$, we do not impose a restrictive relative speed limit between them. Also, no specific estimation method is required. From the proof of Theorem 1, we find that parameter estimation uncertainty for $\beta$ has no impact on the null limit distribution of $\hat{W}_1$ and $\hat{W}_2$, no matter whether $X_{it}$ contains lagged dependent variables or deterministic/stochastic trending variables. This is in contrast to Durbin and Watson (1951) and Box and Pierce (1970), whose test statistics or limit distributions have to be modified when applied to estimated residuals of a stationary dynamic model. If regressors contain deterministic or stochastic trending variables, the limit distributions of these tests will become nonstandard (e.g., Kao and Chiang (2000), Kao and Emerson (2004)). Intuitively, parameter estimation uncertainty for $\beta$ induces an adjustment of a finite number of degrees of freedom for $\hat{W}_1$ and $\hat{W}_2$, but this becomes negligible as $n \to \infty$.

The tests $\hat{W}_1$ and $\hat{W}_2$ are applicable for both small and large $J_i$. When (and only when) $J_i \to \infty$ for all $i = 1, \ldots, n$, we can use the following simplified
versions of test statistics:

\[ W_1 = \frac{\sum_{i=1}^{n} k_{ij} \sum_{k=1}^{2^j} \hat{a}_{jk}^2 - \hat{R}_i^2(0)(2^{j+1} - 1)}{2 \left[ \sum_{i=1}^{n} \hat{R}_i(0)(2^{j+1} - 1) \right]^{1/2}} \]

\[ W_2 = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[ \frac{2 \pi T_i \sum_{j=0}^{l} \sum_{k=1}^{2^j} \hat{a}_{jk}^2 - (2^{j+1} - 1)}{2(2^{j+1} - 1)^{1/2}} \right]. \]

These are the generalizations of Lee and Hong’s (2001) test to estimated residuals of panel models. Theorem 2 shows that they are asymptotically \( N(0, 1) \) under \( H_0 \) if \( J_i \to \infty \) for all \( i \).

**Theorem 2:** Suppose Assumptions 1–5 hold, \( 2^{J_i+1} = a_i T_i^\nu \) for \( a_i \in [c, C] \) and \( \nu \in (0, \frac{1}{2}) \), \( n/T^\nu \log^2 T \to 0 \), \( n/T^{2(r-1) - 2(2r-1)/\nu} \to 0 \) as \( n, T \to \infty \), where \( r \geq \frac{1}{2} \) is as in Assumption 2. If \( (e_{it}) \) in (2.1) is i.i.d. for each \( i \), then \( W_1 \xrightarrow{d} 0 \), \( W_2 \xrightarrow{d} 0 \), \( W_1 \xrightarrow{d} N(0, 1) \), and \( W_2 \xrightarrow{d} N(0, 1) \).

Thus, for large (and only large) \( J_i \), \( \hat{W}_1 \) and \( \hat{W}_2 \) are asymptotically equivalent to \( \hat{W}_1 \) and \( \hat{W}_2 \) respectively. Note that here, \( n \) cannot grow faster than \( T^\nu \), where \( \nu < \frac{1}{2} \).

5. CONSISTENCY

We now show that \( \hat{W}_1 \) and \( \hat{W}_2 \) are consistent against \( H_A \). We assume the following condition.

**Assumption 6:** For each \( i \), \( \{v_{it}\} \) is a fourth-order zero-mean stationary process with \( \sum_{h=\infty}^{\infty} R_i^2(h) \leq C \) and \( \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} |\kappa_i(j, k, l)| \leq C \), where \( \kappa_i(j, k, l) \) is the fourth-order cumulant of the joint distribution of \( \{v_{it}, v_{it+j}, v_{it+k}, v_{it+l}\} \).

Assumption 6 characterizes temporal dependence of \( \{v_{it}\} \). When \( \{v_{it}\} \) is Gaussian, the cumulant condition holds trivially because \( \kappa_i(j, k, l) = 0 \) for all \( j, k, l \in \mathbb{Z} \). If for each \( i \), \( \{v_{it}\} \) is a fourth-order stationary linear process with absolutely summable coefficients and i.i.d. innovations whose fourth order moment exists, the cumulant condition also holds (e.g., Hannan (1970, p. 211)). More primitive conditions (e.g., strong mixing) could be imposed, but such primitive conditions would rule out long memory processes. Assumption 6 allows long memory processes \( I(d) \) with \( d < \frac{1}{2} \) for \( \{v_{it}\} \).

**Theorem 3:** Put \( n_A \equiv \#(N_A) \) and \( c_i \equiv T_i/T \), where \( N_A \equiv \{i: 0 < i \leq n, Q(f_i, f_{i0}) > 0\} \). Suppose Assumptions 1–6 hold, \( (n_A T)^{-1} \sum_{i=1}^{n} 2^i \to 0 \), and
$J_i \to \infty$ for all $i = 1, \ldots, n$ as $n, T \to \infty$. Then (a) $(n_A T)^{-1} \hat{V}^{1/2} \hat{W}_1 - n_A^{-1} \sum_{i \in \mathbb{N}_A} 2\pi c_i Q(f_i, f_{i0}) \Rightarrow 0$; (b) if in addition $2^{J_i+1} = a_i T_i$ for all $i$, where $a_i \in [c, C]$ and $\nu \in (0, 1)$, then $(n_A T^{1-\nu/2})^{-1} \hat{W}_2 - n_A^{-1} \sum_{i \in \mathbb{N}_A} \pi(c_i/a_i)^{1/2} Q(f_i, f_{i0}) \Rightarrow 0$.

As discussed earlier, although serial correlation in $\{v_{it}\}$ may differ from serial correlation in $\{e_{it}\}$, $\mathbb{H}_0$ holds if and only if $\{v_{it}\}$ is serially uncorrelated. Consequently, the index set $\mathbb{N}_A$ is nonempty under $\mathbb{H}_A$, at least for large $n$. It follows that $n_A^{-1} \sum_{i \in \mathbb{N}_A} c_i Q(f_i, f_{i0}) \geq c$ for large $n$. Then $P[\hat{W}_1 > C(n, T)] \to 1$ and $P[\hat{W}_2 > C(n, T)] \to 1$ under $\mathbb{H}_A$ for any sequence of constants $\{C(n, T) = o(n_A T/(\sum_{i=1}^n 2^{J_i})^{1/2})\}$. Thus, $\hat{W}_1$ and $\hat{W}_2$ are consistent against $\mathbb{H}_A$ provided $(n_A T)^{-1} \sum_{i=1}^n 2^{J_i} \to 0$ and $J_i \to \infty$ for all $i$. For simplicity, we let $J_i \to \infty$ for all $i$ to ensure consistency against $\mathbb{H}_A$. This differs from the case under $\mathbb{H}_0$, where $J_i$ can be fixed for all $i$. Our data-driven method below will deliver a finest scale that automatically converges to 0 under $\mathbb{H}_0$ but grows to $\infty$ with $T$ under $\mathbb{H}_A$.

Under $\mathbb{H}_A$, $\hat{W}_1$ and $\hat{W}_2$ diverge to $\infty$ at the rate of $n_A T/(\sum_{i=1}^n 2^{J_i})^{1/2}$. Thus, the larger the set $\mathbb{N}_A$ is, the more powerful $\hat{W}_1$ and $\hat{W}_2$ are. In fact, the power depends on $n_A/n$, the proportion of individuals with serial correlation. For $2^{J_i+1} = a_i T_i$, $n_A T(\sum_{i=1}^n 2^{J_i})^{1/2} \propto (n_A/n)n^{1/2} T^{1-\nu/2}$. This implies that $\hat{W}_1$ and $\hat{W}_2$ have asymptotic power 1 against $\mathbb{H}_A$ even if the proportion $n_A/n \to 0$ at a rate slightly slower than $n^{1/2} T^{1-\nu/2}$. For $\hat{W}_1$ and $\hat{W}_2$, serial correlations from different individuals never cancel each other out when some individuals have positive autocorrelations and some have negative autocorrelations. In contrast, cancellation may occur at least in part for existing tests, leading to low or little power; see Section 7 for more discussion. We emphasize that the ability of our tests to detect serial correlation in the presence of substantial inhomogeneity in serial correlation across $i$ is not due to the use of wavelets, but to the use of the quadratic form in (3.13). On the other hand, we may extend the adaptive procedures of Fan (1996) and Spokoiny (1996) to further improve the power of our tests, as one referee pointed out. We leave this for future research.

Theorems 1 and 3 imply that for large $n$ and $T$, the negative values of $\hat{W}_1$ and $\hat{W}_2$ can occur only under $\mathbb{H}_0$. Thus, upper-tailed $\text{N}(0, 1)$ critical values should be used.

As noted earlier, $\hat{W}_1$ and $\hat{W}_2$ are heteroskedasticity-consistent and heteroskedasticity-corrected tests respectively. An interesting question is, which test, $\hat{W}_1$ or $\hat{W}_2$, is more powerful? For convenience, we assume $2^{J_i+1} = a_i T_i$ for all $i$, where $a_i \in [c, C]$ and $\nu \in (0, 1)$, and assume a larger $a_i$ for processes with stronger serial correlation in terms of a larger $Q(\hat{f}_i, f_{i0})$. With this rule, we have the following theorem.

**Theorem 4:** Suppose Assumptions 1–6 hold, $n = \gamma T^s$ for $\gamma \in (0, \infty)$ and $s \in (0, \infty)$, and $2^{J_i+1} = a_i T_i$ for $a_i \in [c, C]$ and $\nu \in (0, 1)$. If $a_i$ is monotonically
increasing in $Q(f_i, f_{i0})$ and $T_i = T$ for all $i$, then $\hat{W}_1$ is more efficient than $\hat{W}_2$ in terms of Bahadur’s asymptotic efficiency criterion.

Bahadur’s (1960) asymptotic slope criterion is pertinent for power comparison of large sample tests under fixed alternatives. The basic idea is to compare the logarithms of the asymptotic significance levels (i.e., $p$-values) of the tests under a fixed alternative. Bahadur’s asymptotic efficiency is defined as the limit ratio of the sample sizes required by the two tests under comparison to achieve the same asymptotic significance level ($p$-value) under a fixed alternative.

Theorem 4 implies that for hypothesis testing, correcting heteroskedasticity may give poorer power. This is in contrast to the well-known result that correcting heteroskedasticity leads to more efficient estimation. Intuitively, for $\hat{W}_2$, a larger $Q(f_i, f_{i0})$ is more heavily discounted by $\sqrt{V_{i0}} \sim 2(2^{J_i+1} - 1)^{1/2}$ when $J_i$ is larger. Thus, it is less powerful than $\hat{W}_1$, which puts uniform weighting on each $Q(f_i, f_{i0})$. Of course it is possible that $\hat{W}_1$ is asymptotically less powerful than $\hat{W}_2$, as will occur when $a_i$ is monotonically decreasing in $Q(f_i, f_{i0})$. However, sensible data-driven methods usually provide a rule that $a_i$ is increasing in $Q(f_i, f_{i0})$. When $J_i = J$ for all $i$, $\hat{W}_1$ and $\hat{W}_2$ may still not be asymptotically equally efficient, because of heteroskedasticity ($\sigma_i^2 \neq \sigma^2$). We note that the asymptotic power of $\hat{W}_1$ and $\hat{W}_2$ does not depend on mother wavelet $\psi(\cdot)$. All wavelets are asymptotically equally efficient by Bahadur’s criterion. This differs from the kernel method, where the choice of a kernel affects the asymptotic power of tests (Hong (1996)).

6. ADAPTIVE CHOICE OF FINEST SCALE

Theorem 1 implies that the choice of $J_i$ is not important for asymptotic normality of $\hat{W}_1$ and $\hat{W}_2$. Both small and large $J_i$ can be used. However, the choice of $J_i$ may have significant impact on the power. Therefore, it is desirable to choose $J_i$ via suitable data-driven methods. We now develop a data-driven method to select a finest scale. We first justify the use of a data-driven finest scale $\hat{J}$. For simplicity, we consider a common $\hat{J}$ for all $i$ here. We use $\hat{W}_c(\hat{J})$ and $\hat{W}_c(J)$ to denote the $\hat{W}_c$ tests using $\hat{J}$ and $J$ respectively, where $c = 1, 2$.

We impose a condition on the smoothness of $\hat{\psi}(\cdot)$ at 0.

**ASSUMPTION 7:** $|\hat{\psi}(z)| \leq C|z|^q$ for some $q \in (0, \infty)$.

**THEOREM 5:** Suppose Assumptions 1–5 and 7 hold, and $\hat{J}$ is a data-driven finest scale with $2^J / 2^\hat{J} = 1 + o_p(2^{-J/2})$, where $J$ is a nonstochastic finest scale such that $2^{2J} / (n^2 + T) \rightarrow 0$ as $n \rightarrow \infty$ and $T \rightarrow \infty$. If $\{\varepsilon_{it}\}$ in (2.1) is i.i.d. for each $i$, then $W_1(\hat{J}) - W_1(J) \overset{p}{\rightarrow} 0$, $W_2(\hat{J}) - W_2(J) \overset{p}{\rightarrow} 0$, $\hat{W}_1(\hat{J}) \overset{d}{\rightarrow} N(0, 1)$, and $\hat{W}_2(\hat{J}) \overset{d}{\rightarrow} N(0, 1)$. 
Thus, the use of $\hat{J}$ rather than $J$ has no impact on the limit distribution of $\hat{W}_1(\hat{J})$ and $\hat{W}_2(\hat{J})$ provided that $\hat{J}$ converges to $J$ at a suitable rate. The rate condition $2^J/2^j - 1 = o_p(2^{-j/2})$ is mild. If $2^j \propto T^{1/5}$, for example, we require $2^J/2^j = 1 + o_p(T^{-1/10})$. If $J$ is fixed (e.g., $J = 0$), which occurs under $\mathbb{H}_0$ for our data-driven method below, the condition becomes $2^J/2^j \propto 1$.

So far very few data-driven methods to choose $J$ are available in the literature. Walter (1994) and Hall and Patil (1996) consider some data-driven methods in related but different contexts. They cannot be applied directly to our tests. We now develop a data-driven $\hat{J}$ that can satisfy the condition of Theorem 5. We first derive the average asymptotic IMSE formula for $\{\hat{f}_i(\cdot)\}_{i=1}^n$ which was not available in the literature. We impose an additional condition on $\{\nu_{it}\}$.

**ASSUMPTION 8:** $\sum_{h=-\infty}^{\infty} |h|^q |R_i(h)| \leq C$ for all $i$, where $q \in [1, \infty)$ is as in Assumption 7.

This characterizes the smoothness of $f_i(\cdot)$. It rules out long memory processes. Under Assumption 8, we have a well-defined $q$th order generalized spectral derivative of $f_i(\omega)$:

$$f_i^{(q)}(\omega) \equiv (2\pi)^{-1} \sum_{h=-\infty}^{\infty} |h|^q R_i(\omega)e^{-ih\omega}, \quad \omega \in [-\pi, \pi].$$

We also define a measure $\lambda_q \in (0, \infty)$ of the smoothness of $\hat{\psi}(\cdot)$ at 0:

$$\lambda_q \equiv -\frac{(2\pi)^{2q+1}}{1 - 2^{-2q}} \lim_{z \to 0} \frac{\hat{\psi}(z)^2}{|z|^{2q}}.$$

For the Franklin wavelet (3.3), $q = 2$; for the second-order spline wavelet (3.4), $q = 3$.

To state the next result, we define a pseudo spectral density estimator $\tilde{f}_i(\cdot)$ for $f_i(\cdot)$ that is based on the unobservable series $\{\nu_{it}\}_{t=1}^{T_i};$ namely,

$$\tilde{f}_i(\omega) \equiv (2\pi)^{-1} \tilde{R}_i(0) + \sum_{j=1}^{J_i} \sum_{k=1}^{2^j} \tilde{\alpha}_{ijk} \Psi_{jk}(\omega), \quad \omega \in [-\pi, \pi],$$

where $\tilde{R}_i(h) \equiv T_i^{-1} \sum_{t=h+1}^{T_i} \nu_{it} \nu_{it-|h|}$ and $\tilde{\alpha}_{ijk} \equiv (2\pi)^{-1/2} \sum_{h=1}^{T_i-1} \tilde{R}_i(h) \hat{\psi}_{jk}(h)$.

**THEOREM 6:** Suppose Assumptions 1–8 hold, $\lambda_q \in (0, \infty)$, $J_i \to \infty$, $2^J/T_i \to 0$ as $T_i \to \infty$. Then (a) for each $i$, $Q(\hat{f}_i, f_i) = Q(\tilde{f}_i, f_i) + o_p(2^{J_i}/T_i + 2^{-2qJ_i})$,
and

\[ EQ(\tilde{f}_i, f_i) = \frac{2^{J_{i+1}}}{T_i} \int_{-\pi}^{\pi} \tilde{f}_i^2(\omega) \, d\omega + 2^{-2q(J_{i+1})} \lambda_q \int_{-\pi}^{\pi} \left[ f^{(q)}(\omega) \right]^2 \, d\omega \]

\[ + o(2^{J_i}/T_i + 2^{-2qJ}). \]

(b) If in addition \( J_i = J \) for all \( i \) and \( T_i/T = c_i \), then \( n^{-1} \sum_{i=1}^{n} Q(\tilde{f}_i, f_i) = n^{-1} \sum_{i=1}^{n} Q(\tilde{f}_i, f_i) + o_P(2^J/T + 2^{-2qJ}) \), and

\[ n^{-1} \sum_{i=1}^{n} EQ(\tilde{f}_i, f_i) = \frac{2^{J_{i+1}}}{T} n^{-1} \sum_{i=1}^{n} c_i^{-1} \int_{-\pi}^{\pi} \tilde{f}_i^2(\omega) \, d\omega \]

\[ + 2^{-2q(J_{i+1})} \lambda_q n^{-1} \sum_{i=1}^{n} \int_{-\pi}^{\pi} \left[ f^{(q)}(\omega) \right]^2 \, d\omega \]

\[ + o(2^J/T + 2^{-2qJ}). \]

Theorem 6(a) gives the asymptotic IMSE of \( \hat{f}_i(\cdot) \), and Theorem 6(b) gives the average asymptotic IMSE of \( \{\hat{f}_i(\cdot)\}_{i=1}^{n} \). They imply that the optimal convergence rates of \( Q(\hat{f}_i, f_i) \) and \( n^{-1} \sum_{i=1}^{n} Q(\hat{f}_i, f_i) \) are the same as those of \( Q(\tilde{f}_i, f_i) \) and \( n^{-1} \sum_{i=1}^{n} Q(\tilde{f}_i, f_i) \) respectively. Parameter estimation uncertainty in \( \hat{\beta} \) has no impact on the optimal convergence rates of \( Q(\hat{f}_i, f_i) \) and \( n^{-1} \sum_{i=1}^{n} Q(\hat{f}_i, f_i) \).

The optimal finest scale \( J_0 \) that minimizes the average asymptotic IMSE of \( \{\hat{f}_i(\cdot)\}_{i=1}^{n} \) is

\[ 2^{J_0} = [2q\lambda_q \xi_0(q)T]^{1/(2q+1)}, \]

where \( \xi_0(q) \equiv \sum_{i=1}^{n} \int_{-\pi}^{\pi} [f^{(q)}(\omega)]^2 \, d\omega/\sum_{i=1}^{n} c_i^{-1} \int_{-\pi}^{\pi} \tilde{f}_i^2(\omega) \, d\omega \). This is infeasible because \( \xi_0(q) \) is unknown under \( \mathbb{H}_A \). However we can use some estimator \( \hat{\xi}_0(q) \) to obtain a plug-in finest scale \( \hat{J}_0 \):

\[ 2^{\hat{J}_0} = [2q\lambda_q \hat{\xi}_0(q)T]^{1/(2q+1)}. \]

Because \( \hat{J}_0 \) is a nonnegative integer, we should use

\[ \hat{J}_0 = \max \left\{ \left[ \frac{1}{2q+1} \log_2(2q\lambda_q \hat{\xi}_0(q)T) - 1 \right] , 0 \right\}, \]

where the square bracket denotes the integer part. We impose the following condition on \( \hat{\xi}_0(q) \).

**ASSUMPTION 9:** \( \hat{\xi}_0(q) - \xi_0(q) = o_P(T^{-\delta}) \) for some constant \( \xi_0(q) \), where \( \delta = 1/(2q+1) \) if \( \xi_0(q) \in [c, C] \) and \( \delta = 1/(2q+1) \) if \( \xi_0(q) = 0 \).
Note that the condition on $\hat{\xi}_0(q)$ is more stringent when $\xi_0(q) = 0$ than when $\xi_0(q) \neq 0$, but for both cases the conditions are mild. We do not require $p \lim \hat{\xi}_0(2) \equiv \xi_0(q) = \xi_0(q)$, where $\xi_0(q)$ is as in (6.4). When (and only when) $\xi_0(q) = \xi_0(q)$, $\hat{J}_0$ in (6.6) will converge to the optimal $J_0$ in (6.4).

COROLLARY 1: Suppose Assumptions 1–9 hold and $\hat{J}_0$ is given as in (6.6). If $\{\varepsilon_{it}\}$ in (2.1) is i.i.d. for each $i$, then $\hat{W}_1(\hat{J}_0) \overset{d}{\to} N(0, 1)$ and $\hat{W}_2(\hat{J}_0) \overset{d}{\to} N(0, 1)$.

We can use parametric or nonparametric (e.g., local linear smoothing) methods for estimator $\hat{\xi}_0(q)$. The former generally deliver a suboptimal finest scale, but have less variation in finite samples. The latter will deliver an asymptotically optimal finest scale, but are subject to substantial variation. There are also some methods (e.g., cross-validation and AIC) in the literature for selecting the truncation parameter. There is usually a trade-off between Type I and Type II errors in choosing a specific method.

To obtain reasonable power while having a proper rejection probability under $\mathbb{H}_0$ for sample sizes often encountered in economics, we use a parametric AR($p_i$) model for each $i$:

$$\hat{\nu}_{it} = \gamma_{i0} + \sum_{h=1}^{p_i} \gamma_{ih} \hat{\nu}_{it-h} + \varepsilon_{it} \quad (t = 1, \ldots, T_i; \ i = 1, \ldots, n),$$

where $\hat{\nu}_{it} \equiv 0$ if $t < 0$. In practice, one can use AIC or BIC to select $p_i$ for each $i$. Suppose $\hat{\gamma}_i \equiv (\hat{\gamma}_{i0}, \hat{\gamma}_{i1}, \ldots, \hat{\gamma}_{ip_i})'$ is the OLS estimator of $\gamma_i \equiv (\gamma_{i0}, \gamma_{i1}, \ldots, \gamma_{ip_i})'$. We consider $q = 2$; an example is the Franklin wavelet in (3.3), whose $\lambda_2 = \pi^4/45$. We have

$$\hat{\xi}_0(2) \equiv \sum_{i=1}^{n} \int_{-\pi}^{\pi} \left[ \frac{d^2}{d\omega^2} \hat{f}_i(\omega) \right]^2 d\omega // \sum_{i=1}^{n} (T_i/T) \int_{-\pi}^{\pi} \hat{f}_i^2(\omega) d\omega,$$

where $\hat{f}_i(\omega) \equiv (2\pi)^{-1} |1 - \sum_{h=1}^{p_i} \hat{\gamma}_{ih} e^{-i\omega}|^{-2}$. We note that $\hat{\xi}_0(2)$ satisfies Assumption 9 with $q = 2$ because for parametric AR($p_i$) approximations, $\hat{\xi}_0(2) - \xi_0(2) = O_p((nT)^{-1/2})$.

The performance of the plug-in method in (6.6) relies on the specification in (6.7). To further improve the power, one could also consider a data-driven, individual-specific $\hat{J}_i$ using the IMSE criterion of $f_i(\cdot)$ in Theorem 6(a). Such individual-specific $\hat{J}_i$ may more effectively capture spatial nonhomogeneity in the degree of serial correlation across $i$. However, they may have wide variations across $i$, leading to large Type I errors for the tests. A compromise is to develop a data-driven $\hat{J}_c$, where $c$ is an index for some suitable groups such as regions and sectors where all individuals in the same group will have the same finest scale.
In fact, the IMSE criterion is more suitable for estimation than for testing. A better criterion for testing is to maximize power or trade off between level distortion and power improvement. This will, however, require higher-order asymptotic analysis for our tests, which is beyond the scope of this paper and should be pursued in future work. Simulation studies below show that the finest scale chosen via (6.6)–(6.8) gives reasonable rejection probabilities under $H_0$ and gives robust and good power, particularly when the spectrum has distinct peaks or kinks.

7. MONTE CARLO EXPERIMENT

We now compare the performance of $\hat{W}_1$ and $\hat{W}_2$ with three existing tests for serial correlation—Bhargava, Franzini, and Narendranathan’s (1982; BFN) Durbin–Watson type test, Baltagi and Li’s (1995; BL) LM test, and Bera, Sosa-Escudero, and Yoon’s (2001; BSY) modified LM test:

$$BL = \frac{nT^2}{T-1} \left( \frac{\sum_{i=1}^{n} \sum_{t=1}^{T} \hat{v}_{it} \hat{v}_{it-1}}{\sum_{i=1}^{n} \sum_{t=1}^{T} \hat{v}_{it}^2} \right)^2,$$

and

$$BSY = nT^2 \left( \tilde{B} + \frac{\tilde{A}}{T} \right)^2 \left[ (T-1) \left( 1 - \frac{2}{T} \right) \right],$$

where $\hat{v}_{it}$ is the within residual, $\tilde{B} \equiv \frac{\sum_{i=1}^{n} \sum_{t=1}^{T} \tilde{u}_{it} \tilde{u}_{it-1}}{\sum_{i=1}^{n} \sum_{t=1}^{T} \tilde{u}_{it}^2}$, and $\tilde{A} \equiv \left( 1 - \sum_{i=1}^{n} \tilde{u}_i' I_T \tilde{u}_i \right) / \sum_{i=1}^{n} \sum_{t=1}^{T} \tilde{u}_{it}^2$, $I_T$ is a matrix of ones, and $\tilde{u}_i \equiv (\tilde{u}_{i1}, \ldots, \tilde{u}_{iT})'$ is the OLS residual without random effects. All these tests consider balanced panels. Both BL and BSY have an asymptotic $\chi^2_1$ distribution under $H_0$. In contrast, BFN converges to 2 under $H_0$, a degenerate distribution. Bhargava, Franzini, and Narendranathan (1982, p. 436) suggest using a critical value of 2 at the 5% level. We find that BFN rejects $H_0$ up to 67.7%, 66.9%, and 64.8% at the 5% level, when $(n, T) = (25, 32), (50, 64), \text{and} (100, 128)$ respectively. It seems that this test cannot be used, so we drop it from comparison. For $\hat{W}_1$ and $\hat{W}_2$, because the choice of $\psi(\cdot)$ is not important, we only use the Franklin wavelet in (3.3). To examine the impact of the choice of $J$, we consider $J = 0, 1$ and the data-driven $\hat{J}_0$ in (6.6). We also compare $\hat{W}_1$ and $\hat{W}_2$ with the panel versions of Hong’s (1996) kernel-based tests, $\hat{K}_1$ and $\hat{K}_2$, which are obtained by gener-
alizing Hong’s (1996) kernel method to model (2.1). They are based on an individual-specific kernel spectral density estimator. We use the Daniell kernel $k(z) = \sin(\pi z) / \pi z$, $z \in \mathbb{R}$, which has the optimal power over a class of kernels. We choose a data-driven bandwidth using a plug-in method similar to that for $\hat{J}_0$.

We consider the following two data generating processes (DGP): (a) DGP1, a static panel model: $Y_{it} = 5 + .5X_{it} + \mu_i + \varepsilon_{it}$, $X_{it} = .5X_{it-1} + \eta_{it}$, $\eta_{it} \sim$ i.i.d. $U[-.5, .5]$, and (b) DGP2, a dynamic panel model: $Y_{it} = 5 + .5Y_{it-1} + \mu_i + \varepsilon_{it}$. For both DGPs, $\mu_i \sim$ i.i.d. $N(0, \sigma^2_{\mu})$. Let $\tau$ measure the relative strength of random effects (no random effect when $\tau = 0$) and we choose a variety of $\tau$ as in Baltagi, Chang, and Li (1992). To examine the probability of rejecting a correct $H_0$, we set $\varepsilon_{it} = z_{it}$, where $z_{it}$ i.i.d. $N(0, 1)$. To examine the power, we consider the following two error processes:

AR(1) Alternatives:

$$
\begin{align*}
\text{AR}(1)^a: & \quad \varepsilon_{it} = 2\varepsilon_{it-1} + z_{it}, \quad i = 1, \ldots, n, \\
\text{AR}(1)^b: & \quad \varepsilon_{it} = -2\varepsilon_{it-1} + z_{it}, \quad i = 1, \ldots, n, \\
\text{AR}(1)^c: & \quad \varepsilon_{it} = \begin{cases} .2\varepsilon_{it-1} + z_{it}, & i = 1, \ldots, \frac{n}{2}, \\
-2\varepsilon_{it-1} + z_{it}, & i = \frac{n}{2} + 1, \ldots, n. \end{cases}
\end{align*}
$$

(7.3)

ARMA(12, 4) Alternatives:

$$
\begin{align*}
\text{ARMA}(12, 4)^a: & \quad \varepsilon_{it} = -3\varepsilon_{it-12} + z_{it} + z_{it-4}, \quad i = 1, \ldots, n, \\
\text{ARMA}(12, 4)^b: & \quad \varepsilon_{it} = 3\varepsilon_{it-12} + z_{it} - z_{it-4}, \quad i = 1, \ldots, n, \\
\text{ARMA}(12, 4)^c: & \quad \varepsilon_{it} = \begin{cases} -3\varepsilon_{it-12} + z_{it} + z_{it-4}, & i = 1, \ldots, \frac{n}{2}, \\
3\varepsilon_{it-12} + z_{it} - z_{it-4}, & i = \frac{n}{2} + 1, \ldots, n. \end{cases}
\end{align*}
$$

(7.4)

AR(1)^a and AR(1)^b are the full positive and negative AR(1) respectively. We expect BL and BSY to have optimal power against them. Wavelet tests have no advantages because these alternatives have a relatively flat spectrum. AR(1)^c is a mixed AR(1), where the first half individuals have a positive AR coefficient, and the second half have a negative AR coefficient. On the other hand, ARMA(12, 4) can arise from monthly data; it has four distinct spectral peaks.

The top panel of Table I reports rejection probabilities under $H_0$ with $\tau = .4$ for DGP1. When $n < T$, the rejection probabilities of BSY are reasonable and the best among all the tests. BL overrejects $H_0$, while $\hat{W}_1$, $\hat{W}_2$, $\hat{K}_1$, and $\hat{K}_2$ underreject $H_0$ but not excessively. For other values of $\tau$ (not reported), the rejection probability of BSY is sensitive to the choice of $\tau$, displaying severe overrejections when $\tau$ is large. BL still overrejects $H_0$ for all $\tau$. The $\hat{W}_1$, $\hat{W}_2$, $\hat{K}_1$, and $\hat{K}_2$ tests are robust to the choice of $\tau$. The rejection probabilities of $\hat{W}_1$ and $\hat{W}_2$
### TABLE I

PERCENTAGE OF REJECTIONS UNDER THE NULL HYPOTHESIS IN THE STATIC PANEL MODEL

<table>
<thead>
<tr>
<th>$(n, T)$</th>
<th>(5, 8)</th>
<th>(10, 16)</th>
<th>(25, 32)</th>
<th>(50, 64)</th>
<th>(5, 5)</th>
<th>(10, 10)</th>
<th>(25, 25)</th>
<th>(50, 50)</th>
<th>(8, 5)</th>
<th>(16, 10)</th>
<th>(32, 25)</th>
<th>(64, 50)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Level</strong></td>
<td>10%</td>
<td>5%</td>
<td>10%</td>
<td>5%</td>
<td>10%</td>
<td>5%</td>
<td>10%</td>
<td>5%</td>
<td>10%</td>
<td>5%</td>
<td>10%</td>
<td>5%</td>
</tr>
<tr>
<td>(W_1(0)) &amp; 2.5 &amp; 1.6 &amp; 4.5 &amp; 2.8 &amp; 6.0 &amp; 3.1 &amp; 6.6 &amp; 2.9 &amp; .6 &amp; .0 &amp; 3.1 &amp; 1.7 &amp; 4.8 &amp; 2.6 &amp; 6.6 &amp; 3.4 &amp; .4 &amp; .0 &amp; 2.6 &amp; 1.8 &amp; 4.9 &amp; 2.5 &amp; 6.3 &amp; 3.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(W_1(1)) &amp; 1.8 &amp; 1.3 &amp; 3.3 &amp; 2.0 &amp; 5.1 &amp; 2.8 &amp; 5.8 &amp; 2.7 &amp; .3 &amp; .0 &amp; 2.0 &amp; 1.3 &amp; 3.7 &amp; 2.1 &amp; 5.8 &amp; 3.5 &amp; .2 &amp; .1 &amp; 1.7 &amp; .7 &amp; 3.5 &amp; 1.2 &amp; 4.8 &amp; 2.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(W_1(J_0)) &amp; 1.8 &amp; 1.3 &amp; 3.3 &amp; 1.9 &amp; 6.3 &amp; 4.0 &amp; 7.9 &amp; 4.0 &amp; .6 &amp; .0 &amp; 2.0 &amp; 1.3 &amp; 4.5 &amp; 2.5 &amp; 7.8 &amp; 4.3 &amp; .4 &amp; .1 &amp; 1.7 &amp; .7 &amp; 4.1 &amp; 1.7 &amp; 6.9 &amp; 3.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(W_2(0)) &amp; 1.4 &amp; .4 &amp; 4.5 &amp; 2.5 &amp; 5.7 &amp; 3.0 &amp; 6.8 &amp; 3.4 &amp; .1 &amp; .0 &amp; 2.4 &amp; 1.1 &amp; 4.3 &amp; 2.2 &amp; 6.6 &amp; 3.5 &amp; .1 &amp; .0 &amp; 1.4 &amp; .7 &amp; 4.3 &amp; 2.8 &amp; 5.9 &amp; 2.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(W_2(1)) &amp; 1.0 &amp; .4 &amp; 2.8 &amp; 1.4 &amp; 4.6 &amp; 1.9 &amp; 6.0 &amp; 3.0 &amp; .0 &amp; .0 &amp; .8 &amp; .3 &amp; 2.9 &amp; 1.4 &amp; 5.4 &amp; 2.6 &amp; .0 &amp; .0 &amp; .7 &amp; .3 &amp; 2.4 &amp; 1.7 &amp; 4.2 &amp; 2.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(W_2(J_0)) &amp; 1.0 &amp; .4 &amp; 3.0 &amp; 1.1 &amp; 4.7 &amp; 2.3 &amp; 8.2 &amp; 3.7 &amp; .1 &amp; .0 &amp; .8 &amp; .3 &amp; 3.0 &amp; 1.6 &amp; 7.2 &amp; 3.6 &amp; .1 &amp; .0 &amp; .7 &amp; .3 &amp; 2.6 &amp; 1.3 &amp; 6.3 &amp; 3.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(K_1) &amp; 2.5 &amp; 1.5 &amp; 3.6 &amp; 1.7 &amp; 6.6 &amp; 3.9 &amp; 7.2 &amp; 3.9 &amp; .9 &amp; .0 &amp; 2.3 &amp; 1.3 &amp; 5.0 &amp; 2.9 &amp; 7.5 &amp; 4.0 &amp; .4 &amp; .2 &amp; 1.6 &amp; .8 &amp; 4.2 &amp; 2.1 &amp; 6.7 &amp; 3.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(K_2) &amp; 1.4 &amp; .4 &amp; 2.8 &amp; 1.2 &amp; 5.8 &amp; 2.6 &amp; 8.0 &amp; 3.8 &amp; .1 &amp; .0 &amp; 1.2 &amp; .2 &amp; 3.3 &amp; 2.0 &amp; 7.5 &amp; 3.4 &amp; .1 &amp; .0 &amp; .7 &amp; .2 &amp; 3.1 &amp; 1.7 &amp; 7.1 &amp; 3.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>BL</strong> &amp; 30.5 &amp; 19.9 &amp; 22.6 &amp; 14.9 &amp; 23.2 &amp; 14.6 &amp; 22.5 &amp; 14.5 &amp; 45.3 &amp; 32.4 &amp; 35.2 &amp; 23.9 &amp; 30.2 &amp; 20.8 &amp; 28.2 &amp; 19.2 &amp; 57.8 &amp; 46.3 &amp; 48.0 &amp; 35.4 &amp; 34.7 &amp; 24.5 &amp; 35.3 &amp; 23.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>BSY</strong> &amp; 9.0 &amp; 3.0 &amp; 5.7 &amp; 2.1 &amp; 6.7 &amp; 1.8 &amp; 7.7 &amp; 2.2 &amp; 18.3 &amp; 9.3 &amp; 13.5 &amp; 5.1 &amp; 9.9 &amp; 3.4 &amp; 7.3 &amp; 2.8 &amp; 31.9 &amp; 20.7 &amp; 22.0 &amp; 10.3 &amp; 14.2 &amp; 5.9 &amp; 11.4 &amp; 4.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Asymptotic Critical Values**

- \(W_1(0)\)
- \(W_1(1)\)
- \(W_1(J_0)\)
- \(W_2(0)\)
- \(W_2(1)\)
- \(W_2(J_0)\)
- \(K_1\)
- \(K_2\)
- **BL**
- **BSY**

**Bootstrapped Critical Values**

- \(W_1(0)\)
- \(W_1(1)\)
- \(W_1(J_0)\)
- \(W_2(0)\)
- \(W_2(1)\)
- \(W_2(J_0)\)
- \(K_1\)
- \(K_2\)
- **BL**
- **BSY**

**Note:**
(a) DGP: \(Y_{it} = 5 + .5X_{it} + \mu_i + \epsilon_{it}, X_{it} = .5X_{it-1} + \eta_{it}, \eta_{it} \sim \text{i.i.d.} U[-5,5], \mu_i \sim \text{i.i.d.} N(0,0.4), \) and \(\epsilon_{it} \sim \text{i.i.d.} N(0,1).\) (b) \(W_1,\) heteroskedasticity-consistent Franklin wavelet-based test; \(W_2,\) heteroskedasticity-corrected Franklin wavelet-based test; \(J_0,\) data-driven finest scale. (c) \(K_1,\) heteroskedasticity-consistent Daniell kernel-based test; \(K_2,\) heteroskedasticity-corrected Daniell kernel-based test. (d) BL, Baltagi–Li test; BSY, Bera, Sosa-Escudero, and Yoon test. (e) 1000 simulation replications; 500 bootstrap replications.
are better when a smaller $J$ or data-driven $\hat{J}_0$ is used. When $n = T$, $\hat{W}_1$, $\hat{W}_2$, $\hat{K}_1$, and $\hat{K}_2$ all underreject $H_0$, but they improve when $n = T$ increase. On the other hand, BL overrejects severely while BSY performs well. When $n > T$, again, BSY has the best rejection probabilities if $T > 10$, but overrejects if $T$ is small. BL still overrejects $H_0$ for all sample sizes. The $\hat{W}_1$, $\hat{W}_2$, $\hat{K}_1$, and $\hat{K}_2$ tests all underreject $H_0$ but they are substantially improved when $T > 25$, especially when data-driven $\hat{J}_0$ or $J = 0$ is used. The rejection probabilities of $\hat{W}_1$ and $\hat{W}_2$ are similar to those of $\hat{K}_1$ and $\hat{K}_2$ in all cases.

Table I indicates that when using asymptotic critical values, most tests do not perform well when $n, T < 32$. For small samples, we suggest using wild bootstrap (e.g., Härdle and Mammen (1993), Davidson and Flachaire (2001), Horowitz (2001)) as an alternative to asymptotic approximation. The bottom panel of Table I reports bootstrap rejection probabilities, which are based on 1000 replications and 500 bootstrap resamples. The $\hat{W}_1$ and $\hat{W}_2$ tests have reasonable rejection probabilities in small samples for all cases ($n < T$, $n = T$, and $n > T$). It appears that wild bootstrap can remedy the underrejection of $\hat{W}_1$ and $\hat{W}_2$ using asymptotic critical values, especially when the sample size is as small as 5 or 8. However, $\hat{K}_1$ and $\hat{K}_2$ overreject when $(n, T) = (5, 5)$. BL and BSY perform better using wild bootstrap, but they still tend to underreject.

Table II reports the rejection probabilities under DGP$_2$, a dynamic panel process. Again $\hat{W}_1$, $\hat{W}_2$, $\hat{K}_1$, and $\hat{K}_2$ all underreject using asymptotic critical values, but they all perform well using wild bootstrap, though they still underreject compared to Table I. Unlike under DGP$_1$ (a static panel process), BSY and BL perform poorly using either asymptotic or bootstrap critical values, though BL performs slightly better than BSY when $n < T$. BSY overrejects for almost all cases using bootstrap critical values.

The top panel of Table III first reports the Type I error corrected powers against AR(1) alternatives, with $\tau = .4$, using empirical critical values, which provides a fair comparison. Because empirical critical values are not observable in practice, we also report power using bootstrap critical values. Under AR(1)$^a$, BSY is the most powerful, followed by BL. This is expected because both BSY and BL are optimal against AR(1). The $\hat{W}_1$, $\hat{W}_2$, $\hat{K}_1$, and $\hat{K}_2$ tests have nontrivial but substantially lower power when sample sizes are small. This is because AR(1)$^a$ has a relatively flat spectrum and the advantage of consistent testing is not displayed. Under AR(1)$^b$, BL becomes the most powerful. Somewhat surprisingly, $\hat{W}_1$, $\hat{W}_2$, $\hat{K}_1$, and $\hat{K}_2$ have rather high power and dominate BSY, perhaps due to the fact that a negative AR(1) has a less smooth spectrum than a positive AR(1). For example, with $(n, T) = (10, 16)$, the power of BSY is 6.3% while those of BL and $\hat{W}_1(0)$ are 71.4% and 47.1% respectively. Interestingly, both BSY and BL fail to detect AR(1)$^c$, the mixed model. In contrast, $\hat{W}_1$, $\hat{W}_2$, $\hat{K}_1$, and $\hat{K}_2$ are very powerful against AR(1)$^c$, indicating that wavelet and kernel tests are rather effective in
<table>
<thead>
<tr>
<th>$(n, T)$</th>
<th>(5, 8)</th>
<th>(10, 16)</th>
<th>(25, 32)</th>
<th>(50, 64)</th>
<th>(5, 5)</th>
<th>(10, 10)</th>
<th>(25, 25)</th>
<th>(50, 50)</th>
<th>(8, 5)</th>
<th>(16, 10)</th>
<th>(32, 25)</th>
<th>(64, 50)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level</td>
<td>10%</td>
<td>5%</td>
<td>10%</td>
<td>5%</td>
<td>10%</td>
<td>5%</td>
<td>10%</td>
<td>5%</td>
<td>10%</td>
<td>5%</td>
<td>10%</td>
<td>5%</td>
</tr>
<tr>
<td>$\hat{W}_1(0)$</td>
<td>.5</td>
<td>.2</td>
<td>1.3</td>
<td>1.1</td>
<td>4.6</td>
<td>2.3</td>
<td>6.2</td>
<td>2.9</td>
<td>.0</td>
<td>.0</td>
<td>.5</td>
<td>2.7</td>
</tr>
<tr>
<td>$\hat{W}_1(1)$</td>
<td>.4</td>
<td>.2</td>
<td>1.1</td>
<td>.4</td>
<td>2.5</td>
<td>1.3</td>
<td>4.1</td>
<td>2.4</td>
<td>.0</td>
<td>.0</td>
<td>.4</td>
<td>3.7</td>
</tr>
<tr>
<td>$\hat{W}_1(J_0)$</td>
<td>.5</td>
<td>.2</td>
<td>1.1</td>
<td>.4</td>
<td>2.7</td>
<td>1.1</td>
<td>3.2</td>
<td>1.3</td>
<td>.0</td>
<td>.3</td>
<td>2.3</td>
<td>1.5</td>
</tr>
<tr>
<td>$\hat{W}_2(0)$</td>
<td>.1</td>
<td>.0</td>
<td>1.0</td>
<td>.3</td>
<td>3.4</td>
<td>2.0</td>
<td>5.5</td>
<td>2.4</td>
<td>.0</td>
<td>.0</td>
<td>.1</td>
<td>2.3</td>
</tr>
<tr>
<td>$\hat{W}_2(1)$</td>
<td>.0</td>
<td>.0</td>
<td>.4</td>
<td>.2</td>
<td>2.6</td>
<td>1.3</td>
<td>3.5</td>
<td>2.0</td>
<td>.0</td>
<td>.0</td>
<td>.0</td>
<td>1.7</td>
</tr>
<tr>
<td>$\hat{W}_2(J_0)$</td>
<td>.1</td>
<td>.0</td>
<td>.4</td>
<td>.2</td>
<td>3.2</td>
<td>1.1</td>
<td>5.9</td>
<td>2.7</td>
<td>.0</td>
<td>.0</td>
<td>.0</td>
<td>1.2</td>
</tr>
<tr>
<td>$\hat{K}_1$</td>
<td>.6</td>
<td>.3</td>
<td>1.6</td>
<td>.7</td>
<td>3.7</td>
<td>2.2</td>
<td>7.0</td>
<td>3.8</td>
<td>.0</td>
<td>.0</td>
<td>.3</td>
<td>2.8</td>
</tr>
<tr>
<td>$\hat{K}_2$</td>
<td>.1</td>
<td>.0</td>
<td>.8</td>
<td>.2</td>
<td>3.4</td>
<td>2.1</td>
<td>6.7</td>
<td>2.8</td>
<td>.0</td>
<td>.0</td>
<td>.1</td>
<td>1.8</td>
</tr>
<tr>
<td>BL</td>
<td>16.6</td>
<td>9.8</td>
<td>4.8</td>
<td>1.8</td>
<td>2.1</td>
<td>.3</td>
<td>.3</td>
<td>.0</td>
<td>40.2</td>
<td>31.4</td>
<td>12.8</td>
<td>5.8</td>
</tr>
<tr>
<td>BSY</td>
<td>12.9</td>
<td>6.2</td>
<td>29.8</td>
<td>9.9</td>
<td>19.0</td>
<td>12.0</td>
<td>5.5</td>
<td>7.5</td>
<td>4.5</td>
<td>4.5</td>
<td>3.0</td>
<td>7.0</td>
</tr>
<tr>
<td>(Asymptotic Critical Values)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Bootstrapped Critical Values)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: (a) $DGP$: $y_{it} = 5 + .5y_{it-1} + \mu_i + \epsilon_{it}, \mu_i \sim \text{i.i.d. } N(0, .4)$, and $\epsilon_{it} \sim \text{i.i.d. } N(0, 1)$. (b) $\hat{W}_1$: heteroskedasticity-consistent Franklin wavelet-based test; $\hat{W}_2$: heteroskedasticity-corrected Franklin wavelet-based test; $\hat{J}_0$: data-driven finest scale. (c) $\hat{K}_1$: heteroskedasticity-consistent Daniell kernel-based test; $\hat{K}_2$: heteroskedasticity-corrected Daniell kernel-based test. (d) BL, Baltagi–Li test; BSY, Bera, Sosa-Escudero, and Yoon test. (e) 1000 simulation replications; 500 bootstrap replications.
### Table III

**Percentage of Rejections Under the AR(1) Alternative in the Static Panel Model**

<table>
<thead>
<tr>
<th>(n, T)</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(5, 8)</td>
<td>(10, 16)</td>
<td>(25, 32)</td>
</tr>
<tr>
<td></td>
<td>(5, 8)</td>
<td>(10, 16)</td>
<td>(25, 32)</td>
</tr>
<tr>
<td></td>
<td>(5, 8)</td>
<td>(10, 16)</td>
<td>(25, 32)</td>
</tr>
<tr>
<td>Level</td>
<td>10%</td>
<td>5%</td>
<td>10%</td>
</tr>
<tr>
<td></td>
<td>10%</td>
<td>5%</td>
<td>10%</td>
</tr>
<tr>
<td></td>
<td>10%</td>
<td>5%</td>
<td>10%</td>
</tr>
<tr>
<td></td>
<td>10%</td>
<td>5%</td>
<td>10%</td>
</tr>
<tr>
<td></td>
<td>10%</td>
<td>5%</td>
<td>10%</td>
</tr>
<tr>
<td></td>
<td>10%</td>
<td>5%</td>
<td>10%</td>
</tr>
<tr>
<td></td>
<td>10%</td>
<td>5%</td>
<td>10%</td>
</tr>
<tr>
<td></td>
<td>10%</td>
<td>5%</td>
<td>10%</td>
</tr>
<tr>
<td></td>
<td>10%</td>
<td>5%</td>
<td>10%</td>
</tr>
</tbody>
</table>

**Empirical Critical Values**

- $\hat{W}_1(0)$: 3.6, 1.4, 14.4, 6.1, 74.8, 62.3, 99.9, 99.9
- $\hat{W}_1(2)$: 4.3, 1.4, 9.9, 5.8, 55.4, 42.4, 99.9, 99.9
- $\hat{W}_1(J_0)$: 4.2, 1.4, 11.5, 6.9, 56.4, 41.7, 99.9, 99.9
- $\hat{W}_3(0)$: 3.6, .9, 9.3, 3.5, 68.3, 53.7, 99.9, 99.9
- $\hat{W}_3(1)$: 5.6, 1.4, 7.2, 4.0, 46.2, 33.3, 99.9, 99.9
- $\hat{W}_3(J_0)$: 5.6, 1.4, 10.2, 4.9, 47.7, 37.3, 99.9, 99.9
- $\hat{K}_2$: 4.7, 1.6, 15.7, 9.6, 72.0, 60.0, 99.9, 99.9
- $\hat{K}_2$: 3.8, 1.5, 11.6, 5.8, 65.3, 51.1, 99.9, 99.9
- BL: 3.5, 1.1, 19.9, 11.9, 97.8, 96.0, 99.9, 99.9
- BSY: 30.7, 20.9, 74.8, 59.9, 99.9, 99.9, 99.9

**Bootstrapped Critical Values**

- $\hat{W}_1(0)$: 2.5, 2.0, 9.0, 3.0, 70.0, 56.0, 99.5, 99.5
- $\hat{W}_1(1)$: 2.5, 1.5, 7.5, 4.0, 49.5, 37.0, 99.5, 99.5
- $\hat{W}_1(J_0)$: 2.5, 2.0, 7.5, 4.0, 53.5, 41.0, 99.5, 99.5
- $\hat{W}_2(0)$: 4.0, .5, 6.5, 4.0, 46.4, 51.5, 99.5, 99.5
- $\hat{W}_2(1)$: 3.5, 1.6, 6.0, 3.0, 44.5, 35.0, 99.5, 99.5
- $\hat{W}_2(J_0)$: 4.0, .5, 6.0, 3.0, 48.0, 34.0, 99.5, 99.5
- $\hat{K}_2$: 3.0, 2.0, 9.0, 4.0, 70.0, 55.5, 99.5, 99.5
- $\hat{K}_2$: 5.0, 1.0, 9.0, 4.5, 70.0, 53.0, 99.5, 99.5
- BL: 1.5, .0, 11.0, 4.0, 95.0, 90.5, 99.5, 99.5

**Notes:**

- (a) DGP: $Y_{it} = 5 + 5X_{it} + \mu_i + \mu_t + \eta_{it}, X_{it} = 5X_{it-1} + \eta_{it}, \eta_{it} \sim \text{i.i.d.} \text{Unif}[{-1.5, 1.5}], \mu_i \sim \text{i.i.d.} \text{Unif}(0, 4)$, and Model 1: $\epsilon_{it} = 2\epsilon_{it-1} + \epsilon_{it}, i = 1, \ldots, n$; Model 2: $\epsilon_{it} = -2\epsilon_{it-1} + \epsilon_{it}, i = 1, \ldots, n$; and Model 3: $\epsilon_{it} = 2\epsilon_{it-1} + \epsilon_{it}, i = 1, \ldots, n/2$ and $\epsilon_{it} = -2\epsilon_{it-1} + \epsilon_{it}, i = n/2 + 1, \ldots, n$, where $\epsilon_{it} \sim \text{i.i.d.} \text{Unif}(0, 1)$. (b) $\hat{W}_1$, heteroskedasticity-consistent Frankfurt wavelet-based test; $\hat{W}_2$, heteroskedasticity-corrected Frankfurt wavelet-based test; $\hat{J}_0$, data-driven finest scale. (c) $\hat{K}_2$, heteroskedasticity-consistent Daniell kernel-based test; $\hat{K}_2$, heteroskedasticity-corrected Daniell kernel-based test. (d) BL, Baltagi-Li test; BSY, Bera, Sosa-Escudero, and Yoon test. (e) 200 simulation replications; 200 bootstrap replications.
capturing inhomogeneous serial correlations across individuals. The bottom panel of Table III reports bootstrap power. Due to the extensive computation involved, we use 200 bootstrap resamples and 200 replications for bootstrap power. Again, BSY is the best for AR(1) and BL, \( \hat{W}_1, \hat{W}_2, \hat{K}_1, \) and \( \hat{K}_2 \) have low power when \( T < 32 \). We also observe that \( \hat{W}_1 \) and \( \hat{W}_2 \) are much more powerful than \( \hat{K}_1 \) and \( \hat{K}_2 \) under AR(1) for \((n, T) = (10, 16)\), though the advantages diminish as the sample sizes increase. For all AR(1) alternatives, the choice of \( J \) has significant impact on \( \hat{W}_i \) and \( \hat{W}_2 \), and \( J = 0 \) gives the best power for \( \hat{W}_i \) and \( \hat{W}_2 \) against various AR(1) alternatives. The data-driven \( \hat{J}_0 \) delivers reasonable and robust power in all cases.

The top panel of Table IV reports the Type I error corrected powers against ARMA(12, 4) using empirical critical values. All tests have no power when the sample size is \((10, 16)\) since the effective sample size in fact is \((10, 4)\). This is because we remove the first 12 time series observations for each \( i \). Hence the effective samples for the results reported are \((10, 4), (25, 20), (50, 52)\). \( \hat{W}_1(\hat{J}_0) \) and \( \hat{W}_2(\hat{J}_0) \) have the best power and dominate \( \hat{K}_1, \hat{K}_2, \) BL, and BSY against all three ARMA(12, 4). This is apparently due to the fact that ARMA(12, 4) has four sharp spectral peaks, which can be more effectively captured by wavelets than kernels. This confirms our prediction that wavelet-based tests are powerful in capturing spectral modes/peaks in small and finite samples. The bottom panel of Table IV reports bootstrap powers. Both \( \hat{W}_1(\hat{J}_0) \) and \( \hat{W}_2(\hat{J}_0) \) perform the best and clearly dominate \( \hat{K}_1 \) and \( \hat{K}_2 \) for most cases. We note that the clear dominance of \( \hat{W}_1 \) and \( \hat{W}_2 \) over \( \hat{K}_1 \) and \( \hat{K}_2 \) is reduced if bootstrap rather than empirical critical values are used. The choice of \( J \) has significant impact on the power either using empirical or bootstrap critical values of \( \hat{W}_1 \) and \( \hat{W}_2 \). Data-driven \( \hat{J}_0 \) gives better power than \( J = 0, 1 \). Apparently due to the seasonal patterns of ARMA(12, 4), the choice of \( J = 0, 1 \) yields little or no power for \( \hat{W}_1 \) and \( \hat{W}_2 \) against ARMA(12, 4) and ARMA(12, 4). In contrast, \( \hat{J}_0 \) is able to adapt to the unknown serial correlation pattern and gives robust and high power. This highlights the value of our data-driven finest scale \( \hat{J}_0 \).

We also conduct a simulation study on the power under a dynamic panel model (DGP2), which we do not report in the paper for space. The relative ranking between our tests and other tests under a dynamic model remains similar to that under a static model, and in fact the dominance of our wavelet-based tests over the kernel-based tests against ARMA(12, 4) error alternatives becomes more striking. For example, the Type I error corrected powers of \( \hat{W}_1, \hat{W}_2, \hat{K}_1, \) and \( \hat{K}_2 \) are 67.3\%, 82.8\%, 45.7\%, and 58.3\% respectively when \((n, T) = (10, 16)\). On the other hand, the relative ranking between BL and BSY is reversed: unlike under a static model, BSY now has poor power while BL is most powerful against a positive or negative (but not mixed) AR(1) error alternative.
<table>
<thead>
<tr>
<th>(n, T)</th>
<th>Model 1</th>
<th></th>
<th>Model 2</th>
<th></th>
<th>Model 3</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(10, 16)</td>
<td>(25, 32)</td>
<td>(50, 64)</td>
<td>(10, 16)</td>
<td>(25, 32)</td>
<td>(50, 64)</td>
</tr>
<tr>
<td>Level</td>
<td>10%</td>
<td>5%</td>
<td>10%</td>
<td>5%</td>
<td>10%</td>
<td>5%</td>
<td>10%</td>
</tr>
<tr>
<td><strong>Empirical Critical Values</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{W}_1(0)$</td>
<td>0.0</td>
<td>0.0</td>
<td>18.1</td>
<td>11.9</td>
<td>41.8</td>
<td>33.4</td>
<td>0.0</td>
</tr>
<tr>
<td>$\hat{W}_1(1)$</td>
<td>0.0</td>
<td>0.0</td>
<td>46.3</td>
<td>46.6</td>
<td>91.5</td>
<td>84.1</td>
<td>0.0</td>
</tr>
<tr>
<td>$\hat{W}_1(J_0)$</td>
<td>1.0</td>
<td>0.0</td>
<td>43.2</td>
<td>29.9</td>
<td>96.8</td>
<td>95.2</td>
<td>0.1</td>
</tr>
<tr>
<td>$\hat{W}_2(0)$</td>
<td>0.0</td>
<td>0.0</td>
<td>16.1</td>
<td>9.5</td>
<td>44.2</td>
<td>31.4</td>
<td>0.0</td>
</tr>
<tr>
<td>$\hat{W}_2(1)$</td>
<td>0.0</td>
<td>0.0</td>
<td>42.0</td>
<td>30.1</td>
<td>91.3</td>
<td>84.1</td>
<td>0.0</td>
</tr>
<tr>
<td>$\hat{W}_2(J_0)$</td>
<td>0.0</td>
<td>0.0</td>
<td>34.8</td>
<td>26.2</td>
<td>96.8</td>
<td>95.6</td>
<td>0.0</td>
</tr>
<tr>
<td>$\hat{K}_1$</td>
<td>0.0</td>
<td>0.0</td>
<td>35.0</td>
<td>24.1</td>
<td>93.0</td>
<td>90.3</td>
<td>0.0</td>
</tr>
<tr>
<td>$\hat{K}_2$</td>
<td>0.0</td>
<td>0.0</td>
<td>27.7</td>
<td>19.1</td>
<td>94.1</td>
<td>91.5</td>
<td>0.0</td>
</tr>
<tr>
<td>BL</td>
<td>2.7</td>
<td>0.0</td>
<td>29.1</td>
<td>17.7</td>
<td>94.1</td>
<td>12.3</td>
<td>2.0</td>
</tr>
<tr>
<td>BSY</td>
<td>0.0</td>
<td>0.0</td>
<td>9.5</td>
<td>4.7</td>
<td>6.6</td>
<td>2.9</td>
<td>0.0</td>
</tr>
<tr>
<td><strong>Bootstrapped Critical Values</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{W}_1(0)$</td>
<td>3.0</td>
<td>0.5</td>
<td>27.0</td>
<td>16.5</td>
<td>39.5</td>
<td>30.0</td>
<td>3.0</td>
</tr>
<tr>
<td>$\hat{W}_1(1)$</td>
<td>2.0</td>
<td>0.0</td>
<td>57.0</td>
<td>42.5</td>
<td>92.5</td>
<td>85.5</td>
<td>2.0</td>
</tr>
<tr>
<td>$\hat{W}_1(J_0)$</td>
<td>3.0</td>
<td>0.5</td>
<td>55.0</td>
<td>42.5</td>
<td>96.5</td>
<td>95.0</td>
<td>3.0</td>
</tr>
<tr>
<td>$\hat{W}_2(0)$</td>
<td>5.5</td>
<td>3.0</td>
<td>27.5</td>
<td>18.0</td>
<td>42.0</td>
<td>30.0</td>
<td>9.0</td>
</tr>
<tr>
<td>$\hat{W}_2(1)$</td>
<td>3.5</td>
<td>1.0</td>
<td>61.0</td>
<td>48.0</td>
<td>93.5</td>
<td>87.0</td>
<td>6.5</td>
</tr>
<tr>
<td>$\hat{W}_2(J_0)$</td>
<td>5.5</td>
<td>3.0</td>
<td>60.5</td>
<td>45.5</td>
<td>98.0</td>
<td>96.0</td>
<td>9.0</td>
</tr>
<tr>
<td>$\hat{K}_1$</td>
<td>1.5</td>
<td>0.0</td>
<td>47.0</td>
<td>33.5</td>
<td>92.0</td>
<td>90.0</td>
<td>1.5</td>
</tr>
<tr>
<td>$\hat{K}_2$</td>
<td>3.0</td>
<td>1.0</td>
<td>50.5</td>
<td>37.5</td>
<td>92.5</td>
<td>89.5</td>
<td>5.5</td>
</tr>
<tr>
<td>BL</td>
<td>1.0</td>
<td>0.5</td>
<td>17.0</td>
<td>8.5</td>
<td>16.5</td>
<td>10.5</td>
<td>1.5</td>
</tr>
<tr>
<td>BSY</td>
<td>0.5</td>
<td>0.5</td>
<td>0.0</td>
<td>0.0</td>
<td>1.0</td>
<td>0.5</td>
<td>1.0</td>
</tr>
</tbody>
</table>

Notes: (a) DGP: $Y_{it} = 5 + 5X_{it} + \mu_i + \epsilon_{it}, X_{it} = 5X_{it-1} + \eta_{it}, \eta_{it} \sim \text{i.i.d. } U[-5, 5], \mu_i \sim \text{i.i.d. } N(0, 4)$, and Model 1: $\epsilon_{it} = -3\epsilon_{it-2} + \epsilon_{it-1} + \eta_{it}, i = 1, \ldots, n$; Model 2: $\epsilon_{it} = -3\epsilon_{it-2} + \epsilon_{it-1} + \eta_{it}, i = 1, \ldots, n$; and Model 3: $\epsilon_{it} = -3\epsilon_{it-2} + \epsilon_{it-1} + \eta_{it}, i = 1, \ldots, n/2$ and $\epsilon_{it} = -3\epsilon_{it-2} + \epsilon_{it-1} + \eta_{it}, i = n/2 + 1, \ldots, n$, where $\eta_{it} \sim \text{i.i.d. } N(0, 1)$. (b) $\hat{W}_1$, heteroskedasticity-consistent Franklin wavelet-based test; $\hat{W}_2$, heteroskedasticity-corrected Franklin wavelet-based test; $J_0$, data-driven finest scale. (c) $\hat{K}_1$, heteroskedasticity-consistent Daniell kernel-based test; $\hat{K}_2$, heteroskedasticity-corrected Daniell kernel-based test. (d) BL, Baltagi–Li test; BSY, Bera, Sosa-escudero, and Yoon test. (e) 200 simulation replications; 200 bootstrap replications.
8. CONCLUSION

We have proposed a class of generally applicable wavelet-based consistent tests for serial correlation in static and dynamic panel models. Wavelets are powerful for detecting serial correlation where the spectrum has peaks or kinks. The new tests have a convenient limit $N_{\text{ori}}$ distribution, which is not affected by parameter estimation uncertainty, even if regressors contain lagged dependent variables or deterministic/stochastic trending variables. Our tests do not require an alternative model, and are consistent against serial correlation of unknown form even in the presence of substantial inhomogeneity in serial correlation across individuals. They are applicable to unbalanced heterogeneous panel models with one way or two way error components. A data-driven method is developed to select the smoothing parameter—the finest scale in wavelet estimation, making our tests entirely operational in practice. Simulation shows that our tests perform well in finite samples.

Dept. of Economics and Dept. of Statistical Science, Cornell University, 424 Uris Hall, Ithaca, NY 14850, U.S.A., and Department of Economics, Tsinghua University, Beijing 100084, China; yh20@cornell.edu

and

Dept. of Economics and Center for Policy Research, Syracuse University, 426 Eggers Hall, Syracuse, NY, U.S.A.; cdkao@maxwell.syr.edu.

Manuscript received October, 2000; final revision received February, 2004.

APPENDIX A: MATHEMATICAL APPENDIX

To prove Theorems 1–6 and Corollary 1, we will use the following lemma.

LEMMA A.1: Suppose Assumptions 1 and 2 hold, and let $b_{J_i}(h, m)$ be as in (3.15). Then for any $J_i, T_i \in \mathbb{Z}^+$ and a bounded constant $C$ that does not depend on $i, J_i$ and $T_i$,

(i) $b_{J_i}(h, m)$ is real-valued, $b_{J_i}(0, m) = b_{J_i}(h, 0) = 0$, and $b_{J_i}(h, m) = b_{J_i}(m, h)$;

(ii) $\sum_{h_i=1}^{T_i-1} \sum_{m=1}^{T_i-1} h_i |b_{J_i}(h, m)| \leq C 2^{(1+\nu)(J_i+1)}$ for $0 = \nu \leq \frac{1}{2}$;

(iii) $\sum_{h_i=1}^{T_i-1} [\sum_{m=1}^{T_i-1} |b_{J_i}(h, m)|^2] \leq C 2^{J_i+1}$;

(iv) $\sum_{h_i=1}^{T_i-1} [\sum_{m=1}^{T_i-1} b_{J_i}(h_1, m) b_{J_i}(h_2, m)]^2 \leq C (J_i+1) 2^{J_i+1}$;

(v) $|\sum_{h_i=1}^{T_i-1} b_{J_i}(h, h) - (2^{J_i+1} - 1)| \leq C((J_i+1) + 2^{J_i+1} (2^{J_i+1}/T_i) (2^{\tau-1})$, where $\tau$ is in Assumption 2;

(vi) $|\sum_{h_i=1}^{T_i-1} \sum_{m=1}^{T_i-1} b_{J_i}(h, m) - 2(2^{J_i+1} - 1)| \leq C((J_i+1)^2 + 2^{J_i+1} (2^{J_i+1}/T_i) (2^{\tau-1})$;

(vii) $\sup_{1 \leq h \leq T_i, 1 \leq m \leq T_i} |b_{J_i}(h, m)| \leq C(J_i + 1)$;

(viii) $\sup_{1 \leq h \leq T_i, 1 \leq m \leq T_i} |b_{J_i}(h, m)| \leq C(J_i + 1)$.

PROOF OF LEMMA A.1: This lemma extends Lee and Hong (2001, Lemma A.1), who consider the case where $J_i \equiv J \to \infty$ as $T_i \equiv T \to \infty$. See Hong and Kao (2002) for a detailed proof. Q.E.D.

PROOF OF THEOREM 1: We shall show Theorems A.1–A.3.
THEOREM A.1: Let \( \hat{\alpha}_{ijk} \) and \( \bar{\alpha}_{ijk} \) be defined as in (3.12) and (6.3), and \( V_{nT} \equiv \sum_{i=1}^{n} \sigma_i^2 V_{t0} \), where \( V_{t0} \) is as in (3.15). Then \( V_{nT}^{-1/2} \sum_{i=1}^{n} \sum_{j=0}^{J_i} \sum_{k=1}^{2j} (\hat{\alpha}_{ijk}^2 - \bar{\alpha}_{ijk}^2) \to P 0. \)

THEOREM A.2: Put \( M_{nT} \equiv \sum_{i=1}^{n} \sigma_i^2 M_{t0} \), where \( M_{t0} \) is as in (3.14). Then \( V_{nT}^{-1/2} (\sum_{i=1}^{n} 2\pi T_i \times \sum_{j=0}^{J_i} \sum_{k=1}^{2j} (\hat{\alpha}_{ijk}^2 - \bar{\alpha}_{ijk}^2) - M_{nT}) \to N(0, 1). \)

THEOREM A.3: Let \( \hat{M} \) and \( \hat{V} \) be defined as in (3.15). Then \( V_{nT}^{-1/2} (\hat{M} - M_{nT}) \to P 0 \) and \( \hat{V} / V_{nT} \to P 1. \)

PROOF OF THEOREM A.1: Because \( \hat{\alpha}_{ijk}^2 - \bar{\alpha}_{ijk}^2 = (\hat{\alpha}_{ijk} - \bar{\alpha}_{ijk})^2 + 2(\hat{\alpha}_{ijk} - \bar{\alpha}_{ijk})\bar{\alpha}_{ijk} \), we shall show Propositions A.1 and A.2 below under the conditions of Theorem 1.

Q.E.D.

PROPOSITION A.1: \( V_{nT}^{-1/2} \sum_{i=1}^{n} 2\pi T_i \sum_{j=0}^{J_i} \sum_{k=1}^{2j} (\hat{\alpha}_{ijk} - \bar{\alpha}_{ijk})^2 = O_P[V_{nT}^{-1/2} + (n^{-1} + T^{-1})V_{nT}^{-1/2}]. \)

PROPOSITION A.2: \( V_{nT}^{-1/2} \sum_{i=1}^{n} 2\pi T_i \sum_{j=0}^{J_i} \sum_{k=1}^{2j} (\hat{\alpha}_{ijk} - \bar{\alpha}_{ijk})\bar{\alpha}_{ijk} \to P 0. \)

PROOF OF PROPOSITION A.1: By the definition of \( \hat{u}_{it} \) in (2.3), we have \( \hat{u}_{it} = \tilde{u}_{it} - \tilde{X}_{it}(\hat{\beta} - \beta) \), where \( \tilde{X}_{it} \) and \( \tilde{u}_{it} \) are as in Assumption 5. We note that under \( \|\xi_{1i}\|, \|v_{1i}\| \) in (2.4) coincides with the true errors \( \{\xi_{1i}\} \) in (2.1), and so is i.i.d. for each \( i \), and \( \{v_{1i}\} \) and \( \{v_{hi}\} \) are independent for all \( i \neq l \) and all \( t \).

Putting \( \hat{R}_t(h) \equiv T^{-1} \sum_{|t|=|h|+1} \bar{v}_{it} \tilde{u}_{i(t-|h|)} \), we write

\begin{align}
(\text{A.1}) \quad \hat{R}_t(h) - \bar{R}_t(h) &= (\hat{\beta} - \beta) \hat{T}_{ixx}(h)(\hat{\beta} - \beta) - (\hat{\beta} - \beta) \bar{T}_{ixx}(h) - (\hat{\beta} - \beta) \bar{T}_{ixx}(h) \\
&= \sum_{c=1}^{3} \hat{\xi}_{c(r)}(h),
\end{align}

where \( \hat{T}_{ixx}(h) \equiv T_{t}^{-1} \sum_{|t|=|h|+1} \tilde{X}_{it} \tilde{X}_{i(t-|h|)} \), and \( \bar{T}_{ixx}(h) \) and \( \bar{T}_{ixx}(h) \) are as in Assumption 5.

Next, recalling the definition of \( \hat{R}_t(h) \) as in (6.3), we can write

\begin{align}
(\text{A.2}) \quad \hat{R}_t(h) - \bar{R}_t(h) &= T_{t}^{-1} \sum_{|t|=|h|+1} \sum_{i=1}^{p} \left( \tilde{v}_{it} \tilde{u}_{i(t-|h|)} - \tilde{v}_{it} \tilde{u}_{i(t-|h|)} - \tilde{v}_{it} \tilde{u}_{i(t-|h|)} + \tilde{v}_{it} \tilde{u}_{i(t-|h|)} \right) \\
&= \sum_{c=4}^{9} \hat{\xi}_{c(r)}(h).
\end{align}

Given (A1) and (A2), we have \( \hat{\alpha}_{ijk} - \bar{\alpha}_{ijk} = (2\pi)^{-1/2} \sum_{c=1}^{9} T_{t}^{-1} \sum_{h=1}^{T_{t}} \hat{\xi}_{c(r)}(h) \tilde{Y}_{jk}(h) \). It follows that

\begin{align}
(\text{A.3}) \quad \sum_{i=1}^{n} 2\pi T_i \sum_{j=0}^{J_i} \sum_{k=1}^{2j} (\hat{\alpha}_{ijk} - \bar{\alpha}_{ijk})^2 &\leq 2^{5} \sum_{c=1}^{9} \sum_{i=1}^{p} T_{t}^{-1} \sum_{h=1}^{T_{t}} \hat{\xi}_{c(r)}(h) \tilde{Y}_{jk}(h) \\
&= 2^{5} \sum_{c=1}^{9} \hat{A}_{c}.
\end{align}

We shall show that \( V_{nT}^{-1/2} A_{c} \to P 0 \) for \( 1 < c \leq 9 \). We first consider \( A_1 \). From (A1), we have \( \|\hat{\xi}_{1}(h)\| \leq \|\hat{\beta} - \beta\| \|\hat{T}_{ixx}(h)\| \leq \|\hat{\beta} - \beta\| \|\hat{T}_{ixx}(0)\| \). Let \( b_{ij}(h, m) \) be defined as in Lemma A.1.
Then we have
\[ V_{nT}^{-1/2} \hat{A}_1 = V_{nT}^{-1/2} \left| \sum_{i=1}^{n} \sum_{h=1}^{T_i} \sum_{m=1}^{k(h,m)} b_{ij}(h,m) \hat{\xi}_i(h) \hat{\xi}_i(m) \right| \]
\[ \leq V_{nT}^{-1/2} \left\| \hat{\beta} - \beta \right\|^2 \left( \sum_{i=1}^{n} \sum_{h=1}^{T_i} \sum_{m=1}^{k(h,m)} b_{ij}^2(h,m) \right)^{1/2} \left( \sum_{i=1}^{n} T_i^3 ||\hat{\Gamma}_{ix}(0)||^2 \right)^{1/2} \]
\[ = O_p\left(n^{-3/2}\right), \]
given Lemma A.1(vii), \( V_{nT} \leq C \sum_{n=1}^{T} 2^{f_{i+1}} \), Assumptions 3–5, \( T_i \leq CT \), and \( \sigma^2 \in [c, C] \).

Next, we consider the second term \( \hat{A}_2 \) in (A.3). Recalling \( \Gamma_{ix}(h) \equiv p \lim_{T_i \to \infty} \hat{\Gamma}_{ix}(h) \), we have
\[ \hat{\xi}_i(h) = (\hat{\beta} - \beta)'\hat{\Gamma}_{ix}(h) + (\hat{\beta} - \beta)'[\hat{\Gamma}_{ix}(h) - \Gamma_{ix}(h)]. \]
It follows that
\[ \hat{A}_2 \leq 2 \left\| \hat{\beta} - \beta \right\|^2 \sum_{i=1}^{n} T_i \sum_{j=0}^{2^j} \left| \sum_{h=1}^{T_i} \hat{\Gamma}_{ix}(h) \hat{\Psi}_{jk}(h) \right|^2 \]
\[ + \left( \sum_{h=1}^{T_i} \left| \hat{\Gamma}_{ix}(h) - \Gamma_{ix}(h) \right| \hat{\Psi}_{jk}(h) \right)^2 \]
\[ = 2 \left\| \hat{\beta} - \beta \right\|^2 \hat{M}_1 + 2 \left\| \hat{\beta} - \beta \right\|^2 \hat{M}_2, \]
say.

We now consider \( \hat{M}_1 \) in (A.4). Let \( \Lambda_{ij} \equiv f_{\pi} f_{\tau}(\omega) \Psi_{jk}(\omega) d\omega \), where \( f_{\tau}(\omega) \equiv (2\pi)^{-1} \sum_{h=-\infty}^{\infty} \Gamma_{ix}(h)e^{-i\omega h} \). Then \( \Lambda_{ij} \equiv (2\pi)^{-1/2} \sum_{h=-\infty}^{\infty} \Gamma_{ix}(h) \hat{\Psi}_{jk}(h) \) by Parseval’s identity, and
\[ \sum_{h=1}^{T_i} \left( \hat{\Gamma}_{ix}(h) \hat{\Psi}_{jk}(h) - \Gamma_{ix}(h) \hat{\Psi}_{jk}(h) \right)^2 \leq \sum_{j=0}^{2^j} \sum_{h=1}^{T_i} \left| \hat{\Psi}_{jk}(h) \right|^2 = C \left( 2^j + 2 \right) \]
given Assumption 5. We consider \( \hat{M}_2 \) in (A.4), and have
\[ \left\| \hat{\beta} - \beta \right\|^2 \hat{M}_2 \leq \sum_{i=1}^{n} T_i \sum_{h=1}^{T_i} \sum_{m=1}^{k(h,m)} |b_{ij}(h,m)| \left| \hat{\Gamma}_{ix}(h) - \Gamma_{ix}(h) \right| \left| \hat{\Gamma}_{ix}(m) - \Gamma_{ix}(m) \right| \]
\[ = O_p \left( (nT)^{-1} \right) \]
given Lemma A.1(ii), \( V_{nT} \leq C \sum_{i=1}^{T} 2^{f_{i+1}} \), and Assumption 5. Combining (A.4)–(A.6) yields
\[ V_{nT}^{-1/2} \hat{A}_2 = O_p(V_{nT}^{-1/2} + (nT)^{-1/2} \cdot \hat{M}_2). \]
Similarly, we have
\[ V_{nT}^{-1/2} \hat{A}_3 = O_p(V_{nT}^{-1/2} + (nT)^{-1/2} \cdot \hat{M}_2). \]
Now we consider the term $\hat{A}_4$ in (A.3). By the Cauchy–Schwarz inequality and the fact that under $\mathbb{H}_0$, $\{v_{ri}\}$ coincides with $\{e_{ri}\}$ and so is i.i.d. with $E\nu_{it}^8 \leq C$ for each $i$, we have

$$E(\tilde{v}_t^i|T^{-1}_i \sum_{r=i+1}^{T_i} v_{ri}||T^{-1}_i \sum_{r=i+1}^{T_i} v_{ri}) \leq C T_i^{-2}$$

for $h, m > 0$. It follows from Markov’s inequality, Lemma A.1(ii), and $V_{nt} \leq C \sum_{i=1}^{2T_i+1}$ that

$$V_{nt}^{-1/2} \hat{A}_4 \leq V_{nt}^{-1/2} \sum_{i=1}^{n} T_i \sum_{h=1}^{T_i-1} \sum_{m=1}^{T_i-1} (b_{ij}(h, m) |\tilde{v}_t^i| T_i^{-1} \sum_{r=i+1}^{T_i} v_{ri} |T_i^{-1} \sum_{r=m+1}^{T_i} v_{ri}|)$$

$$= O_p(T^{-1}V_{nt}^{1/2}).$$

Similarly, we have $V_{nt}^{-1/2} \hat{A}_5 = O_p(T^{-1}V_{nt}^{1/2}).$

Next, for the term $\hat{A}_6$ in (A.3), noting that $v_{ri}$ and $\tilde{v}_{r-h}$ are independent for $h > 0$ under $\mathbb{H}_0$, we have

$$E(|\tilde{v}_{r-h} \tilde{v}_{r-m}|T^{-1}_i \sum_{r=i+1}^{T_i} v_{ri}||T^{-1}_i \sum_{r=m+1}^{T_i} v_{ri}) \leq C(nT_i^{-1})^{-1}$$

for $h, m > 0$ by the Cauchy–Schwarz inequality and $E\nu_{it}^8 \leq C$. It follows that $V_{nt}^{-1/2} \hat{A}_6 = O_p(n^{-1}V_{nt}^{1/2})$. Similarly, we have $V_{nt}^{-1/2} \hat{A}_7 = O_p(n^{-1}V_{nt}^{1/2}).$

Finally, given $E(\tilde{v}_t^i|T^{-1}_i \sum_{r=i+1}^{T_i} v_{ri}||T^{-1}_i \sum_{r=m+1}^{T_i} v_{ri}) \leq C n^{-1}T_i^{-2}$ for $h, m > 0$ under $\mathbb{H}_0$, we have $V_{nt}^{-1/2} \hat{A}_8 = O_p((nT)^{-1}V_{nt}^{1/2})$ for $c = 8, 9$. We have shown $V_{nt}^{-1/2} \hat{A}_9 \rightarrow 0$ for all $1 \leq c < 9$ given $\max_{1 \leq c \leq 9} 2^{2(j+c)} / (n^2 + T) \rightarrow 0$. Proposition A.1 then follows from (A.3).

**Q.E.D.**

**PROOF OF PROPOSITION A.2:** Recalling $\hat{\alpha}_{ijk} - \tilde{\alpha}_{ijk} = (2\pi)^{-1/2} \sum_{c=1}^{9} \sum_{h=1-T_i}^{T_i-1} \hat{\Psi}^c(h) \hat{\Psi}^c(h)$, we can write

$$\sum_{c=1}^{9} \sum_{h=1-T_i}^{T_i-1} \hat{\Psi}^c(h) \hat{\Psi}^c(h) = \sum_{c=1}^{9} \sum_{h=1-T_i}^{T_i-1} \hat{\Psi}^c(h) \hat{\Psi}^c(h)$$

$$\Rightarrow \sum_{c=1}^{9} \sum_{h=1-T_i}^{T_i-1} \hat{\Psi}^c(h) \hat{\Psi}^c(h)$$

We shall show $V_{nt}^{-1/2} \hat{\delta}_c \rightarrow 0$ for $1 \leq c \leq 9$. First, we have

$$V_{nt}^{-1/2} \hat{\delta}_c \rightarrow 0$$

$$\Rightarrow V_{nt}^{-1/2} \hat{\delta}_c \rightarrow 0$$

where $V_{nt}^{-1/2} \hat{\delta}_c = O_p(1)$ by Lemma A.1(v) and $E\tilde{\alpha}_{ijk}^2 \leq CT_i^{-1} \sum_{h=1-T_i}^{T_i-1} \hat{\Psi}^c(h) \hat{\Psi}^c(h)^2$.

Next, we consider the second term $\hat{\delta}_2$ in (A.8). We write

$$\hat{\delta}_2 = (\hat{\beta} - \beta) \sum_{i=1}^{n} T_i \sum_{j=0}^{T_j} \sum_{k=1}^{T_k-1} \sum_{h=1-T_i}^{T_i-1} \Gamma_{ixx}(h) \hat{\Psi}(h)$$

$$+ \sum_{h=1-T_i}^{T_i-1} \Gamma_{ixx}(h) \hat{\Psi}(h)$$

$$\Rightarrow (\hat{\beta} - \beta) \hat{\delta}_2 = (\hat{\beta} - \beta) \hat{\delta}_2 = (\hat{\beta} - \beta) \hat{\delta}_2$$

say.
For the first term $\hat{M}_3$, noting that $[\hat{\alpha}_{ijk}]$ is a zero-mean sequence independent across $i$, we obtain

$$E\hat{M}_3^2 = \sum_{i=1}^{n} T_i^2 E \left( \sum_{h=1}^{T_i-1} \sum_{m=1}^{T_i-1} b_{ij}(h, m) \hat{\Gamma}_{xx}(h) \hat{R}_i(m) \right)^2$$

$$\leq \sum_{i=1}^{n} \sigma_i^4 T_i \left( \sup_{1 \leq h \leq T_i-1} \left( \sum_{m=1}^{T_i-1} b_{ij}^2(h, m) \right) \right)^2$$

$$= O \left( T \sum_{i=1}^{n} (J_i + 1)^2 \right) = O(TV_{nT})$$

given Assumption 5, Lemma A.1(viii), and $V_{nT} \leq C \sum_{i=1}^{n} 2^{J_i+1}$. It follows that $V_{nT}^{-1/2} (\hat{\beta} - \beta) \hat{M}_3 = O_p(n^{-1/2})$ by Chebyshev’s inequality. For the second term $\hat{M}_4$ in (A.10), we have $V_{nT}^{-1/2} |(\hat{\beta} - \beta) \hat{M}_4| \leq V_{nT}^{-1/2} |\hat{\beta} - \beta| \hat{M}_4^2 = O_p((nT)^{-1}V_{nT}^{1/2})$, where $\hat{M}_4 = O_p((nT)^{-1}V_{nT}^{1/2})$ as shown in (A.6). It follows from (A.10) that $V_{nT}^{-1/2} \hat{\delta}_2 = O_p(n^{-1/2} + (nT)^{-1}V_{nT}^{1/2})$. Similarly, we have $V_{nT}^{-1/2} \hat{\delta}_3 = O_p(n^{-1/2} + (nT)^{-1}V_{nT}^{1/2})$.

We now consider $\hat{\delta}_4$ in (A.8). Write $\hat{\delta}_4 = \sum_{i=1}^{n} T_i \sum_{h=1}^{T_i-1} \sum_{m=1}^{T_i-1} b_{ij}(h, m) \hat{\xi}_u(h) \hat{R}_i(m)$. Using Lemma A.1(ii) and $V_{nT} \leq C \sum_{i=1}^{n} 2^{J_i+1}$, we can obtain

$$E\hat{\delta}_4^2 = \sum_{i=1}^{n} \sum_{h=1}^{n} T_i T_i \sum_{h=1}^{T_i-1} \sum_{m=1}^{T_i-1} b_{ij}(h, m) b_{ij}(h, m)$$

$$\times E[\hat{\xi}_u(h) \hat{R}_i(m_1) \hat{\xi}_u(h_2) \hat{R}_i(m_2)]$$

$$\leq CT^{-1} \sum_{i=1}^{n} \sum_{h=1}^{n} T_i \sum_{m=1}^{T_i-1} \left( \sum_{h=1}^{T_i-1} b_{ij}(h, m) \right)^2$$

$$\leq C(2^{J/T})V_{nT} + C\hat{\delta}_4^2 V_{nT}^{1/2}$$

where the first inequality follows from the facts that (a) $|E[\hat{\xi}_u(h) \hat{R}_i(m_1) \hat{\xi}_u(h_2) \hat{R}_i(m_2)]| \leq CT^{-3/2} T_i^{-3/2}$, and (b) for $i \neq l$, $|E[\hat{\xi}_u(h_1) \hat{R}_i(m_1) \hat{\xi}_u(h_2) \hat{R}_l(m_2)]| \leq CT^{-2} T_i^{-2}$, which can be shown by exploiting the facts that under $H_{0i}$, $\{v_{ih}\}$ coincides with $\{e_{ih}\}$, and so is i.i.d. with $E(v_{ih}) = C$ for each $i$, and $\{v_{ih}\}$ and $\{e_{ih}\}$ are mutually independent for $i \neq l$. Hence, we have $V_{nT}^{-1/2} \hat{\delta}_4 = O_p(2^{J/T} + V_{nT}^{1/2}/T)$ by Chebyshev’s inequality. Similarly, we can obtain $V_{nT}^{-1/2} \hat{\delta}_5 = O_p(2^{J/T} + V_{nT}^{1/2}/T)$.

Next, we consider $\hat{\delta}_6$. Write $\hat{\delta}_6 = \sum_{i=1}^{n} T_i \sum_{h=1}^{T_i-1} \sum_{m=1}^{T_i-1} b_{ij}(h, m) \hat{\xi}_d(h) \hat{R}_i(m)$, where $\hat{\xi}_d(h) = T_i^{-1} \sum_{h=1}^{T_i} v_{ih} \tilde{u}_{i,0,h}$ as in (A.2). Then using Lemma A.1(ii) and $V_{nT} \leq C \sum_{i=1}^{n} 2^{J_i+1}$, we can obtain

$$E\hat{\delta}_6^2 = \sum_{i=1}^{n} \sum_{h=1}^{n} T_i T_i \sum_{h=1}^{T_i-1} \sum_{m=1}^{T_i-1} b_{ij}(h, m) b_{ij}(h, m)$$

$$\times E[\hat{\xi}_d(h) \hat{R}_i(m_1) \hat{\xi}_d(h_2) \hat{R}_i(m_2)]$$

$$\leq Cn^{-1} \sum_{i=1}^{n} \sum_{h=1}^{n} \sum_{m=1}^{T_i-1} \left( \sum_{m=1}^{T_i-1} b_{ij}(h, m) \right)^2$$

$$\leq C(2^{J/T})V_{nT} + Cn^{-2} V_{nT}^{1/2}$$
where for the first inequality, we have used the facts that (a) $|E[\hat{\xi}_n(h_1)\hat{R}_i(m_1)\hat{\xi}_n(h_2)\hat{R}_i(m_2)]| \leq CT^{-n}p^{-1}$; (b) for $i \neq l$, $|E[\hat{\xi}_n(h_1)\hat{R}_i(m_1)\hat{\xi}_n(h_2)\hat{R}_l(m_2)]| \leq CT^{-1}T^{-n}p^{-2}$, which can be shown by exploiting the i.i.d. property of $\{v_{il}\}$ and the independence between $\{v_{il}\}$ and $\{v_{ij}\}$ for $i \neq l$ under $H_0$, via tedious algebra. It follows by Chebyshev's inequality and $2\sqrt{n}/n \to 0$ that $V_n^{-1/2}\delta_0 = O_p(2T^{1/2}/T^{1/2} + V_n^{1/2}/n) \to 0$. Similarly, we have $V_n^{-1/2}\delta_1 = O_p(2T^{-1/2}/T^{1/2} + V_n^{1/2}/n) \to 0$. We have shown $V_n^{-1/2}\delta_c \to 0$ for $1 \leq c \leq 9$ given $\max_{i \leq n} 2^{(J_i+1)}/(n^2 + T) \to 0$. Proposition A.2 then follows from (A.8).

**PROOF OF THEOREM A.2:** Recalling the definition of $\tilde{\alpha}_{ijk}$ in (6.3), we can write $\sum_{i=0}^n \sum_{k=1}^{2^J} \tilde{\alpha}_{ijk} = \sum_{i=1}^n T_i \sum_{h=1}^{T_i-1} \sum_{m=1}^{T_i-1} h_j(h, m)\hat{R}_i(h)\hat{R}_i(m) = \sum_{i=1}^n (\hat{A}_i + \hat{B}_i + \hat{B}_j - \hat{B}_k)$, where

\[
\hat{A}_i = 2T_i^{-1} \sum_{h=1}^{T_i-1} \sum_{m=1}^{T_i-1} b_j(h, m) \sum_{t=1}^{T_i} \sum_{u_t=1}^{T_i} v_{ilt} v_{ilt-m} \quad \text{by symmetry of } b_j(\cdot, \cdot),
\]

\[
\hat{B}_{ii} = T_i^{-1} \sum_{h=1}^{T_i-1} \sum_{m=1}^{T_i-1} b_j(h, m) \sum_{t=1}^{T_i} v_{ilt}^2 v_{ilt-h} v_{ilt-m},
\]

\[
\hat{B}_{ij} = T_i^{-1} \sum_{h=1}^{T_i-1} \sum_{m=1}^{T_i-1} b_j(h, m) \sum_{t=1}^{T_i} v_{ilt} v_{ilt-h} v_{ilt-s},
\]

\[
\hat{B}_{jj} = T_i^{-1} \sum_{h=1}^{T_i-1} \sum_{m=1}^{T_i-1} b_j(h, m) \sum_{t=1}^{T_i} v_{ilt} v_{ilt-h} v_{ilt-m}.
\]

Note again that under $H_0$, $\{v_{il}\}$ coincides with $\{v_{ij}\}$, and so is i.i.d. for each $i$, and $\{v_{il}\}$ and $\{v_{ij}\}$ are independent for $i \neq l$ and all $t, s$.

**PROPOSITION A.3:** $V_n^{-1/2}(\sum_{i=1}^n 2\pi T_i \sum_{j=0}^{J_i} \sum_{k=1}^{2^J} \tilde{\alpha}_{ijk}^2 - M_n) = V_n^{-1/2} \sum_{i=1}^n \hat{A}_i + o_P(1)$.

Next, we decompose $\hat{A}_i$ into the terms with $t - s > q_i$, and $t - s \leq q_i$, for some integer $q_i \in \mathbb{Z}^+$.

(A.11) \[
\hat{A}_i = 2T_i^{-1} \sum_{h=1}^{T_i-1} \sum_{m=1}^{T_i-1} b_j(h, m) \left( \sum_{t=q_i+1}^{T_i} \sum_{s=1}^{T_i} + \sum_{t=2}^{T_i} \sum_{s=\max(t-q_i, 1)}^{T_i-1} \right) v_{ilt} v_{ilt-h} v_{ilt-m}
\]

\[
\equiv \hat{B}_i + \hat{B}_j.
\]

Furthermore, we decompose

(A.12) \[
\hat{B}_i = 2T_i^{-1} \sum_{h=1}^{q_i} \sum_{m=1}^{q_i} \sum_{t=q_i+1}^{T_i} \sum_{s=1}^{T_i} b_j(h, m) \sum_{t=q_i+1}^{T_i} \sum_{s=1}^{T_i} v_{ilt} v_{ilt-h} v_{ilt-m} + \hat{B}_j.
\]

**PROPOSITION A.4:** Suppose Assumptions 1 and 2 hold, $2^{-J}/T \to 0$, $q_i \equiv q_i(T_i) \to \infty$, $q_i/2^J \to \infty$, and $q_i^2/T_i \to 0$, where $J \equiv \max_{i \leq n} (J_i)$. If $\{v_{il}\}$ is i.i.d. for each $i$, then $V_n^{-1/2} \sum_{i=1}^n \hat{A}_i = V_n^{-1/2} \sum_{i=1}^n \hat{U}_i + o_P(1)$.

**PROPOSITION A.5:** Under the conditions of Proposition A.4, $V_n^{-1/2} \sum_{i=1}^n \hat{U}_i d \to N(0, 1)$. 

**Proof of Proposition A.3**: Recall the definition of $M_{n,T}$ in Theorem A.2. We obtain

$$
\sum_{i=1}^{n} \left( 2 \pi T_i \sum_{j=0}^{k} \sum_{k=1}^{2^l} \tilde{\alpha}_{jk}^2 - M_{n,T} \right) = \sum_{i=1}^{n} \tilde{A}_i + \sum_{i=1}^{n} (\tilde{B}_{1i} - \sigma_i^2 M_{i0}) - \sum_{i=1}^{n} \tilde{B}_{2i} - \sum_{i=1}^{n} \tilde{B}_{3i}.
$$

We shall show (a) $V_{nT}^{-1/2} (\sum_{i=1}^{n} \tilde{B}_{1i} - M_{n,T} ) \not\rightarrow 0$; (b) $V_{nT}^{-1/2} \sum_{i=1}^{n} \tilde{B}_{2i} \not\rightarrow 0$; and (c) $V_{nT}^{-1/2} \times \sum_{i=1}^{n} \tilde{B}_{3i} \not\rightarrow 0$.

(a) Observe that $\tilde{B}_{1i}$ has a structure similar to $\tilde{B}_{1i}$ in Lee and Hong (2001). Following Lee and Hong’s (2001) reasoning and using Lemma A.1(ii), we can obtain that for each $i$ and for $T_i$ sufficiently large,

$$
E(\tilde{B}_{1i} - E\tilde{B}_{1i})^2 \leq C T_i^{-1} \left[ \sum_{h=1}^{T_i-1} \sum_{m=1}^{T_i-1} |b_{1j}(h,m)| \right]^2 \leq C^2 (q^2 / T) \sum_{h=1}^{T_i-1} \sum_{m=1}^{T_i-1} |b_{1j}(h,m)|.
$$

Because $\{\tilde{B}_{1i}\}$ is a random sequence independent across $i$ and $\sum_{i=1}^{n} E\tilde{B}_{1i} = M_{n,T}$, we have $E(\sum_{i=1}^{n} \tilde{B}_{1i} - M_{n,T})^2 = \sum_{i=1}^{n} E(\tilde{B}_{1i} - E\tilde{B}_{1i})^2 = O(V_{nT} 2^l / T)$ given Lemma A.1(ii) and $V_{nT} \leq C \sum_{i=1}^{n} 2^{h+1}$. Hence, by Chebyshev’s inequality and $2^l / T \rightarrow 0$, we have $V_{nT}^{-1/2} (\sum_{i=1}^{n} \tilde{B}_{1i} - M_{n,T}) = O_p((2 / T)^{1/2}) = o_p(1)$.

(b) Next, we consider $\tilde{B}_{2i}$. Following Lee and Hong (2001), we have $E\tilde{B}_{2i} \leq C T_i^{-1} \times |\sum_{h=1}^{T_i-1} \sum_{m=1}^{T_i-1} |b_{1j}(h,m)||^2$. Then by the fact that $\tilde{B}_{2i}$ is a zero-mean random sequence independent across $i$, Lemma A.1(ii), and $V_{nT} \leq C \sum_{i=1}^{n} 2^{h+1}$, we have $E(\sum_{i=1}^{n} \tilde{B}_{2i})^2 = \sum_{i=1}^{n} E\tilde{B}_{2i}^2 = O((2^l / T)V_{nT})$. Hence, $V_{nT}^{-1/2} \sum_{i=1}^{n} \tilde{B}_{2i} \not\rightarrow 0$ by Chebyshev’s inequality and $2^{l+1} / T \rightarrow 0$.

(c) By reasoning similar to (b), we can obtain $V_{nT}^{-1/2} \tilde{B}_{3i} \not\rightarrow 0$. Q.E.D.

**Proof of Proposition A.4**: Given (A.11) and (A.12), we have $\hat{A}_i = \hat{U}_i + \hat{B}_{di} + \hat{B}_{si} + \hat{B}_{ui}$. It suffices to show $V_{nT}^{-1/2} \sum_{i=1}^{n} \hat{B}_{di} \not\rightarrow 0$ for $c = 4, 5, 6$. (a) We first consider $\hat{B}_{di}$ in (A.11). From Lee and Hong (2001, Proof of Theorem i), we have for each $i$ and for $T_i$ sufficiently large,

$$
E\hat{B}_{di}^2 \leq C(q / T_i) \left[ \sum_{h=1}^{T_i-1} \sum_{m=1}^{T_i-1} |b_{1j}(h,m)| \right]^2 \leq C^2 (q^2 / T) \sum_{h=1}^{T_i-1} \sum_{m=1}^{T_i-1} |b_{1j}(h,m)|,
$$

where $\bar{q} \equiv \max_{1 \leq i \leq n}(q_i)$ and the last inequality follows from Lemma A.1(ii). Hence, using the fact that $\{\tilde{B}_{di}\}$ is a zero-mean random sequence independent across $i$, Lemma A.1(ii) and $V_{nT} \leq C \sum_{i=1}^{n} 2^{h+1}$, we have $E(\sum_{i=1}^{n} \hat{B}_{di})^2 = \sum_{i=1}^{n} E\hat{B}_{di}^2 = O(V_{nT}q^2 / T)$. This, Chebyshev’s inequality, $q^2 / T \rightarrow 0$, and $2^l / T \rightarrow 0$, imply $V_{nT}^{-1/2} \sum_{i=1}^{n} \hat{B}_{di} \not\rightarrow 0$.

(b) Next, we consider $\hat{B}_{sji}$ as in (A.12). By the definition of $b_{1j}(h,m)$, the Cauchy–Schwarz inequality, and Assumption 2, we obtain

$$
E\tilde{B}_{sji}^2 = \sigma_i^2 \sum_{h=1}^{T_i-1} \sum_{m=0}^{q_i} b_{1j}(h,m) \leq C \sum_{h=1}^{T_i-1} \sum_{m=0}^{q_i} |\psi(2\pi m / q_i)|^2 \leq C^2 2^{2h / q_i^2}.
$$

Therefore, $E(\sum_{i=1}^{n} \hat{B}_{sji})^2 = \sum_{i=1}^{n} E\tilde{B}_{sji}^2 \leq C(2 / q_0)^{2r-1} \sum_{i=1}^{n} 2^{x_i}$, where $q_0 \equiv \min_{1 \leq i \leq n}(q_i)$. It follows by Chebyshev’s inequality and $2^l / q_0 \rightarrow 0$ that $V_{nT}^{-1/2} \sum_{i=1}^{n} \tilde{B}_{sji} = O_p((2 / q_0)^{2r-1}) = o_p(1)$. 


(c) Finally, we consider \( \hat{B}_n \) in (A.12). Following Lee and Hong (2001, Proof of Theorem 1), we obtain

\[
E\hat{B}_n^2 \leq C2^{2\lambda_{[i]}/q_{[i]}^2} + CT_{[i]}^{-1} \left[ \sum_{h=1}^{T_{[i]-1}} \sum_{m=1}^{T_{[i]-1}} |b_{[h,m]}(h,m)| \right]^2
\]

\[
\leq C2^{\lambda_{[i]}/q_{[i]}^2} + C(2^{\lambda_{[i]}/T_{[i]}}) \sum_{h=1}^{T_{[i]-1}} \sum_{m=1}^{T_{[i]-1}} |b_{[h,m]}(h,m)|
\]

by Lemma A.1(ii) and Assumption 2. Thus, \( V_n^{-1/2} \sum_{i=1}^{N} \hat{B}_n = o_P[2^{\lambda_{[i]}/q_{[i]}^2} + (2^{\lambda_{[i]}/T_{[i]}}/2^2)] \) by Chebyshev’s inequality, \( V_n \leq C \sum_{i=1}^{n} 2^{\lambda_{[i]+1}/2} / q_{[i]} / q_{[i]} \to 0 \), and \( 2^{\lambda_{[i]}/T_{[i]}} \to 0 \). This completes the proof of Proposition A.4.

Q.E.D.

PROOF OF PROPOSITION A.5: We write \( \hat{U}_i = T_{i+1}^{1-1} \sum_{t=q_{[i]+2}}^{T_{[i]}} U_{vt} \), where \( U_{vt} \equiv 2v_{vt} \sum_{h=1}^{q_{[i]}} v_{vt-h} \times H_{i-1}^{j-1}(h), H_{i-1}^{j-1}(h) \equiv \sum_{m=1}^{q_{[i]}} b_{[h,m]}(h,m)S_{i-1}^{[j-1]}(m) \), and \( S_{i-1}^{[j-1]}(m) \equiv \sum_{t=1}^{\lambda_{[i]}} v_{vt} v_{vt-m} \). Then

\( \hat{U} \equiv \sum_{t=q_{[i]+2}}^{T} U_{vt} \), where \( \hat{T} \equiv \max_{1 \leq s \leq T_{[i]}} \left( T_{[i]} \right) \), \( U_{[i]} \equiv \sum_{t=1}^{n} U_{vt}(q_{[i]} \leq t \leq T_{[i]}) \), and \( \mathbf{1}(\cdot) \) is the indicator function.

Put \( \mathcal{F}_{vt} \equiv \bigotimes_{t=1}^{n} \mathcal{F}_{vt} \), where \( \mathcal{F}_{vt} \) is the sigma field generated by \( \{v_{vt}, s \leq t\}_{t=1}^{n} \). Because \( \{v_{vt}, v_{vt-h}\} \) is independent of \( H_{i-1}^{j-1}(h) \) for \( 0 < h \leq q_{[i]} \), \( \{U_{vt}, \mathcal{F}_{vt}\} \) is an adapted martingale difference sequence (m.d.s.), with

\[
E\tilde{U}_i^2 = \sum_{t=q_{[i]+2}}^{T} E(U_{vt}^2) = \sum_{t=q_{[i]+2}}^{T} \sum_{l=1}^{n} E(U_{vt}^2) \mathbf{1}(q_{[i]} \leq t \leq T_{[i]}) = V_{vt}[1 + o(1)]
\]

given \( q_{[i]} \to \infty, q_{[i]}/2^{[i]} \to \infty, \tilde{q}_{[i]}/T_{[i]} \to 0 \). We apply Brown’s (1971) martingale limit theorem by verifying his two conditions:

(a) \( \text{var}^{-1/2}(\tilde{U}_i) \sum_{t=q_{[i]+2}}^{T} E(U_{vt}^2) \mathbf{1}(|U_{vt}| \geq \epsilon \text{var}^{1/2}(\tilde{U}_i)) \to 0 \) for all \( \epsilon > 0 \),

(b) \( \text{var}^{-1/2}(\tilde{U}_i) T^{-2} \sum_{t=q_{[i]+2}}^{T} E(U_{vt}^2) \mathbf{1}(\mathcal{F}_{vt}) \to 1 \).

We first verify (a) by showing \( V_{vt}^{-2} \sum_{t=q_{[i]+2}}^{T} E(U_{vt}^2) \to 0 \). Given \( t, \{U_{vt}\} \) is a zero-mean independent sequence across \( i \), so we have \( E(U_{vt}^2) \leq C \sum_{t=1}^{n} T_{[i]}^{-2} (E(U_{vt}^2))^{1/2} \mathbf{1}(q_{[i]} \leq t \leq T_{[i]})^2 \). Moreover, following Lee and Hong (2001, Proof of Theorem 1), we can obtain that for each \( i \) and \( T_{[i]} \) sufficiently large, \( E(U_{vt}^2) \leq CI \sum_{t=1}^{n} T_{[i]}^{-1} b_{[h,m]}(h,m) \). It follows that \( V_{vt}^{-2} \sum_{t=1}^{n} T_{[i]}^{-1} E(U_{vt}^2) \mathbf{1}(q_{[i]} \leq t \leq T_{[i]}) \to 0 \).

Hence, condition (a) holds.

Next, we verify condition (b) by showing \( V_{vt}^{-2} E(\tilde{U}_i - \tilde{E}\tilde{U}_i) \to 0 \), where

\[
\tilde{U}_i \equiv \sum_{t=q_{[i]+2}}^{T} E(U_{vt}^2) \mathbf{1}(\mathcal{F}_{vt}) \]

\[
E(U_{vt}^2) \mathbf{1}(\mathcal{F}_{vt}) \equiv E \left[ \left( T_{[i]}^{-1} \sum_{t=1}^{n} U_{vt} \mathbf{1}(q_{[i]} \leq t \leq T_{[i]}) \right)^2 \mathbf{1}(\mathcal{F}_{vt}) \right]
\]

\[= \sum_{t=1}^{n} T_{[i]}^{-1} E(U_{vt}^2) \mathbf{1}(q_{[i]} \leq t \leq T_{[i]}) \]
where the second equality follows from the facts that for each \(t\), \(\{U_i\}\) is a zero-mean random sequence independent across \(i\), and that for each \(i\), \(\{U_i, F_{i-1}\}\) is an m.d.s. Lee and Hong (2001) show that for each \(i\) and for \(T_i\) sufficiently large,

\[
E \left\{ \sum_{t=\bar{q}+2}^{T_i} [E(U^2_{\bar{q}+2} | F_{i-1}) - E(U^2_{\bar{q}+2})] \right\}^2 \leq C(\bar{q}/T) \sum_{h=1}^{T_{i-1}} \sum_{m=1}^{T_{i-1}} |b_{j_i}(h, m)|^2 + C(J_i + 1)2^j_i .
\]

It follows that

\[
E(\hat{U}^2 - E\hat{U}^2)^2 = \sum_{i=1}^{n} E \left\{ \sum_{t=\bar{q}+2}^{T_i} [E(U^2_{\bar{q}+2} | F_{i-1}) - E(U^2_{\bar{q}+2})] \right\}^2 \leq C(\bar{q}/T) V_{nT}^2 + C(\bar{q}/T)V_{nT}.
\]

for \(T\) sufficiently large, where the equality follows from the fact that \(\{\sum_{t=\bar{q}+2}^{T_i} [E(U^2_{\bar{q}+2} | F_{i-1}) - E(U^2_{\bar{q}+2})]\) is a zero-mean independent sequence across \(i\), and the inequality follows from Lemma A.1(ii) and \(V_{nT} \leq C \sum_{i=1}^{n} 2^{j_i+1}\). Hence, given \(\bar{q}/T \to 0\) and \(2^{j_i}/T \to 0\), we have \(V_{nT}^2 E(\hat{U}^2 - E\hat{U}^2)^2 = O(\bar{q}/T^2) + O(V_{nT}^2 \sum_{i=1}^{n} (J_i + 1)2^{j_i}) \to 0\) as \(n, T \to \infty\). Thus, condition (b) holds, and so \(V^{-1/2}_{nT} \hat{U} \sim N(0, 1)\) by Brown’s theorem. \(Q.E.D.\)

**PROOF OF THEOREM A.3:** (a) Recalling the definition of \(\hat{M}\) and \(M_{nT}\), we have

(A.13) \[
\hat{M} - M_{nT} = \sum_{i=1}^{n} \left[ \left( \hat{R}_i(0) - R_i(0) \right)^2 + 2(\hat{R}_i(0) - R_i(0))R_i(0) \right] \sum_{h=1}^{T_{i-1}} b_{j_i}(h, h)
\]

\[
\equiv \hat{M}_5 + 2\hat{M}_6.
\]

Because \(\hat{R}_i(0) - R_i(0) = [\hat{R}_i(0) - \hat{R}_i(0)] + [\hat{R}_i(0) - \hat{R}_i(0)] + [\hat{R}_i(0) - R_i(0)]\), we can write

(A.14) \[
\hat{M}_5 \leq 4 \sum_{i=1}^{n} \left[ \left( \hat{R}_i(0) - \hat{R}_i(0) \right)^2 + [\hat{R}_i(0) - \hat{R}_i(0)]^2 + [\hat{R}_i(0) - R_i(0)]^2 \right] \sum_{h=1}^{T_{i-1}} b_{j_i}(h, h)
\]

\[
\equiv 4(\hat{M}_{51} + \hat{M}_{52} + \hat{M}_{53}).
\]

Using (A1), the Cauchy–Schwarz inequality, Assumptions 3–5, Lemma A.1(v), and \(V_{nT} \leq C \sum_{i=1}^{n} 2^{j_i+1}\), we have

\[
V^{-1/2}_{nT} \hat{M}_{51} \leq 4V^{-1/2}_{nT} \|\hat{\beta} - \beta\|^2 \sum_{i=1}^{n} \left[ \|\hat{\beta} - \beta\|^2 \|\hat{R}_{i,0}(0)\|^2 + \|\hat{R}_{i,0}(0)\|^2 \right]
\]

\[
\times \sum_{h=1}^{T_{i-1}} b_{j_i}(h, h)
\]

\[
= O_P \left[ (nT)^{-1} V_{nT}^{1/2} \right].
\]

Similarly, using \(V_{nT} \leq C \sum_{i=1}^{n} 2^{j_i+1}\), the Cauchy–Schwarz inequality, Markov’s inequality, the i.i.d. property of \(\{v_{i,j}\}\) for each \(i\), and independence between \(\{v_{i,j}\}\) and \(\{v_{i,l}\}\) for all \(i \neq l\) under \(\mathbb{H}_0\), we can obtain \(E[\hat{R}_i(0) - R_i(0)]^2 \leq C(T_i^{-2} + n^{-1}T_i^{-1})\). It follows from Markov’s inequality, the Cauchy–Schwarz inequality, Lemma A.1(v), and \(V_{nT} \leq C \sum_{i=1}^{n} 2^{j_i+1}\) that \(V^{-1/2}_{nT} \hat{M}_{52} = O_P \left[ (T^{-2} + n^{-1}T^{-1}) V_{nT}^{1/2} \right] \). Using Markov’s inequality, \(E[\hat{R}_i(0) - R_i(0)]^2 \leq C T_i^{-1}\), Lemma A.1(v), and \(V_{nT} \leq C \sum_{i=1}^{n} 2^{j_i+1}\), we have \(V^{-1/2}_{nT} \hat{M}_{53} = O_P \left( T^{-1} V_{nT}^{1/2} \right) \). Hence, we have \(V^{-1/2}_{nT} \hat{M}_5 = O_P \left( T^{-1} V_{nT}^{1/2} \right) = o_P(1)\) given (A.14) and \(V_{nT}/T^2 \to 0\).
Next, we consider the second term $\tilde{M}_d$ in (A.13). We write
\begin{equation}
\tilde{M}_d = \sum_{i=1}^n \left[ [\tilde{R}_i(0) - \tilde{R}_i(0)] + [\tilde{R}_i(0) - \tilde{R}_i(0)] + [\tilde{R}_i(0) - \tilde{R}_i(0)] \right] R_i(0) \sum_{h=1}^{T_n-1} b_j(h, h)
\end{equation}
\[\equiv \tilde{M}_{d1} + \tilde{M}_{d2} + \tilde{M}_{d3}.
\]

Using (A.1) with $h = 0$, Assumptions 3–5, Lemma A.1(v), and $V_{nt} \leq C \sum_{i=1}^n 2^{j_i+1}$, we have
\[V_{nt}^{-1/2} |\tilde{M}_{d3}| \leq V_{nt}^{-1/2} \|\hat{\beta} - \beta\|^2 \sum_{i=1}^n \left[ \|\hat{\beta} - \beta\| \|\tilde{f}_{i1}(0)\| + \|\tilde{f}_{i2}(0)\| + \|\tilde{f}_{i3}(0)\| \right] R_i(0)
\times \sum_{h=1}^{T_n-1} b_j(h, h)
\]
\[= O_P\left( (nT)^{-1/2} V_{nt}^{1/2} \right).
\]

Also, using (A.2), $E[\tilde{R}_i(0) - \tilde{R}_i(0)]^2 \leq C(T_i^{-2} + (nT_i)^{-1})$, the Cauchy–Schwarz inequality, Markov's inequality, Lemma A.1(ii), and $V_{nt} \leq C \sum_{i=1}^n 2^{j_i+1}$, we have $V_{nt}^{-1/2} \tilde{M}_{d2} = O_P(T^{-1} V_{nt}^{1/2} + (nT)^{-1/2} V_{nt}^{1/2})$.

Finally, noting that $\tilde{R}_i(0) - \tilde{R}_i(0)$ is a zero-mean sequence independent across $i$, we have
\[E\tilde{M}_{d3} = \sum_{i=1}^n E[\tilde{R}_i(0) - \tilde{R}_i(0)]^2 \sum_{h=1}^{T_n-1} b_j(h, h)
\]
\[\leq C^2(2^j/T) \sum_{i=1}^n \sum_{h=1}^{T_n-1} b_j(h, h),
\]
where the last inequality follows by Lemma A.1(v). It follows that $V_{nt}^{-1/2} \tilde{M}_{d3} = O_P(2^{j_i}/T^{1/2})$ by Chebyshev’s inequality. Hence, (A.15) and $2^j/T \to 0$ imply $V_{nt}^{-1/2} \tilde{M}_d \xrightarrow{p} 0$. We have shown $V_{nt}^{-1/2} (\tilde{M} - M_{nt}) \xrightarrow{p} 0$.

(b) The proof for $\tilde{V}/V_{nt} \xrightarrow{p} 1$ is analogous to (a).

PROOF OF THEOREM 2: To conserve space, we only show for $\tilde{W}_1$. The proof for $\tilde{W}_2$ is similar. Put $\tilde{M} \equiv \sum_{i=1}^n R_i(0)(2^{j_i+1} - 1)$ and $\tilde{V} \equiv 4 \sum_{i=1}^n \tilde{R}_i(0)(2^{j_i+1} - 1)$. Then we can write $\tilde{W}_1 - \tilde{W}_1 = \tilde{W}_1(\tilde{V}_{1/2}^{1/2} - 1) + \tilde{V}^{1/2}(\tilde{M} - M_{nt})^{(1/2)/2}$, where $\tilde{M}$ and $\tilde{V}$ are as in (3.14). Because $\tilde{W}_1 = O_P(1)$ by Theorem 1, $V_{nt}^{-1/2} (\tilde{M} - M_{nt}) \xrightarrow{p} 0$ and $\tilde{V}/V_{nt} \xrightarrow{p} 1$ by Theorem A.3, it suffices for $\tilde{W}_1 - \tilde{W}_1 \xrightarrow{p} 0$ and $\tilde{W}_1 \xrightarrow{d} N(0, 1)$ if (a) $V_{nt}^{-1/2} (\tilde{M} - M_{nt}) \xrightarrow{p} 0$ and (b) $\tilde{V}/V_{nt} \xrightarrow{p} 1$.

We first show (a). Following reasoning analogous to the proof of Theorem A.3, we obtain $V_{nt}^{-1/2} (\tilde{M} - M_{nt}) \xrightarrow{p} 0$, where $M_{nt} \equiv \sum_{i=1}^n \alpha_i (2^{j_i+1} - 1)$. It remains to show $V_{nt}^{-1/2} (M_{nt} - M_{nt}) \xrightarrow{p} 0$. This follows from Lemma A.1(v), $2^{j_i+1} = a_i T^r, n/T^r \log^2(T) \to 0$, and $n/T^{2(2r-1)-2(2r-1)/2} \to 0$ because
\[V_{nt}^{-1/2} |M_{nt} - M_{nt}^0| \leq CV_{nt}^{-1/2} \left[ \frac{n}{J_i} + 1 + (2^j/T)^{2r-1} \sum_{i=1}^{J_i} 2^{j_i+1} \right] \to 0.
\]

Now we show (b). Put $V_{nt}^0 \equiv \sum_{i=1}^n \alpha_i (2^{j_i+1} - 1)$. Following reasoning analogous to the proof of Theorem A.3, we can obtain $\tilde{V}_{nt}/V_{nt}^0 \xrightarrow{p} 1$. It remains to show $V_{nt}/V_{nt}^0 \xrightarrow{p} 1$. This follows from Lemma A.1(vi) and $J_i \to \infty$, because
\[V_{nt}^{-1/2} V_{nt}^0 - V_{nt}^0 \leq CV_{nt}^{-1/2} \left[ \frac{n}{J_i} + 1 \right] + (2^j/T)^{2r-1} \sum_{i=1}^{J_i} 2^{j_i+1} \to 0,
\]
where $V_{nT}^{-1} \sum_{i=1}^{n} (J_i + 1)^2 \to 0$ given $V_{nT} \leq C \sum_{i=1}^{n} 2^{J_i + 1}$ and $J_i \to \infty$. \hfill Q.E.D.

**Proof of Theorem 3:** Recall the definition of $\hat{M}$ and $\hat{V}$ as in (3.14). Following reasoning analogous to the proof of Theorem A.3, we can obtain $\hat{M} = M_{nT}[1 + o_p(1)]$ and $\hat{V} = V_{nT}[1 + o_p(1)]$. It follows that $(n_A T)^{-1} \hat{V}^{-1/2} \hat{W}_t = (n_A T)^{-1} \sum_{i=1}^{n_A} 2 \pi T_i \sum_{j=1}^{J_i} \sum_{k=1}^{2^j} \alpha_{ijk}^2 + o_p(1)$ given $M_{nT} \leq C \sum_{i=1}^{n_A} 2^{J_i + 1} = O(V_{nT})$, and $V_{nT} / n_A T \to 0$ by $(n_A T)^{-1} \sum_{i=1}^{n_A} 2^{J_i + 1} \to 0$. It remains to show

(a) $(n_A T)^{-1} \sum_{i=1}^{n_A} 2 \pi T_i \sum_{j=0}^{J_i} \sum_{k=1}^{2^j} (\hat{\alpha}_{ijk}^2 - \alpha_{ijk}^2) \to 0$; and

(b) $n_A T^{-1} \sum_{i=1}^{n_A} 2 \pi T_i \sum_{j=0}^{J_i} \sum_{k=1}^{2^j} \alpha_{ijk} = (n_A T)^{-1} \sum_{i=1}^{n_A} 2 \pi \alpha_{ijk} Q(f_i, f_0) + o(1)$, where $\alpha_{ijk}$ is defined in (3.8).

We first show (a). Because

$$\begin{align*}
(n_A T)^{-1} \sum_{i=1}^{n_A} 2 \pi T_i \sum_{j=0}^{J_i} \sum_{k=1}^{2^j} (\hat{\alpha}_{ijk}^2 - \alpha_{ijk}^2) &= (n_A T)^{-1} \sum_{i=1}^{n_A} 2 \pi T_i \sum_{j=0}^{J_i} \sum_{k=1}^{2^j} (\hat{\alpha}_{ijk} - \alpha_{ijk})^2 + 2(\hat{\alpha}_{ijk} - \alpha_{ijk})\alpha_{ijk},
\end{align*}$$

it suffices to show that the first term in (A.16) vanishes in probability. That the second term in (A.16) vanishes in probability then follows by the Cauchy–Schwarz inequality and the fact that $(n_A T)^{-1} \sum_{i=1}^{n_A} 2 \pi T_i \sum_{j=0}^{J_i} \sum_{k=1}^{2^j} (\hat{\alpha}_{ijk}^2 - \alpha_{ijk}^2) \leq C \sup_{i \in \mathbb{N}_A} Q(f_i, f_0) \leq C^2$. Noting $\hat{\alpha}_{ijk} - \alpha_{ijk} = (\hat{\alpha}_{ijk} - \alpha_{ijk}) + (\hat{\alpha}_{ijk} - \alpha_{ijk})$, we obtain

$$\begin{align*}
\sum_{i=1}^{n_A} 2 \pi T_i \sum_{j=0}^{J_i} \sum_{k=1}^{2^j} (\hat{\alpha}_{ijk} - \alpha_{ijk})^2 &\leq 2 \sum_{i=1}^{n_A} 2 \pi T_i \sum_{j=0}^{J_i} \sum_{k=1}^{2^j} [(\hat{\alpha}_{ijk} - \alpha_{ijk})^2 + (\alpha_{ijk} - \alpha_{ijk})^2]
\equiv 2(\hat{M}_{71} + \hat{M}_{72}).
\end{align*}$$

Following reasoning analogous to the proof of Proposition A.1, we can obtain

$$\begin{align*}
(n_A T)^{-1} \hat{M}_{72} &= O_p((n_A T)^{-1} + (n_A T)^{-1} V_{nT})
\end{align*}$$

under Assumptions 1–6 and $\mathbb{H}_4$. Note that we have obtained a slower rate for $\hat{M}_{72}$ under $\mathbb{H}_4$ than under $\mathbb{H}_0$. For the second term in (A.17), we further decompose

$$\begin{align*}
\hat{M}_{72} &\leq 2 \sum_{i=1}^{n_A} 2 \pi T_i \sum_{j=0}^{J_i} \sum_{k=1}^{2^j} [(\hat{\alpha}_{ijk} - E\hat{\alpha}_{ijk})^2 + (E\hat{\alpha}_{ijk} - \alpha_{ijk})^2] \equiv 2(\hat{M}_{721} + \hat{M}_{722}).
\end{align*}$$

We now consider the first term in (A.19). We have $\sup_{1 \leq h \leq T_i - 1} \var[\hat{R}_i(h)] \leq C T_i^{-1}$, which follows from Assumption 6 and

$$\var[\hat{R}_i(h)] = T_i^{-1} \sum_{l=1}^{T_i} (1 - |l|/T_i)[R_i^2(l) + R_i(l - h)R_i(l + h) + \kappa_i(l, h, l + h)].$$

Cf. Hannan (1970, p. 209). Therefore, we have

$$\begin{align*}
\hat{M}_{721} &\leq \sum_{i=1}^{n_A} T_i \sup_{1 \leq h \leq T_i - 1} \var[\hat{R}_i(h)] \sum_{h=1}^{T_i - 1} \sum_{m=1}^{T_i - h - 1} |b_{f_i}(h, m)| = O(V_{nT}).
\end{align*}$$
For the second term in (A.19), noting that \( |E\bar{\alpha}_{ijk} - \alpha_{ijk}| \leq (2\pi)^{-1/2} T_i^{-1} \sum_{h=-\infty}^{\infty} |hR_i(h)\hat{\Psi}_j(h)| \), we have

\[
\hat{M}_{22} \leq \sum_{i=1}^{n} \sum_{j=1}^{J_i} \sum_{k=1}^{2^j} \left[ \sum_{h=1}^{T^{-1}} R_i^2(h) \right] \left[ \sum_{h=-\infty}^{\infty} h^2 |\hat{\Psi}_j(h)|^2 \right] 
\]

\[
= O \left( \frac{2^j}{T} \sum_{i=1}^{n} 2^{j+1} \right) = o(V_{nT})
\]

given Assumption 2 and \( 2^j / T \to 0 \). It follows by Markov’s inequality that \( (n_T^{-1}) \hat{M}_{22} = O_p[(n_T^{-1}) V_{nT}] \). This, (A.17), (A.18), and \( V_{nT} / (n_T^{-1}) \to 0 \) imply (a).

Next, we show (b). This follows because

\[
(n_T)^{-1} 2\pi T_i \sum_{j=0}^{2^j} C_i^2 - (n_T)^{-1} 2\pi T_i \sum_{j=0}^{2^j} C_i^2 \leq C \sup_{i \in N_T} \sum_{j=0}^{2^j} C_i^2 \to 0
\]

as \( \min_{i \leq j \leq a} (J_i) \to \infty \) and \( Q_t(f_i, f_0) = \sum_{j=0}^{\infty} \sum_{k=1}^{2^j} \alpha_{ijk}^2 \leq C \). This completes the proof for \( \hat{W}_1 \).

Q.E.D.

**Proof of Theorem 4:** Given \( T_i = c_i T \) and \( 2^j+1 = b_i T_i^r \), we have \( 2^j+1 = b_i T_i^r \), where \( b_i \equiv a_i c_i^r \). When \( T \to \infty \), we have

\[
V_{n_T} = n^{-1} \sum_{i=1}^{n} \alpha_i^2 (2^j+1 - 1) = T^r \left(n^{-1} \sum_{i=1}^{n} \alpha_i^2 b_i \right) [1 + o(1)] = \bar{b} T^r [1 + o(1)],
\]

where \( \bar{b} \equiv n^{-1} \sum_{i=1}^{n} \alpha_i^2 b_i \). It follows from Theorem 2 and \( \hat{V} / V_{n_T} \to 1 \) that

\[
(n_T)^{-1} (b T_i^r)^{1/2} \hat{W}_1 = n^{-1} \sum_{i \in N_T} c_i Q_i(f_i, f_0) + o_p(1) \quad \text{and}
\]

\[
(n_T)^{-1} (b T_i^r)^{1/2} \hat{W}_2 = n^{-1} \sum_{i \in N_T} (\bar{b} / \alpha_i^2 b_i)^{1/2} c_i Q_i + o_p(1).
\]

For \( c = 1, 2 \), we put \( S_{n_T}^{(c)} \equiv -2 \ln[1 - \Phi(\hat{W}_c)] \), where \( \Phi(\cdot) \) is the N(0, 1) CDF. Because \( \ln[1 - \Phi(z)] = -\frac{1}{2} z^2 [1 + o(1)] \) as \( z \to +\infty \), we have

\[
(n_T)^{-1/2} \hat{b} T_i^r S_{n_T}^{(1)} = \left[n^{-1} \sum_{i \in N_T} c_i Q_i(f_i, f_0) \right]^2 + o_p(1)
\]

and

\[
(n_T)^{-1/2} \hat{b} T_i^r S_{n_T}^{(2)} = \left[n^{-1} \sum_{i \in N_T} (\bar{b} / \alpha_i^2 b_i)^{1/2} c_i Q_i(f_i, f_0) \right]^2 + o_p(1).
\]

Suppose \( \{T_i^{(1)}\}_{i=1}^{n_1} \) and \( \{T_i^{(2)}\}_{i=1}^{n_2} \) are two sequences of sample sizes used for \( \hat{W}_1 \) and \( \hat{W}_2 \) respectively so that \( S_{n_T}^{(1)} / n_1^{(1)} T_i^{(1)} \to 1 \) as \( n_1^{(1)} \to \infty, T^{(1)} \to \infty \), and \( T^{(2)} \to \infty \). Then Bahadur’s relative
efficiency of \( \hat{W}_1 \) to \( \hat{W}_2 \)

\[
BE(\hat{W}_1 : \hat{W}_2) = \lim_{n \to \infty} \frac{\sum_{i=1}^{n} T_i^{(2)}}{\sum_{i=1}^{n} T_i^{(1)}} = \lim_{n \to \infty} \frac{\frac{1}{n} \sum_{i=1}^{n} \sigma_i^{(2)} n_i^{(2)} T_i^{(2)}}{\frac{1}{n} \sum_{i=1}^{n} \sigma_i^{(1)} n_i^{(1)} T_i^{(1)}}
\]

\[
= \left( \frac{\lim_{n \to \infty} n^{-1} \sum_{i=1}^{n} \sqrt{n} B_i / (\sigma_i b_i) c_i Q(f_i, f_0)}{\lim_{n \to \infty} n^{-1} \sum_{i=1}^{n} c_i Q(f_i, f_0)} \right)^{(1+c) / (3-c)}
\]

where the last equality follows from \( n^{(c)} = \gamma (T^{(c)})^c \) for \( c = 1, 2 \). Hence, \( BE(\hat{W}_1 : \hat{W}_2) > 1 \) if \( a_i \) is a monotonically increasing function of \( Q(f_i, f_0) \) and \( c_i = c \) (i.e., \( T_i = T \)) for all \( i \).

Q.E.D.

PROOF OF THEOREM 5: We only consider \( \hat{W}_1(\hat{J}) \); the proof for \( \hat{W}_2(\hat{J}) \) is similar. We write

\[
\hat{W}_1(\hat{J}) - \hat{W}_2(\hat{J}) = \hat{V}(\hat{J})^{-1/2} \left\{ \sum_{i=1}^{n} 2\pi T_i \sum_{j=0}^{j} \sum_{k=0}^{k} \hat{\alpha}_{ijk}^2 - [\hat{M}(\hat{J}) - \hat{M}(J)] \right\} - [\hat{V}(\hat{J})^{1/2} / \hat{V}(J)^{1/2} - 1] \hat{W}_1(J).
\]

Given \( \hat{W}_1(J) = O_p(1) \) by Theorem 1 and \( \hat{V}(J)/V_J \rightarrow 1 \) by Theorem A.3, it suffices for \( \hat{W}_1(\hat{J}) - \hat{W}_1(J) \rightarrow 0 \) and \( \hat{W}_2(\hat{J}) \rightarrow N(0, 1) \) if (a) \( V_{nT}^{-1/2} \left( \sum_{i=1}^{n} 2\pi T_i \sum_{j=0}^{j} \sum_{k=0}^{k} \hat{\alpha}_{ijk}^2 - [\hat{M}(\hat{J}) - \hat{M}(J)] \right) \rightarrow 0 \) and (b) \( \hat{V}(J)/\hat{V}(J) \rightarrow 1 \).

We first show (a). Decompose

\[
(A.20) \quad \sum_{i=1}^{n} 2\pi T_i \sum_{j=0}^{j} \sum_{k=0}^{k} \hat{\alpha}_{ijk}^2 = \sum_{i=1}^{n} 2\pi T_i \sum_{j=0}^{j} \sum_{k=0}^{k} (\hat{\alpha}_{ijk} - \bar{\alpha}_{ijk})^2 + \hat{\alpha}_{ijk}^2 + 2(\hat{\alpha}_{ijk} - \bar{\alpha}_{ijk})\bar{\alpha}_{ijk}
\]

\[
\equiv \hat{G}_1 + \hat{G}_2 + 2\hat{G}_3.
\]

For the first term in \( (A.20) \), we write \( \hat{G}_1 = \sum_{i=1}^{n} 2\pi T_i (\sum_{j=0}^{j} \sum_{k=0}^{k} \hat{\alpha}_{ijk} - \bar{\alpha}_{ijk})^2 = \hat{G}_{11} - \hat{G}_{12} \). By Proposition A.1, we have \( V_{nT}^{-1/2} \hat{G}_{12} \rightarrow 0 \). Next, for any given constants \( M > 0 \) and \( \epsilon > 0 \), we have

\[
P(\hat{G}_{11} > \epsilon) \leq P(\hat{G}_{11} > \epsilon, 2^{1/2} M | 2^{1/2} - 1| \leq \epsilon) + P(2^{1/2} M | 2^{1/2} - 1| > \epsilon),
\]

where the second term vanishes to \( 0 \) as \( n, T \to \infty \) given \( 2^{1/2} | 2^{1/2} - 1| \rightarrow 0 \). For the first probability, given \( 2^{1/2} M | 2^{1/2} - 1| \leq \epsilon \), we have for all \( n \) and \( T \) sufficiently large, \( V_{nT}^{-1/2} \hat{G}_{11} \leq V_{nT}^{-1/2} \sum_{i=1}^{n} 2\pi T_i (\sum_{j=0}^{j} \sum_{k=0}^{k} \hat{\alpha}_{ijk} - \bar{\alpha}_{ijk})^2 \rightarrow 0 \) by Proposition A.1. Therefore, \( V_{nT}^{-1/2} \hat{G}_1 = o_p(1) \).

Next, we consider \( \hat{G}_2 \). By symmetry of \( b_j(\cdot, \cdot) \), we write

\[
(A.21) \quad \hat{G}_2 = \sum_{i=1}^{n} \sum_{h=1}^{T_i - 1} \sum_{h=1}^{T_i} [b_j(h, h) - b_j(h, h)] + \sum_{h=1}^{n} \sum_{h=1}^{T_i - 1} [\sum_{h=1}^{T_i - 1} \hat{R}_i(h) - \sigma_i^2 b_j(h, h) - b_j(h, h)]
\]

\[
+ \sum_{h=1}^{n} T_i \sum_{m=1}^{T_i - 1} \hat{R}_i(h) R_i(m) [b_j(h, m) - b_j(h, m)]
\]

\[
\equiv \hat{G}_{21} + \hat{G}_{22} + 2\hat{G}_{23}, \quad \text{say}.
\]
For the last term \( \hat{G}_{23} \) in (A.21), we have for any constants \( M > 0 \) and \( \epsilon > 0 \),
\[
P(\nu_{-1/2}^{V_n} | \hat{G}_{23}| > \epsilon) \\
\leq P(\nu_{-1/2}^{V_n} | \hat{G}_{23}| > \epsilon, 2^{1/2} M | 2^j / 2^j - 1 | \leq \epsilon) + P(2^{1/2} M | 2^j / 2^j - 1 | > \epsilon),
\]
where, again, the second term vanishes to 0 as \( n, T \to \infty \). We now show the first probability vanishes. Put \( \tilde{T} = \max_{1 \leq j \leq n(T)} \), as before. Given \( 2^{1/2} M | 2^j / 2^j - 1 | \leq \epsilon \) and the definition of \( a_j(h, m) \) as in (3.14), we have
\[
|a_j(h, m) - a_j(h, m)| = 2 \pi \sum_{j=\log_2 2^{(1+\epsilon/2^j/2^j)}}^{\log_2 2^{(1-\epsilon/2^j/2^j)}} |c_j(h, m)\psi(2\pi h/2^j)\psi^*(2\pi m/2^j)|,
\]
where
\[
(A.22) \quad c_j(h, m) = 2^j \sum_{k=1}^{2^j} e^{2\pi i (m - h) k / 2^j} = \begin{cases} 1 & \text{if } m = h + 2^j r \text{ for any integer } r, \\ 0 & \text{otherwise.} \end{cases}
\]
Cf. Priestley (1981, p. 392, (6.19)). It follows that we have for all \( n \) and \( T \) sufficiently large,
\[
E[\hat{G}_{23}] \leq \sum_{h=1}^{\tilde{T}-1} \sum_{m=1}^{\tilde{T}-1} \sum_{i=1}^{n} |1(h \leq T_i)1(m \leq T_i) \tilde{T}_i \tilde{T}_i(h) \tilde{R}(h) \tilde{R}(m)| \\
\times 2 \pi \sum_{j=\log_2 2^{(1+\epsilon/2^j/2^j)}}^{\log_2 2^{(1-\epsilon/2^j/2^j)}} |c_j(h, m)\psi(2\pi h/2^j)\psi^*(2\pi m/2^j)| \\
\leq 2\pi n^{1/2} \sum_{j=\log_2 2^{(1+\epsilon/2^j/2^j)}}^{\log_2 2^{(1-\epsilon/2^j/2^j)}} 2^{2j} \left[ \sum_{h=1}^{\tilde{T}-1} |\psi(2\pi h/2^j)| \right] \\
\times \left[ \sum_{r=-\infty}^{\infty} |\psi(2\pi h/2^j + 2\pi r)| \right] \\
\leq C n^{1/2} 2^{j+1} \epsilon^{1/2} M = O(\epsilon V_n^{1/2} / M)
\]
where the second inequality follows given Assumption 2. Therefore, \( P(\nu_{-1/2}^{V_n} | \hat{G}_{23}| > \epsilon, 2^{1/2} M | 2^j / 2^j - 1 | \leq \epsilon) \) also vanishes to 0. Consequently, we have \( \nu_{-1/2}^{V_n} \hat{G}_{23} \xrightarrow{p} 0 \). Similarly, we can also obtain \( \nu_{-1/2}^{V_n} \hat{G}_{22} \xrightarrow{p} 0 \) and
\[
\nu_{-1/2}^{V_n} \{ \hat{G}_{21} - [\hat{M}(\hat{J}) - \hat{M}(J)] \} = \nu_{-1/2}^{V_n} \sum_{i=1}^{n} |\alpha_i^* - \hat{R}(0)| [b_j(h, h) - b_j(h, h)] \xrightarrow{p} 0,
\]
where \( n^{-1} \sum_{i=1}^{n} [\hat{R}(0) - \alpha_i^*] = o_p((nT)^{-1/2}) \) given Assumptions 3–5, and that under \( \mathbb{H}_0 \), \{\nu_i\} coincides with \{\varepsilon_i\}, and so is i.i.d. for each \( i \). It follows from (A.21) that \( \nu_{-1/2}^{V_n} \{ \hat{G}_2 - [\hat{M}(\hat{J}) - \hat{M}(J)] \} \xrightarrow{p} 0 \).

Next, by the Cauchy–Schwarz inequality, we have
\[
\nu_{-1/2}^{V_n} |\hat{G}_3| \leq (\nu_{-1/2}^{V_n} \hat{G}_1)^{1/2} (\nu_{-1/2}^{V_n} \hat{G}_2)^{1/2} \\
= O_p\left( \nu_{-1/2}^{V_n} + (T^{-1/2} + n^{-1/2} V_n^{1/4}) \right) o_p(n^{1/4}) = o_p(1)
\]
by Proposition A.1 and the fact that \( n^{-1/2}V_{nT}^{-1/2} \hat{G}_2 = n^{-1/2}V_{nT}^{-1/2} (\hat{G}_{21} + \hat{G}_{22} + \hat{G}_{23}) \overset{p}{\to} 0 \), as can be shown using reasoning similar to that for \( \hat{G}_{23} \). Result (a) then follows from (A.20).

(b) To show \( \hat{V}(\hat{J})/V_{nT} = 1 + o_p(1) \), it suffices to show \( \hat{V}(\hat{J})/V_{nT} \overset{p}{\to} 1 \) given \( \hat{V}(J)/V_{nT} \overset{p}{\to} 1 \) by Theorem A.3. Recalling the definitions of \( \hat{V}(\hat{J}) \) and \( V_{nT} \), we can use reasoning analogous to that for \( \hat{G}_{23} \) to obtain

\[
[\hat{V}(\hat{J}) - V_{nT}] / V_{nT} = n \sum_{i=1}^{n} \left[ \hat{R}_i(0) - \alpha_i^2 \right] \sum_{h=1}^{T_i-1} \sum_{m=1}^{T_i-1} [b_f(h, m) - b_f(h, m)] \overset{p}{\to} 0,
\]

where we used the fact that \( n^{-1} \sum_{i=1}^{n} [\hat{R}_i(0) - \alpha_i^2] = O_p[(nT)^{-1/2}] \). Thus, \( \hat{V}(\hat{J})/\hat{V}(J) \overset{p}{\to} 1 \). It follows that \( [\hat{V}(\hat{J})/\hat{V}(J) - 1] \hat{W}_1(J) \overset{p}{\to} 0 \) given \( \hat{W}_1(J) = O_p(1) \) by Theorem 1. Therefore, result (b) holds, and we have \( \hat{W}_1(\hat{J}) - \hat{W}_1(J) \overset{p}{\to} 0 \), and \( \hat{W}_1(\hat{J}) \overset{d}{\to} N(0, 1) \). This completes the proof. Q.E.D.

PROOF OF THEOREM 6: We shall prove for Theorem 6(b) only; the proof for Theorem 6(a) is similar and simpler. (a) We first show \( n^{-1} \sum_{i=1}^{n} Q(\hat{f}_i, \hat{f}_i) = n^{-1} \sum_{i=1}^{n} Q(\hat{f}_i, \hat{f}_i) + o_p(2^J/T + 2^{-2\psi^J}) \).

Write

\[
A(23) \quad n^{-1} \sum_{i=1}^{n} [Q(\hat{f}_i, \hat{f}_i) - Q(\tilde{f}_i, \tilde{f}_i)]
\]

\[
= n^{-1} \sum_{i=1}^{n} Q(\hat{f}_i, \hat{f}_i) + 2n^{-1} \sum_{i=1}^{n} \int_{-\pi}^{\pi} [\hat{f}_i(\omega) - \tilde{f}_i(\omega)][\hat{f}_i(\omega) - \tilde{f}_i(\omega)] d\omega
\]

\[
= \hat{Q}_1 + 2\hat{Q}_2.
\]

For \( \hat{Q}_1 \) in (A.23), by Parseval’s identity, (A.18), and \( V_{nT} \propto n2^{J+1} \), we have

\[
\hat{Q}_1 = n^{-1} \sum_{i=1}^{n} \sum_{j=0}^{J} \sum_{k=1}^{2^J} (\hat{\alpha}_{ijk} - \tilde{\alpha}_{ijk})^2 = O_p([nT]^{-1} + 2^J/nT) = O_p(2^J/T)
\]

as \( n, T \to \infty \). Next, we have \( \hat{Q}_2 = o_p(2^J/T + 2^{-2\psi^J}) \) by the Cauchy–Schwarz inequality, \( \hat{Q}_1 = o_p(2^J/T) \), \( n^{-1} \sum_{i=1}^{n} \sum_{j=0}^{J} \sum_{k=1}^{2^J} (\hat{\alpha}_{ijk} - \tilde{\alpha}_{ijk})^2 = O_p([nT]^{-1} + 2^J/nT) = O_p(2^J/T) \), which follows by Markov’s inequality and \( n^{-1} \sum_{i=1}^{n} Q(\hat{f}_i, \hat{f}_i) = O(2^J/T + 2^{-2\psi^J}) \). The latter is to be shown below.

(b) To compute \( n^{-1} \sum_{i=1}^{n} Q(\hat{f}_i, \hat{f}_i) \), we write

\[
A(24) \quad n^{-1} \sum_{i=1}^{n} E Q(\hat{f}_i, \hat{f}_i) = n^{-1} \sum_{i=1}^{n} E Q(\tilde{f}_i, \tilde{f}_i) + n^{-1} \sum_{i=1}^{n} Q(E \tilde{f}_i, \tilde{f}_i).
\]

We first consider the second term in (A.24). By the orthonormality of \( \{\Psi_{jk}(\cdot)\} \), we obtain

\[
A(25) \quad n^{-1} \sum_{i=1}^{n} Q(E \tilde{f}_i, \tilde{f}_i) = n^{-1} \sum_{i=1}^{n} \sum_{j=0}^{J+1} \sum_{k=1}^{2^J} \sum_{m=0}^{\infty} \sum_{h=0}^{\infty} \sum_{\psi^J=0}^{\infty} (E \tilde{\alpha}_{ijk} - \alpha_{ijk})^2.
\]

For the first term in (A.24), using (3.9) and (A.22), we have

\[
\sum_{j=J+1}^{\infty} \sum_{k=1}^{2^J} \sum_{h=0}^{\infty} \sum_{m=0}^{\infty} R(h)R(m) c_j(h, m) \hat{\psi}(2\pi h/2^J) \hat{\psi}^*(2\pi m/2^J)
\]

\[
= \sum_{j=J+1}^{\infty} \sum_{h=0}^{\infty} \sum_{\psi^J=0}^{\infty} R(h)R(h + 2^Jr) \hat{\psi}(2\pi h/2^J) \hat{\psi}^*(2\pi h/2^J + 2\pi r).
\]
We now evaluate the terms corresponding to $r = 0$ and $r \neq 0$ respectively. For $r = 0$, we have
\[
\sum_{j=J+1}^{\infty} \sum_{h=-\infty}^{\infty} R_j^2(h)|\hat{\phi}(2\pi h/2)|^2 = \sum_{j=J+1}^{\infty} (2\pi/2)^{\eta} \sum_{h=-\infty}^{\infty} |\hat{\phi}(2\pi h/2)|^2 h^2 R_j^2(h)
\]
\[
= \lim_{z \to \infty} \frac{|\hat{\phi}(z)|^2}{|2|^2q} \sum_{j=J+1}^{\infty} (2\pi/2)^{\eta} \sum_{h=-\infty}^{\infty} h^2 R_j^2(h) \left[ 1 + o(1) \right]
\]
\[
= \lambda q 2^{-2q(J+1)} \int_{-\pi}^{\pi} |f_i^{(q)}(\omega)|^2 d\omega \left[ 1 + o(1) \right],
\]
where $f_i^{(q)}(\cdot)$ and $\lambda_q$ are defined in (6.1) and (6.2), and $o(1)$ is uniform in $i$ and $\omega \in [-\pi, \pi]$. For the term corresponding to $r \neq 0$, it can be shown to be $O(2^{-2q(J+1)})$. It follows that for the first term in (A.25),
\[
(A.26) \quad n^{-1} \sum_{i=1}^{n} \sum_{j=J+1}^{2^l} \alpha_{ijk} = 2^{-2q(J+1)} \lambda_q^2 n^{-1} \sum_{i=1}^{n} \int_{-\pi}^{\pi} |f_i^{(q)}(\omega)|^2 d\omega + O(2^{-2q(J+1)}).
\]

For the second term in (A.25), by Lemma A.1(vii) and Assumption 8, we have
\[
(A.27) \quad n^{-1} \sum_{i=1}^{n} \sum_{j=0}^{J} \sum_{k=1}^{2^l} (E\alpha_{ijk} - \alpha_{ijk})^2 \leq 4Cn^{-1} \sum_{i=1}^{n} T_i^{-2} \sum_{h=1}^{T_i} \sum_{m=1}^{T_i} |hR_i(h)R_i(h)\hat{\phi}_{jk}(2\pi h)| = O((J+1)/T^2).
\]

Finally, we consider the first term in (A.24), the variance factor. We write
\[
n^{-1} \sum_{i=1}^{n} EQ(\hat{f}_i, \hat{f}_i) = n^{-1} \sum_{i=1}^{n} \sum_{h=1}^{T_i-1} \sum_{m=1}^{T_i-1} b_{ij}(h, m) \text{cov}[R_i(h), \hat{R}_i(m)]
\]
\[
= n^{-1} \sum_{i=1}^{n} \sum_{h=1}^{T_i-1} \sum_{m=1}^{T_i-1} b_{ij}(h, m)T_i^{-1} \sum_{l=1}^{\infty} \left[ 1 - \frac{\eta(l) + m - \eta(l)}{T_i} \right]
\]
\[
\times [R(l)R_{l}(l + m - h) + R(l + m)R_{l}(l - h) + \kappa(l, h, m - h)]
\]
\[
= \Omega_{1nT} + \Omega_{2nT} + \Omega_{3nT}, \quad \text{say},
\]
where $\eta(l) = l$ if $l > 0$, $\eta(l) = 0$ if $h - m \leq l \leq 0$, and $\eta(l) = -l + h - m$ if $-(T_i - h) + 1 \leq l \leq h - m$. Cf. Priestley (1981, p. 326). Given Assumption 6 and Lemma A.1(vii), we have $|\Omega_{2nT}| \leq C(J+1)$ and $|\Omega_{3nT}| \leq C(J+1)$. For the first term $\Omega_{1nT}$, we can write
\[
\Omega_{1nT} = n^{-1} \sum_{i=1}^{n} \sum_{h=1}^{T_i-1} b_{ij}(h, h)T_i^{-1} \sum_{l=1}^{\infty} \left[ 1 - \frac{h}{T_i} \right] R_l^2(l)
\]
\[
+ n^{-1} \sum_{i=1}^{n} \sum_{h=1}^{T_i-1} \sum_{r=1}^{T_i-1} b_{ij}(h, h + r)T_i^{-1} \sum_{l=1}^{\infty} \left[ 1 - \frac{\eta(l) + h - r}{T_i} \right] R_l(l)R_{l}(l + r)
\]
\[ n^{-1} \sum_{i=1}^{n} T_i^{-1} (2^{i+1} - 1) \sum_{h=-\infty}^{\infty} R_i^2(h) + O((J+1)/T), \]

where we have used Lemma A.1(v) for the first term, which corresponds to \( h = m \); the second term corresponds to \( h \neq m \) and it is \( O((J+1)/T) \) uniformly in \( i \) given \( \sum_{h=-\infty}^{\infty} |R_i(h)| \leq C \) and Lemma A.1(v). It follows that as \( J \rightarrow \infty \),

\[
(A.28) \quad n^{-1} \sum_{i=1}^{n} E Q(f_i, E f_i) = \frac{2 \lambda^{i+1}}{T} n^{-1} \sum_{i=1}^{n} \int_{-\pi}^{\pi} f_i^2(\omega) d\omega + o(2^J/T).
\]

Collecting (A.24)–(A.28) and \( J \rightarrow \infty \), we obtain

\[
n^{-1} \sum_{i=1}^{n} E Q(\hat{f}_i, f_i) = \frac{2 \lambda^{i+1}}{T} n^{-1} \sum_{i=1}^{n} \int_{-\pi}^{\pi} f_i^2(\omega) d\omega + 2^{-2q(i+1)} \lambda_q n^{-1} \sum_{i=1}^{n} \int_{-\pi}^{\pi} [f_i^{(q)}(\omega)]^2 d\omega + o(2^J/T + 2^{-2q})
\]

This completes the proof of Theorem 6. \( Q.E.D. \)

**Proof of Corollary 1:** The result follows from Theorem 5 because Assumption 9 implies

\[ 2^J/2^J - 1 = o_P(T^{-1/(2(2q+1))}) = o_P(2^{-J/2}), \]

where the nonstochastic finest scale \( J \) is given by \( 2^{J+1} \equiv \max\{2\alpha \lambda_q^{2\kappa}(q) T^{1/2(2q+1)}, 0\} \). The latter satisfies the conditions of Theorem 5. \( Q.E.D. \)

**REFERENCES**


