Syracuse University
SURFACE

Center for Policy Research

Maxwell School of Citizenship and Public Affairs

2000

Wavelet-Based Testing for Serial Correlation of Unknown Form in Panel Models

Chihwa Kao
Syracuse University. Center for Policy Research, cdkao@maxwell.syr.edu

Follow this and additional works at: https://surface.syr.edu/cpr

Part of the Mathematics Commons

Recommended Citation

Kao, Chihwa, "Wavelet-Based Testing for Serial Correlation of Unknown Form in Panel Models" (2000). *Center for Policy Research*. 123.
https://surface.syr.edu/cpr/123

This Working Paper is brought to you for free and open access by the Maxwell School of Citizenship and Public Affairs at SURFACE. It has been accepted for inclusion in Center for Policy Research by an authorized administrator of SURFACE. For more information, please contact surface@syr.edu.

ISSN 1525-3066

Center for Policy Research
Working Paper No. 32

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

Yongmiao Hong and Chihwa Kao

Center for Policy Research
Maxwell School of Citizenship and Public Affairs
Syracuse University
426 Eggers Hall
Syracuse, New York 13244-1020
(315) 443-3114 | Fax (315) 443-1081
e-mail: ctrpol@syr.edu

October 2000

$5.00

Up-to-date information about CPR's research projects and other activities is available from our World Wide Web site at **www-cpr.maxwell.syr.edu**. All recent working papers and Policy Briefs can be read and/or printed from there as well.

CENTER FOR POLICY RESEARCH – Fall 2000

Timothy M. Smeeding, Director
Professor of Economics & Public Administration

Associate Directors

Margaret M. Austin
Associate Director,
Budget and Administration

Douglas Holtz-Eakin
Chair, Professor of Economics
Associate Director, Center for Policy Research

Douglas Wolf
Professor of Public Administration
Associate Director, Aging Studies Program

John Yinger
Professor of Economics and Public Administration
Associate Director, Metropolitan Studies Program

SENIOR RESEARCH ASSOCIATES

Scott Allard .... Public Administration
Dan Black .... Economics
Stacy Dickert-Conlin .... Economics
William Duncombe .... Public Administration
Gary Engelhardt .... Economics
Deborah Freund .... Public Administration
Vernon Greene .... Public Administration
Leah Gutierrez .... Public Administration
Madonna Harrington Meyer .... Sociology
Christine Himes .... Sociology
Jacqueline Johnson .... Sociology
Bernard Jump .... Public Administration
Duke Kao .... Economics
Eric Kingson .... Social Work
Thomas Kniesner .... Economics
Jeff Kubik .... Economics
Jerry Miner .... Economics
John Moran .... Economics
Jan Ondrich .... Economics
John Palmer .... Public Administration
Lori Ploutz-Snyder .... Health and Physical Education
Grant Reeher .... Political Science
Stuart Rosenthal .... Economics
Jodi Sandfort .... Public Administration
Michael Wasylenko .... Economics
Assata Zerai .... Sociology

GRADUATE ASSOCIATES

Reagan Baughman .... Economics
Robert Bifulco .... Public Administration
Caroline Bourdeaux .... Public Administration
Christine Caffrey .... Sociology
Christopher Cunningham .... Economics
Tae Ho Eom .... Public Administration
Seth Giertz .... Economics
Andrzej Grodner .... Economics
Rain Henderson .... Public Administration
Pam Herd .... Sociology
Lisa Hotchkiss .... Public Administration
Peter Howe .... Economics
Benjamin Johns .... Public Administration
Kwangho Jung .... Public Administration
James Laditka .... Public Administration
Xiaoli Liang .... Economics
Donald Marples .... Economics
Neddy Matshalaga .... Sociology
Megan Myers-Hayes .... Public Administration
Chada Phuengphinij .... Economics
Suzanne Plourde .... Economics
Nora Ranney .... Public Administration
Mehmet Serkan Tosun .... Economics
Mark Trembley .... Public Administration
Rebecca Walker .... Economics
James Williamson .... Economics

STAFF

JoAnna Berger .... Receptionist
Martha Bonney .... Publications and Events Coordinator
Karen Cimilluca .... Librarian/Office Coordinator
Kati Foley .... Administrative Assistant, LIS
Esther Gray .... Administrative Secretary
Kitty Nasto .... Administrative Secretary
Denise Paul .... Editorial Assistant, *NTJ*
Mary Santy .... Administrative Secretary
Amy Storfer-Isser .... Computer Support, Microsim Project
Debbie Tafel .... Secretary to the Director
Ann Wicks .... Administrative Secretary
Lobrenzo Wingo .... Computer Consultant

ABSTRACT

Wavelet analysis is a new mathematical tool developed as a unified field of science over the last decade. As spatially adaptive analytic tools, wavelets are useful for capturing serial correlation where the spectrum has peaks or kinks, as can arise from persistent/strong dependence, seasonality or use of seasonal data such as quarterly and monthly data, business cycles, and other kinds of periodicity. This paper proposes a new class of wavelet-based tests for serial correlation of unknown form in the estimated residuals of an error component model, where the error components can be one-way or two-way, the individual and time effects can be fixed or random, the regressors may contain lagged dependent variables or deterministic/stochastic trending variables. The proposed tests are applicable to unbalanced heterogenous panel data. They have a convenient null limit N(0,1) distribution. No formulation of an alternative is required, and the tests are consistent against serial correlation of unknown form. We propose and justify a data-driven method to choose the finest scale parameter—the smoothing parameter in wavelet spectral estimation, making the test procedure completely operational for any finite sample. The data-driven finest scale, in an automatic manner, converges to zero under the null hypothesis of no serial correlation and grows to infinity as the sample size increases under the alternative, ensuring the consistency of the proposed tests. Simulation studies show that the new tests perform rather well in small and finite samples in comparison with some existing popular tests for panel models and can be used as an effective evaluation procedure for panel models.

Key Words: Error component, Panel model, Hypothesis testing, Serial correlation of unknown form, Spectral peak, Unbalanced panel data, Wavelet

1 Introduction

Increasingly available panel data have been widely used in economics and econometrics. They often provide insights not available in pure time-series or cross-sectional data (e.g., Baltagi 1995, Granger 1996, Hsiao 1986). This paper proposes a new class of wavelet-based consistent tests for serial correlation of unknown form in panel data models. It is important to test serial correlation of unknown form in panel models because the existence of serial correlation will invalidate conventional tests such as t-tests and F-tests which use standard covariance estimators of parameter estimators, and will indicate model misspecification when the regressors include lagged dependent variables. Moreover, the choice of estimation methods may depend upon whether there exists serial correlation in the errors of panel models. When the errors are serially correlated in panel data models, for example, the computation of MLE (e.g., Anderson and Hsiao 1982, Hsiao 1986, Binder, Hsiao and Pesaran 1999) and GMM (e.g., Blundell and Bond 1998) could be rather complicated, and the feasible GLS estimator will be invalid or have to be modified substantially (e.g., Baltagi and Li 1991). Some inference procedures, such as Breusch and Pagan's (1980) tests for random effects, also assume serial uncorrelatedness in the errors of panel data models.

There have been some tests for serial correlation in panel models. Bhargava, Franzini, and Narendranathan (1982) extended Durbin and Watson's (1951) test to static panel models. Breusch and Pagan (1980) propose an LM test for first order serial correlation, assuming no random effects in an error component model. Baltagi and Li (1991) propose a joint LM test for first order serial correlation and random effects. Baltagi and Li (1995) further propose a class of LM tests for first order serial correlation, allowing the presence of random effects or fixed individual effects. Bera, Sosa-Escudero and Yoon (2000) also propose a convenient OLS-based test for first order serial dependence by modifying Baltagi and Li's (1995) LM tests. Li and Hsiao (1998) propose tests for first order and higher order serial correlation for a semiparametric partially linear panel model.

All of the existing tests for serial correlation in panel models assume a known form of serial correlation, e.g., an AR(1) or MA(1) model. These tests have optimal power when the data generating process coincides with the assumed model. They also have good power against many other alternatives. However, they are not consistent (i.e., do not have asymptotic power 1) against serial correlation of unknown form. From both theoretical and practical points of view, it is useful to test serial correlation of unknown form because prior information about the alternative is usually not available in practice. This is true particularly in the panel context because there may exist significant nonhomogeneity in the degree of serial correlation across individuals (e.g., Choi 1999). Moreover, as Granger and Newbold (1977, p. 92) pointed out, the first few lags of OLS residuals of linear dynamic models often appear to behave like a white noise even under model misspecification, due to the very nature of the OLS estimation. It is therefore important to check serial correlation at higher order lags. Little effort has been made on specification and evaluation of dynamic panel models (cf. Granger 1996). Our tests can be used as an evaluation procedure for dynamic linear panel models. A recent attempt at linear dynamic panel model specification is Hjellvik and Tjøstheim's (1999) order determination procedure.

Wavelets are newly developed mathematical tools alternative to the Fourier transform. They are spatially adaptive analytic tools particularly useful for capturing nonsmooth features such as singularities and nonhomogeneity (e.g., Donoho and Johnstone 1994,1995a,1995b, Donoho, Johnstone and Kerkyacharian 1996, Gao 1997, Hong and Lee 2000, Jensen 2000, Neumann 1996, Lee and Hong 2000, Ramsey 1999 and the references therein, and Wang 1995). Many economic and financial time series have a spectrum with peaks and kinks, as can arise from, for example, persistent/strong dependence, business cycles, seasonality or use of seasonal data such as quarterly and monthly data, as well as other kinds of periodicity (e.g., Bizer and Durlauf 1992, Granger 1969, Watson 1993). In the panel context, there may also exist significant nonhomogeneity in the degree of serial correlation across different individuals. Wavelets are ideal tools in these contexts. In this paper we use wavelets to test serial correlation in the estimated residuals of panel data models. The panel model, which can be one-way or two-way, covers both balanced and unbalanced panel data; the individual and time effects can be fixed or random; the regressors may contain lagged dependent variables or deterministic/stochastic trending variables; and there is no need to require a specific method for parameter estimation or to know the limiting distribution of parameter estimators. The proposed tests have a convenient limit N(0,1) distribution, no matter whether the 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, model parameter estimation has no impact on the limit distribution of the proposed test statistics when applied to dynamic panel models. One can proceed as if model parameters were known and were equal to the estimates. Unlike the existing popular tests for serial correlation in panel models, we do not require formulation of an alternative model (e.g., AR(1) or MA(1)), and our tests are consistent against serial correlation of unknown form. We note that no consistent test for serial correlation of unknown form was available for panel models.

This paper is a substantive extension, in asymptotic analysis, context and results, of Lee and Hong (2000), who consider a wavelet test for serial correlation in observed raw pure time-series data (i.e., not the estimated residual of a time-series model). First, as is well known in the panel literature (cf. Phillips and Moon 1999, Hahn and Kuersteiner 2000), asymptotic analysis in the double-indexed panel context is much more involved than in pure time-series analysis. Our asymptotic theory holds for both large n and large T, where n is the number of individuals and T is an index for the number of time-series observations. Increasing effort has been devoted to the study of panel models with both large n and large T, due to the growing use of cross-country data over time to study growth convergence, international R&D spillover and purchasing power parity. A distinct feature of our asymptotic analysis is that we treat both $n \to \infty$ and $T \to \infty$ *simultaneously*, which complements Phillips and Moon (1999) and Hahn and Kuersteiner's (2000) joint limit theory for panel models. Our main theory does not require the ratio n/T goes to 0 or a constant. As noted earlier, our asymptotic theory shows that the use of the estimated residuals from a possibly nonstationary panel model rather than the unobservable error series has no impact on the limit distribution of the proposed test statistics. In addition, we find several interesting features that are not available in pure time-series analysis. Most remarkably, the limit N(0,1) distribution of the proposed test statistics is obtained without having to require the finest scale parameter—the smoothing parameter 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 (2000), where it is required that $J \to \infty$ as $T \to \infty$ to achieve asymptotic normality under the null hypothesis. We further develop and justify a data-driven method to choose a suitable finest scale, making the proposed tests completely operational in practice. The data-driven finest scale, in an automatic manner, converges to 0 under the null hypothesis of no serial correlation and grows to ∞ under the alternative, ensuring consistency against serial correlation of unknown form. This method is not available elsewhere in the literature and has its own rights in wavelet spectral estimation. We also find that a heteroskedasticity-corrected test may be less powerful than a heteroskedasticity-consistent test. This is in contrast to the well-known result in the context of estimation that heteroskedasticity-corrected estimators (e.g., feasible GLS) are more efficient than heterokedasticity-consistent estimators (e.g., OLS). Our tests work well for sample sizes often encountered in economics, but we emphasize that they are best viewed as complements to rather than competitors of the existing popular tests for serial correlation in panel data models, because each test has its own attractive merits.

The organization of the paper is as follows. We describe the model and hypotheses of interest in Section 2, introduce wavelets and test statistics in Section 3, derive the asymptotic distributions for these tests in Section 4, and establish their consistency in Section 5. Section 6 proposes and justifies a data-driven method to choose a finest scale. Section 7 presents a simulation study on the finite sample performance of the proposed tests in comparison with some existing popular tests for panel models. Section 8 concludes. All proofs are in the Appendix.

Throughout, we use $\|A\|$ to denote the Euclidean norm $[tr(A'A)]^{1/2}$; $\xrightarrow{d}$ and $\xrightarrow{p}$ to denote the convergence in distribution and in probability; A^* and $\mathrm{Re}(A)$ to denote the complex conjugate and the real part of A; $\mathbb{Z} \equiv \{0, \pm 1, ...\}$ and $\mathbb{Z}^+ \equiv \{0, 1, ...\}$ to denote the set of integers and the set of nonnegative integers; and c and C to denote some generic bounded constants that do not depend on any other index, with $0 < c < C < \infty$. Unless indicated explicitly, all limits are taken as both $n \to \infty$ and $T \to \infty$. A GAUSS program for computing the proposed test statistics is available from the authors upon request. The user only needs to supply estimated residuals.

2 The Framework

Consider a panel data model

$$Y_{it} = \alpha + X_{it}^{'}\beta + \mu_i + \lambda_t + v_{it}, \qquad t = 1, ..., T_i, i = 1, ..., n, \quad n, T_i \in \mathbb{Z}^+, \tag{2.1}$$

where Y_{it} is a scalar, X_{it} is a $p \times 1$ vector of explanatory variables that may contain lagged dependent variables Y_{it-h} $(p, h \in \mathbb{Z}^+)$, α is an intercept, β is a $p \times 1$ vector of the slope parameters, μ_i is the individual effect, λ_t is the time effect, and v_{it} is the error term. We allow fixed effects or random effects. Throughout, we assume $T_i = c_i T$ for some integer T and $c_i \in [c, C]$. Thus, we permit unbalanced panel data. Moreover, we allow Y_{it}, X_{it}, α and β to depend on both n and T. (For notational simplicity, we suppress such dependence.)

Throughout, we assume the following conditions on (2.1):

Assumption 1 $\{Y_{it}, X_{it}'\}'$ *are stochastic processes such that (i) for each* i, $\{v_{it}\}$ *is covariance-stationary with* $E(v_{it}) = 0$, $E(v_{it}^2) = \sigma_i^2 \in [c, C]$ *and* $E(v_{it}^8) \in [c, C]$; *(ii) there is no spatial dependence in* $\{v_{it}\}$, *i.e.,* v_{it} *and* v_{js} *are independent for all* $i \neq j$ *and all* t, s; *(iii) the individual and time effects,* μ_i *and* λ_t, *can be stochastic (random effects) or deterministic (fixed effects).*

No dependence assumptions on $\{\mu_i\}$ and $\{\lambda_t\}$ are imposed, because they will be differenced out in the construction of our test statistics. We thus allow $\{\lambda_t\}$ to be serially correlated if λ_t is random, and $\{\mu_i\}$ to be spatially correlated if μ_i is random. We also allow a certain degree of heterogeneity in panel data—$\{Y_{it}, X_{it}'\}'$ need not be stationary for each i, and the errors $\{v_{it}\}$ may have different variances across i. In particular, we allow some nonstationary processes. One example of nonstationary panel time series is the deterministic trend process (e.g., Kao and Emerson 1999)

$$Y_{it} = \alpha + \gamma_1 t + \gamma_2 t^2 + \cdots + \gamma_p t^p + \mu_i + \lambda_t + v_{it}. \tag{2.2}$$

This is covered by (2.1) with $X_{it} \equiv [t/T, (t/T)^2, ..., (t/T)^p]'$ and $\beta \equiv (T\gamma_1, ..., T^p\gamma_p)'$. Note that X_{it} and β depend on T. Another example is the panel cointegration process (e.g., Phillips and Moon 1999, Kao and Chiang 2000):

$$Y_{it} = \alpha + \gamma Z_{it} + \mu_i + \lambda_t + v_{it}, \tag{2.3}$$

where $Z_{it} = Z_{it-1} + \varepsilon_{it}, \{\varepsilon_{it}\}$ is $I(0)$ for each i, and $\{\varepsilon_{it}\}$ may or may not be correlated with $\{v_{it}\}$. This process is also covered by (2.1) with $X_{it} \equiv T^{-1}Z_{it}$ and $\beta \equiv T\gamma$. We will provide regularity conditions on transformed variables $\{X_{it}\}$ and transformed parameters β.

The parameter vector β in (2.1) can be estimated by the popular within estimator

$$\begin{aligned}\hat{\beta} \equiv & \left[\sum_{i=1}^{n}\sum_{t=1}^{T_i}\left(X_{it} - \bar{X}_i - \bar{X}_t + \bar{X}\right)\left(X_{it} - \bar{X}_i - \bar{X}_t + \bar{X}\right)'\right]^{-1} \\ & \times \left[\sum_{i=1}^{n}\sum_{s=1}^{T_i}\left(X_{it} - \bar{X}_i - \bar{X}_t + \bar{X}\right)\left(Y_{it} - \bar{Y}_i - \bar{Y}_t + \bar{Y}\right)\right],\end{aligned} \tag{2.4}$$

where $\bar{X}_i \equiv T_i^{-1}\sum_{t=1}^{T_i} X_{it}, \bar{X}_t \equiv n^{-1}\sum_{i=1}^{n} X_{it}$ and $\bar{X} \equiv (nT_i)^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T_i} X_{it}$. The variables $\bar{Y}_i, \bar{Y}_t$ and $\bar{Y}$ are defined in the same ways. In empirical applications one often uses the following standard covariance estimator of $\hat{\beta}$ for confidence interval estimation and hypothesis testing:

$$\hat{\Omega}_{\hat{\beta}} \equiv \sigma_v^2 \left[\sum_{i=1}^{n}\sum_{t=1}^{T_i}\left(X_{it}-\bar{X}_i-\bar{X}_t+\bar{X}\right)\left(X_{it}-\bar{X}_i-\bar{X}_t+\bar{X}\right)'\right]^{-1},$$

where $\sigma_v^2 \equiv \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n}\sigma_i^2$. This covariance estimator is valid when $\{v_{it}\}$ is homoskedastic and serially uncorrelated, among other things. The existence of serial correlation, of any form, will generally invalidate the covariance estimator and thus the inferences based on it. In particular, the conventional t-tests and F-tests will be misleading. (New procedures using heteroskedasticity and autocorrelation consistent covariance estimators of parameter estimators are now available, but these tests are usually over-sized in finite samples even when there exists no serial correlation. Thus one may like to first check serial correlation to see if conventional tests can be used.) On the other hand, when the regressors X_{it} contain lagged dependent variables Y_{it-h} for $h > 0$, serial correlation will render inconsistent the within estimator $\hat{\beta}$ for β, because the orthogonality condition $E(X_{it}v_{it}) = 0$ will not hold in general.

In this paper we are interested in testing whether $\{v_{it}\}$ is serially correlated. Suppose that the covariance-stationary process $\{v_{it}\}$ has the autocovariance function $R_i(h) \equiv E(v_{it}v_{it-|h|})$, $h \in \mathbb{Z}$ and $i \in \mathbb{Z}^+$, and power spectrum

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

The hypotheses of interest are

$$\mathbb{H}_0: \quad R_i(h) = 0 \text{ for all } h \neq 0 \text{ and all } i$$

versus

$$\mathbb{H}_A: \quad R_i(h) \neq 0 \text{ at least for some } h \neq 0 \text{ and some } i.$$

The alternative hypothesis $\mathbb{H}_A$ allows some (but not all) of the individual series to be white noises. More generally, there may exist substantial nonhomogeneity in the degree of serial correlation across i under $\mathbb{H}_A$. It is highly desirable to develop powerful procedures against $\mathbb{H}_A$, because prior information about the alternative is usually not available.

Both the autocovariance function $R_i(h)$ and the spectral density $f_i(\omega)$ are Fourier transforms of each other; they contain the same amount of information on serial correlation of $\{v_{it}\}$. One can use $R_i(h)$ or $f_i(\omega)$ to test $\mathbb{H}_0$ versus $\mathbb{H}_A$. All the existing tests for serial correlation for panel models are based on $R_i(h)$, assuming a common model with some prespecified lags h (e.g., AR(1) and MA(1)). In this paper we use a spectral approach. Spectral analysis is often used in economic and econometric analysis (e.g., Bizer and Durlauf 1990, Durlauf 1991, Granger 1969, Watson 1993). It is a natural and convenient approach to testing serial correlation of unknown form, because $f_i(\omega)$ contains information on serial correlation at *all* lags. Under $\mathbb{H}_0$, $f_i(\omega)$ becomes $f_{i0}(\omega) \equiv (2\pi)^{-1}\sigma_i^2$ for all $\omega \in [-\pi, \pi]$. Under $\mathbb{H}_A$, we have $f_i(\omega) \neq (2\pi)^{-1}\sigma_i^2$ at least for some i. Thus, a consistent test for $\mathbb{H}_0$ versus $\mathbb{H}_A$ can be formed by comparing consistent estimators for $f_i(\omega)$ and $f_{i0}(\omega)$. We will use wavelets to estimate $f_i(\omega)$, which are particularly suitable for economic and financial time series with spectral peaks and kinks.

3 Wavelet Method

3.1 Wavelets

We first review wavelet analysis briefly. The essence of wavelet analysis is to expand a given 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, Ramsey 1999). Many economic and financial time series have spectral peaks or kinks, due to strong dependence, business cycles, seasonality or use of seasonal data such as quarterly and monthly data, and other kinds of periodicity. Wavelets are quite suitable in these contexts.

We first impose a standard condition on the mother wavelet $\psi(\cdot)$.

Assumption 2 *$\psi : \mathbb{R} \to \mathbb{R}$ 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

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

constitutes an orthonormal basis for $L_2(\mathbb{R})$, the space of square-integrable functions on $\mathbb{R}$ (cf. 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). This simultaneous time-frequency decomposition is the key to wavelet analysis, explaining why it is attractive for approximating nonsmooth functions.

Assumption 2 ensures that the Fourier transform of $\psi(\cdot)$,

$$\hat{\psi}(z) \equiv (2\pi)^{-1/2} \int_{-\infty}^{\infty} \psi(x)e^{-\mathbf{i}zx}dx, \qquad z \in \mathbb{R}, \tag{3.2}$$

exists and is continuous in z almost everywhere. Note that $\hat{\psi}(0) = (2\pi)^{-1/2} \int_{-\infty}^{\infty} \psi(x)dx = 0$, which implies that $\psi(\cdot)$ must have alternating signs. This is one of the characteristic properties of wavelets and one reason why wavelets are sensitive to changes or singularities.

The mother wavelet $\psi(\cdot)$ can have bounded or unbounded support. A well-known compactly supported wavelet is the Haar wavelet,

$$\psi(x) = \begin{cases} 1, & 0 \leq x < 1/2 \\ -1, & 1/2 \leq x < 1 \\ 0 & \text{otherwise.} \end{cases} \tag{3.3}$$

An example of a wavelet with unbounded support is the Shannan wavelet

$$\psi(x) = -2\frac{\sin(2\pi x) + \cos(\pi x)}{\pi(2x+1)}, \qquad x \in \mathbb{R}. \tag{3.4}$$

We impose a condition to ensure that $\hat{\psi}(z) \to 0$ sufficiently fast as $z \to \infty$.

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

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

$$\hat{\psi}(z) = e^{\mathbf{i}z/2}(2\pi)^{-1/2}\frac{\sin^4(z/4)}{(z/4)^2}\left[\frac{P_3(z/4+\pi/4)}{P_3(z/2)P_3(z/4)}\right]^{1/2}, \tag{3.5}$$

where $P_3(z) \equiv \frac{2}{3} + \frac{1}{3}\cos(2z)$. Another member is the second order spline wavelet, with

$$\hat{\psi}(z) = -\mathbf{i}e^{\mathbf{i}\omega/2}(2\pi)^{-1/2}\frac{\sin^6(z/4)}{(z/4)^3}\left[\frac{P_5(z/4+\pi/4)}{P_5(z/2)P_5(z/4)}\right]^{1/2}. \tag{3.6}$$

where $P_5(z) \equiv \frac{1}{30}\cos^2(2z) + \frac{13}{30}\cos(2z) + \frac{8}{15}$. Both the Franklin wavelet and the second order spline wavelet have compact support in the time domain and an exponential decay in the frequency domain (e.g., Henández and Weiss 1996, p.149). In fact, the Harr wavelet in (3.3) is the 0-th order spline wavelet. However, it does not satisfy Assumption 3 because its $\hat{\psi}(z) = -\mathbf{i}e^{\mathbf{i}z/2}(2\pi)^{-1/2}\sin^2(z/4)/(z/4)$ decays to 0 as $|z| \to \infty$ at the rate of $|z|^{-1}$ only.

3.2 Wavelet Representation of a Spectrum

We now consider wavelet representation of spectral density $f_i(\cdot)$ of $\{v_{it}\}$. Since $f_i(\cdot)$ is 2π-periodic, it is not square-integrable over $\mathbb{R}$. We need to construct a wavelet basis $\{\Psi_{jk}(\cdot)\}$ for $L_2[-\pi,\pi]$, the space of 2π-periodic functions on $[-\pi,\pi]$. Given an orthonormal wavelet basis $\{\psi_{jk}(\cdot)\}$ for $L_2(\mathbb{R})$, we can construct an orthonormal wavelet basis $\{\Psi_{jk}(\cdot)\}$ for $L_2[-\pi,\pi]$, where

$$\Psi_{jk}(\omega) \equiv (2\pi)^{-1/2}\sum_{m=-\infty}^{\infty}\psi_{jk}\left(\frac{\omega}{2\pi}+m\right), \qquad \omega \in [-\pi,\pi]. \tag{3.7}$$

See, e.g., Daubechies (1992, Ch.9) or Hernández and Weiss (1996, Ch.4) for more discussion. Since (3.7) is an infinite sum, it is convenient to use compactly supported wavelets so that only a finite number of terms are nonzero. Alternatively, if $\hat{\psi}(\cdot)$ has bounded support, one can compute $\Psi_{jk}(\cdot)$ from its Fourier transform via the formula

$$\Psi_{jk}(\omega) = (2\pi)^{-1/2} \sum_{h=-\infty}^{\infty} \hat{\Psi}_{jk}(h) e^{\mathbf{i}h\omega}, \tag{3.8}$$

where $\hat{\Psi}_{jk}(h) \equiv (2\pi)^{-1/2} \int_{-\pi}^{\pi} \Psi_{jk}(\omega) e^{-\mathbf{i}h\omega} d\omega$. By (3.7) and change of variables, we have

$$\hat{\Psi}_{jk}(h) = (2\pi)^{1/2} \hat{\psi}_{jk}(2\pi h) = e^{-\mathbf{i}2\pi hk/2^j} (2\pi/2^j)^{1/2} \hat{\psi}(2\pi h/2^j). \tag{3.9}$$

Note that the dilation parameter j varies dyadically and the translation parameter k varies as the modulation.

Lee and Hong (2000) show that the spectral density $f_i(\cdot)$ in (2.3) can be expressed as

$$f_i(\omega) = (2\pi)^{-1}\sigma_i^2 + \sum_{j=0}^{\infty} \sum_{k=1}^{2^j} \alpha_{ijk} \Psi_{jk}(\omega), \qquad \omega \in [-\pi, \pi], \tag{3.10}$$

where the wavelet coefficient

$$\alpha_{ijk} \equiv \int_{-\pi}^{\pi} f_i(\omega) \Psi_{jk}(\omega) d\omega. \tag{3.11}$$

The real-valued coefficients $\{\alpha_{ijk}\}$ are the orthogonal projections of $f_i(\cdot)$ on wavelet bases $\{\Psi_{jk}(\cdot)\}$. Unlike the Fourier transforms, the wavelet coefficient α_{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 a powerful tool for capturing nonsmooth features.

By Parseval's identity and (3.8), we can also express α_{ijk} in time domain, namely,

$$\alpha_{ijk} = (2\pi)^{-1/2} \sum_{h=-\infty}^{\infty} R_i(h) \hat{\Psi}_{jk}^*(h) = \sum_{h=-\infty}^{\infty} R_i(h) \hat{\psi}_{jk}^*(2\pi h), \tag{3.12}$$

where $\{\hat{\psi}_{jk}(\cdot)\}$ is given in (3.9). Note that $\{\alpha_{ijk}\}$ do not represent autocorrelations at different lags. They are weighted averages of autocorrelations centered at varying locations.

3.3 Wavelet Spectral Density Estimator

Suppose that we have an $\sqrt{nT}$-consistent estimator $\hat{\beta}$ for β. Put

$$\hat{v}_{it} \equiv \hat{u}_{it} - \bar{u}_i - \bar{u}_t + \bar{u}, \qquad t = 1, ..., T_i, i = 1, ..., n, \tag{3.13}$$

where $\hat{u}_{it} \equiv Y_{it} - X_{it}'\hat{\beta}, \bar{u}_i \equiv T_i^{-1} \sum_{t=1}^{T_i} \hat{u}_{it}, \bar{u}_t \equiv n^{-1} \sum_{i=1}^{n} \hat{u}_{it}, \bar{u} \equiv (nT_i)^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T_i} \hat{u}_{it}$. Note that it is not necessary to center $\hat{u}_{it}$ using an intercept estimator $\hat{\alpha}$. The demeaned residual $\hat{v}_{it}$ has a zero sample mean. When $\hat{\beta}$ is the within estimator in (2.4), $\hat{v}_{it}$ is the well-known within residual. Of course, we do not requite $\hat{\beta}$ to be the within estimator.

Now, we define the sample autocovariance function of $\{\hat{v}_{it}\}$

$$\hat{R}_i(h) \equiv T_i^{-1} \sum\nolimits_{t=|h|+1}^{T_i} \hat{v}_{it}\hat{v}_{it-|h|}, \qquad h = 0, \pm 1, ..., \pm(T_i - 1). \tag{3.14}$$

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

$$\hat{f}_i(\omega) \equiv (2\pi)^{-1}\hat{R}_i(0) + \sum_{j=0}^{J_i}\sum_{k=1}^{2^j} \hat{\alpha}_{ijk}\Psi_{jk}(\omega), \qquad \omega \in [-\pi, \pi], \tag{3.15}$$

where the empirical wavelet coefficient

$$\hat{\alpha}_{ijk} \equiv \sum_{h=1-T_i}^{T_i-1} \hat{R}_i(h)\hat{\psi}^*_{jk}(2\pi h), \tag{3.16}$$

and $J_i \equiv J_i(T_i)$ is the finest scale corresponding to the highest resolution level used in (3.15). The degree of approximation (or bias) depends on J_i. The larger J_i is, the smaller the bias of $\hat{f}_i(\cdot)$. We must let $J_i \to \infty$ as $T_i \to \infty$ to achieve consistency of $\hat{f}_i(\cdot)$ for $f_i(\cdot)$. On the other hand, the larger J_i is, the larger the variance of $\hat{f}_i(\cdot)$. To ensure that the variance of $\hat{f}_i(\cdot)$ is asymptotically negligible, J_i cannot grow too fast. Given each T_i, a suitable J_i should be chosen to balance the variance and the squared bias so that $\hat{f}_i(\cdot)$ will be consistent for $f_i(\cdot)$ as $T_i \to \infty$. There are totally $\sum_{j=0}^{J_i} 2^j = 2^{J_i+1} - 1$ empirical wavelet coefficients $\{\hat{\alpha}_{ijk}\}$ in (3.15). Thus, the finest scale J_i should be smaller than $\log_2(T_i + 1) - 1$. Proper conditions on J_i will be given to ensure that the proposed test statistics have a well-defined limit distribution. We allow a different J_i for each i. This is useful from a theoretical point of view because the degree of serial correlation may vary substantially across i. We will also propose an automatic data-driven method to choose J_i, which lets data themselves determine a proper J_i given each finite T_i. The data-driven finest scale, in an automatic manner, converges to 0 under $\mathbb{H}_0$ and grows to ∞ under $\mathbb{H}_A$, thus ensuring consistency against $\mathbb{H}_A$. See Section 6 for more discussion.

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 2π-periodic functions $f_1(\cdot)$ and $f_2(\cdot)$. To construct our tests, we use the quadratic form

$$Q(\hat{f}_i; \hat{f}_{i0}) = \sum_{j=0}^{J_i}\sum_{k=1}^{2^j} \hat{\alpha}^2_{ijk}, \tag{3.17}$$

where $\hat{f}_{i0}(\omega) \equiv (2\pi)^{-1}\hat{R}_i(0)$ and the second equality follows by Parseval's identity and the orthonormality of wavelet bases $\{\Psi_{jk}(\cdot)\}$. Many other divergence measures could also be used, but the quadratic form $Q(\hat{f}_i, \hat{f}_{i0})$ is convenient to compute. In particular, there is no need to calculate numerical integration over $\omega \in [-\pi, \pi]$.

We first consider the test statistic

$$\hat{W}_1 \equiv \left(\sum_{i=1}^{n} 2\pi T_i \sum_{j=0}^{J_i} \sum_{k=1}^{2^j} \hat{\alpha}_{ijk}^2 - \hat{M} \right) / \hat{V}^{1/2}, \tag{3.18}$$

where $\hat{M} \equiv \sum_{i=1}^{n} \hat{R}_i^2(0) M_{i0}$, $\hat{V} \equiv \sum_{i=1}^{n} \hat{R}_i^4(0) V_{i0}$,

$$\begin{aligned} M_{i0} &\equiv \sum_{h=1}^{T_i-1} (1 - h/T_i) b_{J_i}(h,h), \\ V_{i0} &\equiv 4 \sum_{h=1}^{T_i} \sum_{m=1}^{T_i} (1 - h/T_i)(1 - m/T_i) b_{J_i}^2(h,m), \end{aligned}$$

$b_{J_i}(h,m) \equiv 2\,\mathrm{Re}[a_{J_i}(h,m) + a_{J_i}(h,-m)]$, $a_{J_i}(h,m) \equiv 2\pi \sum_{j=0}^{J_i} \sum_{k=1}^{2^j} \hat{\psi}_{jk}(2\pi h) \hat{\psi}_{jk}^*(2\pi m)$ and $\hat{\psi}_{jk}(\cdot)$ is as in (3.9). The standardization factors $\hat{M}$ and $\hat{V}$ are the estimators for the mean and variance of $\sum_{i=1}^{n} 2\pi T_i Q(\hat{f}_i, \hat{f}_{i0})$. The factors $(1 - h/T_i)$ and $(1 - m/T_i)$ are finite sample corrections. Note that $b_{J_i}(h,m)$ is readily computable given function $\hat{\psi}(\cdot)$ and finest scale J_i.

Alternatively, we can also use the test statistic:

$$\hat{W}_2 \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(2\pi T_i \sum_{j=0}^{J_i} \sum_{k=1}^{2^j} \hat{a}_{ijk}^2 - M_{i0} \right) / V_{i0}^{1/2}, \tag{3.19}$$

where $\hat{a}_{ijk} \equiv \sum_{h=1-T_i}^{T_i-1} \hat{\rho}_i(h) \hat{\psi}_{jk}(2\pi h)$ is the normalized empirical wavelet coefficient and $\hat{\rho}_i(h) \equiv \hat{R}_i(h)/\hat{R}_i(0)$ is the sample autocorrelation function of $\{v_{it}\}$.

Intuitively, $\hat{W}_2$ can be viewed as a heteroskedasticity-corrected test while $\hat{W}_1$ is a heteroskedasticity-consistent test, where heteroskedasticity arises from two sources: unequal individual variances σ_i^2 and different finest scales J_i. In $\hat{W}_2$, these two forms of heteroskedasticity are corrected first for each i. As will be shown below, both $\hat{W}_1$ and $\hat{W}_2$ are asymptotically $\mathrm{N}(0,1)$ under $\mathbb{H}_0$. Their power properties, however, generally differ. The heteroskedasticity-robust test $\hat{W}_1$ may be more powerful than the heteroskedasticity-consistent test $\hat{W}_2$ (cf. Section 5). This differs from the well-known result in the context of estimation that correcting heteroskedasticity leads to more efficient estimation (e.g., the feasible GLS is more efficient than OLS).

Our tests $\hat{W}_1$ and $\hat{W}_2$ apply to both one-way and two-way error component models. For one-way component models, however, if one knows $\lambda_t = 0$, then one can use $\hat{v}_{it} \equiv \hat{u}_{it} - \bar{u}_i$. And if one knows $\mu_i = 0$, then one can use $\hat{v}_{it} \equiv \hat{u}_{it} - \bar{u}_t$. The asymptotic distributions of the test statistics remain unchanged, although their finite sample performances may differ.

4 Asymptotic Distribution

We now derive the asymptotic distribution of $\hat{W}_1$ and $\hat{W}_2$ under $\mathbb{H}_0$. We impose the following additional assumptions.

Assumption 4 $\sqrt{nT}(\hat{\beta}-\beta)=O_P(1)$.

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* $h < 0$, $\Gamma_{ixv}(h) \equiv p\lim \tilde{\Gamma}_{ixv}(h)$, $\tilde{X}_{it} \equiv X_{it}-\bar{X}_i-\bar{X}_t+\bar{X}$ *and* $\tilde{v}_{it} \equiv v_{it}-\bar{v}_i-\bar{v}_t+\bar{v}$. *Then (i)* $\sup_{1\leq i\leq n}\sup_{1\leq h<T_i} T_i^{-1}\sum_{t=1}^{T_i}E||\tilde{X}_{it}||^4 \leq C$; *(ii)* $\sup_{1\leq i\leq n}E||\tilde{\Gamma}_{ixv}(h)-\Gamma_{ixv}(h)||^2 \leq CT_i^{-1}$; *(iii)* $\sum_{h=-\infty}^{\infty}||\Gamma_{ixv}(h)|| \leq C$.

We permit but do not require using the popular within estimator $\hat{\beta}$ in (2.4). Other estimators such as OLS, feasible GLS and MLE are allowed as well. Note that the parameter β may be a transformation of the original parameters of interest, as it is in the case with deterministic and stochastic trending regressors in (2.2) and (2.3). Thus, Assumption 4 implies that the estimators of original parameters may converge at a rate faster than $(nT)^{-1/2}$. See Phillips and Moon (1999) and Kao and Chiang (2000) for more discussion.

Theorem 1 *Suppose that Assumptions 1–5 hold and* $\max_{1\leq i\leq n}(2^{2J_i})/(n^2+T)\to 0$ *as* $n\to\infty$ *and* $T\to\infty$. *If* $\{v_{it}\}$ *is i.i.d. for each* i, *then* $\hat{W}_1 \xrightarrow{d} N(0,1)$ *and* $\hat{W}_2 \xrightarrow{d} N(0,1)$.

A remarkable feature of Theorem 1 is that we permit but do not require $J_i\to\infty$ for *any* i; that is, all J_i can be fixed as $n,T\to\infty$ under $\mathbb{H}_0$. This is in sharp contrast to Lee and Hong (2000), who consider a wavelet-based test for serial correlation in observed raw time-series data (i.e., not the estimated residual of a time series model), where it is required that $J\to\infty$ as $T\to\infty$ to achieve asymptotic normality. The reason that all J_i can be fixed in the present panel context is that the additional smoothing provided by n ensures asymptotic normality of $\hat{W}_1$ and $\hat{W}_2$ even if J_i is fixed for all i. 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 with proper mean and variance as $n\to\infty$. This occurs no matter whether $J_i\to\infty$. (Our proof, of course, does not rely on this simple intuition. Instead, we treat both $n\to\infty$ and $T\to\infty$ simultaneously.) In the pure time-series or pure cross-sectional nonparametric literature it is often found (e.g., Skaug and Tjøstheim 1993, Hjellvik, Yao and Tjøstheim 1998) that the normal approximation is inadequate for the finite sample distributions of quadratic forms such as $Q(\hat{f}_i,\hat{f}_{i0})$ with kernel estimators. The latter are usually significantly skewed toward the right tail even when the sample size is rather large. This occurs because the asymptotic normality of quadratic forms for pure time-series data or pure cross-sectional data requires the smoothing parameter to grow or vanish at suitable rates (neither too fast nor too slow) as the sample size increases 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 reasonably well in the panel context. Indeed, our simulation studies show that when $J=0$, both $\hat{W}_1$ and $\hat{W}_2$ have reasonable sizes in finite samples. Most importantly, the fact that J_i may be fixed for all i allows use of data-driven methods that may deliver fixed finest scales under $\mathbb{H}_0$. Sensible data-driven methods may have this feature because the optimal finest scale under $\mathbb{H}_0$ is $J_0 = 0$. We will propose and justify a plug-in method to select a data-driven finest scale for $\hat{W}_1$ and $\hat{W}_2$, which, in an automatic manner, converges to 0 under $\mathbb{H}_0$ and grows to ∞ under $\mathbb{H}_A$, thus ensuring consistency against serial correlation of unknown form. We note that such a data-driven method could not be used for Lee and Hong's (2000) test in pure time series, as it requires $J \to \infty$ under $\mathbb{H}_0$.

Although we require both T and n grow to ∞, we do not impose a restrictive relative speed limit between them. On the other hand, from the proof of Theorem 1 (cf. Theorem A.1 in the appendix), we find that the parameter estimation for β has no impact on the limit distribution of $\hat{W}_1$ and $\hat{W}_2$, no matter whether the regressors X_{it} contain lagged dependent variables or deterministic/stochastic trending variables. Thus, there is no need to use a specific method to estimate β or to know the limit distribution of $\hat{\beta}$. This is in contrast to the tests of Durbin and Watson (1951) and Box and Pierce (1970), whose test statistics or limit distributions should be modified when applied to the estimated residuals of a covariance-stationary dynamic regression model. If the regressors contain deterministic or stochastic trending variables, the limit distributions of these tests will become nonstandard (e.g., Kao and Emerson 1999, Kao and Chiang 2000). Intuitively, although the parameter estimation for β may induce an adjustment of a finite number of degrees of freedom for $\hat{W}_1$ and $\hat{W}_2$, such an adjustment is asymptotically 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 $i = 1, ..., n$, we can use the following simplified versions of test statistics:

$$\tilde{W}_1 = \frac{\sum_{i=1}^{n}\left[2\pi T_i \sum_{j=0}^{J_i}\sum_{k=1}^{2^j}\hat{\alpha}_{ijk}^2 - \hat{R}_i^2(0)(2^{J_i+1}-1)\right]}{2\left[\sum_{i=1}^{n}\hat{R}_i^4(0)(2^{J_i+1}-1)\right]^{1/2}} \tag{4.1}$$

and

$$\tilde{W}_2 = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\frac{2\pi T_i \sum_{j=0}^{J_i}\sum_{k=1}^{2^j}\hat{a}_{ijk}^2 - (2^{J_i+1}-1)}{2(2^{J_i+1}-1)^{1/2}}\right]. \tag{4.2}$$

These tests can be viewed as the generalizations of Lee and Hong's (2000) test to the estimated residuals of panel models. The following theorem shows that they are asymptotically N(0,1) under $\mathbb{H}_0$, but under the condition that $J_i \to \infty$ for all i.

Theorem 2 *Suppose that Assumptions 1–5 hold, $2^{J_i+1} = a_i T_i^{\nu}$ for $a_i \in [c, C]$ and $\nu \in (0, \frac{1}{2})$, $n/T^{\nu}\log_2^2 T \to 0$, $n/T^{2(2\tau-1)-2(2\tau-\frac{1}{2})\nu} \to 0$ as $n, T \to \infty$, where $\tau \geq \frac{3}{2}$ is as in Assumption 3. If $\{v_{it}\}$ is i.i.d. for each i, then $\tilde{W}_1 - \hat{W}_1 \xrightarrow{p} 0, \tilde{W}_2 - W_2 \xrightarrow{p} 0$ $\tilde{W}_1 \xrightarrow{d} N(0,1)$ and $\tilde{W}_2 \xrightarrow{d} N(0,1)$.*

Thus, for large J_i, $\tilde{W}_1$ and $\tilde{W}_2$ are asymptotically equivalent to $\hat{W}_1$ and $\hat{W}_2$ respectively, and both are asymptotically $N(0,1)$ under $\mathbb{H}_0$. However, we now have to impose a restrictive condition that n cannot grow faster than T^{ν}, where $\nu < \frac{1}{2}$. As a consequence, the finite sample performance of $\tilde{W}_1$ and $\tilde{W}_2$ may not be as good as $\hat{W}_1$ and $\hat{W}_2$ respectively in finite samples.

5 Consistency

We now show that $\hat{W}_1$ and $\hat{W}_2$ are consistent against $\mathbb{H}_A$. We assume the following conditions:

Assumption 6 *For each i, $\{v_{it}\}$ is a mean 0, fourth order 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 $\{u_{it}, u_{it+j}, u_{it+k}, u_{it+l}\}$.*

The fourth order cumulant is defined as

$$\kappa_i(j,k,l) \equiv E(v_{it}v_{it+j}v_{it+k}v_{it+l}) - E(e_{it}e_{it+j}e_{it+k}e_{it+l}), \qquad j,k,l \in \mathbb{Z},$$

where $\{e_{it}\}$ is a Gaussian process with the same mean and covariances as $\{v_{it}\}$. The cumulant condition in Assumption 6 characterizes the temporal dependence of $\{v_{it}\}$. It is a standard condition in time-series analysis. 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 $d < \frac{1}{4}$.

Theorem 3 *Put $n_A \equiv \#(\mathbb{N}_A)$ and $c_i \equiv T_i/T$, where $\mathbb{N}_A \equiv \{i : 0 \leq i \leq n, Q(f_i, f_{i0}) > 0\}$. Suppose that Assumptions 1–6 hold, $(n_AT)^{-1}\sum_{i=1}^{n} 2^{J_i} \to 0$ and $J_i \to \infty$ for all $i = 1, ..., n$ as $n, T \to \infty$. Then (a)*

$$(n_AT)^{-1}\hat{V}^{1/2}\hat{W}_1 - n_A^{-1}\sum_{i\in\mathbb{N}_A} 2\pi c_i Q(f_i, f_{i0}) \xrightarrow{p} 0.$$

(b) If in addition $2^{J_i+1} = a_i T_i^{\nu}$ for all i, where $a_i \in [c, C]$ and $\nu \in (0,1)$, then

$$(n_AT^{1-\nu/2})^{-1}\hat{W}_2 - n_A^{-1}\sum_{i\in\mathbb{N}_A} \pi(c_i/a_i)^{1/2} Q(f_i, f_{i0}) \xrightarrow{p} 0.$$

Under $\mathbb{H}_A$, the index set $\mathbb{N}_A$ is nonempty, at least for n sufficiently large. It follows that $n_A^{-1}\sum_{i\in\mathbb{N}_A} c_i Q(f_i, f_{i0}) \geq c > 0$ for n sufficiently large. Then Theorem 3 implies $P[\hat{W}_1 > C(n,T)] \to 1$ and $P[\hat{W}_2 > C(n,T)] \to 1$ under $\mathbb{H}_A$ for any sequence of constants $\{C(n,T) = o[n_AT/(\sum_{i=1}^{n} 2^{J_i})^{1/2}]\}$. Thus, $\hat{W}_1$ and $\hat{W}_2$ are consistent against $\mathbb{H}_A$ provided $(n_AT)^{-1}\sum_{i=1}^{n} 2^{J_i} \to 0$ and $J_i \to \infty$ for all i. Note that to ensure consistency against $\mathbb{H}_A$, we let $J_i \to \infty$ for all i here. This differs from the situation under $\mathbb{H}_0$, where J_i can be fixed for all i. The data-driven method we develop in Section 6 will deliver a data-dependent finest scale that, in an automatic manner, converges to 0 under $\mathbb{H}_0$ but grows with T under $\mathbb{H}_A$, thus ensuring consistency of $\hat{W}_1$ and $\hat{W}_2$ against $\mathbb{H}_A$.

Under $\mathbb{H}_A$, both $\hat{W}_1$ and $\hat{W}_2$ diverge to ∞ at the rate of $n_AT/(\sum_{i=1}^{n} 2^{J_i})^{1/2}$. 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 $2^{J_i+1} = a_i T_i^{\nu}$, the rate $n_A T(\sum_{i=1}^n 2^{J_i})^{1/2} \propto (n_A/n) n^{1/2} T^{1-\nu/2}$. 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 $n^{1/2}T^{1-\nu/2}$. We further note that for $\hat{W}_1$ and $\hat{W}_2$, serial correlations from different individuals never cancel each other out when some individuals have positive autocorrelation and some have negative autocorrelation, thanks to the use of the L_2-norm. In contrast, for some existing popular tests, serial correlations from different individuals may cancel each other out at least in part when some individuals have positive autocorrelation and some have negative autocorrelation, leading to low or little power. See Section 7 for examples and more discussion.

Theorems 1 and 3 imply that for all n and T sufficiently large, the negative values of $\hat{W}_1$ and $\hat{W}_2$ can occur only under $\mathbb{H}_0$. Thus, it is appropriate to use the upper-tailed N(0,1) critical values for inference. The asymptotic critical value at the 5% level, for example, is 1.645.

As noted earlier, the tests $\hat{W}_1$ and $\hat{W}_2$ are heteroskedasticity-consistent and heteroskedasticity-corrected versions respectively. An interesting question is which test, $\hat{W}_1$ or $\hat{W}_2$, is more powerful? Without loss of generality, we assume $2^{J_i+1} = a_i T_i^{\nu}$ for all $i = 1, ..., n$, where $a_i \in [c, C]$ and $\nu \in (0, 1)$. For processes with stronger serial correlation, there is a sharper spectral peak for $f_i(\cdot)$. Consequently, a larger a_i will be appropriate. In contrast, for processes with weaker serial correlation, there is a smoother spectral peak for $f_i(\cdot)$. In this case, a smaller a_i will be appropriate. With this rule, we obtain the following:

Theorem 4 *Suppose that Assumptions 1–6 hold, $n = \gamma T^{\varsigma}$ for some $\gamma \in (0, \infty)$ and $\varsigma \in (0, \infty)$, and $2^{J_i+1} = a_i T_i^{\nu}$ for $a_i \in [c, C]$ and $\nu \in (0, 1)$. If a_i is a monotonically increasing function of $Q(f_i, f_{i0})$ and $T_i = T$ for all $i = 1, ..., n$, 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. Geweke (1981), among others, has used this criterion in econometrics.

Theorem 4 implies that in the context of hypothesis testing, correcting heteroskedasticity does not necessarily lead to better power. This is in contrast to the well-known result that correcting heteroskedasticity leads to more efficient estimation. Intuitively, for the test $\hat{W}_2$, a larger $Q(f_i, f_{i0})$ is more heavily discounted by $\sqrt{V_{i0}} = 2(2^{J_i+1} - 1)^{1/2}[1 + o(1)]$ when J_i is larger. Thus, it is less powerful than $\hat{W}_1$, which puts uniform weighting to each $Q(f_i, f_{i0})$. It is of course 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 such a rule that a_i is monotonically increasing in $Q(f_i, f_{i0})$. Note that $\hat{W}_1$ and $\hat{W}_2$ may not be asymptotically equally efficient when $J_i = J$ for all i, because heteroskedasticity ($\sigma_i^2 \neq \sigma^2$) may exist.

It is also important to note that the asymptotic power of $\hat{W}_1$ and $\hat{W}_2$ does not depend on wavelet function $\psi(\cdot)$. In other words, all wavelets satisfying Assumptions 2 and 3 are asymptotically equally efficient in terms of Bahadur's criterion. Thus, the choice of $\psi(\cdot)$ is not important. This differs from the kernel method, where the choice of the kernel function affects the asymptotic power of the tests (e.g., Hong 1996).

6 Adaptive Choice of Finest Scale

Theorem 1 implies that the choice of J_i is not important for the 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. If J_i is fixed and does not grow with T_i for all i, for example, $\hat{W}_1$ and $\hat{W}_2$ will not be consistent against serial correlation of unknown form. Therefore, it will be highly desirable to choose J_i via suitable data-driven methods, which let data speak for proper finest scales.

We will develop a data-driven method to select a suitable finest scale. Before discussing a specific method, we first justify the use of a data-driven finest scale $\hat{J}$ say. For simplicity and convenience, 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 that Assumptions 1–5 and 7 hold, and $\hat{J}$ is a data-driven finest scale with* $2^{\hat{J}}/2^J = 1 + o_P(2^{-J/2})$, *where J is a nonstochastic finest scale such that* $2^{2J}/(n^2 + T) \to 0$ *as* $n \to \infty$ *and* $T \to \infty$. *If* $\{v_{it}\}$ *is i.i.d. for each i, then* $W_1(\hat{J}) - W_1(J) \xrightarrow{p} 0, W_2(\hat{J}) - W_2(J) \xrightarrow{p} 0$, $\hat{W}_1(\hat{J}) \xrightarrow{d} N(0,1)$ *and* $\hat{W}_2(\hat{J}) \xrightarrow{d} N(0,1)$.

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 $2^{\hat{J}}/2^J - 1 = o_P(2^{-J/2})$ is mild. If $2^J \propto T^{1/5}$, for example, we require $2^{\hat{J}}/2^J = 1 + o_P(T^{-1/10})$. If J is fixed (e.g., $J = 0$), which occurs under $\mathbb{H}_0$ for our data-driven method below, the condition becomes $2^{\hat{J}}/2^J \xrightarrow{p} 1$.

So far very few data-driven methods to choose J are available in the wavelet literature. To the best of our knowledge, only Walter (1994) proposes a data-driven method to choose a finest scale for probability density estimation, using an integrated mean square error (IMSE) criterion. The method can be adapted to choose $\hat{J}$ in spectral density estimation. It is based on the fact that the change in the average IMSE of $\{\hat{f}_i(\cdot)\}_{i=1}^n$ from $J-1$ to J is proportional to $n^{-1}\sum_{i=1}^n \sum_{k=1}^{2^{J_i}} \hat{\alpha}_{iJk}^2$, where $\hat{\alpha}_{iJk}$ is the empirical wavelet coefficient of $f_i(\cdot)$ at scale J. One starts from $J = 0$ and checks how IMSE changes from $J = 0$ to $J = 1$. The grid search is iterated until one gets the scale $\hat{J}$

at which the average IMSE increases most rapidly. Then, one obtains the finest scale $\hat{J}$, which gives an average IMSE that one cannot improve practically by further increasing J. This method might be suitable here because it is based on the information of $\{f_i(\cdot)\}_{i=1}^n$ over $[-\pi, \pi]$. However, no formal results on the rate of Walter's $\hat{J}$ is available and it is unknown whether it satisfies the condition of Theorem 5. Below we develop a data-driven method to choose $\hat{J}$ that will satisfy the condition of Theorem 5. For this purpose, 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 the following additional condition on $\{v_{it}\}$.

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 assumption characterizes the smoothness of $f_i(\cdot)$. It rules out long memory processes, because it implies $\sum_{h=-\infty}^{\infty} R_i(h) \leq C$. Under Assumption 8, the q-th order generalized spectral derivative of $f_i(\omega)$,

$$f_i^{(q)}(\omega) \equiv (2\pi)^{-1} \sum_{h=-\infty}^{\infty} |h|^q R_i(\omega) e^{-\mathbf{i}h\omega}, \qquad \omega \in [-\pi, \pi], \tag{6.1}$$

exists and is continuous on $[-\pi, \pi]$.

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

$$\lambda_q \equiv -\frac{(2\pi)^q}{1 - 2^{-q}} \lim_{z \to 0} \lambda(z)/|z|^q, \tag{6.2}$$

where $\lambda(z) \equiv 2\pi \hat{\psi}^*(z) \sum_{r=-\infty}^{\infty} \hat{\psi}(z + 2\pi r)$. Given Assumption 7, we have $\lambda_q < \infty$. We will also assume $\lambda_q > 0$. For the Franklin wavelet (3.5), $q = 2$; for the second order spline wavelet (3.6), $q = 3$. For the Harr wavelet (3.3), $q = 1$, but Assumption 3 rules out the Harr wavelet.

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

$$\bar{f}_i(\omega) \equiv (2\pi)^{-1} \bar{R}_i(0) + \sum_{j=1}^{J_i} \sum_{k=1}^{2^j} \bar{\alpha}_{ijk} \Psi_{jk}(\omega), \qquad \omega \in [-\pi, \pi], \tag{6.3}$$

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

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

$$EQ(\bar{f}_i, f_i) = \frac{2^{J_i+1}}{T_i} \int_{-\pi}^{\pi} f_i^2(\omega) d\omega + 2^{-2q(J_i+1)} \lambda_q^2 \int_{-\pi}^{\pi} [f^{(q)}(\omega)]^2 d\omega + o(2^{J_i}/T_i + 2^{-2qJ_i}).$$

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

$$n^{-1} \sum_{i=1}^n EQ(\bar{f}_i, f_i) = \frac{2^{J+1}}{T} n^{-1} \sum_{i=1}^n c_i^{-1} \int_{-\pi}^{\pi} f_i^2(\omega) d\omega + 2^{-2q(J+1)} \lambda_q^2 n^{-1} \sum_{i=1}^n \int_{-\pi}^{\pi} [f_i^{(q)}(\omega)]^2 d\omega + o(2^J/T + 2^{-2qJ}).$$

Theorem 6(a) gives the asymptotic IMSE of $\bar{f}_i(\cdot)$, and Theorem 6(b) gives the average asymptotic IMSE of $\{\bar{f}_i(\cdot)\}_{i=1}^n$. These results 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 $Q(\bar{f}_i, f_i)$ and $n^{-1}\sum_{i=1}^n Q(\bar{f}_i, f_i)$ respectively. The parameter estimation $(\hat{\beta})$ has no impact on the optimal convergence rates of $Q(\hat{f}_i, f_i)$ and $n^{-1}\sum_{i=1}^n Q(\hat{f}_i, f_i)$.

To obtain the optimal finest scale that minimizes the average asymptotic IMSE of $\{\bar{f}_i(\cdot)\}_{i=1}^n$, we differentiate the average asymptotic IMSE in Theorem 6(b) with respect to J and set the derivative equal to 0. This yields

$$2^{J_0+1} = \left[2q\lambda_q^2\xi_0(q)T\right]^{1/(2q+1)}, \tag{6.4}$$

where

$$\xi_0(q) \equiv \sum_{i=1}^n \int_{-\pi}^{\pi} [f_i^{(q)}(\omega)]^2 d\omega / \sum_{i=1}^n c_i^{-1} \int_{-\pi}^{\pi} f_i^2(\omega) d\omega.$$

This optimal finest scale, J_0, is infeasible because $\xi_0(q)$ is unknown under $\mathbb{H}_A$. However, we can use some estimator $\hat{\xi}_0(q)$ and plug it in (6.4). This gives a data-driven finest scale $\hat{J}_0$:

$$2^{\hat{J}_0+1} \equiv \left[2q\lambda_q^2\hat{\xi}_0(q)T\right]^{1/(2q+1)}. \tag{6.5}$$

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

$$\hat{J}_0 \equiv \max\left\{\left[\frac{1}{2q+1}\log_2\left(2q\lambda_q^2\hat{\xi}_0(q)T\right) - 1\right], 0\right\}, \tag{6.6}$$

where the square bracket denotes the integer part.

We impose the following condition on $\hat{\xi}_0(q)$:

Assumption 9 $\hat{\xi}_0(q) - \zeta_0(q) = o_P(T^{-\delta})$, *where* $\delta = 1/2(2q+1)$ *if* $\zeta_0(q) \in [c, C]$ *and* $\delta = 1/(2q+1)$ *if* $\zeta_0(q) = 0$.

Note that the condition on $\hat{\xi}_0(q)$ is more stringent when $\zeta_0(q) = 0$ than when $\zeta_0(q) \neq 0$, but for both cases the conditions are mild. We do not require $p\lim \hat{\xi}_0(2) \equiv \zeta_0(q) = \xi_0(q)$, where $\xi_0(q)$ is as in (6.4). When (and only when) $\zeta_0(q) = \xi_0(q)$, $\hat{J}_0$ in (6.6) will converge to the optimal J_0 in (6.4).

Corollary 7 *Suppose that Assumptions 1–9 hold and* $\hat{J}_0$ *is given as in (6.6). If* $\{v_{it}\}$ *is i.i.d. for each* i, *then* $\hat{W}_1(\hat{J}_0) \xrightarrow{d} N(0,1)$ *and* $\hat{W}_2(\hat{J}_0) \xrightarrow{d} N(0,1)$.

For the estimator $\hat{\xi}_0(q)$, we can use parametric or nonparametric methods. Such methods are popular in choosing narrow-band bandwidths in kernel-based spectral density estimation at frequency 0 (cf. Andrews 1991, Newey and West 1994). The nonparametric method will deliver asymptotically optimal finest scales, but may be subject to substantial variation in finite samples and thus leads to less accurate sizes for the tests. The parametric method generally delivers suboptimal finest scales, but is subject to less variation in finite samples, which lead to better sizes for the tests. Here, we use Andrew's (1991) type parametric plug-in approach and consider a parametric AR(p_i) model for each i:

$$\hat{v}_{it} = \gamma_{i0} + \sum_{h=1}^{p_i} \gamma_{ih}\hat{v}_{it-h} + \varepsilon_{it}, \qquad t = 1, ..., T_i, i = 1, ..., n, \tag{6.7}$$

where $\hat{v}_{it} \equiv 0$ if $t \leq 0$. The lag order p_i is fixed but may differ across i. In practice, one can use AIC, BIC or the order determination procedure by Hjellvik and Tjøstheim (1999) to determine p_i. Suppose $\hat{\gamma}_i \equiv (\hat{\gamma}_{i0}, \hat{\gamma}_{i1}, ..., \hat{\gamma}_{ip})'$ is the OLS estimator of $\gamma_i \equiv (\gamma_{i0}, \gamma_{i1}, ..., \gamma_{ip})'$. For concreteness, we consider $q = 2$ here. An example is the Franklin wavelet in (3.5), whose $\lambda_2 = (8/3)\sum_{l=0}^{\infty}(2l+1)^{-2}$ by direct calculation. We have

$$\hat{\xi}_0(2) \equiv \sum_{i=1}^{n} \int_{-\pi}^{\pi} \left[\frac{d^2}{d\omega^2}\hat{f}_i(\omega)\right]^2 d\omega / \sum_{i=1}^{n}(T_i/T) \int_{-\pi}^{\pi} \hat{f}_i^2(\omega)d\omega, \tag{6.8}$$

where $\hat{f}_i(\omega) \equiv (2\pi)^{-1}|1 - \sum_{h=1}^{p_i}\hat{\gamma}_{ih}e^{-ih\omega}|^{-2}$. Note that we have used the fact that for $q = 2$ the generalized spectral derivative $f_i^{(2)}(\omega) = -\frac{d^2}{d\omega^2}f_i(\omega)$. Also, for convenience we have set the estimator for var(ε_{it}) equal to 1 in $\hat{f}_i(\cdot)$. This has no impact at all because the variance estimators will cancel in the numerator and denominator of (6.8). The estimator $\hat{\xi}_0(2)$ incorporates information of $\{f_i(\cdot)\}_{i=1}^n$ over $[-\pi, \pi]$ rather than at frequency 0 only. We can use one-dimensional numerical integrations to compute $\hat{\xi}_0(2)$. We note that $\hat{\xi}_0(2)$ satisfies Assumption 9 with $q = 2$ because for parametric AR(p_i) approximations, $\hat{\xi}_0(2) - \zeta_0(2) = O_P((nT)^{-1/2})$.

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 an individual-specific $\hat{J}_i$ may effectively capture spatial nonhomogeneity in the degree of serial correlation across i. However, more stringent rate conditions on individual-specific $\{\hat{J}_i\}_{i=1}^n$ would be required to ensure that use of them has no impact on the limit distribution of the test statistics. In particular, these conditions would impose restrictive relative speed limits on n and T. Moreover, for small and finite samples, $\{\hat{J}_i\}_{i=1}^n$ may have wide variations across i, leading to poor sizes for the tests. Our simulation (not reported) shows that such individual-specific $\{\hat{J}_i\}$ lead to strong overrejection for $\hat{W}_1$ (but not for $\hat{W}_2$), and somewhat surprisingly, they may not necessarily deliver better power than $\hat{J}_0$. Thus, we recommend using $\hat{J}_0$. Perhaps a compromise is to develop a data-driven finest $\hat{J}_c$, 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. We leave this for future research.

7 Monte Carlo Experiment

We now compare the finite sample properties of $\hat{W}_1$ and $\hat{W}_2$ with three existing tests for serial correlation in the panel literature, namely, the Durbin-Watson type test of Bhargava *et al.* (1982, BFN), the LM test of Baltagi and Li (1995, BL), and the modified LM test of Bera *et al.* (2000, BSY). These test statistics are derived for balanced panels and are given below:

$$\text{BFN} = \frac{\sum_{i=1}^{n}\sum_{t=1}^{T}(\hat{v}_{it}-\hat{v}_{it-1})^2}{\sum_{i=1}^{n}\sum_{t=1}^{T}\hat{v}_{it}^2}, \tag{7.1}$$

$$\text{BL} = \frac{nT^2}{T-1}\left(\frac{\sum_{i=1}^{n}\sum_{t=1}^{T}\hat{v}_{it}\hat{v}_{it-1}}{\sum_{i=1}^{n}\sum_{t=1}^{T}\hat{v}_{it}^2}\right)^2, \tag{7.2}$$

$$\text{BSY} = \frac{nT^2\left(B+\frac{A}{T}\right)^2}{(T-1)\left(1-\frac{2}{T}\right)}, \tag{7.3}$$

where $\hat{v}_{it}$ is the within residual,

$$A = \frac{1-\sum_{i=1}^{n}\widetilde{u}_i^{\prime}J_T\widetilde{u}_i}{\sum_{i=1}^{n}\sum_{t=1}^{T}\widetilde{u}_{it}^2}, \qquad B = \frac{\sum_{i=1}^{n}\sum_{t=1}^{T}\widetilde{u}_{it}\widetilde{u}_{it-1}}{\sum_{i=1}^{n}\sum_{t=1}^{T}\widetilde{u}_{it}^2},$$

J_T is a $T\times T$ matrix of ones, and $\tilde{u}_i \equiv (\tilde{u}_{i1},...,\tilde{u}_{iT})'$ is the OLS residual vector of individual i from (2.1) without random effects. Baltagi and Li (1995) and Bera *et al.* (2000) show respectively that BL and BSY are asymptotically distributed as χ_1^2 under $\mathbb{H}_0$. Under $\mathbb{H}_0$, BFN converges to 2, a degenerate distribution. Bhargava *et al.* (1982) argue that for large n there is a positive serial correlation if BFN < 2 and suggest using a critical value of 2 for the 5% level (Bhargava *et al.* 1982, p. 436). We note that for computational simplicity we have used Baltagi and Li's (1995) LM test for first order serial correlation assuming the presence of fixed individual effects. This test can be used for both fixed and random effects models because the within transformation will wipe out the individual effects even if they are random.

We consider the following three DGPs for the panel data model:

DGP1: X_{it} is $I(0)$:

$$\begin{aligned} Y_{it} &= \alpha+\beta X_{it}+\mu_i+v_{it}, \\ X_{it} &= 0.5X_{it-1}+\eta_{it}; \end{aligned}$$

DGP2: X_{it} is $I(1)$:

$$\begin{aligned} Y_{it} &= \alpha+\beta X_{it}+\mu_i+v_{it}, \\ X_{it} &= X_{it-1}+\eta_{it}; \end{aligned}$$

DGP3: X_{it} is time trend:

$$Y_{it} = \alpha+\beta t+\mu_i+v_{it},$$

where $\eta_{it} \overset{iid}{\sim} U[-0.5,0.5]$, $\mu_i \overset{iid}{\sim} N(0,\sigma_\mu^2)$, $\alpha=5$ and $\beta=0.5$. The initial values X_{i0} were chosen as in Baltagi *et al.* (1992). We let $\sigma^2 \equiv \sigma_\mu^2+\sigma_v^2=20$ and $\tau \equiv \frac{\sigma_\mu^2}{\sigma^2}$ take five different values, $(0, 0.05, 0.2, 0.4, 0.6, 0.8)$. The value of τ measures the relative strength of random effects (when $\tau=0$, there is no random effect). We note that a combination of DGP1 and DGP 3 has been used in Baltagi and Li (1995) and Bera *et al.* (2000). We consider three sample size combinations: $(n,T) = (25,32), (50,64), (100,128)$.

Because the choice of wavelet is not important for our tests $\hat{W}_1$ and $\hat{W}_2$ (cf. Section 5), we only use the Franklin wavelet in (3.5). To examine the impact of the choice of finest scale J, we consider $J = 0,1,2$ and the data-driven finest scale $\hat{J}_0$ in (6.6).

To examine the size of the tests, we consider $v_{it} = \varepsilon_{it}$, where $\varepsilon_{it} \overset{iid}{\sim} N(0,1)$. To examine the power of the tests, we consider the following processes for $\{v_{it}\}$:

AR(1) Alternatives:

$$\begin{cases} \text{AR}(1)^a: & v_{it} = 0.2v_{it-1} + \varepsilon_{it}, \\ \text{AR}(1)^b: & v_{it} = -0.2v_{it-1} + \varepsilon_{it}, \\ \text{AR}(1)^c: & v_{it} = \begin{cases} 0.2v_{it-1} + \varepsilon_{it} & i = 1,...,\frac{n}{2}, \\ \varepsilon_{it} & i = \frac{n}{2}+1,...,n, \end{cases} \\ \text{AR}(1)^d: & v_{it} = \begin{cases} -0.2v_{it-1} + \varepsilon_{it} & i = 1,...,\frac{n}{2}, \\ \varepsilon_{it} & i = \frac{n}{2}+1,...,n, \end{cases} \\ \text{AR}(1)^e: & v_{it} = \begin{cases} 0.2v_{it-1} + \varepsilon_{it} & i = 1,...,\frac{n}{2}, \\ -0.2v_{it-1} + \varepsilon_{it} & i = \frac{n}{2}+1,...,n. \end{cases} \end{cases} \tag{7.4}$$

AR(12) Alternatives:

$$\begin{cases} \text{AR}(12)^a: & v_{it} = 0.1\sum_{h=1}^{12} \frac{13-h}{12} v_{it-h} + \varepsilon_{it}, \quad i = 1,...,n, \\ \text{AR}(12)^b: & v_{it} = -0.1\sum_{h=1}^{12} \frac{13-h}{12} v_{it-h} + \varepsilon_{it}, \quad i = 1,...,n, \\ \text{AR}(12)^c: & v_{it} = \begin{cases} 0.1\sum_{h=1}^{12} \frac{13-h}{12} v_{it-h} + \varepsilon_{it}, & i = 1,...,\frac{n}{2}, \\ \varepsilon_{it} & i = \frac{n}{2}+1,...,n, \end{cases} \\ \text{AR}(12)^d: & v_{it} = \begin{cases} -0.1\sum_{h=1}^{12} \frac{13-h}{12} v_{it-h} + \varepsilon_{it}, & i = 1,...,\frac{n}{2}, \\ \varepsilon_{it} & i = \frac{n}{2}+1,...,n, \end{cases} \\ \text{AR}(12)^e: & v_{it} = \begin{cases} 0.1\sum_{h=1}^{12} \frac{13-h}{12} v_{it-h} + \varepsilon_{it}, & i = 1,...,\frac{n}{2}, \\ -0.1\sum_{h=1}^{12} \frac{13-h}{12} v_{it-h} + \varepsilon_{it}, & i = \frac{n}{2}+1,...,n. \end{cases} \end{cases} \tag{7.5}$$

ARMA(4,4) Alternatives:

$$\begin{cases} \text{ARMA(4,4)}^a: & v_{it} = -0.4v_{it-4} + \varepsilon_{it} + \varepsilon_{it-4}, \quad i = 1,...,n, \\ \text{ARMA(4,4)}^b: & v_{it} = 0.4v_{it-4} + \varepsilon_{it} - \varepsilon_{it-4}, \quad i = 1,...,n, \\ \text{ARMA(4,4)}^c: & v_{it} = \begin{cases} -0.4v_{it-4} + \varepsilon_{it} + \varepsilon_{it-4}, & i = 1,...,\frac{n}{2}, \\ \varepsilon_{it}, & i = \frac{n}{2}+1,...,n, \end{cases} \\ \text{ARMA(4,4)}^d: & v_{it} = \begin{cases} 0.4v_{it-4} + \varepsilon_{it} - \varepsilon_{it-4}, & i = 1,...,\frac{n}{2}, \\ \varepsilon_{it}, & i = \frac{n}{2}+1,...,n, \end{cases} \\ \text{ARMA(4,4)}^e: & v_{it} = \begin{cases} -0.4v_{it-4} + \varepsilon_{it} + \varepsilon_{it-4}, & i = 1,...,\frac{n}{2}, \\ 0.4v_{it-1} + \varepsilon_{it} - \varepsilon_{it-4}, & i = \frac{n}{2}+1,...,n. \end{cases} \end{cases} \tag{7.6}$$

The first block contains five AR(1) alternatives. For $\text{AR}(1)^a$ and $\text{AR}(1)^b$, all n individuals follow a positive AR(1) and a negative AR(1) respectively. We call $\text{AR}(1)^a$ and $\text{AR}(1)^b$ the "full positive and full negative AR(1)." For these two alternatives, BL and BSY have optimal power by design. The wavelet tests do not have advantages because these alternatives have a relatively flat spectrum. Alternatives $\text{AR}(1)^c$ and $\text{AR}(1)^d$ are called the "half positive and half negative AR(1)" respectively, where the first half of the n individuals follow an AR(1) while the second half are white noises. Alternative $\text{AR}(1)^e$ is a mixed AR(1), where the first half of the n individuals follow a positive AR(1) while the second half follow a negative AR(1). As will be seen shortly, the wavelet tests are powerful in detecting such an alternative while BL and BSY will fail to detect it.

The second block contains various AR(12) alternatives, which can arise from monthly panel data. Although the autoregressive coefficients are very small at each lag and decay to 0 linearly as the lag order increase, the AR(12) alternative has a distinct spectral mode at frequency 0. We expect that the wavelet tests will be powerful. Both BL and BSY will also have power because they are based on the first and largest autoregressive coefficient, but they are expected to be less powerful.

The third block contains various ARMA(4,4) alternatives, which can arise from quarterly panel data. Such an ARMA(4,4) alternative does not have a spectral peak at frequency 0 because its autocorrelation decays to 0 exponentially as the lag order increases. However, it displays a seasonal pattern and has a spectral spike at some nonzero frequencies. We expect that wavelet tests will also perform well here.

The simulations were performed by an Ultra Enterprise 3000 computer. Random numbers for error terms, $\{\eta_{it}, \varepsilon_{it}\}$, were generated by the GAUSS 3.2.31 random number generators RNDUS and RNDNS. At each replication, we generated an $n(T+1000)$ length of random numbers and then split it into n series so that each series had the same mean and variance. The first 1000 observations were discarded for each series to reduce the impact of the initial observations and to diminish the dependence between the samples. The number of replication is 1000 for each experiment.

As our simulation studies are rather extensive, we only report the results from DGP1 at the 10% and 5% levels. The results from DGP2 and SGP3 and for the 1% level are similar; they are available from the authors upon request. We first examine the empirical sizes of the tests. Using the critical value suggested by Bhargava *et al.* (1982, p.436), the BFN test strongly over-rejects $\mathbb{H}_0$. It rejects $\mathbb{H}_0$ up to 67.7%, 66.9% and 64.8% at the 5% level, for example, when $(n, T) = (25, 32), (50, 64)$ and $(100, 128)$ respectively, and $\tau = 0.2$. It appears that the BFN test could not be used in practice. For this reason, we drop it from comparison. Table 1 presents the sizes of the tests at the 10% and 5% levels under $\mathbb{H}_0$ for $(n, T) = (25, 32)$ and six values of τ. When there is no random effect ($\tau = 0$), the size of the BSY test is quite reasonable and the best among all the tests. The BL test over-rejects $\mathbb{H}_0$. Both $\hat{W}_1$ and $\hat{W}_2$ under-reject $\mathbb{H}_0$. For other values of τ, as expected from Bera *et al.*'s (2000) theory, the size of BSY is sensitive to the choice of τ. When τ increases, the size distortion (either underrejection or overrejection) of the BSY test increases. The BL test still over-rejects $\mathbb{H}_0$ for all τ. It should be emphasized here that Baltagi and Li (1995, p.16) has pointed out their LM test for first order serial correlation assuming the presence of random individual effects has better size than BL in (7.2). Thus, we expect that the size of their test will improve if the former version is used. We do not use this version here for computational simplicity, because it would require computation of MLE and hence lose it simplicity as Bera *et al.* (2000) point out. On the other hand, $\hat{W}_1$ and $\hat{W}_2$ are generally robust to the choice of τ. They have the best sizes when τ is large. The sizes of $\hat{W}_1$ and $\hat{W}_2$ are better when smaller J or data-driven $\hat{J}_0$ is used.

Tables 2 and 3 report the sizes for $(n, T) = (50, 64)$ and $(100, 128)$ respectively. Again, BSY has the best size among all the tests when $\tau = 0$, but it is sensitive to the choice of τ. BL still over-rejects $\mathbb{H}_0$ for all τ. Now, the sizes of $\hat{W}_1$ and $\hat{W}_2$ are substantially improved and reasonable, especially when data-driven $\hat{J}_0$ or $J = 0$ is used. Both $\hat{W}_1$ and $\hat{W}_2$ have better sizes than BL and BSY except for $\tau = 0$. Finally, we note that the empirical sizes of $\hat{W}_1$ and $\hat{W}_2$ are very similar for all cases.

We now turn to examine the power of the tests. For a fair comparison, we consider size-corrected power using empirical critical values. Tables 4 and 5 report the power of the tests at the 5% level against various AR(1) alternatives for $(n, T) = (25, 32)$ and $(50, 64)$, and three values of τ. Table 4 shows the power for $(n, T) = (25, 32)$. Under $\text{AR}(1)^a$, the full positive AR(1), BSY is most powerful, followed very closely by BL. This is expected because both BSY and BL are optimal against AR(1) by design. The $\hat{W}_1$ and $\hat{W}_2$ tests have nontrivial but substantial lower power. This is because $\text{AR}(1)^a$ has a relatively flat spectrum and the advantage of wavelets is not displayed. Under $\text{AR}(1)^b$, the full negative AR(1), BL becomes the most powerful. Somewhat surprisingly, $\hat{W}_1$ and $\hat{W}_2$ have rather high power and dominate BSY for $\tau = 0.4, 0.6$. This perhaps is because $\text{AR}(1)^b$ has a less smooth spectrum than $\text{AR}(1)^a$. The power patterns of the tests under $\text{AR}(1)^c$ and $\text{AR}(1)^d$, the half positive and the half negative AR(1) alternatives, are similar to those under $\text{AR}(1)^a$ and $\text{AR}(1)^b$ respectively, except that the power of $\hat{W}_1$ and $\hat{W}_2$ is getting closer to the most powerful test (BSY under $\text{AR}(1)^c$ and BL under $\text{AR}(1)^d$). Interestingly, both BSY and BL fail to detect $\text{AR}(1)^e$, the mixture of positive AR(1) and negative AR(1). The $\hat{W}_1$ and $\hat{W}_2$ tests are very powerful against $\text{AR}(1)^e$, indicating that wavelets are rather effective in capturing nonhomogeneous serial correlations across individuals. It seems that $\hat{W}_1$ is slightly more powerful than $\hat{W}_2$ in most cases.

Table 5 reports the power for $(n, T) = (50, 64)$. Now, the power of $\hat{W}_1$ and $\hat{W}_2$ increase to 1 or almost 1 against all five AR(1) alternatives, consistent with their consistency property. The BL and BSY tests have power equal to or close to 1 except for $\text{AR}(1)^e$, against which BL and BSY have virtually no power.

For all various AR(1) alternatives, the choice of J has significant impact on the power of $\hat{W}_1$ and $\hat{W}_2$. The choice of $J = 0$ gives the best power for $\hat{W}_1$ and $\hat{W}_2$ against various AR(1) alternatives.

The data-driven $\hat{J}_0$ delivers reasonable and robust power in all cases.

Tables 6 and 7 report the power of the tests against various AR(12) alternatives. Again, the choice of J has significant impact on the power of $\hat{W}_1$ and $\hat{W}_2$. Unlike under AR(1) alternatives, now a larger J yields better power. The data-driven $\hat{J}_0$ yields better power than $J = 0, 1$, and 2. Among all the tests, the $\hat{W}_1$ and $\hat{W}_2$ tests with data-driven $\hat{J}_0$ have the best power and dominate BL and BSY against all five AR(12) alternatives. Wavelets are indeed rather powerful in capturing spectral modes/peaks. Note that $\hat{W}_2$ is more powerful than $\hat{W}_1$ that is not predicted from our theory in Theorem 4. In contrast, BL has low or no power for all cases. BSY has some power against AR(12)^a and AR(12)^c, but has low or little power against AR(12)^b, AR(12)^d and AR(12)^e. We note that under AR(12)^d and AR(12)^e, the power of $\hat{W}_1$ and $\hat{W}_2$ when $(n, T) = (50, 64)$ is somewhat lower than when $(n, T) = (25, 32)$. This is a small sample phenomenon because each individual autoregressive coefficient is very small. We examine the power of $\hat{W}_1$ and $\hat{W}_2$ for $(n, T) = (100, 128)$ and find that their power increases to 1 or close to 1 for all AR(12) alternatives.

Tables 8 and 9 report the power of the tests against various ARMA(4,4) alternatives. The $\hat{W}_1$ and $\hat{W}_2$ tests with data-driven $\hat{J}_0$ have similar power and are more powerful than BSY and BL. The BSY test has some power against ARMA(4,4)^b and ARMA(4,4)^d and BL has no lower virtually for all cases. The choice of J has significant impact on the power of $\hat{W}_1$ and $\hat{W}_2$. The choice of $J =$ 2 gives better power than $J = 0, 1$. Apparently due to the seasonal patterns of the ARMA(4,4) alternatives, the choice of $J = 0, 1$ yields little or no power for $\hat{W}_1$ and $\hat{W}_2$ against ARMA(4,4)^b and ARMA(4,4)^d. This gives a warning of the possible consequence of an arbitrary choice of J. The data-driven $\hat{J}_0$, in contrast, is able to adapt to different serial correlation patterns and gives robust power. This highlights the value of the data-driven finest scale $\hat{J}_0$.

In summary, we conclude:

1. The $\hat{W}_1$ and $\hat{W}_2$ tests with data-driven finest scale $\hat{J}_0$ or $J = 0$ have better sizes in all scenarios except when there is no random effect ($\tau = 0$). The data-driven finest scale $\hat{J}_0$ yields reasonable and robust sizes for $\hat{W}_1$ and $\hat{W}_2$ under various cases. The BSY test has the best size among all the tests under comparison when there is no random effect ($\tau = 0$), but may under-reject or over-reject $\mathbb{H}_0$ for other cases; The BL test displays over-rejections in all cases.

2. The BSY and BL tests are the most powerful against full AR(1) alternatives. They should be used if the user has the prior information that the alternative is a first order AR(1)/MA(1) process that is common for all individuals. On the other hand, both BSY and BL are dominated by the $\hat{W}_1$ and $\hat{W}_2$ tests with data-driven finest scale $\hat{J}_0$ under the mixture of positive AR(1) and negative AR(1). The $\hat{W}_1$ and $\hat{W}_2$ tests with data-driven $\hat{J}_0$ also dominate BSY and BL under various AR(12) and ARMA(4,4) alternatives, where the spectral densities have distinct modes/peaks and are less smooth than AR(1). Wavelets are indeed a powerful tool

in capturing serial correlation with nonsmooth spectrum. They are useful when the user has no prior information about the alternative, or when the alternative has significantly nonhomogeneous spectrum across different frequencies or across different individuals.

3. The choice of the finest scale parameter has some impact on the size and a significant impact on the power of $\hat{W}_1$ and $\hat{W}_2$ in small and finite samples. Smaller finest scales yield better size but may yield better or poorer power, depending on the alternatives. The data-driven finest scale $\hat{J}_0$ is able to adapt to the various serial correlation patterns and delivers reasonable sizes and robust power in most cases.

8 Conclusion

Testing for serial correlation of unknown form for both static and dynamic panel models is important. The existence of serial correlation, of any form, will generally invalidate statistical procedures involving using the standard covariance estimator of parameter estimators. It also indicates dynamic model misspecification when the regressors contain lagged dependent variables. This paper proposes a new class of wavelet-based consistent tests for serial correlation of unknown form for the estimated residuals of the panel data models. Wavelets are particularly useful for detecting serial correlation where the spectrum has peaks or kinks, as can arise from persistent/strong dependence, business cycles, seasonalities or use of seasonal data such as quarterly and monthly data, and other kinds of periodicity. The new tests have a convenient limit N(0,1) distribution. The limit distribution of the test statistic is not affected by parameter estimation, even if the regressors contain lagged dependent variables or deterministic/stochastic trending variables. The proposed tests do not require formulation of an alternative model, and are consistent against serial correlation of unknown form. They are applicable to unbalanced heterogeneous panel models. A data-driven method is developed to select finest scales—the smoothing parameters in wavelet estimation, making the test procedure entirely operational in practice. The data-driven finest scale, in an automatic manner, converges to 0 under the null hypothesis of no serial correlation but grows to infinity under the alternative, ensuring consistency of the proposed tests against serial correlation of unknown form. We examine the finite sample properties of the proposed tests as compared to some popular existing tests of serial correlation using Monte Carlo experiments. The results show that the proposed tests have good size and power in various cases, and can be used as an evaluation procedure for panel models.

Anderson, T. W., and Hsiao, C. (1982), "Formulation and Estimation of Dynamic Models Using Panel Data," *Journal of Econometrics*, 18, 47-82.

Andrews, D.W.K. (1991), "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica* 59, 817-858.

Baltagi, B. H. (1995), *Econometric Analysis of Panel Data*, Wiley: New York.

Baltagi, B. H., and Li, Q. (1991), "A Joint Test for Serial Correlation and Random Individual Effects," *Statistical and Probability Letters,* 11, 277-280.

Baltagi, B. H., and Li, Q. (1995), "Testing AR(1) against MA(1) disturbances in an Error Component Model," *Journal of Econometrics*, 68, 133-151.

Baltagi, B. H., Chang, Y.-J., and Li, Q. (1992), "Monte Carlo Results on Several New and Existing Tests for the Error Component Model," *Journal of Econometrics*, 54, 95-120.

Bahadur, R. R. (1960), "Stochastic Comparison of Tests," *Annals of Mathematical Statistics,* 31, 276-295.

Bera, A.K., Sosa-Escudero, W. and Yoon, M. (2000), "Tests for the Error Component Model in the Presence of Local Misspecification", *Journal of Econometrics*, forthcoming.

Bhargava, A., Franzini, L., and Narendranathan, W. (1982), "Serial Correlation and the Fixed Effects Model," *Review of Economic Studies*, 49, 533-549.

Binder, M., Hsiao, C., and Pesaran, M. H. (1999), "Likelihood Based Inference for Panel Vector Autoregressions: Testing for Unit Roots and Cointegration in Short Panels," Working Paper, Department of Economics, University of Southern California.

Bizer, D. S. and Durlauf, S, N. (1992), "Testing the Positive Theory of Government Finance," *Journal of Monetary Economics,* 26, 123-141.

Blundell, R. and Bond, S. (1998), "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models," *Journal of Econometrics*, 87, 115-143.

Box, G., and Pierce, D. (1970), "Distribution of Residual Autocorrelations in Autoregressive Integrated Moving Average Time Series Models," *Journal of the American Statistical Association*, 65, 1509-1526.

Breusch, T. S., and Pagan, A. (1980), "The Lagrange Multiplier Test and Its Applications to Model Specification in Econometrics," *Review of Economic Studies,* 47, 239-253.

Brown, B. M. (1971), "Martingale Central Limit Theorems," *Annals of Mathematical Statistics*, 42, 59-66.

Choi, I. (1999), "Instrumental Variables Estimation of a Nearly Nonstationary Error Component Model," Working paper, Department of Economics, Kookmin University, Korea.

Daubechies, I. (1992), *Ten Lectures on Wavelets*, SIAM: Philadelphia.

Donoho, D. L., and Johnstone, I. M. (1994), "Ideal Spatial Adoption by Wavelet Shrinkage," *Biometrika*, 81, 425-455.

Donoho, D. L., and Johnstone, I. M. (1995a), "Adapting to Unknown Smoothness via Wavelet

Shrinkage," *Journal of the American Statistical Association*, 90, 1200-1224.

Donoho, D. L., and Johnstone, I. M. (1995b), "Wavelet Shrinkage: Asymptopia," *Journal of the Royal Statistical Society*, Serial B, 57, 301-369.

Donoho, D. L., and Johnstone, I. M., and Kerkyacharian, I. M. (1996), "Density Estimation by Wavelet Thresholding," *Annals of Statistics,* 24, 508-539.

Durbin, J., and Watson, G. S. (1951), "Testing for Serial Correlation in Least Squares Regression: II," *Biometrika*, 38, 159-178.

Durlauf, S. (1991), "Spectral Based Testing for the Martingale Hypothesis," *Journal of Econometrics*, 50, 1-19.

Gao, H.-Y. (1997), "Choice of Thresholds for Wavelet Shrinkage Estimate of the Spectrum," *Journal of Time Series Analysis*, 18, 231-251.

Geweke, J. (1981), "The Approximate Slopes of Econometric Tests," *Econometrica*, 49, 1427-1442.

Granger, C.W.J. (1969), "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods," *Econometrica,* 37, 424-438.

Granger, C.W.J. (1996), "Evaluation of Panel Models: Some Suggestions from Time Series", Working Paper, Department of Economics, University of California, San Diego.

Granger, C. W. J., and Newbold, P. (1977), *Forecasting Economic Time Series*, Academic Press, New York.

Hahn, J. and Kuersteiner, G. (2000), "Asymptotically Unbiased Inference for a Dynamic Panel Model with Fixed Effects When Both n and T are large," Working Paper, Department of Economics, Brown University.

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

Hernandez, E., and Weiss, G. (1996), *A First Course on Wavelets*, CRC Press, Boca Raton.

Hjellvik, V. and Tjøstheim, D. (1999), "Residual Variance Estimates and Order Determination," in *Cointegration, Causality, and Forecasting: A Festschrift in Honor of Clive W. J. Granger*, (eds.) R. Engle and H. White, Oxford University Press: London, pp.385-409.

Hjellvik, V., Yao, Q. and Tjøstheim, D. (1998), "Linearity Testing Using Local Polynomial Approximation", *Journal of Statistical Planning and Inference,* 68, 295–321

Hong, Y. (1996), "Consistent Testing for Serial Correlation of Unknown Form," *Econometrica*, 64, 837-864.

Hong, Y. and Lee, J. (2000), "One-Sided Testing for ARCH Effects," *Econometric Theory,* forthcoming.

Hsiao, C. (1986), *Analysis of Panel Data,* Cambridge University Press: Cambridge.

Jensen, M. J. (2000), "An Alternative Maximum Likelihood Estimator of Long-Memory Processes Using Compactly Supported Wavelets," *Journal of Economic Dynamics and Control*, 24, 361-387.

Kao, C., and Chiang, M.-H. (2000), "On the Estimation and Inference of a Cointegrated Regression in Panel Data." *Advances in Econometrics*, Volume 15, forthcoming.

Kao, C., and Emerson, J. (1999), "On the Estimation of a Linear Time Trend Regression with a One-Way Error Component Model in the Presence of Serially Correlated Errors," Working Paper, Department of Economics, Syracuse University.

Lee, J., and Hong, Y. (2000), "Testing for Serial Correlation of Unknown Form Using Wavelet Methods," *Econometric Theory*, forthcoming.

Li, Q., and Hsiao, C. (1998), "Testing Serial Correlation in Semiparametric Panel Data Models," *Journal of Econometrics*, 87, 207-237.

Neumann, M. H. (1996), "Spectral Density Estimation Via Nonlinear Wavelet Methods for Stationary Non-Gaussian Time Series," *Journal of Time Series Analysis*, 17, 601-633.

Newey, W. and K. West (1994), "Automatic Lag Selection in Covariance Matrix Estimation," *Reviews of Economic Studies* 61, 631-653.

Phillips, P.C. and Moon, H. R. (1999), "Linear Regression Limit Theory for Nonstationary Panel Data," *Econometrica*, 67, 1057-1112.

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

Ramsey, J. (1999), "The Contribution of Wavelets to the Analysis of Economic and Financial Data", *Phil. Trans. Royal Society London, Series A*, 537, 2593-2606.

Skaug, H. J. and Tjøstheim, D. (1993), "A Nonparametric Test of Serial Independence Based on the Empirical Distribution Function," *Biometrika*, 80, 591–602.

Skaug, H. J., and Tjøstheim, D. (1996), "Measures of Distance Between Densities With Application to Testing for Serial Independence," in *Time Series Analysis in Memory of E. J. Hannan*, eds, P. Robinson and M. Rosenblatt, Springer: New York, pp. 363-377.

Walter, G. (1994), *Wavelets and Other Orthogonal Systems with Applications*, CRC Press: Boca Raton.

Wang, Y. (1995), "Jump and Sharp Cusp Detection by Wavelets," *Biometrika* 82, 385-397.

Watson, M. W. (1993), "Measures of Fit for Calibrated Models", *Journal of Political Economy*, 101, 1011-1041.

Appendix

To prove Theorems 1–6 and Corollary 7, we will use the following lemma:

Lemma A.1: *Suppose that Assumptions 2 and 3 hold. Let $b_{J_i}(h,m)$ be defined as that used in (3.18). Then for any $J_i, T_i \in \mathbb{Z}^+$ and a bounded constant $C \in (0,\infty)$ 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) $\sum_{h=1}^{T_i-1}\sum_{m=1}^{T_i-1} h^{\upsilon}|b_{J_i}(h,m)| \le C2^{(1+\upsilon)(J_i+1)}$ *for* $0 \le \upsilon \le \frac{1}{2}$;

(iii) $\sum_{h=1}^{T_i-1}[\sum_{m=1}^{T_i-1}|b_{J_i}(h,m)|]^2 \le C2^{(J_i+1)}$;

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

(v) $|\sum_{h=1}^{T_i-1} b_{J_i}(h,h) - (2^{J_i+1}-1)| \le C[(J_i+1) + 2^{J_i+1}(2^{J_i+1}/T_i)^{(2\tau-1)}]$, *where* τ *is in Assumption 3;*

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

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

(viii) $\sup_{1\le h\le T_i-1}\sum_{m=1}^{T_i-1}|b_{J_i}(h,m)| \le C(J_i+1)$;

Proof of Lemma A.1: Parts (i)–(vi) of this lemma extends Lemma A.1 of Lee and Hong (2000), who consider the case where $J_i \equiv J \to \infty$ as $T_i \equiv T \to \infty$ (i.e., the case with only one individual). As a consequence, Lemma A.1 of Lee and Hong (2000) cannot be applied to the present context because the $O(\cdot)$ and $o(\cdot)$ orders in Lemma A.1 of Lee and Hong (2000) would depend on i and because we allow both fixed and growing J_i. By carefully examining the proof of Lee and Hong (2000, Appendix B), however, we can replace the $O(\cdot)$ and $o(\cdot)$ orders by the upper-bounds in the right hand sides of each of parts (ii)–(vi) here. The proof of part (i) is identical to that of Lemma A.1(i) of Lee and Hong (2000).

For parts (vii) and (viii), there are no counterparts in Lee and Hong (2000). To prove (vii), it suffices to show $\max_{1-T_i\le h,m\le T_i-1}|a_{J_i}(h,m)| \le C(J_i+1)$, where $a_j(h,m)$ is as that used in (3.18). As in Lee and Hong (2000), we put $c_j(h,m) \equiv 2^{-j}\sum_{k=1}^{2^j} e^{\mathbf{i}2\pi(m-h)k/2^j}$, where $j \in \mathbb{Z}^+$. Then

$$c_j(h,m) = \begin{cases} 1 & \text{if } m-h = 2^j r \text{ for some } r \in \mathbb{Z}, \\ 0 & \text{otherwise.} \end{cases} \tag{A1}$$

Cf. Priestley (1981, (6.19), p.392). Thus by the definition of $a_{J_i}(h,m)$, we have

$$\begin{aligned} |a_{J_i}(h,m)| &= \left|2\pi\sum_{j=0}^{J_i} c_j(h,m)\hat{\psi}(2\pi h/2^j)\hat{\psi}^*(2\pi hm/2^j)\right| \\ &\le 2\pi\sum_{j=0}^{J_i}|\hat{\psi}(2\pi h/2^j)\hat{\psi}(2\pi hm/2^j)| \le C(J_i+1) \end{aligned}$$

given $|\hat{\psi}(\cdot)| \le C$ as in Assumption 3. For the proof of part (viii), using (A1), we have

$$\begin{aligned}\sum_{m=1-T_i}^{T_i-1} |a_{J_i}(h,m)| &\le \sum_{r=-\infty}^{\infty}\sum_{j=0}^{J_i} |\hat{\psi}(2\pi h/2^j)\hat{\psi}(2\pi h/2^j+2\pi r)| \\ &= \sum_{j=0}^{J_i} |\hat{\psi}(2\pi h/2^j)\sum_{j=0}^{J_i} |\hat{\psi}(2\pi h/2^j+2\pi r)| \le C(J_i+1)\end{aligned}$$

given $|\hat{\psi}(\cdot)| \le C$ and $\sup_{z\in\mathbb{R}}\sum_{r=-\infty}^{\infty} |\hat{\psi}(z+2\pi r)| \le C$ as implied by Assumption 3. This completes the proof. ■

A Proof of Theorem 1

Let $\tilde{v}_{it} \equiv v_{it} - \overline{v}_i - \overline{v}_t + \overline{v}$ be as in Assumption 5. We define

$$\tilde{\alpha}_{ijk} \equiv \sum_{h=1-T_i}^{T_i-1} \tilde{R}_i(h)\hat{\psi}_{jk}(2\pi h), \tag{A2}$$

where $\tilde{R}_i(h) \equiv T^{-1}\sum_{t=|h|+1}^{T} \tilde{v}_{it}\tilde{v}_{it-|h|}$, $h = 0, \pm1, ..., \pm(T_i - 1)$. We show Theorem 1 by proving Theorems A.1–A.3 below under the conditions of Theorem 1.

Theorem A.1: *Let $\hat{\alpha}_{ijk}$ and $\bar{\alpha}_{ijk}$ be defined as in (3.16) and (6.3), and $V_{nT} \equiv \sum_{i=1}^{n} \sigma_i^8 V_{i0}$, where V_{i0} is as in (3.18). Then*

$$V_{nT}^{-1/2}\sum_{i=1}^{n} 2\pi T_i \sum_{j=0}^{J_i}\sum_{k=1}^{2^j} (\hat{\alpha}_{ijk}^2 - \bar{\alpha}_{ijk}^2) \xrightarrow{p} 0.$$

Theorem A.2: *Put $M_{nT} \equiv \sum_{i=1}^{n} \sigma_i^4 M_{i0}$, where M_{i0} is as in (3.18). Then*

$$V_{nT}^{-1/2}\left(\sum_{i=1}^{n} 2\pi T_i \sum_{j=0}^{J_i}\sum_{k=1}^{2^j} \bar{\alpha}_{ijk}^2 - M_{nT}\right) \xrightarrow{d} N(0,1).$$

Theorem A.3: *Let $\hat{M}$ and $\hat{V}$ be defined as in (3.18). Then $V_{nT}^{-1/2}(\hat{M} - M_{nT}) \xrightarrow{p} 0$ and $\hat{V}/V_{nT} \xrightarrow{p} 1$.*

Note that for any $J_i \in \mathbb{Z}^+$ such that $2^{J_i+1}/T_i \le 1$, we have by Lemma A.1(v,vi)

$$c\sum_{i=1}^{n}(2^{J_i+1}-1) \le V_{nT} \le C\sum_{i=1}^{n}(2^{J_i+1}-1), \tag{A3}$$

$$c\sum_{i=1}^{n}(2^{J_i+1}-1) \le M_{nT} \le C\sum_{i=1}^{n}(2^{J_i+1}-1). \tag{A4}$$

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.

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

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

Proof of Proposition A.1: By straightforward algebra and the definition of $\hat{v}_{it}$ in (3.13), we have $\hat{v}_{it} = \tilde{v}_{it} - \tilde{X}_{it}'(\hat{\beta} - \beta)$. Recalling the definitions of $\hat{R}_i(h)$ in (3.14) and $\tilde{R}_i(h)$ as used in (A2), we can write

$$\begin{aligned} \hat{R}_i(h) - \tilde{R}_i(h) &= (\hat{\beta} - \beta)' \tilde{\Gamma}_{ixx}(h) (\hat{\beta} - \beta) - (\hat{\beta} - \beta)' \tilde{\Gamma}_{ixv}(h) - (\hat{\beta} - \beta)' \tilde{\Gamma}_{ivx}(h) \\ &\equiv \sum_{c=1}^{3} \hat{\xi}_{ci}(h), \text{ say,} \end{aligned} \tag{A5}$$

where $\tilde{\Gamma}_{ixx}(h) \equiv T_i^{-1} \sum_{t=|h|+1}^{T_i} \tilde{X}_{it} \tilde{X}_{it-|h|}'$, and as in Assumption 5, $\tilde{\Gamma}_{ixv}(h) \equiv T_i^{-1} \sum_{t=|h|+1}^{T_i} \tilde{X}_{it} \tilde{v}_{it-|h|}$ and $\tilde{\Gamma}_{ivx}(h) \equiv T_i^{-1} \sum_{t=|h|+1}^{T_i} \tilde{X}_{it-|h|} \tilde{v}_{it}$.

Next, recalling $\tilde{v}_{it} \equiv v_{it} - \bar{v}_i - \bar{v}_t + \bar{v}$ and the definition of $\bar{R}_i(h)$ as used in (6.3), we can write

$$\begin{aligned} \tilde{R}_i(h) - \bar{R}_i(h) &= T_i^{-1} \sum_{t=|h|+1}^{T_i} \left(-\bar{v}_i v_{it} - \bar{v}_i \tilde{v}_{it-|h|} - v_{it} \bar{v}_{t-|h|} - \bar{v}_t \tilde{v}_{it-|h|} + \bar{v} v_{it} + \bar{v} \tilde{v}_{it-|h|} \right) \\ &\equiv \sum_{c=4}^{9} \hat{\xi}_{ci}(h), \text{ say.} \end{aligned} \tag{A6}$$

Given $\hat{R}_i(h) - \bar{R}_i(h) = [\hat{R}_i(h) - \tilde{R}_i(h)] + [\tilde{R}_i(h) - \bar{R}_i(h)]$, we have

$$\hat{\alpha}_{ijk} - \bar{\alpha}_{ijk} = \sum_{c=1}^{9} \sum_{h=1-T_i}^{T_i-1} \hat{\xi}_{ci}(h) \hat{\psi}_{jk}(2\pi h). \tag{A7}$$

It follows from the C_r-inequality that

$$\begin{aligned} \sum_{i=1}^{n} 2\pi T_i \sum_{j=0}^{J_i} \sum_{k=1}^{2^j} (\hat{\alpha}_{ijk} - \bar{\alpha}_{ijk})^2 &\leq 2^8 \sum_{c=1}^{9} \left\{ \sum_{i=1}^{n} 2\pi T_i \sum_{j=0}^{J_i} \sum_{k=1}^{2^j} \left[\sum_{h=1-T_i}^{T_i-1} \hat{\xi}_{ci}(h) \hat{\psi}_{jk}(2\pi h) \right] \right\} \\ &\equiv 2^8 \sum_{c=1}^{9} \hat{A}_c, \text{ say.} \end{aligned} \tag{A8}$$

We shall show that $V_{nT}^{-1/2} \hat{A}_c \xrightarrow{p} 0$ for $1 \leq c \leq 9$.

We first consider $\hat{A}_1$. From (A3) and using the Cauchy-Schwarz inequality twice, we have

$$|\hat{\xi}_{1i}(h)| \leq ||\hat{\beta} - \beta||^2 ||\tilde{\Gamma}_{ixx}(h)|| \leq ||\hat{\beta} - \beta||^2 ||\tilde{\Gamma}_{ixx}(0)||. \tag{A9}$$

Let $b_{J_i}(h, m)$ be defined as in Lemma A.1 (or (3.18)). Then we have

$$\begin{aligned} V_{nT}^{-1/2} |\hat{A}_1| &= V_{nT}^{-1/2} \left| \sum_{i=1}^{n} T_i \sum_{h=1}^{T_i-1} \sum_{m=1}^{T_i-1} b_{J_i}(h, m) \hat{\xi}_{1i}(h) \hat{\xi}_{1i}(m) \right| \\ &\leq V_{nT}^{-1/2} ||\hat{\beta} - \beta||^4 \left[\sum_{i=1}^{n} T_i \sum_{h=1}^{T_i-1} \sum_{m=1}^{T_i-1} b_{J_i}^2(h, m) \right]^{1/2} \left[\sum_{i=1}^{n} T_i^3 ||\tilde{\Gamma}_{ixx}(0)||^4 \right]^{1/2} \\ &= O_P(n^{-3/2}), \end{aligned} \tag{A10}$$

given Lemma A.1(vi), (A3), Assumptions 3 and 4, $T_i \leq CT$ and $\sigma_i^2 \in [c, C]$.

Next, we consider the second term $\hat{A}_2$ in (A8). Recalling $\Gamma_{ixv}(h) \equiv p\lim_{T_i \to \infty} \tilde{\Gamma}_{ixv}(h)$, we have

$$\hat{\xi}_{2i}(h) = (\hat{\beta} - \beta)'\Gamma_{ixv}(h) + (\hat{\beta} - \beta)'[\tilde{\Gamma}_{ixv}(h) - \Gamma_{ixv}(h)]. \tag{A11}$$

It follows that

$$\begin{aligned}
\hat{A}_2 &\leq 2||\hat{\beta} - \beta||^2 \sum_{i=1}^{n} 2\pi T_i \sum_{j=0}^{J_i} \sum_{k=1}^{2^j} \left\| \sum_{h=1-T}^{T-1} \Gamma_{ixv}(h)\hat{\psi}_{jk}(2\pi h) \right\|^2 \\
&\quad + 2||\hat{\beta} - \beta||^2 \sum_{i=1}^{n} 2\pi T_i \sum_{j=0}^{J_i} \sum_{k=1}^{2^j} \left\| \sum_{h=1-T}^{T-1} [\tilde{\Gamma}_{ixv}(h) - \Gamma_{ixv}(h)]\hat{\psi}_{jk}(2\pi h) \right\|^2 \\
&\equiv 2||\hat{\beta} - \beta||^2 \hat{M}_1 + 2||\hat{\beta} - \beta||^2 \hat{M}_2, \text{ say.}
\end{aligned} \tag{A12}$$

We now consider $\hat{M}_1$ in (A12). Let $\Lambda_{ijk}^{xv} \equiv \int_{-\pi}^{\pi} f_{ixv}(\omega)\Psi_{jk}(\omega)d\omega$ be the wavelet coefficient of the cross-spectral density $f_{xv}(\omega) \equiv (2\pi)^{-1}\sum_{h=-\infty}^{\infty} \Gamma_{xv}(h)e^{-\mathbf{i}j\omega}$. Then $\Lambda_{ijk}^{xv} = \sum_{h=-\infty}^{\infty} \Gamma_{ixv}(h)\hat{\psi}_{jk}(2\pi h)$ by Parsavel's identity, and

$$\sum_{h=1-T_i}^{T_i-1} \Gamma_{ixv}(h)\hat{\psi}_{jk}(2\pi h) = \Lambda_{ijk}^{xv} + \sum_{|h| \geq T_i} \Gamma_{ixv}(h)\hat{\psi}_{jk}(2\pi h),$$

It follows by the Cauchy-Schwarz inequality that

$$\begin{aligned}
\hat{M}_1 &\leq 2\sum_{i=1}^{n} 2\pi T_i \sum_{j=0}^{J_i} \sum_{k=1}^{2^j} ||\Lambda_{ijk}^{xv}||^2 + 2\sum_{i=1}^{n} 2\pi T_i \sum_{|h| \geq T_i} ||\Gamma_{ixv}(h)||^2 \sum_{j=0}^{J_i} \sum_{k=1}^{2^j} \sum_{|h| \geq T_i} |\hat{\psi}_{jk}(2\pi h)|^2 \\
&= O(nT) + o[nT(2^{\bar{J}}/T)^{2\tau}] = O(nT)
\end{aligned} \tag{A13}$$

given $2^{\bar{J}}/T \to 0$, where $\bar{J} \equiv \max_{1 \leq i \leq n}(J_i)$ and we used the facts that (i) $\sum_{j=0}^{J_i} \sum_{k=1}^{2^j} ||\Lambda_{ijk}^{xv}||^2 = (2\pi)^{-1}\sum_{h=-\infty}^{\infty} ||\Gamma_{ixv}(h)||^2 \leq C$ by Parseval's identity and Assumption 5; (ii) $\sum_{|h| \geq T_i} ||\Gamma_{ixv}(h)||^2 = o(T^{-1})$ given Assumption 5 and $T_i = c_i T \geq cT$; and (iii) given Assumption 3,

$$\sum_{j=0}^{J_i} \sum_{k=1}^{2^j} \sum_{h \geq |T_i|} |\hat{\psi}_{jk}(2\pi h)|^2 \leq 2\pi \sum_{j=0}^{J_i} \sum_{|h| \geq T_i} |\hat{\psi}(2\pi h/2^j)|^2 \leq C\sum_{j=0}^{J_i} \sum_{|h| \geq T_i} (2\pi h/2^j)^{-2\tau} \leq C^2 2^{2\tau J_i}/T_i^{2\tau-1}.$$

For the second term $\hat{M}_2$ in (A12), we have

$$\begin{aligned}
||\hat{\beta} - \beta||^2 \hat{M}_2 &\leq ||\hat{\beta} - \beta||^2 \sum_{i=1}^{n} T_i \sum_{h=1}^{T_i-1} \sum_{m=1}^{T_i-1} |b_{J_i}(h,m)| \left\| \tilde{\Gamma}_{ixv}(h) - \Gamma_{ixv}(h) \right\| \left\| \tilde{\Gamma}_{ixv}(m) - \Gamma_{ixv}(m) \right\| \\
&= O_P\left[(nT)^{-1} \sum_{i=1}^{n} \sum_{h=1}^{T_i-1} \sum_{m=1}^{T_i-1} |b_{J_i}(h,m)| \right] = O_P[(nT)^{-1}V_{nT}]
\end{aligned} \tag{A14}$$

given Lemma A.1(ii), (A3) and Assumption 5. Combining (A12)–(A14) yields

$$V_{nT}^{-1/2}\hat{A}_2 = O_P(V_{nT}^{-1/2}) + O_P[(nT)^{-1}V_{nT}^{1/2}]. \tag{A15}$$

Using more tedious but analogous reasoning, we have for the third term $\hat{A}_3$ in (A6),

$$V_{nT}^{-1/2}\hat{A}_3 = O_P(V_{nT}^{-1/2}) + O_P[(nT)^{-1}V_{nT}^{1/2}]. \tag{A16}$$

Now we consider the fourth term $\hat{A}_4$ in (A8). By the Cauchy-Schwarz inequality and the fact that for each i, $\{v_{it}\}$ is i.i.d. with $Ev_{it}^8 \leq C$, we have $E(\bar{v}_i^2|T_i^{-1}\sum_{t=h+1}^{T_i} v_{it}||T_i^{-1}\sum_{t=m+1}^{T_i} v_{it}|) \leq CT_i^{-2}$ for $h, m > 0$. It follows from Markov's inequality, Lemma A.1(ii) and (A3) that

$$\begin{aligned} V_{nT}^{-1/2}\hat{A}_4 &\leq V_{nT}^{-1/2}\sum_{i=1}^{n} T_i \sum_{h=1}^{T_i-1}\sum_{m=1}^{T_i-1} |b_{J_i}(h,m)|\bar{v}_i^2 \left|T_i^{-1}\sum_{t=h+1}^{T_i} v_{it}\right|\left|T_i^{-1}\sum_{t=m+1}^{T_i} v_{it}\right| \\ &= O_P\left[V_{nT}^{-1/2}T^{-1}\sum_{i=1}^{n}\sum_{h=1}^{T_i-1}\sum_{m=1}^{T_i-1}|b_{J_i}(h,m)|\right] = O_P(T^{-1}V_{nT}^{1/2}). \end{aligned} \tag{A17}$$

Similarly, using more tedious but analogous reasoning to that for $\hat{A}_4$, we have

$$V_{nT}^{-1/2}\hat{A}_5 = O_P(T^{-1}V_{nT}^{1/2}). \tag{A18}$$

Next, for the sixth term $\hat{A}_6$ in (A8), noting that v_{it} and $\bar{v}_{t-h}$ are independent for $h > 0$ under $\mathbb{H}_0$, we have $E(|\bar{v}_{t-h}\bar{v}_{t-m}||T_i^{-1}\sum_{t=h+1}^{T_i} v_{it}||T_i^{-1}\sum_{t=m+1}^{T_i} v_{it}|) \leq Cn^{-1}T_i^{-1}$ for $h, m > 0$ by the Cauchy-Schwarz inequality and $Ev_{it}^8 \leq C$. It follows that

$$V_{nT}^{-1/2}\hat{A}_6 = O_P(n^{-1}V_{nT}^{1/2}). \tag{A19}$$

Similarly, we can also obtain

$$V_{nT}^{-1/2}\hat{A}_7 = O_P(n^{-1}V_{nT}^{1/2}). \tag{A20}$$

Finally, because $E(\bar{v}^2|T_i^{-1}\sum_{t=h+1}^{T_i} v_{it}||T_i^{-1}\sum_{t=m+1}^{T_i} v_{it}|) \leq Cn^{-1}T_i^{-2}$ for $h, m > 0$ under $\mathbb{H}_0$, we have

$$V_{nT}^{-1/2}\hat{A}_c = O_P\left[(nT)^{-1}V_{nT}^{1/2}\right], \qquad c = 8, 9. \tag{A21}$$

Collecting (A8), (A10) and (A15)-(A21), we obtain the desired result that

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

Proof of Proposition A.2: Using (A7), we can write

$$\begin{aligned} \sum_{i=1}^{n} 2\pi T_i \sum_{j=0}^{J_i}\sum_{k=1}^{2^j}(\hat{\alpha}_{ijk} - \bar{\alpha}_{ijk})\bar{\alpha}_{ijk} &= \sum_{c=1}^{9}\left\{\sum_{i=1}^{n} 2\pi T_i \sum_{j=0}^{J_i}\sum_{k=1}^{2^j}\left[\sum_{h=1-T_i}^{T_i-1}\hat{\xi}_{ci}(h)\hat{\psi}_{jk}(2\pi h)\right]\bar{\alpha}_{ijk}\right\} \\ &= \sum_{c=1}^{9}\hat{\delta}_c, \text{ say.} \end{aligned} \tag{A22}$$

From the Cauchy-Schwarz inequality, (A10) and (A21), we have

$$\begin{aligned} V_{nT}^{-1/2}\left|\hat{\delta}_1+\hat{\delta}_8+\hat{\delta}_9\right| &\leq V_{nT}^{-1/2}\left(\hat{A}_1+\hat{A}_8+\hat{A}_9\right)^{1/2}\left(\sum_{i=1}^{n}2\pi T_i\sum_{j=0}^{J_i}\sum_{k=1}^{2^j}\bar{\alpha}_{ijk}^2\right)^{1/2} \\ &= O_P[n^{-3/4}V_{nT}^{1/4}+(V_{nT}/nT)^{1/2}], \qquad \text{(A23)} \end{aligned}$$

where $V_{nT}^{-1}\sum_{i=1}^n 2\pi T_i\sum_{j=0}^{J_i}\sum_{k=1}^{2^j}\bar{\alpha}_{ijk}^2=O_P(1)$ by Markov's inequality, Lemma A.1(v), (A4) and the fact that $E\bar{\alpha}_{ijk}^2\leq CT_i^{-1}\sum_{h=1-T_i}^{T_{i-1}-1}|\hat{\psi}_{jk}(2\pi h)|^2$.

Next, we consider the second term $\hat{\delta}_2$ in (A22). Using (A11), we write

$$\begin{aligned} \hat{\delta}_2 &= (\hat{\beta}-\beta)'\sum_{i=1}^n 2\pi T_i\sum_{j=0}^{J_i}\sum_{k=1}^{2^j}\left[\sum_{h=1-T_i}^{T_i-1}\Gamma_{ixv}(h)\hat{\psi}_{jk}(2\pi h)\right]\bar{\alpha}_{ijk} \\ &\quad +(\hat{\beta}-\beta)'\sum_{i=1}^n 2\pi T_i\sum_{j=0}^{J_i}\sum_{k=1}^{2^j}\left\{\sum_{h=1-T_i}^{T_i-1}[\tilde{\Gamma}_{ixv}(h)-\Gamma_{ixv}(h)]\hat{\psi}_{jk}(2\pi h)\right\}\bar{\alpha}_{ijk} \\ &\equiv (\hat{\beta}-\beta)'\hat{M}_3+(\hat{\beta}-\beta)'\hat{M}_4,\ \text{say.} \qquad \text{(A24)} \end{aligned}$$

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

$$\begin{aligned} E\hat{M}_3^2 &= \sum_{i=1}^n(2\pi T_i)^2E\left\|\sum_{j=0}^{J_i}\sum_{k=1}^{2^j}\left[\sum_{h=1-T_i}^{T_i-1}\Gamma_{ixv}(h)\hat{\psi}_{jk}(2\pi h)\right]\bar{\alpha}_{ijk}\right\|^2 \\ &= \sum_{i=1}^n T_i^2E\left\|\sum_{h=1}^{T_i-1}\sum_{m=1}^{T_i-1}b_{J_i}(h,m)\Gamma_{ixv}(h)\bar{R}_i(m)\right\|^2 \\ &= \sum_{i=1}^n T_i^2\sum_{h_1=1}^{T_i-1}\sum_{h_2=1}^{T_i-1}\sum_{m=1}^{T_i-1}b_{J_i}(h_1,m)b_{J_i}(h_2,m)\mathrm{tr}\left[\Gamma_{ixv}(h_1)\Gamma_{ixv}'(h_2)\right]E\bar{R}_i^2(m) \\ &\leq \sum_{i=1}^n\sigma_i^4T_i\left[\sum_{h=1}^{T_i-1}||\Gamma_{ixv}(h)||\right]^2\sup_{1\leq h\leq T_i-1}\left[\sum_{m=1}^{T_i-1}b_{J_i}^2(h,m)\right]^2 \\ &= O\left[T\sum_{i=1}^n(J_i+1)^2\right]=O(TV_{nT}) \end{aligned}$$

given Assumption 5, Lemma A.1(viii) and (A3). It follows from Chebyshev's inequality that

$$V_{nT}^{-1/2}(\hat{\beta}-\beta)'\hat{M}_3=O_P(n^{-1/2}). \qquad \text{(A25)}$$

For the second term $\hat{M}_4$ in (A24), we have

$$V_{nT}^{-1/2}|(\hat{\beta}-\beta)'\hat{M}_4|\leq V_{nT}^{-1/2}||\hat{\beta}-\beta||\hat{M}_2^{1/2}\left(\sum_{i=1}^n 2\pi T_i\sum_{j=0}^{J_i}\sum_{k=1}^{2^j}\bar{\alpha}_{ijk}^2\right)^{1/2}=O_P[(nT)^{-1}V_{nT}^{1/2}]. \qquad \text{(A26)}$$

where $\hat{M}_2=O_P[(nT)^{-1}V_{nT}]$ as shown in (A14). Collecting (A24)–(A26), we have

$$V_{nT}^{-1/2}\hat{\delta}_2=O_P(n^{-1/2})+O_P[(nT)^{-1}V_{nT}^{1/2}]. \qquad \text{(A27)}$$

Similarly, we have

$$V_{nT}^{-1/2}\hat{\delta}_3 = O_P(n^{-1/2}) + O_P[(nT)^{-1}V_{nT}^{1/2}]. \tag{A28}$$

We now consider $\hat{\delta}_4$ in (A22). We write $\hat{\delta}_4 = \sum_{i=1}^n T_i \sum_{h=1}^{T_i-1}\sum_{m=1}^{T_i-1} b_{J_i}(h,m)\hat{\xi}_{4i}(h)\bar{R}_i(m)$. It follows that

$$E\hat{\delta}_4^2 = \sum_{i=1}^n\sum_{l=1}^n T_iT_l \sum_{h_1=1}^{T_i-1}\sum_{h_2=1}^{T_i-1}\sum_{m_1=1}^{T_i-1}\sum_{m_2=1}^{T_i-1} b_{J_i}(h_1,m_1)b_{J_l}(h_2,m_2)E[\hat{\xi}_{4i}(h_1)\bar{R}_i(m_1)\hat{\xi}_{4l}(h_2)\bar{R}_l(m_2)]. \tag{A29}$$

Put $\bar{v}_i(h) \equiv -T_i^{-1}\sum_{i=h+1}^n v_{it}$. Then given the definition of $\hat{\xi}_{4i}(h)$ in (A4), we can write $\hat{\xi}_{4i}(h) = \bar{v}_i\bar{v}_i(h)$. Thus, for $i = l$, we have

$$\left|E[\hat{\xi}_{4i}(h_1)\bar{R}_i(m_1)\hat{\xi}_{4i}(h_2)\bar{R}_i(m_2)]\right| \le CT_i^{-3/2}T_l^{-3/2}, \tag{A30}$$

by the Cauchy-Schwarz inequality and the fact that when $\{v_{it}\}$ is i.i.d. for each i,

$$\left|E[\hat{\xi}_{4i}^2(h)\bar{R}_i^2(m)]\right| \le (E\bar{v}_i^8)^{1/2}[E\bar{v}_i^8(h)]^{1/2}[E\bar{R}_i^4(m)]^{1/2} \le CT_i^{-3}.$$

For $i \ne l$, $\hat{\xi}_{4i}(h_1)\bar{R}_i(m_1)$ is independent of $\hat{\xi}_{4l}(h_2)\bar{R}_l(m_2)$, so we have

$$\left|E[\hat{\xi}_{4i}(h_1)\bar{R}_i(m_1)\hat{\xi}_{4l}(h_2)\bar{R}_l(m_2)]\right| = \left|E[\hat{\xi}_{4i}(h_1)\bar{R}_i(m_1)]E[\hat{\xi}_{4l}(h_2)\bar{R}_l(m_2)]\right| \le CT_i^{-2}T_l^{-2}, \tag{A31}$$

where we used the fact that for $h, m > 0$

$$\left|E[\hat{\xi}_{4i}(h)\bar{R}_i(m)]\right| = \left|E[\bar{v}_i\bar{v}_i(h)\bar{R}_i(m)]\right| = \left|T_i^{-3}\sum_{t=1}^{T_i}\sum_{s=h+1}^{T_i}\sum_{\tau=m+1}^{T_i} E[v_{it}v_{is}v_{i\tau}v_{i\tau-m}]\right| \le CT^{-2}$$

given that $\{v_{it}\}$ is i.i.d. with $E(v_{it}^8) \le C$ for each i. Combining (A29)–(A31) and using Lemma A.1(ii) and (A3), we obtain

$$\begin{aligned} E\hat{\delta}_4^2 &\le CT^{-1}\sum_{i=1}^n\left[\sum_{h=1}^{T_i-1}\sum_{m=1}^{T_i-1}|b_{J_i}(h,m)|\right]^2 + CT^{-2}\left[\sum_{i=1}^n\sum_{h=1}^{T_i-1}\sum_{m=1}^{T_i-1}|b_{J_i}(h,m)|\right]^2 \\ &\le C(2^{\bar{J}}/T)V_{nT} + CT^{-2}V_{nT}^2, \end{aligned}$$

where, as before, $\bar{J} \equiv \max_{1\le i\le n}(J_i)$. Hence, by Chebyshev's inequality, we have

$$V_{nT}^{-1/2}\hat{\delta}_4 = O_P(2^{\bar{J}/2}/T^{1/2} + V_{nT}^{1/2}/T). \tag{A32}$$

Similarly, we can obtain

$$V_{nT}^{-1/2}\hat{\delta}_5 = O_P(2^{\bar{J}/2}/T^{1/2} + V_{nT}^{1/2}/T). \tag{A33}$$

Next, we consider $\hat{\delta}_6$. Write $\hat{\delta}_6 = \sum_{i=1}^n T_i\sum_{h=1}^{T_i-1}\sum_{m=1}^{T_i-1} b_{J_i}(h,m)\hat{\xi}_{6i}(h)\bar{R}_i(m)$, where $\hat{\xi}_{6i}(h) = T_i^{-1}\sum_{t=h+1}^{T_i} v_{it}\bar{v}_{t-h}$ as before. Then we have

$$E\hat{\delta}_6^2 = \sum_{i=1}^n\sum_{l=1}^n T_iT_l \sum_{h_1=1}^{T_i-1}\sum_{h_2=1}^{T_i-1}\sum_{m_1=1}^{T_i-1}\sum_{m_2=1}^{T_i-1} b_{J_i}(h_1,m_1)b_{J_l}(h_2,m_2)E[\hat{\xi}_{6i}(h_1)\bar{R}_i(m_1)\hat{\xi}_{6l}(h_2)\bar{R}_l(m_2)]. \tag{A34}$$

For $i = l$, we have

$$\left|E[\hat{\xi}_{6i}(h_1)\bar{R}_i(m_1)\hat{\xi}_{6i}(h_2)\bar{R}_i(m_2)]\right| \leq \left\{E[\hat{\xi}_{6i}^2(h_1)\bar{R}_i^2(m_1)]E[\hat{\xi}_{6i}^2(h_2)\bar{R}_i^2(m_2)]\right\}^{1/2} \leq CT_i^{-2}n^{-1}, \quad \text{(A35)}$$

where $E[\hat{\xi}_{6i}^2(h)\bar{R}_i^2(m)] \leq [E\hat{\xi}_{6i}^4(h)E\bar{R}_i^4(m)]^{1/2} \leq CT_i^{-2}n^{-1}$ given Assumption 1(b) and that $\{v_{it}\}$ is i.i.d. for each i (so that $\bar{v}_{it}$ and $\bar{v}_{t-h}$ are independent for $h > 0$). For $i \neq l$, we write $\bar{v}_{t-h} = \bar{v}_{t-h}(i,l) + n^{-1}(v_{it-h} + v_{lt-h})$, where $\bar{v}_{t-h}(i,l) \equiv n^{-1}\sum_{c=1,c\neq i,l}^{n} v_{ct-h}$ is independent of (v_{it-h}, v_{lt-h}). Then we have for $i \neq l$,

$$\begin{aligned}
& \left|E[\hat{\xi}_{6i}(h_1)\bar{R}_i(m_1)\hat{\xi}_{6l}(h_2)\bar{R}_l(m_2)]\right| \\
= & \left|T_i^{-1}T_l^{-1}\sum_{t_1=h_1+1}^{T_i}\sum_{t_2=h_2+1}^{T_i} E[v_{it_1}\bar{R}_i(m_1)v_{lt_2}\bar{R}_l(m_2)\bar{v}_{t_1-h_1}\bar{v}_{t_2-h_2}]\right| \\
= & \left|T_i^{-1}T_l^{-1}\sum_{t_1=h_1+1}^{T_i}\sum_{t_2=h_2+1}^{T_i} n^{-2}E[v_{it_1}\bar{R}_i(m_1)v_{lt_2}\bar{R}_l(m_2)(v_{it_1-h_1}+v_{lt_1-h_1})(v_{it_2-h_2}+v_{lt_2-h_2})]\right| \\
= & \left|T_i^{-1}T_l^{-1}\sum_{t_1=h_1+1}^{T_i}\sum_{t_2=h_2+1}^{T_i} n^{-2}E[v_{it_1}v_{it_1-h_1}\bar{R}_i(m_1)]E[v_{lt_2}v_{lt_2-h_2}\bar{R}_l(m_2)]\right| \\
& + \left|T_i^{-1}T_l^{-1}\sum_{t_1=h_1+1}^{T_i}\sum_{t_2=h_2+1}^{T_i} n^{-2}E[v_{it_1}v_{it_2-h_2}\bar{R}_i(m_1)]E[v_{lt_2}v_{lt_1-h_1}\bar{R}_l(m_2)]\right| \\
\leq & \; CT_i^{-1}T_l^{-1}n^{-2}, \qquad \text{(A36)}
\end{aligned}$$

where $|E[v_{it_1}v_{it_2-h_2}\bar{R}_i(m_1)]| = |T_i^{-1}\sum_{s=m_1+1}^{T_i} E(v_{it}v_{it_2-h_2}v_{is}v_{is-m_1})| \leq CT_i^{-1}$.

Combining (A34)–(A36) and Lemma A.1(ii) and (A3), we have

$$\begin{aligned}
E\hat{\delta}_6^2 & \leq Cn^{-1}\sum_{i=1}^{n}\left[\sum_{h=1}^{T_i-1}\sum_{m=1}^{T_i-1}|b_{J_i}(h,m)|\right]^2 + Cn^{-2}\left[\sum_{i=1}^{n}\sum_{h=1}^{T_i-1}\sum_{m=1}^{T_i-1}|b_{J_i}(h,m)|\right]^2 \\
& \leq C(2^{\bar{J}}/T)V_{nT} + Cn^{-2}V_{nT}^2.
\end{aligned}$$

It follows by Chebyshev's inequality and $2^{\bar{J}}/n \to 0$ that

$$V_{nT}^{-1/2}\hat{\delta}_6 = O_P(2^{\bar{J}/2}/T^{1/2} + V_{nT}^{1/2}/n) \xrightarrow{p} 0. \qquad \text{(A37)}$$

Similarly, we have

$$V_{nT}^{-1/2}\hat{\delta}_7 = O_P(2^{\bar{J}/2}/T^{1/2} + V_{nT}^{1/2}/n) \xrightarrow{p} 0 \qquad \text{(A38)}$$

Collecting (A23), (A27), (A28), (A32), (A33), (A37) and (A38) yields the desired result of Proposition A.2. ■

Proof of Theorem A.2: Recalling the definition of $\bar{\alpha}_{ijk}$ in (6.3), we can write

$$\sum_{i=1}^{n} 2\pi T_i \sum_{j=0}^{J_i}\sum_{k=1}^{2^j}\bar{\alpha}_{ijk}^2 = \sum_{i=1}^{n} T_i \sum_{h=1}^{T-1}\sum_{m=1}^{T-1} b_{J_i}(h,m)\bar{R}_i(h)\bar{R}_i(m) \equiv \sum_{i=1}^{n}(\hat{A}_i + \hat{B}_{1i} - \hat{B}_{2i} - \hat{B}_{3i}), \quad \text{(A39)}$$

where we have

$$
\begin{aligned}
\hat{A}_i &\equiv T_i^{-1}\sum_{h=1}^{T_i-1}\sum_{m=1}^{T_i-1} b_{J_i}(h,m)\left(\sum_{t=2}^{T_i}\sum_{s=1}^{t-1}+\sum_{s=2}^{T_i}\sum_{t=1}^{s-1}\right) v_{it}v_{it-j}v_{is}v_{is-m} \\
&= 2T_i^{-1}\sum_{h=1}^{T_i-1}\sum_{m=1}^{T_i-1} b_{J_i}(h,m)\sum_{t=2}^{T_i}\sum_{s=1}^{t-1} v_{it}v_{it-j}v_{is}v_{is-m} \quad \text{(by symmetry of } b_{J_i}(\cdot,\cdot)) \\
\hat{B}_{1i} &\equiv T_i^{-1}\sum_{h=1}^{T_i-1}\sum_{m=1}^{T_i-1} b_{J_i}(h,m)\sum_{t=1}^{T_i} v_{it}^2 v_{it-h}v_{it-m}, \\
\hat{B}_{2i} &\equiv T_i^{-1}\sum_{h=1}^{T-1}\sum_{m=1}^{T-1} b_{J_i}(h,m)\sum_{t=1}^{h}\sum_{s=m+1}^{T_i} v_{it}v_{it-h}v_{is}v_{is-m}, \\
\hat{B}_{3i} &\equiv T_i^{-1}\sum_{h=1}^{T-1}\sum_{m=1}^{T-1} b_{J_i}(h,m)\sum_{t=1}^{T_i}\sum_{s=1}^{m} v_{it}v_{it-h}v_{is}v_{is-m}.
\end{aligned}
$$

Proposition A.3 shows that the statistic $\sum_{i=1}^n \hat{A}_i$ dominates the other terms in (A39).

Proposition A.3: $V_{nT}^{-1/2}(\sum_{i=1}^n 2\pi T_i \sum_{j=0}^{J_i}\sum_{k=1}^{2^j}\bar{\alpha}_{ijk}^2 - M_{nT}) = V_{nT}^{-1/2}\sum_{i=1}^n \hat{A}_i + o_P(1)$.

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

$$
\begin{aligned}
\hat{A}_i &= 2T_i^{-1}\sum_{h=1}^{T_i-1}\sum_{m=1}^{T_i-1} b_{J_i}(h,m)\left(\sum_{t=q_i+2}^{T_i}\sum_{s=1}^{t-q_i-1}+\sum_{t=2}^{T_i}\sum_{s=\max(t-q_i,1)}^{t-1}\right) v_{it}v_{it-h}v_{is}v_{is-m} \\
&\equiv \hat{B}_i+\hat{B}_{4i}, \text{ say.}
\end{aligned} \tag{A40}
$$

Furthermore, we decompose

$$
\begin{aligned}
\hat{B}_i &= 2T_i^{-1}\left(\sum_{h=1}^{q_i}\sum_{m=1}^{q_i}+\sum_{h=1}^{q_i}\sum_{m=q_i+1}^{n-1}+\sum_{h=q_i+1}^{n-1}\sum_{m=1}^{n-1}\right) b_{J_i}(h,m)\sum_{t=q_i+2}^{T_i}\sum_{s=1}^{t-q_i-1} v_{it}v_{it-h}v_{is}v_{is-m} \\
&\equiv \hat{U}_i+\hat{B}_{5i}+\hat{B}_{6i}, \text{ say.}
\end{aligned} \tag{A41}
$$

where B_{5i} and $\hat{B}_{6i}$ are the contributions from $m>q_i$ and $h>q_i$ respectively.

Proposition A.4 below shows that $\sum_{i=1}^n \hat{A}_i$ can be approximated arbitrarily well by $\sum_{i=1}^n \hat{U}_i$ under proper conditions on q_i.

Proposition A.4: *Suppose that Assumptions 2–3 hold,* $2^{2\bar{J}}/T\to 0, q_i\equiv q_i(T_i)\to\infty, q_i/2^{J_i}\to\infty$ *and* $q_i^2/T_i\to 0$, *where* $\bar{J}\equiv\max_{1\le i\le n}(J_i)$. *If* $\{v_{it}\}$ *is i.i.d. for each* i, *then* $V_{nT}^{-1/2}\sum_{i=1}^n\hat{A}_i = V_{nT}^{-1/2}\sum_{i=1}^n\hat{U}_i + o_P(1)$.

It is much easier to show the asymptotic normality of $\sum_{i=1}^n \hat{U}_i$ than of $\sum_{i=1}^n \hat{A}_i$, because for $\hat{U}_i$, $\{v_{it}v_{it-h}\}$ and $\{v_{is}v_{is-m}\}$ are independent given $t-s>q_i$ and $0<h,m\le q_i$.

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

Propositions A.3–A.5 and Slutsky Theorem imply Theorem A.2. We now prove Propositions A.3–A.5.

Proof of Proposition A.3: Recalling the definition of M_{nT} in Theorem A.2. By (A39), we obtain

$$\sum_{i=1}^{n}\left(2\pi T_i\sum_{j=0}^{J_i}\sum_{k=1}^{2^j}\hat{\alpha}_{ijk}^2-M_{nT}\right)=\sum_{i=1}^{n}\hat{A}_i+\sum_{i=1}^{n}(\hat{B}_{1i}-\sigma_i^4M_{i0})-\sum_{i=1}^{n}\hat{B}_{2i}-\sum_{i=1}^{n}\hat{B}_{3i},$$

We shall show (i) $V_{nT}^{-1/2}(\sum_{i=1}^n\hat{B}_{1i}-M_{nT})\xrightarrow{p}0$; (ii) $V_{nT}^{-1/2}\sum_{i=1}^n\hat{B}_{2i}\xrightarrow{p}0$; (iii) $V_{nT}^{-1/2}\sum_{i=1}^n\hat{B}_{3i}\xrightarrow{p}0$.

(i) Observe that $\hat{B}_{1i}$ has the similar structure as $\hat{B}_{1n}$ in Lee and Hong (2000). (Note that the sample size n in Lee and Hong (2000) corresponds to our T_i here.) Following Lee and Hong's (2000) reasoning and using Lemma A.1(ii), we can obtain that for each i and for T_i sufficiently large

$$E(\hat{B}_{1i}-E\hat{B}_{1i})^2\leq CT_i^{-1}\left[\sum_{h=1}^{T_i-1}\sum_{m=1}^{T_i-1}|b_{J_i}(h,m)|\right]^2\leq C^2(2^{\bar{J}}/T)\sum_{h=1}^{T_i-1}\sum_{m=1}^{T_i-1}|b_{J_i}(h,m)|.$$

Also, because $\hat{B}_{1i}$ is a sequence independent across i, we have

$$E\left[\sum_{i=1}^{n}(\hat{B}_{1i}-E\hat{B}_{1i})\right]^2=\sum_{i=1}^{n}E(\hat{B}_{1i}-E\hat{B}_{1i})^2\leq C(2^{\bar{J}}/T)\sum_{i=1}^{n}\sum_{h=1}^{T_i-1}\sum_{m=1}^{T_i-1}|b_{J_i}(h,m)|=O(V_{nT}2^{\bar{J}}/T)$$

given (A3). Hence, by Chebyshev's inequality, $\sum_{i=1}^n E\hat{B}_{1i}=M_{nT}$ and $2^{2\bar{J}}/T\to0$, we have $V_{nT}^{-1/2}(\sum_{i=1}^n\hat{B}_{1i}-M_{nT})=O_P[(2^{\bar{J}}/T)^{1/2}]=o_P(1)$. This completes the proof for (i).

(ii) Next, we consider $\hat{B}_{2i}$. Following Lee and Hong's (2000) reasoning, we have $E\hat{B}_{2i}^2\leq CT_i^{-1}[\sum_{h=1}^{T_i-1}\sum_{m=1}^{T_i-1}|b_{J_i}(h,m)|]^3$. Then by the fact that $\hat{B}_{2i}$ is a zero-mean sequence independent across i, Lemma A.1(ii) and (A3), we have

$$E\left(\sum_{i=1}^{n}\hat{B}_{2i}\right)^2=\sum_{i=1}^{n}E\hat{B}_{2i}^2\leq C(2^{2\bar{J}}/T)\sum_{i=1}^{n}\sum_{h=1}^{T_i-1}\sum_{m=1}^{T_i-1}|b_{J_i}(h,m)|=O[(2^{2\bar{J}}/T)V_{nT}].$$

Hence, by Chebyshev's inequality and $2^{2\bar{J}}/T\to0$, we obtain $V_{nT}^{-1/2}\sum_{i=1}^n\hat{B}_{2i}\xrightarrow{p}0$.

(iii) By reasoning similar to (ii), we can obtain $V_{nT}^{-1/2}\hat{B}_{3i}=O_P(2^{\bar{J}}/T^{1/2})=o_P(1)$. ■

Proof of Proposition A.4: Given (A39) and (A41), we have $\hat{A}_i=\hat{U}_i+\hat{B}_{4i}+\hat{B}_{5i}+\hat{B}_{6i}$. It suffices to show $V_{nT}^{-1/2}\sum_{i=1}^n\hat{B}_{ci}\xrightarrow{p}0$ for $c=4,5,6$.

(i) We first consider $\hat{B}_{4i}$ as in (A40). From Lee and Hong's (2000, proof of Theorem 1) reasoning, we have for each i and for T_i sufficiently large

$$E\hat{B}_{4iT_i}^2\leq C(q_i/T_i)\left[\sum_{h=1}^{T_i-1}\sum_{m=1}^{T_i-1}|b_{J_i}(h,m)\right]^2\leq C^2(\bar{q}2^{\bar{J}}/T)\sum_{h=1}^{T_i-1}\sum_{m=1}^{T_i-1}|b_{J_i}(h,m)|,$$

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 $\{\hat{B}_{4i}\}$ is a zero-mean sequence independent across i, Lemma A.1(ii) and A(3), we have for all

T_i sufficiently large,

$$E\left(\sum_{i=1}^{n}\hat{B}_{4i}\right)^2=\sum_{i=1}^{n}E\hat{B}_{4i}^2\le C(\bar{q}2^{\bar{J}}/T)\sum_{i=1}^{n}\sum_{h=1}^{T_i-1}\sum_{m=1}^{T_i-1}|b_{J_i}(h,m)|=O(V_{nT}\bar{q}2^{\bar{J}}/T).$$

This, Chebyshev's inequality, $\bar{q}^2/T\to 0$ and $2^{2\bar{J}}/T\to 0$, imply $V_{nT}^{-1/2}\sum_{i=1}^{n}\hat{B}_{4i}\xrightarrow{p}0$.

(ii) Next, we consider $\hat{B}_{5i}$ as in (A41). By the definition of $b_{J_i}(h,m)$, the Cauchy-Schwarz inequality and Assumption 3, we obtain

$$\begin{aligned}E\hat{B}_{5i}^2 &= \sigma_i^4\sum_{h=1}^{T-1}\sum_{m>q_i}^{T-1}b_{J_i}^2(h,m)\le C\sum_{j=0}^{J_i}\sum_{m>q_i}|\hat{\psi}(2\pi m/2^j)|^2\\ &\le C^2\sum_{j=0}^{J_i}\sum_{m>q_i}|2\pi m/2^j|^{-2\tau}\le C^2 2^{2\tau J_i}/q_i^{2\tau-1}.\end{aligned}$$

Therefore, $E(\sum_{i=1}^{n}\hat{B}_{5i})^2=\sum_{i=1}^{n}E\hat{B}_{5i}^2\le C(2^{\bar{J}}/q_0)^{2\tau-1}\sum_{i=1}^{n}2^{J_i}$, where $q_0\equiv\min_{1\le i\le n}(q_i)$. It follows by Chebyshev's inequality and $2^{\bar{J}}/q_0\to 0$ that $V_{nT}^{-1/2}\sum_{i=1}^{n}\hat{B}_{5i}=O_P[(2^{\bar{J}}/q_0)^{2\tau-1}]=o_P(1)$.

(iii) Finally, we consider $\hat{B}_{6i}$, as in (A41). Following Lee and Hong's (2000, proof of Theorem 1) reasoning and using Lemma A.1(ii), we obtain

$$\begin{aligned}E\hat{B}_{6i}^2 &\le C2^{2\tau J_i}/q_i^{2\tau-1}+CT_i^{-1}\left[\sum_{h=1}^{T_i-1}\sum_{m=1}^{T_i-1}|b_{J_i}(h,m)|\right]^2\\ &\le C2^{J_i}(2^{J_i}/q_0)^{2\tau-1}+C(2^{\bar{J}}/T)\sum_{h=1}^{T_i-1}\sum_{m=1}^{T_i-1}|b_{J_i}(h,m)|.\end{aligned}$$

Thus, $V_{nT}^{-1/2}\sum_{i=1}^{N}\hat{B}_{6i}=O_P[(2^{\bar{J}}/q_0)^{2\tau-1}+(2^{\bar{J}}/T)^{1/2}]\xrightarrow{p}0$ by Chebyshev's inequality, (A3), $2^{\bar{J}}/q_0\to 0$ and $2^{2\bar{J}}/T\to 0$, This completes the proof. ■

Proof of Proposition A.5: We write $\hat{U}_i=T_i^{-1}\sum_{t=q_i+2}^{T_i}U_{it}$, where

$$U_{it}\equiv 2v_{it}\sum_{h=1}^{q_i}v_{it-h}H_{i,t-q_i-1}(h), \tag{A42}$$

$H_{i,t-q_i-1}(h)\equiv\sum_{m=1}^{q_i}b_{J_i}(h,m)S_{i,t-q_i-1}(m)$ and $S_{i,t-q_i-1}(m)\equiv\sum_{s=1}^{t-q_i-1}v_{is}v_{is-m}$. Then we have $\hat{U}=\sum_{t=q_0+2}^{\bar{T}}U_t$, where $\bar{T}\equiv\max_{1\le i\le n}(T_i), U_t\equiv\sum_{i=1}^{n}U_{it}\mathbf{1}(q_i\le t\le T_i)$, and $\mathbf{1}(\cdot)$ is the indicator function.

Put $\mathcal{F}_t\equiv\bigotimes_{i=1}^{n}\mathcal{F}_{it}$, where $\mathcal{F}_{it}$ is the sigma field generated by $\{v_{is},s\le t\}_{i=1}^{n}$. Because $\{v_{it}v_{it-h}\}$ is independent of $H_{i,t-q_i-1}(h)$ for $0<h\le q_i$, $\{U_t,\mathcal{F}_{t-1}\}$ is an adapted martingale difference sequence, with

$$E\hat{U}^2 = \sum_{t=q_0+2}^{\bar{T}}EU_t^2=\sum_{t=q_0+2}^{\bar{T}}\sum_{i=1}^{n}E(U_{it}^2)\mathbf{1}(q_i\le t\le T_i)$$

$$
\begin{aligned}
&= 2\sum_{i=1}^{n}\sigma_i^8(1-q_i/T_i)(1-(q_i+1)/T_i)\sum_{h=1}^{q_i}\sum_{m=1}^{q_i}b_{J_i}^2(h,m) \\
&= 4\left[\sum_{i=1}^{n}\sigma_i^8\sum_{h=1}^{T_i-1}\sum_{m=1}^{T_i-1}b_{J_i}^2(h,m)\right][1+o(1)] = V_{nT}[1+o(1)] \qquad \text{(A43)}
\end{aligned}
$$

given $q_i \to \infty, q_i/2^{J_i} \to \infty, \bar{q}^2/T \to 0$. It follows that $V_{nT}/E\hat{U}^2 \to 1$.

We now apply Brown's (1971) martingale limit theorem by verifying his two conditions: (i) $var^{-2}(\hat{U})\sum_{t=q_0+2}^{\bar{T}} E\{U_t^2\mathbf{1}[|U_t| \geq \epsilon var^{1/2}(\hat{U})]\} \to 0$ for all $\epsilon > 0$; (ii) $var^{-2}(\hat{U}^2)T^{-2}\sum_{t=q_0+2}^{\bar{T}} E(U_t^2|\mathcal{F}_{t-1})$ $\xrightarrow{p} 1$. Given (A41), we first verify (i) by showing that $V_{nT}^{-2}\sum_{t=q_0+2}^{\bar{T}} EU_t^4 \to 0$. Given t, $\{U_{it}\}$ is a zero-mean independent sequence across i, so we have

$$
EU_t^4 \leq C\left[\sum_{i=1}^{n} T_i^{-2}(EU_{it}^4)^{1/2}1(q_i \leq t \leq T_i)\right]^2.
$$

Moreover, following Lee and Hong's (2000, proof of Theorem 1) reasoning, we can obtain that for each i and for T_i sufficiently large, $EU_{it}^4 \leq Ct^2\sum_{h=1}^{q_i}\sum_{m=1}^{q_i} b_{J_i}^2(h,m)$. It follows that $V_{nT}^{-2}\sum_{t=1}^{\bar{T}} EU_t^4 = O(T^{-1}) \to 0$. Hence, condition (i) holds.

Next, we verify (ii) by showing that $V_{nT}^{-2}E(\tilde{U}^2 - E\hat{U}^2) \to 0$, where $\tilde{U}^2 \equiv \sum_{t=q_0+2}^{\bar{T}} E(U_t^2|\mathcal{F}_{t-1})$, $E(U_t^2|\mathcal{F}_{t-1}) \equiv E\{[T_i^{-1}\sum_{i=1}^{n} U_{it}\mathbf{1}(q_i \leq t \leq T_i)]^2|\mathcal{F}_{t-1}\} = \sum_{i=1}^{n} T_i^{-1}E(U_{it}^2|\mathcal{F}_{it-1})\mathbf{1}(q_i \leq t \leq T_i)$, where the second equality follows from the facts that for each t, $\{U_{it}\}$ is a zero-mean sequence independent across i, and that for each i, $\{U_{it}, \mathcal{F}_{it-1}\}$ is a martingale difference sequence. Lee and Hong (2000, p. 27) show that for each i and for T_i sufficiently large,

$$
E\left\{\sum_{t=q_i+2}^{T_i}[E(U_{it}^2|\mathcal{F}_{i,t-1}) - EU_{it}^2]\right\}^2 \leq C(\bar{q}/T)\left[\sum_{h=1}^{T_i-1}\sum_{m=1}^{T_i-1}|b_{J_i}(h,m)|\right]^2 + C(J_i+1)2^{J_i}.
$$

It follows that for T sufficiently large,

$$
\begin{aligned}
E\left(\tilde{U}^2 - E\tilde{U}^2\right)^2 &= E\left[\sum_{i=1}^{n}\sum_{t=q_i+2}^{T_i}[E(U_{it}^2|\mathcal{F}_{i,t-1}) - EU_{it}^2]\right]^2 = \sum_{i=1}^{n}E\left\{\sum_{t=q_i+2}^{T_i}[E(U_{it}^2|\mathcal{F}_{i,t-1}) - EU_{it}^2]\right\}^2 \\
&\leq C(\bar{q}/T)\sum_{i=1}^{n}\left[\sum_{h=1}^{T_i-1}\sum_{m=1}^{T_i-1}|b_{J_i}(h,m)|\right]^2 + C\sum_{i=1}^{n}(J_i+1)2^{J_i} \\
&\leq C(\bar{q}/T)V_{nT}^2 + C(\bar{J}+1)V_{nT},
\end{aligned}
$$

where the second equality follows from the fact that $\{\sum_{t=q_i+2}^{T_i}[E(U_{it}^2|\mathcal{F}_{i,t-1}) - EU_{it}^2]\}$ is a zero-mean independent sequence across i, and the last inequality follows from Lemma A.1(ii) and (A3). Consequently, given $\bar{q}^2/T \to 0$ and $2^{2\bar{J}}/T \to 0$, we have $V_{nT}^{-2}E(\tilde{U}^2 - E\tilde{U}^2)^2 = O(\bar{q}/T) + O[V_{nT}^{-2}\sum_{i=1}^{n}(J_i+1)2^{J_i}] \to 0$ as $n, T \to \infty$. Thus, condition (ii) hold, and so $V_{nT}^{-1/2}\hat{U}_n \xrightarrow{d} N(0,1)$ by Brown's Theorem. ■

Proof of Theorem A.3: (i) Recalling the definition of $\hat{M}$ and M_{nT}, we have

$$\begin{aligned}\hat{M}-M_{nT} &= \sum_{i=1}^{n}[\hat{R}_i(0)-R_i(0)]^2\sum_{h=1}^{T_i-1}b_{J_i}(h,h)+2\sum_{i=1}^{n}[\hat{R}_i(0)-R_i(0)]R_i(0)\sum_{h=1}^{T_i-1}b_{J_i}(h,h)\\ &\equiv \hat{M}_5+2\hat{M}_6,\ \text{say.} \end{aligned} \tag{A44}$$

Because $\hat{R}_i(0)-R_i(0)=[\hat{R}_i(0)-\tilde{R}_i(0)]+[\tilde{R}_i(0)-\bar{R}_i(0)]+[\bar{R}_i(0)-R_i(0)]$, we have

$$\begin{aligned}\hat{M}_5 &\leq 4\sum_{i=1}^{n}[\hat{R}_i(0)-\tilde{R}_i(0)]^2\sum_{h=1}^{T_i-1}b_{J_i}(h,h)+4\sum_{i=1}^{n}[\tilde{R}_i(0)-\bar{R}_i(0)]^2\sum_{h=1}^{T_i-1}b_{J_i}(h,h)\\ &\quad +4\sum_{i=1}^{n}[\bar{R}_i(0)-R_i(0)]^2\sum_{h=1}^{T_i-1}b_{J_i}(h,h)\equiv 4\hat{M}_{51}+4\hat{M}_{52}+4\hat{M}_{53},\ \text{say.}\end{aligned} \tag{A45}$$

Using (A3), the Cauchy-Schwarz inequality, Assumptions 4 and 5, Lemma A.1(v) and (A4), we have

$$\begin{aligned}V_{nT}^{-1/2}\hat{M}_{51} &\leq 4V_{nT}^{-1/2}||\hat{\beta}-\beta||^4\sum_{i=1}^{n}||\tilde{\Gamma}_{ixx}(0)||^4\sum_{h=1}^{T_i-1}b_{J_i}(h,h)\\ &\quad +4V_{nT}^{-1/2}||\hat{\beta}-\beta||^2\sum_{i=1}^{n}||\tilde{\Gamma}_{ixv}(0)||^2\sum_{h=1}^{T_i-1}b_{J_i}(h,h)\\ &\quad +4V_{nT}^{-1/2}||\hat{\beta}-\beta||^2\sum_{i=1}^{n}||\tilde{\Gamma}_{ivx}(0)||^2\sum_{h=1}^{T_i-1}b_{J_i}(h,h)=O_P[(nT)^{-1}V_{nT}^{1/2}].\end{aligned} \tag{A46}$$

Similarly, using (A4), the Cauchy-Schwarz inequality, and Markov's inequality, the i.i.d. property of $\{v_{it}\}$ for each i, and the spatial independence between $\{v_{it}\}$ and $\{v_{js}\}$ for all $i\neq j$, we can obtain $E[\tilde{R}_i(0)-\bar{R}_i(0)]^2\leq C(T_i^{-2}+n^{-1}T_i^{-1})$. It follows from Markov's inequality, the Cauchy-Schwarz inequality, Lemma A.1(v) and (A4) that

$$V_{nT}^{-1/2}\hat{M}_{52}=O_P[(T^{-2}+n^{-1}T^{-1})V_{nT}^{1/2}]. \tag{A47}$$

Using Markov's inequality, $E[\bar{R}_i(0)-R_i(0)]^2\leq CT_i^{-1}$, Lemma A.1(v) and (A4), we have

$$V_{nT}^{-1/2}\hat{M}_{53}=O_P(T^{-1}V_{nT}^{1/2}). \tag{A48}$$

Combining (A46)–(A48) and $V_{nT}/T^2\to 0$, we obtain

$$V_{nT}^{-1/2}\hat{M}_5=O_P(T^{-1}V_{nT}^{1/2})=o_P(1). \tag{A49}$$

Next, we consider the second term $\hat{M}_6$ in (A44). We write

$$\begin{aligned}\hat{M}_6 &= \sum_{i=1}^{n}[\hat{R}_i(0)-\tilde{R}_i(0)]R_i(0)\sum_{h=1}^{T_i-1}b_{J_i}(h,h)+\sum_{i=1}^{n}[\tilde{R}_i(0)-\bar{R}_i(0)]R_i(0)\sum_{h=1}^{T_i-1}b_{J_i}(h,h)\\ &\quad +\sum_{i=1}^{n}[\bar{R}_i(0)-R_i(0)]R_i(0)\sum_{h=1}^{T_i-1}b_{J_i}(h,h)\equiv \hat{M}_{61}+\hat{M}_{62}+\hat{M}_{63},\ \text{say.}\end{aligned} \tag{A50}$$

Using (A5) with $h=0$, Assumptions 4 and 5, Lemma A.1(v) and (A3), we have

$$\begin{aligned} V_{nT}^{-1/2}|\hat{M}_{61}| &\leq V_{nT}^{-1/2}||\hat{\beta}-\beta||^2 \sum_{i=1}^{n} ||\tilde{\Gamma}_{ixx}(0)||R_i(0) \sum_{h=1}^{T_i-1} b_{J_i}(h,h) \\ &\quad + V_{nT}^{-1/2}||\hat{\beta}-\beta|| \sum_{i=1}^{n} ||\tilde{\Gamma}_{ixv}(0)||R_i(0) \sum_{h=1}^{T_i-1} b_{J_i}(h,h) \\ &\quad + V_{nT}^{-1/2}||\hat{\beta}-\beta|| \sum_{i=1}^{n} ||\tilde{\Gamma}_{ivx}(0)||R_i(0) \sum_{h=1}^{T_i-1} b_{J_i}(h,h) = O_P[(nT)^{-1/2}V_{nT}^{1/2}]. \quad \text{(A51)} \end{aligned}$$

Also, using (A6) with $h=0$, $E[\tilde{R}_i(0)-\bar{R}_i(0)]^2 \leq C(T_i^{-2}+(nT_i)^{-1})$, the Cauchy-Schwarz inequality and Markov's inequality, Lemma A.1(ii) and (A4), we have

$$V_{nT}^{-1/2}|\hat{M}_{62}| = O_P[T^{-1}V_{nT}^{1/2} + (nT)^{-1/2}V_{nT}]. \quad \text{(A52)}$$

Finally, noting that $\bar{R}_i(0)-R_i(0)$ is a zero-mean sequence independent across i, we have

$$\begin{aligned} E\hat{M}_{63}^2 &= \sum_{i=1}^{n} E[\bar{R}_i(0)-R_i(0)]^2 \left[\sum_{h=1}^{T_i-1}\sum_{m=1}^{T_i-1} b_{J_i}(h,h)\right]^2 \leq CT^{-1}\sum_{i=1}^{n}\left[\sum_{h=1}^{T_i-1}\sum_{m=1}^{T_i-1} b_{J_i}(h,h)\right]^2 \\ &\leq C^2(2^{\bar{J}}/T)\sum_{i=1}^{n}\sum_{h=1}^{T_i-1}\sum_{m=1}^{T_i-1} b_{J_i}(h,h), \end{aligned}$$

where the last two equalities follows by Lemma A.1(v) and (A4). It follows by Chebyshev's inequality that

$$V_{nT}^{-1/2}\hat{M}_{63} = O_P(2^{\bar{J}/2}/T^{1/2}). \quad \text{(A53)}$$

Collecting (A50)–(A53) and $2^{2\bar{J}}/T \to 0$ yields $V_{nT}^{-1/2}\hat{M}_6 \xrightarrow{p} 0$. This, together with (A44) and (A49), implies $V_{nT}^{-1/2}(\hat{M}-M_{nT}) \xrightarrow{p} 0$. This proves (i).

(ii) The proof for $\hat{V}/V_{nT} \xrightarrow{p} 1$ is analogous to (i). We thus omit it here. ■

B Proof of Theorem 2

To conserve space, we only show for $\tilde{W}_1$. The proof for $\tilde{W}_2$ is similar. Put $\tilde{M} \equiv \sum_{i=1}^{n} \hat{R}_i^2(0)(2^{J_i+1}-1)$ and $\tilde{V} \equiv 4\sum_{i=1}^{n} \hat{R}_i^4(0)(2^{J_i+1}-1)$. Then we can write

$$\tilde{W}_1 - \hat{W}_1 = \hat{W}_1\left(\sqrt{\hat{V}}/\sqrt{\tilde{V}}-1\right) + \frac{\hat{M}-\tilde{M}}{\sqrt{\hat{V}}}\left(\sqrt{\hat{V}}/\sqrt{\tilde{V}}\right),$$

where $\hat{M}$ and $\hat{V}$ are as in (3.18). Because $\hat{W}_1 = O_P(1)$ by Theorem 1, $V_{nT}^{-1/2}(\hat{M}-M_{nT}) \xrightarrow{p} 0$ and $\hat{V}/V_{nT} \xrightarrow{p} 1$ by Theorem A.3, it suffices for $\tilde{W}_1-\hat{W}_1 \xrightarrow{p} 0$ and $\tilde{W}_1 \xrightarrow{d} N(0,1)$ if (i) $V_{nT}^{-1/2}(\tilde{M}-M_{nT}) \xrightarrow{p} 0$ and (ii) $\tilde{V}/V_{nT} \xrightarrow{p} 1$.

We first show (i). Following the reasoning analogous to the proof of Theorem A.3, we obtain $V_{nT}^{-1/2}(\tilde{M}-M_{nT}^0) \xrightarrow{p} 0$, where $M_{nT}^0 \equiv \sum_{i=1}^{n}\sigma_i^4(2^{J_i+1}-1)$. Thus, it remains to show $V_{nT}^{-1/2}(M_{nT}^0 -$

$M_{nT}) \to 0$. This follows from Lemma A.1(v), $2^{J_i+1} = a_i T^\nu, n/T^\nu \log_2^2 T \to 0$ and $n/T^{2(2\tau-1)-2(2\tau-\frac{1}{2})\nu} \to 0$ because

$$\begin{aligned} V_{nT}^{-1/2} \left| M_{nT} - M_{nT}^0 \right| &\leq CV_{nT}^{-1/2} \left[\sum_{i=1}^{n} (J_i+1) + (2^{\bar{J}}/T)^{(2\tau-1)} \sum_{i=1}^{n} 2^{J_i} \right] \\ &\leq C\nu n^{1/2} T^{-\nu/2} \log_2(T) + Cn^{1/2}/T^{[(2\tau-1)-(2\tau-\frac{1}{2})\nu]} \to 0. \end{aligned}$$

Now we show (ii). Put $V_{nT}^0 \equiv \sum_{i=1}^n \sigma_i^8 (2^{J_i+1}-1)$. Following reasoning analogous to that for Theorem A.3, we can obtain $\tilde{V}_{nT}/V_{nT}^0 \xrightarrow{p} 1$. It remains to show $V_{nT}/V_{nT}^0 \to 1$. This follows from Lemma A.1(vi) and $J_i \to \infty$, because

$$V_{nT}^{-1} \left| V_{nT}^0 - V_{nT} \right| \leq CV_{nT}^{-1} \left[\sum_{i=1}^{n} (J_i+1)^2 + (2^{\bar{J}}/T)^{(2\tau-1)} \sum_{i=1}^{n} 2^{J_i} \right] \to 0$$

where $V_{nT}^{-1} \sum_{i=1}^n (J_i+1)^2 \to 0$ given (A3) and $J_i \to \infty$. Note that we only require $J_i \to \infty$ for (ii). ■

C Proof of Theorem 3

Recall the definition of $\hat{M}$ and $\hat{V}$ as in (3.18). Following reasoning analogous to that 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

$$(n_A T)^{-1} \hat{V}^{1/2} \hat{W}_1 = (n_A T)^{-1} \sum_{i=1}^{n} 2\pi T_i \sum_{j=0}^{J_i} \sum_{k=1}^{2^j} \hat{\alpha}_{ijk}^2 + o_P(1), \tag{A54}$$

given $M_{nT} = O(V_{nT})$ by (A3) and A(4), and $V_{nT}/n_A T \to 0$ by (A3) and $(n_A T)^{-1} \sum_{i=1}^n 2^{J_i} \to 0$. It remains to show (i) $(n_A T)^{-1} \sum_{i=1}^n 2\pi T_i \sum_{j=0}^{J_i} \sum_{k=1}^{2^j} (\hat{\alpha}_{ijk}^2 - \alpha_{ijk}^2) \xrightarrow{p} 0$; (ii) $n_A^{-1} \sum_{i=1}^n 2\pi T_i \sum_{j=0}^{J_i} \sum_{k=1}^{2^j} \alpha_{ijk}^2 = (n_A T)^{-1} \sum_{i=1}^n 2\pi c_i Q(f_i, f_{i0}) + o(1)$, where α_{ijk} is defined in (3.11) or (3.12).

We first show (i). Because

$$\begin{aligned} (n_A T)^{-1} \sum_{i=1}^{n} 2\pi T_i \sum_{j=0}^{J_i} \sum_{k=1}^{2^i} (\hat{\alpha}_{ijk}^2 - \alpha_{ijk}^2) &= (n_A T)^{-1} \sum_{i=1}^{n} 2\pi T_i \sum_{j=0}^{J_i} \sum_{k=1}^{2^i} (\hat{\alpha}_{ijk} - \alpha_{ijk})^2 \\ &\quad + 2(n_A T)^{-1} \sum_{i=1}^{n} 2\pi T_i \sum_{j=0}^{J_i} \sum_{k=1}^{2^i} (\hat{\alpha}_{ijk} - \alpha_{ijk}) \alpha_{ijk}, \end{aligned} \tag{A55}$$

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

$$\begin{aligned} \sum_{i=1}^{n} 2\pi T_i \sum_{j=0}^{J_i} \sum_{k=1}^{2^j} (\hat{\alpha}_{ijk} - \alpha_{ijk})^2 &\leq 2 \sum_{i=1}^{n} 2\pi T_i \sum_{j=0}^{J_i} \sum_{k=1}^{2^j} (\hat{\alpha}_{ijk} - \bar{\alpha}_{ijk})^2 + 2 \sum_{i=1}^{n} 2\pi T_i \sum_{j=0}^{J_i} \sum_{k=1}^{2^j} (\bar{\alpha}_{ijk} - \alpha_{ijk})^2 \\ &\equiv 2\hat{M}_{71} + 2\hat{M}_{72}, \text{ say.} \end{aligned} \tag{A56}$$

For the first term in (A56), we note that Proposition A.1 continues to hold under Assumptions 1–6 (the proof is similar but more tedious than under the condition that $\{v_{it}\}$ is i.i.d. for each i.) It follows that

$$\hat{M}_{71} = O_P[(n_A T)^{-1} + (n_A n T)^{-1} V_{nT} + (n_A T^2)^{-1} V_{nT}]. \tag{A57}$$

For the second term in (A56), we further decompose

$$\begin{aligned} \hat{M}_{72} &\leq 2\sum_{i=1}^{n} 2\pi T_i \sum_{j=0}^{J_i}\sum_{k=1}^{2^j} (\bar{\alpha}_{ijk} - E\bar{\alpha}_{ijk})^2 + 2\sum_{i=1}^{n} 2\pi T_i \sum_{j=0}^{J_i}\sum_{k=1}^{2^j} (E\bar{\alpha}_{ijk} - \alpha_{ijk})^2 \\ &\equiv 2\hat{M}_{721} + 2\hat{M}_{722}, \text{ say.} \end{aligned} \tag{A58}$$

We now consider the first term in (A58). Under Assumption 6, we have $\sup_{1\leq h\leq T_i-1} \text{var}[\bar{R}_i(h)] \leq CT_i^{-1}$, which follows from $\sum_{h=1}^{\infty} R_i^2(h) \leq C, \sum_{j=-\infty}^{\infty}\sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty} |\kappa_i(j,k,l)| \leq C$, and

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

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

$$\begin{aligned} \hat{M}_{721} &= \sum_{i=1}^{n} T_i \sum_{h=1}^{T_i-1}\sum_{m=1}^{T_i-1} b_{J_i}(h,m) E\left\{\left[\bar{R}_i(h) - R_i(h)\right]\left[\bar{R}_i(m) - R_i(m)\right]\right\} \\ &\leq \sum_{i=1}^{n} T_i \sup_{1\leq h\leq T_i-1} var[\bar{R}_i(h)] \sum_{h=1}^{T_i-1}\sum_{m=1}^{T_i-1} |b_{J_i}(h,m)| = O(V_{nT}). \end{aligned} \tag{A59}$$

For the second term in (A58), noting that $|E\bar{\alpha}_{ijk} - \alpha_{ijk}| \leq T_i^{-1}\sum_{h=-\infty}^{\infty} |hR_i(h)\hat{\psi}_{jk}(2\pi h)|$, we have

$$\begin{aligned} \hat{M}_{722} &\leq \sum_{i=1}^{n}\sum_{j=1}^{J_i}\sum_{k=1}^{2^j} 2\pi T_i^{-1}\left[\sum_{h=1-T_i}^{T_i-1} R_i^2(h)\right]\left[\sum_{h=-\infty}^{\infty} h^2|\hat{\psi}_{jk}(2\pi h)|^2\right] \\ &= O\left[(2^{2\bar{J}}/T)\sum_{i=1}^{n} 2^{J_i}\right] = o(V_{nT}) \end{aligned} \tag{A60}$$

given $2^{2\bar{J}}/T \to 0$ and Assumption A.3. The latter ensures

$$\sum_{j=0}^{J_i}\sum_{k=1}^{2^j}\sum_{h=1-T_i}^{T_i-1} h^2|\hat{\psi}_{jk}(2\pi h)|^2 \leq \sum_{j=0}^{J_i} 2^{3j}\left[(2\pi/2^j)\sum_{h=-\infty}^{\infty}(2\pi h/2^j)^2|\hat{\psi}(2\pi h/2^j)|^2\right] = C2^{3J_i},$$

where $(2\pi/2^j)\sum_{h=-\infty}^{\infty}(2\pi h/2^j)^2|\hat{\psi}(2\pi h/2^j)|^2 \leq C\int_{-\infty}^{\infty} z^2/(1+|z|)^{2\tau}dz < \infty$ given $\tau > \frac{3}{2}$.

It follows by Markov's inequality and (A59)–(A60) that

$$(n_A T)^{-1}\hat{M}_{72} = O_P[(n_A T)^{-1} V_{nT}]. \tag{A61}$$

Combining (A57)–(A61) and $V_{nT}/(n_A T) \to 0$, we then obtain (i).

Next, we show (ii). This follows immediately because the cumulative squared bias

$$(n_AT)^{-1}\sum_{i=1}^{n}2\pi T_i\sum_{j=0}^{\infty}\sum_{k=1}^{2^j}\alpha_{ijk}^2-(n_AT)^{-1}\sum_{i=1}^{n}2\pi T_i\sum_{j=0}^{J_i}\sum_{k=1}^{2^j}\alpha_{ijk}^2$$

$$=(n_AT)^{-1}\sum_{i=1}^{n}2\pi T_i\sum_{j=J_i+1}^{\infty}\sum_{k=1}^{2^j}\alpha_{ijk}^2\leq C\sup_{i\in\mathbb{N}_A}\sum_{j=J_i+1}^{\infty}\sum_{k=1}^{2^j}\alpha_{ijk}^2\to 0$$

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

D Proof of Theorem 4

Given $T_i=c_iT$ and $2^{J_i+1}=a_iT_i^{\nu}$, we have $2^{J_i+1}=b_iT^{\nu}$, where $b_i\equiv a_ic_i^{\nu}$. Then as $T\to\infty$,

$$V_{nT}^0\equiv n^{-1}\sum_{i=1}^{n}\sigma_i^8(2^{J_i+1}-1)=T^{\nu}\left(n^{-1}\sum_{i=1}^{n}\sigma_i^8b_i\right)[1+o(1)]=\bar{b}T^{\nu}[1+o(1)],$$

where $\bar{b}\equiv n^{-1}\sum_{i=1}^{n}\sigma_i^8b_i$. It follows from Theorem 2 and $\hat{V}/V_{nT}^0\to^p 1$ that

$$(n_AT)^{-1}(\bar{b}T^{\nu})^{1/2}\hat{W}_1 = n_A^{-1}\sum_{i\in\mathbb{N}_A}c_iQ(f_i,f_{i0})+o_P(1), \quad \text{(A62)}$$

$$n_AT)^{-1}(\bar{b}T^{\nu})^{1/2}\hat{W}_2 = n_A^{-1}\sum_{i\in\mathbb{N}_A}\sqrt{\frac{\bar{b}}{\sigma_i^8b_i}}c_iQ_i+o_P(1). \quad \text{(A63)}$$

For $c=1,2$, we put $S_{nT}^{(c)}\equiv-2\ln[1-\Phi(\hat{W}_c)]$, the minus twice the logarithm of the asymptotic p-value of the test statistic W_c. Because $\ln[1-\Phi(z)]=-\frac{1}{2}z^2[1+o(1)]$ as $z\to+\infty$, where $\Phi(\cdot)$ is the N$(0,1)$ CDF (cf. Bahadur 1960, Section 5), it follows from (A62) and A(63) that

$$(n_AT)^{-2}\bar{b}T^{\nu}S_{nT}^{(1)} = \left[n_A^{-1}\sum_{i\in\mathbb{N}_A}c_iQ(f_i,f_{i0})\right]^2+o_P(1), \quad \text{(A64)}$$

$$(n_AT)^{-2}\bar{b}T^{\nu}S_{nT}^{(2)} = \left[n_A^{-1}\sum_{i\in\mathbb{N}_A}\sqrt{\frac{\bar{b}}{\sigma_i^8b_i}}c_iQ(f_i,f_{i0})\right]^2+o_P(1). \quad \text{(A65)}$$

Suppose that $\{T_i^{(1)}\}_{i=1}^{n^{(1)}}$ 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 $S_{n^{(1)}T^{(1)}}^{(1)}/S_{n^{(2)}T^{(2)}}^{(2)}\to 1$ as $n^{(1)},n^{(2)},T^{(1)}$ and $T^{(2)}\to\infty$. Then Bahadur's asymptotic relative efficiency of $\hat{W}_1$ to $\hat{W}_2$

$$BE\left(\hat{W}_1:\hat{W}_2\right) \equiv \lim\frac{\sum_{i=1}^{n^{(2)}}T_i^{(2)}}{\sum_{i=1}^{n^{(1)}}T_i^{(1)}}=\lim\frac{(\frac{1}{n^{(2)}}\sum_{i=1}^{n^{(2)}}c_i)n^{(2)}T^{(2)}}{(\frac{1}{n^{(1)}}\sum_{i=1}^{n^{(1)}}c_i)n^{(1)}T^{(1)}}$$

$$= \lim\left[\frac{T^{(2)}}{T^{(1)}}\right]^{1+\kappa}=\left(\frac{\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\sqrt{\frac{\bar{b}}{\sigma_i^8b_i}}c_iQ(f_i,f_{i0})}{\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}c_iQ(f_i,f_{i0})}\right)^{\frac{1+\kappa}{3-\nu}},$$

where the third equality follows from $n^{(c)} = \gamma[T^{(c)}]^{\kappa}$ for $c = 1, 2$, and the last equality follows from (A64) and (A65). It follows that $BE(\hat{W}_1 : \hat{W}_2) > 1$ if $Q(f_i, f_{i0})/\sqrt{b_i} > Q(f_j, f_{j0})/\sqrt{b_j} > 1$, which occurs when a_i is a monotonically increasing function of $Q(f_i, f_{i0})$ and $c_i = c$ (i.e., $T_i = T$) for all i. In this case, therefore, $\hat{W}_1$ is asymptotically more powerful than $\hat{W}_2$. ■

E Proof of Theorem 5

For space, we only consider $\hat{W}_1(\hat{J})$ here. The proof for $\hat{W}_2(\hat{J})$ is similar. We write

$$\hat{W}(\hat{J}) - \hat{W}(J) =$$

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

Given $\hat{W}_1(J) = O_P(1)$ by Theorem 1 and $\hat{V}(J)/V_{nT} \to^p 1$ by Theorem A.3, it suffices for $\hat{W}(\hat{J}) - \hat{W}(J) \xrightarrow{p} 0$ and $W(\hat{J}) \xrightarrow{d} N(0,1)$ if (i) $V_{nT}^{-1/2}\{\sum_{i=1}^{n} 2\pi T_i \sum_{j=J}^{\hat{J}} \sum_{k}^{2^j} \hat{\alpha}_{ijk}^2 - [\hat{M}(\hat{J}) - \hat{M}(J)]\} \xrightarrow{p} 0$ and (ii) $\hat{V}(J)/\hat{V}(J) \xrightarrow{p} 1$.

We first show (i). Decompose

$$\begin{aligned}
\sum_{i=1}^{n} 2\pi T_i \sum_{j=J}^{\hat{J}} \sum_{k=1}^{2^j} \hat{\alpha}_{ijk}^2 &= \sum_{i=1}^{n} 2\pi T_i \sum_{j=J}^{\hat{J}} \sum_{k=1}^{2^j} (\hat{\alpha}_{ijk} - \bar{\alpha}_{ijk})^2 + \sum_{i=1}^{n} 2\pi T_i \sum_{j=J}^{\hat{J}} \sum_{k=1}^{2^j} \bar{\alpha}_{ijk}^2 \\
&\quad + 2\sum_{i=1}^{n} 2\pi T_i \sum_{j=J}^{\hat{J}} \sum_{k=1}^{2^j} (\hat{\alpha}_{ijk} - \bar{\alpha}_{ijk})\bar{\alpha}_{ijk} \\
&\equiv \hat{G}_1 + \hat{G}_2 + 2\hat{G}_3, \text{ say.} \qquad \text{(A66)}
\end{aligned}$$

For the first term in (A66), we write

$$\hat{G}_1 = \sum_{i=1}^{n} 2\pi T_i \sum_{j=0}^{\hat{J}} \sum_{k=1}^{2^j} (\hat{\alpha}_{ijk} - \bar{\alpha}_{ijk})^2 - \sum_{i=1}^{n} 2\pi T_i \sum_{j=0}^{J} \sum_{k=1}^{2^j} (\hat{\alpha}_{ijk} - \bar{\alpha}_{ijk})^2 \equiv \hat{G}_{11} - \hat{G}_{12}. \qquad \text{(A67)}$$

By Proposition A.1, we have $V_{nT}^{-1/2}\hat{G}_{12} \xrightarrow{p} 0$. For the first term in (A67), we have for any given constants $M > 0$ and $\epsilon > 0$,

$$P\left(\hat{G}_{11} > \epsilon\right) \leq P\left(\hat{G}_{11} > \epsilon, M2^{J/2}|2^{\hat{J}}/2^J - 1| \leq \epsilon\right) + P\left(M2^{J/2}|2^{\hat{J}}/2^J - 1| > \epsilon\right). \qquad \text{(A68)}$$

For any given constants $M, \epsilon > 0$, the second term in (A68) vanishes to 0 as $n, T \to \infty$ given $2^{J/2}|2^{\hat{J}}/2^J - 1| \to^p 0$. For the first term, given $M2^{J/2}|2^{\hat{J}}/2^J - 1| \leq \epsilon$, we have for all n and all T_i sufficiently large,

$$V_{nT}^{-1/2}\hat{G}_{11} \leq V_{nT}^{-1/2} \sum_{i=1}^{n} 2\pi T_i \sum_{j=0}^{[\log_2 2^J(1+\epsilon/M2^{J/2})]} \sum_{k=1}^{2^j} (\hat{\alpha}_{ijk} - \bar{\alpha}_{ijk})^2$$

$$\leq \quad V_{nT}^{-1/2}\sum_{i=1}^{n}2\pi T_i\sum_{j=0}^{J+1}\sum_{k=1}^{2^j}(\hat{\alpha}_{ijk}-\bar{\alpha}_{ijk})^2 = o_P(1)$$

by Proposition A.1. Therefore, we have

$$V_{nT}^{-1/2}\hat{G}_1 = o_P(1). \tag{A69}$$

Next, we consider $\hat{G}_2$ in (A66). We write

$$\begin{aligned}
\hat{G}_2 &= \sum_{i=1}^{n}T_i\sum_{h=1}^{T_i-1}\sum_{m=1}^{T_i-1}\bar{R}_i(h)\bar{R}_i(m)\left[b_{\hat{J}}(h,m)-b_J(h,m)\right] \\
&= \sum_{i=1}^{n}\sigma_i^4\sum_{h=1}^{T_i-1}\left[b_{\hat{J}}(h,h)-b_J(h,h)\right]+\sum_{i=1}^{n}\sum_{h=1}^{T_i-1}\left[T_i^{-1}\bar{R}_i^2(h)-\sigma_i^4\right]\left[b_{\hat{J}}(h,h)-b_J(h,h)\right] \\
&\quad +2\sum_{i=1}^{n}T_i\sum_{h=1}^{T_i-1}\sum_{m=1}^{h-1}\bar{R}_i(h)\bar{R}_i(m)\left[b_{\hat{J}}(h,m)-b_J(h,m)\right] \qquad \text{by symmetry of } b_J(\cdot,\cdot) \\
&\equiv \hat{G}_{21}+\hat{G}_{22}+2\hat{G}_{23},\ \text{say.}
\end{aligned} \tag{A70}$$

For the last term $\hat{G}_{23}$ in (A70) we have for any constants $M>0$ and $\epsilon>0$

$$\begin{aligned}
P\left(V_{nT}^{-1/2}|\hat{G}_{23}|>\epsilon\right) &\leq P(V_{nT}^{-1/2}|\hat{G}_{23}|>\epsilon, M2^{J/2}|2^{\hat{J}}/2^J-1|\leq\epsilon) \\
&\quad +P\left(M2^{J/2}|2^{\hat{J}}/2^J-1|>\epsilon\right).
\end{aligned} \tag{A71}$$

Again, the second term here vanishes to 0 as $n,T\to\infty$. Put $\bar{T}\equiv\max_{1\leq i\leq n}(T_i)$, as before. For the first term, given $M2^{J/2}|2^{\hat{J}}/2^J-1|\leq\epsilon$, we have for all n and T sufficiently large,

$$\begin{aligned}
E|\hat{G}_{23}| &\leq \sum_{h=1-\bar{T}}^{\bar{T}-1}\sum_{m=1-\bar{T}}^{\bar{T}-1}E\left|\sum_{i=1}^{n}\mathbf{1}(h\leq T_i)\mathbf{1}(m\leq T_i)T_i\bar{R}_i(h)\bar{R}_i(m)\right|\sum_{j=[\log_2 2^J(1-\epsilon/M2^{J/2})]}^{[\log_2 2^J(1+\epsilon/M2^{J/2})]}|a_j(h,m)| \\
&\leq 2Cn^{1/2}2^{J+1}\epsilon/M2^{J/2}=o(V_{nT}^{1/2})
\end{aligned}$$

as $M\to\infty$, where $a_j(h,m)$ is defined as that used in (3.18) and we used the fact that for all n and T sufficiently large,

$$\begin{aligned}
&\sum_{h=1-\bar{T}}^{\bar{T}-1}\sum_{m=1-\bar{T}}^{\bar{T}-1}\sum_{j=[\log_2 2^J(1-\epsilon/M2^{J/2})]}^{[\log_2 2^J(1+\epsilon/M2^{J/2})]}|a_j(h,m)| \\
&= \sum_{h=1-\bar{T}}^{\bar{T}-1}\sum_{m=1-\bar{T}}^{\bar{T}-1}\sum_{j=[\log_2 2^{J_i}(1-\epsilon/M2^{J/2})]}^{[\log_2 2^{J_i}(1+\epsilon/M2^{J/2})]}\left|c_j(h,m)\hat{\psi}(2\pi h/2^j)\hat{\psi}^*(2\pi m/2^j)\right| \\
&\leq C\sum_{j=[\log_2 2^J(1-\epsilon/M2^{J/2})]}^{[\log_2 2^J(1+\epsilon/M2^{J/2})]}2^j\left[2^{-j}\sum_{h=1-\bar{T}}^{\bar{T}-1}|\hat{\psi}(2\pi h/2^j)|\right]\left[\sum_{r=-\infty}^{\infty}|\hat{\psi}(2\pi h/2^j+2\pi r)|\right] \\
&\leq C2^J\epsilon/M2^{J/2}
\end{aligned} \tag{A72}$$

given (A1) and Assumption 3, where $c_j(h,m)$ is as that used in (A1). Therefore, the first term in (A71) also vanishes to 0. Consequently, we have

$$V_{nT}^{-1/2}\hat{G}_{23} = o_P(1). \tag{A73}$$

Similarly, we can also obtain

$$V_{nT}^{-1/2}\hat{G}_{22} = o_P(1) \tag{A74}$$

and

$$V_{nT}^{-1/2}\left\{\hat{G}_{21} - [\hat{M}(\hat{J}) - \hat{M}(J)]\right\} = V_{nT}^{-1/2}\sum_{i=1}^{n}[\sigma_i^4 - \hat{R}_i^2(0)]\left[b_{\hat{J}}(h,h) - b_J(h,h)\right] = o_P(1). \tag{A75}$$

where $n^{-1}\sum_{i=1}^{n}[\hat{R}_i^2(0) - \sigma_i^4] = O_P[(nT)^{-1/2}]$ given Assumptions 1, 4 and 5, and that $\{v_{it}\}$ is i.i.d. for each i. It follows from (A70) and (A73)–(A75) that

$$V_{nT}^{-1/2}\left\{\hat{G}_2 - [\hat{M}(\hat{J}) - \hat{M}(J)]\right\} \xrightarrow{p} 0. \tag{A76}$$

Next, by the Cauchy-Schwarz inequality and (A69), we have

$$\begin{aligned} V_{nT}^{-1/2}|\hat{G}_3| &\le \left(V_{nT}^{-1/2}\hat{G}_1\right)^{1/2}\left(V_{nT}^{-1/2}|\hat{G}_2|\right)^{1/2} \\ &= O_P[V_{nT}^{-1/4} + (T^{-1/2} + n^{-1/2}V_{nT}^{1/4})]o_P(n^{1/4}) = o_P(1) \qquad \text{(A77)} \end{aligned}$$

by Proposition A.1 and the fact that $n^{-1/2}V_{nT}^{-1/2}\hat{G}_2 = n^{-1/2}V_{nT}^{-1/2}(\hat{G}_{21} + \hat{G}_{22} + \hat{G}_{23}) \xrightarrow{p} 0$ given (A73), (A74) and $n^{-1/2}V_{nT}^{-1/2}\hat{G}_{21} \xrightarrow{p} 0$ (as can be shown using reasoning similar to that for $\hat{G}_{23}$). Combining (A66), (A69), (A76) and (A77), we obtain result (i):

$$V_{nT}^{-1/2}\left\{\sum_{i=1}^{n} 2\pi T_i \sum_{j=J}^{\hat{J}}\sum_{k}^{2^j}\hat{\alpha}_{ijk}^2 - [\hat{M}(\hat{J}) - \hat{M}(J)]\right\} = o_P(1).$$

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

$$\left[\hat{V}(\hat{J}) - V_{nT}\right]/V_{nT} = V_{nT}^{-1}\sum_{i=1}^{n}[\hat{R}_i^4(0) - \sigma_i^8]\sum_{h=1}^{T_i-1}\sum_{m=1}^{T_i-1}\left[b_{\hat{J}}(h,m) - b_J(h,m)\right] = o_P(1),$$

where we used the facts that $n^{-1}\sum_{i=1}^{n}[\hat{R}_i^4(0) - \sigma_i^8] = O_P[(nT)^{-1/2}]$ (as can be shown given Assumptions 1, 4 and 5 and that $\{v_{it}\}$ is i.i.d.), and (A72). Thus, $\hat{V}(\hat{J})/\hat{V}(J) - 1 \xrightarrow{p} 0$. It follows that $[\hat{V}(\hat{J})/\hat{V}(J) - 1]\hat{W}_1(J) \xrightarrow{p} 0$ given $\hat{W}_1(J) = O_P(1)$ by Theorem 1. Therefore, (ii) holds, and we have $\hat{W}_1(\hat{J}) - \hat{W}_1(J) \xrightarrow{p} 0$, and $\hat{W}_1(\hat{J}) \xrightarrow{d} N(0,1)$. This completes the proof. ■

F Proof of Theorem 6

To conserve space, we shall prove for (b) only; the proof for (a) is similar and simpler. (i) We first show $n^{-1}\sum_{i=1}^{n} Q(\hat{f}_i, f_i) = n^{-1}\sum_{i=1}^{n} Q(\bar{f}_i, f_i) + o_P(2^J/T + 2^{-2qJ})$. Write

$$n^{-1}\sum_{i=1}^{n}\left[Q(\hat{f}_i, f_i) - Q(\bar{f}_i, f_i)\right] = n^{-1}\sum_{i=1}^{n} Q(\hat{f}_i, \bar{f}_i) + 2n^{-1}\sum_{i=1}^{n}\int_{-\pi}^{\pi}[\hat{f}_i(\omega) - \bar{f}_i(\omega)][\bar{f}_i(\omega) - f_i(\omega)]d\omega$$
$$\equiv \hat{Q}_1 + 2\hat{Q}_2, \text{ say.} \tag{A78}$$

For the first term in (A78), by Parseval's identity, Proposition A.1 (which, as noted earlier, continues to hold given Assumptions A.1–A.6), and $V_{nT} \propto n2^{J+1}$, we have

$$\hat{Q}_1 = n^{-1}\sum_{i=1}^{n}\sum_{j=0}^{J}\sum_{k=1}^{2^j}(\hat{\alpha}_{ijk} - \bar{\alpha}_{ijk})^2 = O_P[(nT)^{-1} + 2^J/nT + 2^J/T^2] = o_P(2^J/T) \tag{A79}$$

as $n, T \to \infty$. For the second term, we have $\hat{Q}_2 = o_P(2^J/T + 2^{-2qJ})$ by the Cauchy-Schwarz inequality, (A79) and the fact that $n^{-1}\sum_{i=1}^{n} Q(\bar{f}_i, f_i) = O_P(2^J/T + 2^{-2qJ})$, which follows by Markov's inequality and $n^{-1}\sum_{i=1}^{n} EQ(\bar{f}_i, f_i) = O(2^J/T + 2^{-2qJ})$. The latter is to be shown below.

To compute $n^{-1}\sum_{i=1}^{n} EQ(\bar{f}_i, f_i)$, we write

$$n^{-1}\sum_{i=1}^{n} EQ(\bar{f}_i, f_i) = n^{-1}\sum_{i=1}^{n} EQ(\bar{f}_i, E\bar{f}_i) + n^{-1}\sum_{i=1}^{n} Q(E\bar{f}_i, f_i). \tag{A80}$$

We first consider the second term in (A80). Put $B_i(\omega) \equiv \sum_{j=J+1}^{\infty}\sum_{k=1}^{2^j}\alpha_{ijk}\Psi_{jk}(\omega)$. Then

$$n^{-1}\sum_{i=1}^{n} Q(E\bar{f}_i, f_i) = n^{-1}\sum_{i=1}^{n}\int_{-\pi}^{\pi} B_i^2(\omega)d\omega + n^{-1}\sum_{i=1}^{n}\sum_{j=0}^{J}\sum_{k=1}^{2^j}(E\bar{\alpha}_{ijk} - \alpha_{ijk})^2, \tag{A81}$$

Using (3.10), (3.11) and (3.14), and recalling the definition of $c_j(h,m)$ as used in (A1), we have

$$\begin{aligned}
B_i(\omega) &= \sum_{j=J+1}^{\infty}\sum_{k=1}^{2^j}\left[\sum_{h=-\infty}^{\infty} R_i(h)\hat{\psi}_{jk}(2\pi h)\right]\left[\sum_{m=-\infty}^{\infty}\hat{\psi}_{jk}(2\pi m)e^{\mathbf{i}\omega}\right]^* \\
&= \sum_{j=J+1}^{\infty}\sum_{h=-\infty}^{\infty}\sum_{m=-\infty}^{\infty} R_i(h)c_j(h,m)\hat{\psi}(2\pi h/2^j)\hat{\psi}^*(2\pi m/2^j)e^{-\mathbf{i}m\omega} \\
&= \sum_{j=J+1}^{\infty}\sum_{h=-\infty}^{\infty}\sum_{r=-\infty}^{\infty} R_i(h)\hat{\psi}(2\pi h/2^j)\hat{\psi}(2\pi h/2^j + 2\pi r)e^{-\mathbf{i}(h+2^i r)\omega} \text{ by (A1)} \\
&= (2\pi)^{-1}\sum_{j=J+1}^{\infty}\sum_{h=-\infty}^{\infty} R_i(h)\lambda(2\pi h)e^{-\mathbf{i}h\omega} \\
&= (2\pi)^{-1}(1-2^{-q})\sum_{j=J+1}^{\infty} 2^{-jq}\sum_{h=-\infty}^{\infty}\frac{(2\pi)^q}{1-2^{-q}}\frac{\lambda(2\pi h)}{|2\pi h/2^j|^q}|h|^q R_i(h)e^{-\mathbf{i}h\omega} \\
&= 2^{-q(J+1)}\lambda_q(2\pi)^{-1}\sum_{h=-\infty}^{\infty}|h|^q R_i(h)e^{-\mathbf{i}h\omega}[1+o(1)] \\
&= 2^{-q(J+1)}\lambda_q f_i^{(q)}(\omega)[1+o(1)],
\end{aligned}$$

where $f^{(q)}(\cdot)$ is defined in (6.1) and $o(1)$ is uniform in i and $\omega \in [-\pi, \pi]$. It follows that

$$n^{-1}\sum_{i=1}^{n}\int_{-\pi}^{\pi} B_i^2(\omega)d\omega = 2^{-2q(J+1)}\lambda_q^2 n^{-1}\sum_{i=1}^{n}\int_{-\pi}^{\pi}\left[f_i^{(q)}(\omega)\right]^2 d\omega + o(2^{-2qJ}). \quad \text{(A82)}$$

For the second term in (A79), we have

$$\begin{aligned}
n^{-1}\sum_{i=1}^{n}\sum_{j=0}^{J}\sum_{k=1}^{2^j}(E\bar{\alpha}_{ijk}-\alpha_{ijk})^2 &= n^{-1}\sum_{i=1}^{n}\sum_{j=0}^{J}\sum_{k=1}^{2^j}\left[T_i^{-1}\sum_{h=1-T_i}^{T_i-1}|h|R_i(h)\hat{\psi}_{jk}(2\pi h) + \sum_{|h|\geq T_i} R_i(h)\hat{\psi}_{jk}(2\pi h)\right]^2 \\
&\leq 4Cn^{-1}\sum_{i=1}^{n}T_i^{-2}\sum_{h=1}^{T_i-1}\sum_{m=1}^{T_i-1}|hR_i(h)mR_i(m)b_J(h,m)| \\
&= O[(J+1)/T^2) \qquad \text{(A83)}
\end{aligned}$$

given Lemma A.1(vii) and $\sum_{h=-\infty}^{\infty}|hR_i(h)| \leq C$ as implied by Assumption 8.

Finally, we consider the first term in (A80), the variance factor. We write

$$\begin{aligned}
n^{-1}\sum_{i=1}^{n}EQ(\bar{f}_i, E\bar{f}_i) &= n^{-1}\sum_{i=1}^{n}\sum_{h=1-T_i}^{T_i-1}\sum_{m=1-T_i}^{T_i-1} b_{J_i}(h,m)Cov[\bar{R}_i(h), \bar{R}_i(m)] \\
&= n^{-1}\sum_{i=1}^{n}\sum_{h=1-T_i}^{T_i-1}\sum_{m=1-T_i}^{T_i-1} b_{J_i}(h,m)T_i^{-1}\sum_{l}\left[1-\frac{\eta(l)+m}{T_i}\right] \\
&\quad \times [R_i(l)R_i(l+m-h) + R_i(l+m)R_i(l-h) + \kappa_i(l,h,m-h)] \\
&\equiv V_{i1}+V_{i2}+V_{i3}, \text{ say,}
\end{aligned}$$

where the function

$$\eta(l) \equiv \begin{cases} l, & \text{if } l > 0 \\ 0, & \text{if } h-m \leq l \leq 0 \\ -l+h-m, & \text{if } -(T_i-h)+1 \leq l \leq h-m. \end{cases}$$

Cf. Priestley (1981, p.326). Given Assumption 6 and Lemma A.1(vii), we have $|V_{2i}| \leq C(J+1)$ and $|V_{3i}| \leq C(J+1)$. For the first term V_{1i}, we can write

$$\begin{aligned}
V_{1i} &= \sum_{h=1-T_i}^{T_i-1} b_{J_i}(h,h)T_i^{-1}\sum_{l}(1-h/T_i)R_i^2(h) + \sum_{h}\sum_{|r|=1}^{T_i-1} b_{J_i}(h,h+r)T_i^{-1}\sum_{l}R_i(l)R_i(l+r) \\
&= T_i^{-1}(2^{J_i+1}-1)\sum_{h=-\infty}^{\infty}R_i^2(h) + O[(J+1)/T_i],
\end{aligned}$$

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 $J \to \infty$

$$n^{-1}\sum_{i=1}^{n}EQ(\bar{f}_i, E\bar{f}_i) = \frac{2^{J+1}}{T}n^{-1}\sum_{i=1}^{n}\int_{-\pi}^{\pi} f_i^2(\omega)d\omega + o(2^J/T). \quad \text{(A84)}$$

Collecting (A82)–(A84) and $J \to \infty$, we obtain

$$n^{-1}\sum_{i=1}^{n} EQ(\hat{f}_i, f_i) = \frac{2^{J+1}}{T} n^{-1} \sum_{i=1}^{n} \int_{-\pi}^{\pi} f_i^2(\omega) d\omega + 2^{-2qJ} \lambda_q^2 n^{-1} \sum_{i=1}^{n} \int_{-\pi}^{\pi} [f_i^{(q)}(\omega)]^2 d\omega$$
$$+o(2^J/T + 2^{-2qJ}). \quad \blacksquare$$

G Proof of Corollary 7

The result follows immediately from Theorem 5 because Assumption 9 implies $2^{\hat{J}}/2^J - 1 = o_P(T^{-1/2(2q+1)}) = o_P(2^{-J/2})$, where the nonstochastic finest scale J is given by $2^{J+1} \equiv \max\{[2\alpha\lambda_q^2\zeta_0(q)T]^{1/(2q+1)}$ The latter satisfies the conditions of Theorem 5. ■

Table 1: Empirical Size of Tests at the 10% and 5% Levels : (n, T) = (25, 32)

	J	τ	$\hat{W}_1$	$\hat{W}_2$	BL	BSY	τ	$\hat{W}_1$	$\hat{W}_2$	BL	BSY
	0	0	6.4	5.5	24.4	11.0	0.05	6.0	5.7	23.2	8.2
	1		5.7	4.7				5.1	4.6		
10%	2		5.0	4.2				6.3	4.5		
	$\hat{J}_0$		5.4	5.5				6.3	4.7		
	0		4.3	3.5	16.0	5.1		3.1	3.0	14.6	3.5
	1		3.1	2.5				2.8	1.9		
5%	2		2.7	2.6				3.4	2.4		
	$\hat{J}_0$		3.0	2.6				4.0	2.3		
	0	0.2	4.1	5.2	26.6	6.0	0.4	5.7	6.0	23.2	6.5
	1		4.1	4.1				5.5	4.9		
10%	2		3.4	3.7				5.9	5.1		
	$\hat{J}_0$		5.4	5.5				6.3	4.7		
	0		2.1	3.1	16.3	2.7		3.3	2.9	14.6	2.5
	1		1.9	1.6				3.5	2.3		
5%	2		2.0	1.7				3.5	2.9		
	$\hat{J}_0$		3.0	2.6				4.0	2.3		
	0	0.6	6.2	5.2	26.6	9.7	0.8	5.5	6.0	25.6	20.6
	1		4.5	3.5				4.9	3.9		
10%	2		4.3	3.4				3.9	3.8		
	$\hat{J}_0$		4.8	4.9				6.4	5.6		
	0		3.2	2.6	16.3	2.0		3.3	2.8	16.2	15.0
	1		1.7	1.6				2.2	2.4		
5%	2		2.1	1.7				1.8	2.2		
	$\hat{J}_0$		2.4	2.6				3.6	2.6		

Note:

(a) $\hat{W}_1$ and $\hat{W}_2$ are given in (3.18) and (3.19) using the Franklin wavelet;

(b) J denotes the finest scale; $\hat{J}_0$ denotes the data-driven finest scale given in (6.6);

(c) BL denotes the Baltagi and Li test; BSY denotes the Bera *et al.* test;

(d) The number of iterations is 1000.

Table 2: Empirical Size of Tests at the 10% and 5% Levels : (n, T) = (50, 64)

	J	τ	$\hat{W}_1$	$\hat{W}_2$	BL	BSY	τ	$\hat{W}_1$	$\hat{W}_2$	BL	BSY
	0	0	8.1	7.9	2.4	10.8	0.05	6.6	6.8	22.5	8.9
	1		6.1	5.7				5.8	6.0		
10%	2		5.2	5.3				4.9	5.7		
	$\hat{J}_0$		8.8	7.9				7.9	8.2		
	0		4.2	3.8	15.4	5.2		2.9	3.4	14.5	3.5
	1		3.3	3.3				2.7	3.0		
5%	2		2.6	2.5				2.8	2.7		
	$\hat{J}_0$		4.6	4.3				4.0	3.7		
	0	0.2	7.6	7.2	23.1	6.1	0.4	6.8	6.7	24.4	5.6
	1		7.6	6.2				6.0	6.5		
10%	2		6.0	5.4				6.0	6.2		
	$\hat{J}_0$		8.8	7.9				7.9	8.2		
	0		4.2	3.7	15.7	2.8		3.7	3.9	15.4	1.8
	1		3.6	3.2				3.3	3.5		
5%	2		2.9	2.6				3.4	2.9		
	$\hat{J}_0$		4.6	4.3				4.0	3.7		
	0	0.6	8.5	8.6	25.0	9.3	0.8	7.7	7.9	22.3	17.1
	1		8.3	8.0				7.5	6.0		
10%	2		6.5	5.2				6.1	5.5		
	$\hat{J}_0$		9.1	8.5				8.8	8.8		
	0		5.0	5.5	16.7	1.7		3.9	3.8	13.3	0.5
	1		5.2	4.6				3.6	2.6		
5%	2		3.5	2.4				2.5	2.5		
	$\hat{J}_0$		4.7	4.0				5.3	5.3		

Notes:

(a) $\hat{W}_1$ and $\hat{W}_2$ are given in (3.18) and (3.19) using the Franklin wavelet;

(b) J denotes the finest scale; $\hat{J}_0$ denotes the data-driven finest scale given in (6.6);

(c) BL denotes the Baltagi and Li test; BSY denotes the Bera *et al.* test;

(d) The number of iterations is 1000.

Table 3: Empirical Size of Tests at the 10% and 5% Levels : (n, T) = (100, 128)

	J	τ	$\hat{W}_1$	$\hat{W}_2$	BL	BSY	τ	$\hat{W}_1$	$\hat{W}_2$	BL	BSY
	0	0	8.5	8.7	23.1	9.7	0.05	7.1	6.8	23.5	6.8
	1		6.9	7.5				6.9	6.0		
10%	2		5.6	4.9				5.1	5.3		
	$\hat{J}_0$		8.6	9.4				8.2	7.5		
	0		5.0	5.1	15.3	4.9		3.7	3.6	15.8	3.1
	1		4.2	4.0				3.1	2.9		
5%	2		2.3	2.2				3.2	3.2		
	$\hat{J}_0$		5.1	5.1				4.1	4.0		
	0	0.2	8.5	8.5	21.9	6.0	0.4	7.6	8.0	25.5	4.5
	1		7.2	7.4				7.0	7.1		
10%	2		5.3	5.3				6.7	6.8		
	$\hat{J}_0$		8.6	9.4				8.2	7.5		
	0		4.0	3.6	14.3	2.8		3.9	3.7	16.8	1.1
	1		3.3	4.0				3.6	3.7		
5%	2		2.8	2.3				3.9	3.3		
	$\hat{J}_0$		5.1	5.1				4.1	4.0		
	0	0.6	8.3	8.1	24.0	8.2	0.8	8.8	7.7	23.9	14.9
	1		8.0	7.0				6.0	5.1		
10%	2		6.8	6.2				5.3	5.4		
	$\hat{J}_0$		8.9	8.7				9.1	9.6		
	0		4.5	4.1	15.2	1.9		3.9	3.7	14.9	0.9
	1		3.4	4.1				2.4	2.4		
5%	2		3.2	3.6				2.6	2.5		
	$\hat{J}_0$		4.2	3.8				4.8	4.9		

Notes:

(a) $\hat{W}_1$ and $\hat{W}_2$ are given in (3.18) and (3.19) using the Franklin wavelet;

(b) J denotes the finest scale; $\hat{J}_0$ denotes the data-driven finest scale given in (6.6);

(c) BL denotes the Baltagi and Li test; BSY denotes the Bera *et al.* test;

(d) The number of iterations is 1000.

Table 4: Size-Corrected Power of Tests at the 5% Level Under AR(1): (n, T) = (25, 32)

	J	$\hat{W}_1$	$\hat{W}_2$	BL	BSY	$\hat{W}_1$	$\hat{W}_2$	BL	BSY	$\hat{W}_1$	$\hat{W}_2$	BL	BSY
	τ		0.2				0.4				0.6		
	0	69.2	56.9	95.5	100.0	65.0	53.5	95.6	100.0	63.0	55.6	96.0	100.0
	1	45.8	34.9			40.5	32.5			44.9	37.0		
AR(1)^a	2	43.9	32.7			34.2	28.0			39.1	34.8		
	$\widehat{J}_0$	45.2	34.5			41.7	37.3			67.3	57.7		
	τ		0.2				0.4				0.6		
	0	98.6	98.0	100.0	98.8	98.4	97.7	100.0	57.5	98.2	97.8	100	5.1
	1	91.9	89.1			89.2	88.1			91.4	89.8		
AR(1)^b	2	83.2	76.8			75.1	72.7			79.4	78.5		
	$\hat{J}_0$	84.1	83.3			82.8	84.0			67.3	57.7		
	τ		0.2				0.4				0.6		
	0	34.0	25.2	18.0	81.3	29.8	23.3	19.7	81.6	28.8	24.8	21.2	78.6
	1	23.1	17.3			19.0	15.6			21.8	18.9		
AR(1)^c	2	23.4	14.9			15.8	13.0			19.0	16.4		
	$\hat{J}_0$	21.5	15.9			18.9	16.9			23.1	16.5		
	τ		0.2				0.4				0.6		
	0	79.7	70.9	85.7	47.2	75.6	69.2	87.0	6.9	74.8	70.5	87.6	0.4
	1	60.6	53.5			55.2	50.6			59.1	54.9		
AR(1)^d	2	50.1	39.2			42.0	35.2			45.8	41.6		
	$\hat{J}_0$	56.5	48.0			53.4	49.7			67.3	57.7		
	τ		0.2				0.4				0.6		
	0	91.3	89.2	11.2	5.2	89.2	89.0	11.8	4.4	89.1	88.8	12.3	3.0
	1	76.7	70.4			72.4	67.8			75.7	71.8		
AR(1)^e	2	67.6	57.7			57.8	53.5			62.8	60.3		
	$\hat{J}_0$	71.0	65.0			67.8	66.6			67.3	57.7		

Notes:

(a) Alternatives AR(1)^a - AR(1)^e are given in (7.4);

(b) $\hat{W}_1$ and $\hat{W}_2$ are given in (3.18) and (3.19) using the Franklin wavelet;

(c) J denotes the finest scale; $\hat{J}_0$ denotes the data-driven finest scale given in (6.6);

(d) BL denotes the Baltagi and Li test; BSY denotes the Bera *et al.* test;

(e) The number of iterations is 1000.

Table 5: Size-Corrected Power of Tests at the 5% Level Under AR(1): (n, T) = (50,64)

	J	$\hat{W}_1$	$\hat{W}_2$	BL	BSY	$\hat{W}_1$	$\hat{W}_2$	BL	BSY	$\hat{W}_1$	$\hat{W}_2$	BL	BSY
	τ		0.2				0.4				0.6		
	0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	1	100.0	100.0			100.0	100.0			100.0	100.0		
$AR(1)^a$	2	99.4	99.4			99.8	98.7			99.2	99.4		
	$\hat{J}_0$	100.0	100.0			100.0	100.0			99.8	100.0		
	τ		0.2				0.4				0.6		
	0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	1	100.0	100.0			100.0	100.0			100.0	100.0		
$AR(1)^b$	2	100.0	100.0			100.0	100.0			100.0	100.0		
	$\hat{J}_0$	100.0	100.0			100.0	100.0			100.0	100.0		
	τ		0.2				0.4				0.6		
	0	96.3	95.9	98.5	100.0	97.1	96.1	98.5	100.0	96.1	94.9	98.1	100
	1	90.2	88.0			90.9	86.0			88.7	88.1		
$AR(1)^c$	2	76.5	72.1			77.8	68.9			75.0	64.9		
	$\hat{J}_0$	90.9	89.7			90.4	89.7			75.0	72.4		
	τ		0.2				0.4				0.6		
	0	100.0	99.8	100.0	98.9	100.0	99.8	100.0	65.7	100.0	99.8	100.0	0.9
	1	98.5	98.2			98.7	98.0			98.3	98.2		
$AR(1)^d$	2	92.7	91.8			93.2	89.7			91.8	87.5		
	$\hat{J}_0$	98.2	98.1			98.0	98.1			98.4	98.5		
	τ		0.2				0.4				0.6		
	0	100.0	100.0	7.3	6.8	100.0	100.0	8.0	8.1	100.0	100.0	6.4	6.6
	1	100.0	100.0			100.0	100.0			100.0	100.0		
$AR(1)^e$	2	99.7	99.8			100.0	99.8			98.7	99.4		
	$\hat{J}_0$	100.0	100.0			100.0	100.0			100.0	100.0		

Notes:

(a) Alternatives $AR(1)^a$ - $AR(1)^e$ are given in (7.4);

(b) $\hat{W}_1$ and $\hat{W}_2$ are given in (3.18) and (3.19) using the Franklin wavelet;

(c) J denotes the finest scale; $\hat{J}_0$ denotes the data-driven finest scale given in (6.6);

(d) BL denotes the Baltagi and Li test; BSY denotes the Bera *et al.* test;

(e) The number of iterations is 1000.

Table 6: Size-Corrected Power of Tests at the 5% Level Under AR(12): (n, T) = (25, 32)

	J	$\hat{W}_1$	$\hat{W}_2$	BL	BSY	$\hat{W}_1$	$\hat{W}_2$	BL	BSY	$\hat{W}_1$	$\hat{W}_2$	BL	BSY
	τ		0.2				0.4				0.6		
	0	71.6	73.6	23.7	74.3	62.0	69.0	25.0	73.7	66.5	75.7	25.8	70.1
	1	77.5	80.1			72.8	77.6			76.4	81.1		
$AR(12)^a$	2	82.8	84.9			75.9	83.1			79.3	86.2		
	$\hat{J}_0$	68.8	76.1			65.7	83.1			70.0	77.4		
	τ		0.2				0.4				0.6		
	0	39.8	42.0	20.7	5.9	36.6	40.4	21.6	2.8	35.1	39,7	22.8	2.6
	1	54.7	61.9			49.9	59.4			53.9	63.5		
$AR(12)^b$	2	68.0	73.7			56.9	69.9			61.8	75.6		
	$\hat{J}_0$	66.7	77.7			63.4	79.2			68.4	78.8		
	τ		0.2				0.4				0.6		
	0	52.6	55.8	4.2	38.3	48.0	53.6	4.5	38.4	46.7	52.8	5.1	36.6
	1	61.9	69.7			57.3	66.8			61.0	71.6		
$AR(12)^c$	2	73.0	78.2			63.3	74.4			67.6	79.0		
	$\hat{J}_0$	65.4	74.1			62.9	75.6			67.7	75.4		
	τ		0.2				0.4				0.6		
	0	38.4	40.9	13.3	7.3	33.5	39.4	14.7	5.7	32.6	39.4	15.8	4.6
	1	49.9	57.7			44.5	54.7			48.7	59.4		
$AR(12)^d$	2	65.5	71.2			54.2	66.5			59.9	72.7		
	$\hat{J}_0$	63.7	74.6			61.3	76.6			65.3	75.8		
	τ		0.2				0.4				0.6		
	0	40.0	41.0	22.3	5.9	36.8	39.3	24.0	2.2	35.1	39.0	25.2	1.7
	1	55.3	62.9			48.6	58.7			54.3	64.9		
$AR(12)^e$	2	70.3	75.3			58.2	71.8			63.7	76.6		
	$\hat{J}_0$	67.6	77.7			65.1	78.6			68.7	78.3		

Notes:

(a) Alternatives $AR12(1)^a$ - $AR12(1)^e$ are given in (7.5);

(b) $\hat{W}_1$ and $\hat{W}_2$ are given in (3.18) and (3.19) using the Franklin wavelet;

(c) J denotes the finest scale; $\hat{J}_0$ denotes the data-driven finest scale given in (6.6);

(d) BL denotes the Baltagi and Li test; BSY denotes the Bera *et al.* test;

(e) The number of iterations is 1000.

Table 7: Size-Corrected Power of Tests at the 5% Level Under AR(12): (n, T) = (50,64)

	J	$\hat{W}_1$	$\hat{W}_2$	BL	BSY	$\hat{W}_1$	$\hat{W}_2$	BL	BSY	$\hat{W}_1$	$\hat{W}_2$	BL	BSY
	τ		0.2				0.4				0.6		
	0	99.7	99.8	99.4	100.0	99.8	99.8	99.4	99.9	99.7	99.8	99.3	100.0
	1	100.0	100.0			100.0	100.0			100.0	100.0		
AR(12)^a	2	100.0	100.0			100.0	100.0			100.0	100.0		
	$\hat{J}_0$	100.0	100.0			100.0	100.0			100.0	100.0		
	τ		0.2				0.4				0.6		
	0	96.8	96.7	8.0	44.6	97.8	96.8	9.0	49.7	96.6	96.0	7.1	44.4
	1	99.5	99.6			99.6	99.6			99.3	99.0		
AR(12)^b	2	100.0	100.0			100.0	100.0			100.0	100.0		
	$\hat{J}_0$	99.8	99.7			99.8	99.7			99.6	99.7		
	τ		0.2				0.4				0.6		
	0	93.3	92.1	53.0	92.8	94.6	92.2	53.8	94.3	93.1	90.6	50.3	92.6
	1	98.2	98.1			98.2	98.0			97.6	87.4		
AR(12)^c	2	99.0	99.1			99.1	99.0			99.0	99.1		
	$\hat{J}_0$	98.5	98.1			98.6	98.1			98.3	98.1		
	τ		0				0.2				0.4		
	0	28.1	30.7	25.9	4.3	33.6	31.0	26.9	1.7	27.0	26.2	23.6	0.9
	1	41.2	46.4			43.1	45.9			36.9	39.2		
AR(12)^d	2	56.2	59.8			57.8	56.6			53.4	59.9		
	$\hat{J}_0$	47.5	51.1			47.8	52.0			46.0	50.2		
	τ		0				0.2				0.4		
	0	35.5	39.3	58.4	17.2	41.0	40.0	61.2	0.8	34.5	34.6	54.9	0.0
	1	53.4	57.3			55.4	57.2			49.8	51.6		
AR(12)^e	2	68.8	72.8			69.7	69.6			67.0	73.0		
	$\hat{J}_0$	59.0	62.4			59.3	62.2			57.9	62.2		

Note:

(a) Alternatives AR12(1)^a - AR12(1)^e are given in (7.5);

(b) $\hat{W}_1$ and $\hat{W}_2$ are given in (3.18) and (3.19) using the Franklin wavelet;

(c) J denotes the finest scale; $\hat{J}_0$ denotes the data-driven finest scale given in (6.6);

(d) BL denotes the Baltagi and Li test; BSY denotes the Bera *et al.* test;

(e) The number of iterations is 1000.

Table 8: Size Corrected Power of Tests at the 5% Level Under ARMA(4, 4): (n, T) = (25, 32)

	J	$\hat{W}_1$	$\hat{W}_2$	BL	BSY	$\hat{W}_1$	$\hat{W}_2$	BL	BSY	$\hat{W}_1$	$\hat{W}_2$	BL	BSY
	τ		0.2				0.4				0.6		
	0	32.3	27.3	9.0	5.8	28.6	25.8	10.0	8.7	28.0	26.7	10.8	4.0
	1	61.5	61.7			55.1	58.9			59.4	63.5		
ARMA(4,4)[a]	2	78.3	74.4			69.9	70.9			74.2	76.0		
	$\hat{J}_0$	73.9	74.3			72.5	75.3			75.2	75.0		
	τ		0.2				0.4				0.6		
	0	11.6	7.8	3.1	22.7	9.4	7.3	3.2	28.4	8.4	7.5	3.9	18.6
	1	2.0	1.1			1.2	0.9			1.7	1.2		
ARMA(4,4)[b]	2	77.7	74.2			68.7	68.2			72.9	76.7		
	$\hat{J}_0$	63.0	62.3			60.1	63.8			64.0	64.8		
	τ		0.2				0.4				0.6		
	0	25.2	19.1	7.7	9.3	21.6	17.9	8.6	12.4	21.2	18.9	9.4	7.2
	1	41.1	35.1			35.7	31.7			39.7	33.5		
ARMA(4,4)[c]	2	57.0	44.1			46.0	39.5			50.8	46.6		
	$\hat{J}_0$	50.8	42.8			48.0	44.0			52.3	43.6		
	τ		0.2				0.4				0.6		
	0	10.8	8.2	4.1	17.7	9.3	7.7	4.6	23.6	8.8	8.0	5.0	13.3
	1	3.6	3.9			2.3	3.4			3.4	4.3		
ARMA(4,4)[d]	2	57.6	45.7			47.0	39.0			51.8	47.4		
	$\hat{J}_0$	43.9	35.4			40.7	33.8			45.0	36.9		
	τ		0.2				0.4				0.6		
	0	20.3	16.5	4.0	12.0	17.3	14.3	5.0	17.2	16.6	16.1	5.2.	9.2
	1	21.0	20.4			16.9	18.4			20.2	22.0		
ARMA(4,4)[e]	2	79.4	75.5			69.5	71.2			74.3	77.3		
	$\hat{J}_0$	66.8	68.1			64.3	69.7			67.3	67.8		

Notes:

(a) Alternatives ARMA(4,4)[a] - ARMA(4,4)[e] are given in (7.6);

(b) $\hat{W}_1$ and $\hat{W}_2$ are given in (3.18) and (3.19) using the Franklin wavelet;

(c) J denotes the finest scale; $\hat{J}_0$ denotes the data-driven finest scale given in (6.6);

(d) BL denotes the Baltagi and Li test; BSY denotes the Bera *et al.* test;

(e) The number of iterations is 1000.

Table 9: Size-Corrected Power of Tests at the 5% Level Under ARMA(4, 4): (n, T) = (50,64)

	J	$\hat{W}_1$	$\hat{W}_2$	BL	BSY	$\hat{W}_1$	$\hat{W}_2$	BL	BSY	$\hat{W}_1$	$\hat{W}_2$	BL	BSY
	τ		0.2				0.4				0.6		
	0	40.8	45.3	12.1	5.0	45.6	45.9	12.9	9.8	39.8	40.4	10.0	4.5
	1	89.7	92.0			90.5	92.0			87.6	89.8		
ARMA(4,4)^a	2	100.0	100.0			100.0	100.0			100.0	100.0		
	$\hat{J}_0$	99.0	99.2			99.0	99.2			99.0	99.1		
	τ		0.2				0.4				0.6		
	0	8.8	9.6	2.4	23.7	10.7	9.9	2.7	35.9	8.4	7.3	1.5	22.5
	1	0.3	0.3			0.3	0.3			0.2	0.2		
ARMA(4.4)^b	2	100.0	100.0			100.0	100.0			100.0	100.0		
	$\hat{J}_0$	98.3	63.1			98.3	98.3			98.2	98.3		
	τ		0.2				0.4				0.6		
	0	27.0	24.2	9.7	6.7	32.0	24.6	10.4	14.0	25.9	21.2	8.2	6.5
	1	65.1	59.6			66.6	59.3			62.0	52.2		
ARMA(4,4)^c	2	100.0	100.0			100.0	99.8			100.0	100.0		
	$\hat{J}_0$	90.0	88.8			91.0	89.0			90.8	88.8		
	τ		0.2				0.4				0.6		
	0	7.6	8.7	4.0	16.4	11.1	9.2	4.3	29.5	7.1	5.9	2.8	15.6
	1	1.2	2.0			1.4	2.0			0.9	1.3		
ARMA(4,4)^d	2	100.0	99.7			100.0	99.5			100.0	99.7		
	$\hat{J}_0$	91.7	91.4			91.7	91.4			91.7	91.4		
	τ		0.2				0.4				0.6		
	0	22.4	24.7	5.5	12.0	25.9	25.2	5.7	21.7	21.8	21.8	4.6	10.7
	1	29.3	33.5			31.0	33.1			25.2	28.6		
ARMA(4,4)^e	2	100.0	100.0			100.0	100.0			100.0	100.0		
	$\hat{J}_0$	98.1	98.1			98.1	98.1			98.0	98.1		

Notes:

(a) Alternatives ARMA(4,4)^a - ARMA(4,4)^e are given in (7.6);

(b) $\hat{W}_1$ and $\hat{W}_2$ are given in (3.18) and (3.19) using the Franklin wavelet;

(c) J denotes the finest scale; $\hat{J}_0$ denotes the data-driven finest scale given in (6.6);

(d) BL denotes the Baltagi and Li test; BSY denotes the Bera *et al.* test;

(e) The number of iterations is 1000.