Econometric Theory, **37**, 2021, 82–137.
doi:10.1017/S0266466620000079

NEARLY OPTIMAL TEST FOR LONG-RUN PREDICTABILITY WITH NEARLY INTEGRATED REGRESSORS

NATALIA SIZOVA*
Microsoft

We develop a method for long-run predictability testing in series Y by a persistent series X. We consider a class of tests based on the long-run behavior of these series that are robust to short-run dynamics and attempt to attain the highest possible power. The test is based on the Whittle approximation to the likelihood ratio that is adjusted to remain accurate across a range of persistence in X. We verify the properties of this test in small simulations and compare this test against a group of recently proposed methods.

1. INTRODUCTION

Consider the problem of determining if x_t predicts y_t and finding the sign of this relationship. We are interested in the equilibrium relation between x_t and y_t that is realized in the long run, rather than in a regression coefficient of y_t on one lag of x_t. From the perspective of empirical studies, long-run predictability is important because although violations of economic restrictions can be explained in the short run, their violation in the long run invalidates the corresponding economic models. For example, the risk-return trade-off between stock returns and financial volatility is not observed (and the correlation is even negative) at monthly frequencies. However, financial volatility does predict returns with the positive sign in the long run; see Bandi and Perron (2008) and Sizova (2013). Similarly, the coefficients in the uncovered interest rate parity are different in simple and long-horizon regressions; see Boudoukh, Richardson and Whitelaw (2016).

In long-run predictability studies, x_t is often a persistent process such as interest rates, inflation, unemployment, and exchange rates. If this persistence is modeled as a local-to-unity root[1] in x_t, then the long-run relation between y_t and x_t becomes akin to the cointegration. However, in contrast to classic cointegration examples, in long-run predictability studies, the cointegrating relation between x_t and y_t may not be immediately obvious, as y_t contains a large noise component. Examples of this include asset returns, macroeconomic growth variables, and price changes.

*I am thankful to Michael Jansson for providing the code for the optimal inference method, as well as to Michalis P. Stamatogiannis and Tassos Magdalinos for providing the code for the IVX estimator. Address correspondence to Natalia Sizova, Microsoft, Bellevue, WA 98005, USA; e-mail: Natalia.Sizova@microsoft.com.

[1] That is, x_t is nearly integrated.

© Cambridge University Press 2020

https://doi.org/10.1017/S0266466620000079 Published online by Cambridge University Press

The presence of the large noise component in y_t calls for efficient methods that determine long-run relations and, at the same time, are robust to short-run dynamics in model errors—especially, robust to the *short-run endogeneity*.[2] Developing such a method is a goal of this study. To this end, we apply spectral methods to achieve the highest possible test power while maintaining the correct size over a large set of data-generating processes. In particular, we do not impose serial independence or martingale difference sequence (m.d.s.) assumptions on model (cointegration) errors. Before presenting details, it is important to define the scope of the mature and expansive literature on long-run predictability testing.

In the econometric literature, there are three interrelated branches that address the problem of long-run predictability estimation. One strain of the literature applies the so-called long-horizon regressions, which seems to be a straightforward extension of a simple regression to the study of long-term relations. For example, estimations of cumulative multipliers in empirical macroeconomic studies use a variant of this method, see Ramey and Zubairy (2018). However, the simplicity of long-horizon regressions is misleading, as the asymptotic normality of the estimator does not hold under an appropriate econometric framework, see Valkanov (2003). Moreover, long-horizon regressions have insufficient power to detect long-term predictability, see Sizova (2014).

The low power of existing long-run predictability tests triggered a search for methods with certain optimality qualities. As a result, a second strain of the literature, which is concerned with efficient testing, emerged. See Stock and Watson (1996), Jansson and Moreira (2006), Elliott, Müller, and Watson (2015), and Campbell and Yogo (2006), as well as discussions in Phillips (2014a,b) and Hjalmarsson (2007). To obtain efficiency results, these studies rely on simple assumptions about model errors, which are often assumed to be independent identically distributed (i.i.d.) and m.d.s. errors. Although the methods proposed by this group of authors offer certain paths for extensions to the short-run endogeneity case, the required corrections do not necessarily inherit good statistical properties from their parent methods. Specifically, our simulations show that intuitive corrections to the aforementioned methods result in small-sample size distortions. Further work on appropriate methods to extend these methods is warranted but is not undertaken in this study.

The third strain of the literature is the large body of research on cointegration testing, which started from unit-root and fractionally integrated processes but now extends to the $I(0)$ case, see Magdalinos and Phillips (2009). Spectral (frequency-domain) methods are often suggested as approaches to cointegration testing, for example, Robinson (1994), Christensen and Nielsen (2006), Robinson (1994), Marmol and Velasco (2004), Nielsen and Frederiksen (2011) and Marinucci and Robinson (2003). The attractiveness of the spectral methods lies in their ability to focus on lower frequency ranges, while remaining robust to short-run dynamics.

[2]Short-run endogeneity arises when model errors are correlated with past shocks to x_t, but this correlation becomes negligible over longer intervals of time.

https://doi.org/10.1017/S0266466620000079 Published online by Cambridge University Press

Müller and Watson (2013) applied low-frequency data transformations to robust cointegration testing. A closely related research on trend IV regressions avoids the explicit use of the frequency-domain transformations, but achieves the same purpose by using sinusoidal instruments for trending regressors, see Phillips (2014b) and Hwang and Sun (2018). In agreement with this strain of the literature, we consider a frequency-domain approach[3] to long-run predictability testing but try to achieve the highest possible test power. To this end, we revisit the literature on low-frequency estimation and show that methods relying on a popular tool in frequency-domain analysis—the so-called Whittle approximation to the Gaussian likelihood—can also be justified in the near local-to-unity case (see Lemma 1). We offer a correction to the conditional part of the Whittle approximation that makes it uniform in the persistence parameter of x_t, ρ. The result is a nearly optimal method for long-run predictability testing that yields a power bound for other spectral methods.

We offer three contributions to the literature. First, we develop an adjusted Whittle likelihood that approximates the Gaussian likelihood uniformly in the persistence parameter ρ. Second, we devise a nearly optimal test based on this likelihood and study the resulting optimal trade-offs between accuracy and robustness to short-run dynamics. Finally, we compare the small-sample performance of our method to that of several tests currently proposed in the literature. In contrast to prior studies, our focus is on cases with short-run endogeneity.

The details of our method are as follows. We consider tests that maximize the asymptotic weighted-average power (WAP) and are subject to restrictions on the asymptotic level. The suggested statistic is the likelihood ratio that uses the Whittle approximation for the conditional distribution of y_t given x_t at low frequencies. Proposition 1 derives the conditional part of the Whittle approximation. Based on this proposition, we reduce the data on y_t to q transformations $d_{y,j}$, $j = 1, \ldots, q$, whose conditional distribution is robust to short-run dynamics. The likelihood under the null hypothesis is calculated at the approximate least favorable distribution (ALFD)[4] of the nuisance parameter ρ, see Elliott et al. (2015). To ensure the correct size for moderate values of ρ, the method includes a switch to a standard test if the persistence in x_t is low. This test yields the WAP that is an epsilon close to the asymptotic power bound in the Gaussian case; see Lemma 2. Importantly, we can exactly assess this epsilon; the value of epsilon in our application is 0.5%.

Note that our method transforms the y_t data into complex-valued observations $d_{y,j}$ such that the distribution of a subsample of $d_{y,j}$ does not depend asymptotically on the form of the short-run endogeneity. Because we use only a subsample of $d_{y,j}$,

[3] Our study is closely related to the existing literature on the application of the spectral methods to estimation of the regression coefficients; see Phillips (1991), Robinson (1994), Corbae, Ouliaris, and Phillips (2002) and Marinucci and Robinson (2003). For example, our method is related to the spectral regression of Phillips (1991) and band spectral regression of Corbae et al. (2002). However, the spectral regressions concern parameters that do not directly define long-run predictability, unless x_t is a unit-root process.

[4] The ALFD is, generally, not an approximation to the least favorable distribution; rather, it is a distribution that nearly maximizes the WAP when used in place of the least favorable distribution.

https://doi.org/10.1017/S0266466620000079 Published online by Cambridge University Press

we intentionally give up some test power. However, data reduction is common in long-run predictability literature. For example, long-run predictability studies are often based on sparsely sampled data sets, in effect removing in-between observations. In this study, the choice of the data reduction technique follows from the form of the likelihood in the Whittle-approximation form.

In Monte Carlo simulations, our focus is on the trade-off between the loss of information due to retaining only low-frequency data and the robustness to short-run dynamics. We consider a model with no short-run endogeneity to evaluate the loss of information and a model with short-run endogeneity to evaluate robustness. For comparison, we provide Monte Carlo evidence for the methods suggested by Magdalinos and Phillips (2009), Campbell and Yogo (2006), Jansson and Moreira (2006) and Müller and Watson (2013), including versions of these methods corrected for short-run endogeneity whenever such corrections are warranted. Our first conclusion is that the methods developed under i.i.d. and m.d.s. assumptions suffer from size distortions when short-run endogeneity corrections are applied. Such corrections rely on accurate estimates of long-run predictability and short-run dynamics, both of which are challenging to obtain in the presence of confounding noise. Our second conclusion is that our nearly optimal test yields noticeably lower power when X is only moderately persistent and model errors are uncorrelated with past shocks to X. A trade-off between efficiency and robustness is similar to what occurs in other methods, such as the trade-off between least squares estimators and the IV approach. To strike a balance between the two objectives, a comprehensive empirical study of long-run predictability should include an analysis of full information and low frequency methods.

This article is organized as follows. Section 2 sets up the problem in the Gaussian case and provides justification for the Whittle approximation for nearly integrated regressors. Section 3 derives the asymptotic likelihood for the four sufficient statistics in the Whittle approximation and describes the steps of the nearly optimal test for the corresponding limiting problem. Sections 4 and 5 explain how the test is applied in the general non-Gaussian case. Section 6 reports the results of the Monte Carlo simulations. Section 7 remarks on the case with the number of frequencies q increasing to infinity at a slow rate. Section 8 summarizes application steps.

2. MODEL IN THE GAUSSIAN CASE

Let y_t denote the variable that we are forecasting and let x_t denote the explanatory variable. We observe a bivariate process (x_{t-1}, y_t) for $t = 1, \ldots, T$ whose dynamics can be represented by the following system:

$$\begin{aligned} y_t &= \mu_y + \beta_x(x_{t-1} - \mu_x) + \varepsilon_t, \\ x_t &= \mu_x + \rho(x_{t-1} - \mu_x) + u_t, \end{aligned} \tag{1}$$

where the mean-zero innovations $e_t = (u_t, \varepsilon_t)$ are normally distributed. The initial value x_0 comes from a distribution that does not depend on the rest of the dynamics parameters. Parameter ρ characterizes serial correlations of x_t at longer

https://doi.org/10.1017/S0266466620000079 Published online by Cambridge University Press

lags, that is, the impulse response $\partial u_{t+k}/\partial u_t$ is negligible relative to $\partial x_{t+k}/\partial u_t$ as $k \to \infty$. For example, if x_t is an autoregressive process, then ρ is the largest root of its characteristic polynomial. Similarly, parameter β_x defines the long-run predictability in y_t, that is, the impulse response $\partial \varepsilon_{t+k}/\partial u_t$ is negligible relative to $\partial x_{t+k-1}/\partial u_t$ as $k \to \infty$. Let us collect all y_t observations in one vector Y and all x_t observations in another vector X.

The goal is to test a null hypothesis $H_0 : \beta_x = \beta_0$ against a composite alternative $H_a : \beta_x \in \mathrm{B}$, where $\beta_0 \notin \mathrm{B}$. As noted by Jansson and Moreira (2006), the uniformly most powerful (UMP) test generally does not exist for this problem and the asymptotically UMP test does not exist if a set of possible ρ includes a neighborhood of unity; see Jeganathan (1995). One way around this is to consider tests that maximize a weighted-average power (WAP) over alternatives; see Stock and Watson (1996).

A convenient starting point in the analysis of the stated problem is the case in which all parameters except β_x are known. In this case, H_0 is simple and the WAP is obtained by averaging across $\beta_x \in \mathrm{B}$. Denote a probability measure that assigns importance, that is, weights, to different alternatives by $F(\beta_x)$. Then, as in Elliott et al. (2015), H_a can be replaced by a simple hypothesis $H_{a,F}$, according to which the probability density of vector (X, Y) is $\int_{\mathrm{B}} \mathcal{L}(Y, X|\beta_x) dF(\beta_x)$, where $\mathcal{L}(Y, X|\beta_x)$ is the distribution given β_x.

The most powerful test for deciding between the two simple hypotheses is the likelihood ratio test, as follows from the Neyman–Pearson Lemma (see Lehmann and Romano, 2005, Thm. 3.2.1). Therefore, the optimal test rejects for low values of the ratio:

$$\frac{\mathcal{L}(Y, X|\beta_0)}{\int_{\mathrm{B}} \mathcal{L}(Y, X|\beta_x) dF(\beta_x)}.$$

However, if the other parameters $\theta_{-\beta} \in \Theta_{-\beta}$ are unknown, then the null hypothesis is composite $H_0 : \beta_x = \beta_0, \theta_{-\beta} \in \Theta_{-\beta}$. In this case, a usual approach is to further simplify model (1); see Jansson and Moreira (2006), Elliott et al. (2015) and Campbell and Yogo (2006). In particular, the serial correlation in e_t is ruled out. Under this assumption, we can restrict the set of possible tests to the class that depends on a small set of minimal statistics R. Furthermore, the dependence on the shift μ_y is removed by invoking the invariance principle. Finally, the variance of the shocks e_t is assumed to be known and the intercept μ_x is set to zero.

Importantly, the parameter ρ remains the only nuisance parameter that defines the nonstandard behavior of tests in a neighborhood of the unit root. As noted by many authors, it is important to acknowledge the effect of ρ being close to one in small samples; see Cavanagh, Elliott, and Stock (1995). It is known that, even for moderate values of $\rho > 0$, small-sample behavior of statistics can be drastically different from the asymptotic limit $\mathcal{R}$ due to nonuniform convergence over $\rho \in \Theta_\rho$ where Θ_ρ includes unity; see Mikusheva (2007).

Once the tests with optimal properties are established in the Gaussian i.i.d. case, the next usual step is to generalize the result to the non-Gaussian serially

https://doi.org/10.1017/S0266466620000079 Published online by Cambridge University Press

correlated shocks. It is normally possible to suggest statistics R′ that approximate the asymptotic behavior of the minimal statistics R under more general conditions. Therefore, one can replicate the asymptotic profile of optimal tests derived under i.i.d. assumptions. The condition $E(\varepsilon_t|u_{t-1}, u_{t-2}, \dots) = 0$ remains in place under more general assumptions.

This article follows a different route. First, we replace the Gaussian likelihood by an asymptotic equivalent. In particular, we represent the likelihood in terms of X and new series[5] $D_y = (d_{y,1}, \dots, d_{y,\lfloor T/2 \rfloor})$. The new series D_y are asymptotically independent conditional on X. The subsequent analysis is based on the likelihood of X and the conditional distribution of the first q elements of D_y. The rest of the analysis follows the usual blueprint: we derive minimal statistics in the Gaussian case and generalize our results to non-Gaussian distributions in Section 4.

Remarks. Note that the first equation in (1) is not a predictive equation because e_t is not a martingale difference sequence. Nevertheless, parameter β_x is directly related to the problem of long-run predictability testing. In particular, if predictability in ε_t rapidly vanishes with the time horizon, then $E\left(y_{t+k} - \mu_y | x_t, x_{t-1}, \dots\right) \approx \beta_x \rho^k (x_t - \mu_x)$ for a large horizon $k > 0$. Similarly, because $E\left(\sum_{i=1}^{k} (y_{t+i} - \mu_y) | x_t, x_{t-1}, \dots\right) \approx \beta_x \sum_{i=1}^{k} \rho^i (x_t - \mu_x)$, the null hypothesis of no long-run predictability in long-horizon regressions coincides with hypothesis $H_0 : \beta_x = 0$. Naturally, one approach to testing the long-run predictability can be based on long-horizon regressions. However, Monte Carlo simulations in Sizova (2014) demonstrate the low power of such tests in small samples. Another approach is suggested by our referee and involves adding lags of $\Delta x_t = x_t - x_{t-1}$ to the first equation in (1). We consider the latter method in our Monte Carlo study in Section 6.

2.1. Asymptotically Equivalent Likelihood Representation

Notation. For a complex-valued matrix A, A^* denotes its conjugate transpose. Character i is reserved to denote the imaginary unit. Operator Re returns the real part of complex numbers and matrices.

Let us start by defining the multivariate Whittle approximation, as introduced by Whittle (1953) for the stationary case. Consider a sample of the Gaussian stationary vector process $\mathcal{Z}_T = \{Z_t, t = 1, \dots, T\}$ with a mean μ and covariance structure $\Gamma = \{\Gamma_k\}$, $k \in \mathbb{Z}$, where $\Gamma_k = \mathbb{E}(Z_t Z_{t-k}')$. Its Gaussian likelihood is denoted by $\mathcal{L}_G(\mathcal{Z}_T; \Gamma, \mu)$. If Γ_k are absolutely summable, we can define the population spectrum of $\mathcal{Z}_T$ at frequency $\omega \in \mathbb{R}$ as follows: $s(\omega; \Gamma) = (2\pi)^{-1} \sum_{k=-\infty}^{\infty} \Gamma_k \exp\{-i\omega k\}$, which depends only on the covariance structure Γ. The population spectrum $s(\omega; \Gamma)$ shows which part of the process with covariance structure Γ is attributable to the cyclical components at frequency ω in the frequency-domain representation of the process.

[5] Symbol $\lfloor\ \rfloor$ denotes rounding toward minus infinity.

https://doi.org/10.1017/S0266466620000079 Published online by Cambridge University Press

Separately, we can define Fourier transformations of the demeaned data $d_\mu(\omega;\mathcal{Z}_T,\mu) = (2\pi T)^{-1/2}\sum_{t=1}^{T}(Z_t-\mu)\exp\{-i\omega t\}$ and Fourier transformations $d(\omega;\mathcal{Z}_T) = (2\pi T)^{-1/2}\sum_{t=1}^{T} Z_t\exp\{-i\omega t\}$. Note that $d_\mu(\omega_j;\mathcal{Z}_T,\mu) = d(\omega_j;\mathcal{Z}_T)$ for $\omega_j = 2\pi j/T, j = 1,2,\ldots,T-1$. From these Fourier transformations, we can construct a sample analog of $s(\omega;\Gamma)$, referred to as the sample periodogram $\mathcal{I}(\omega;\mathcal{Z}_T,\mu) = d_\mu(\omega;\mathcal{Z}_T,\mu)\times d_\mu^*(\omega;\mathcal{Z}_T,\mu)$.

The Whittle approximation calculated for the sample $\mathcal{Z}_T$, postulated mean μ, and covariance structure Γ is defined as follows:

$$\ln\mathcal{L}_W(\mathcal{Z}_T;\Gamma,\mu) = -T\ln 2\pi - \frac{1}{2}\sum_{j=1}^{T}\Big\{\ln\big[\det|s(\omega_j;\Gamma)|\big] + \mathrm{tr}\big[s(\omega_j;\Gamma)^{-1}\mathcal{I}(\omega_j;\mathcal{Z}_T,\mu)\big]\Big\},$$

where $\omega_j = 2\pi j/T$, $j = 1,2,\ldots,T$, are so-called natural frequencies. Operators det( ) and tr( ) denote determinants and traces of matrices, respectively.

In the following, we consider the Whittle approximation for the likelihood of the bivariate process $Z_t = (x_{t-1}, y_t)'$ with the covariance structure $\Gamma_{x,y}$. To extend the notion of the Whittle approximation to the nonstationary case, it is convenient to work with a detailed formula of the spectrum $s\left(\omega;\Gamma_{x,y}\right)$ that explicitly depends on ρ, β_x, and the covariance structure of the shocks $e_t = (u_t,\varepsilon_t)'$ denoted by Γ_e. The details of this derivation are rather standard and can be found in Hamilton (1996). We obtain that $s\left(\omega;\Gamma_{x,y}(\Gamma_e,\beta_x,\rho)\right)$ is

$$\begin{pmatrix} \frac{s_{uu}(\omega;\Gamma_e)}{|1-\rho\exp\{-i\omega\}|^2} & \frac{\beta_x s_{uu}(\omega;\Gamma_e)}{|1-\rho\exp\{-i\omega\}|^2} + \frac{s_{u\varepsilon}(\omega;\Gamma_e)\exp\{-i\omega\}}{1-\rho\exp\{-i\omega\}} \\ \frac{\beta_x s_{uu}(\omega;\Gamma_e)}{|1-\rho\exp\{-i\omega\}|^2} + \frac{s_{\varepsilon u}(\omega;\Gamma_e)\exp\{i\omega\}}{1-\rho\exp\{i\omega\}} & \frac{\beta_x^2 s_{uu}(\omega;\Gamma_e)}{|1-\rho\exp\{-i\omega\}|^2} + \mathrm{Re}\left[\frac{2\beta_x s_{u\varepsilon}(\omega;\Gamma_e)\exp\{-i\omega\}}{1-\rho\exp\{-i\omega\}}\right] \\ & +s_{\varepsilon\varepsilon}(\omega;\Gamma_e) \end{pmatrix}, \quad \textbf{(2)}$$

where $s_{uu}(\omega;\Gamma_e)$, $s_{u\varepsilon}(\omega;\Gamma_e)$, $s_{\varepsilon u}(\omega;\Gamma_e)$, and $s_{\varepsilon\varepsilon}(\omega;\Gamma_e)$ give a natural partition of the matrix $s(\omega;\Gamma_e)$. Under assumptions of this article, the spectrum of shocks $s(\omega;\Gamma_e)$ is always defined. Therefore, we extend the definition of the matrix $s(\omega;\Gamma_{x,y})$ to the nonstationary case using the above formula for $\omega_j = 2\pi j/T$ if $j = 1,2,\ldots,T-1$ and $\rho\in\mathbb{R}$, as well as for $\omega_T = 2\pi$, or $\omega_0 = 0$, and $\rho\neq 1$.

The representation of the Gaussian likelihood in terms of Fourier transformations of the data was first suggested by Whittle (1951). See also Hannan (1973a), who considers the approximation in the non-Gaussian case, and Dahlhaus (2000), who generalizes the Whittle approximation to locally stationary series. The Whittle approximation is fruitfully applied to the estimation of fractionally integrated processes; see references in Robinson (2003). To the best of our knowledge, although the Whittle approximation is often used to motivate estimators with nonstationary data (e.g., Velasco 1999, and Hurvih, Moulines and Soulier, 2005), a result that proves the asymptotic equivalence of the Whittle approximation and Gaussian likelihood in the nonstationary case does not exist. For example, the Whittle approximation is not defined for the unit root because the spectrum $s^{x,y}(\omega)$ has a singularity at $\rho = 1$ and $\omega = 2\pi$. Even if ρ is less than one but is close to unity, then the accuracy of the Whittle approximation remains under question due to

https://doi.org/10.1017/S0266466620000079 Published online by Cambridge University Press

nonuniform convergence around the discontinuity point $\rho = 1$. Consequently, the Whittle approximation $\mathcal{L}_W(X, Y; \Gamma_{x,y}, \mu)$ can be far from the Gaussian likelihood $\mathcal{L}_G(X, Y; \Gamma_{x,y}, \mu)$ in finite samples for moderately large ρ.

Our approach addresses the above concerns as follows. First, we divide the Whittle approximation into the part that corresponds to the conditional distribution of Y given X and the part that corresponds to the marginal distribution of X. Our method modifies only the conditional part. Although the joint density of Y and X implied by the Whittle approximation is not defined for $\rho = 1$, the corresponding conditional density can be extended by continuity to $\rho = 1$. Second, we find that, even with this simplification, the conditional part of the Whittle likelihood includes an asymptotically large error for the local-to-unity sequences $\rho_T = 1 + c/T$. As a solution, we suggest a modification to the Whittle approximation that yields the conditional distribution of Y that converges uniformly over ρ to the true conditional distribution of Y. This result is summarized in Lemma 1.

LEMMA 1. *Consider model (1) under the assumption that the zero-mean process $e_t = (u_t, \varepsilon_t)$ is Gaussian with serial covariances $\Gamma_{e,k} = E(e_t e'_{t-k})$ that satisfy the condition*[6] $\sum_{k=-\infty}^{\infty} k||\Gamma_{e,k}||^2 < \infty$. *We assume that the spectrum of shocks $s(\omega; \Gamma_e)$ is positive definite with the leading principal minors bounded away from zero for all $-\pi \le \omega \le \pi$. Define a real-valued $T \times T$ matrix $\tilde{A}_{y,y}(\Gamma_e)$ with elements (t, s):*

$$\tilde{A}_{yy}(\Gamma_e) = \left[\frac{1}{2\pi T} \sum_{j=1}^{T} \frac{s_{uu}(\omega_j; \Gamma_e)}{\det[s(\omega_j; \Gamma_e)]} e^{-i\omega_j(t-s)} \right]_{t,s},$$

where $\omega_j = 2\pi j/T$, and a $T \times 1$ complex vector $T_2(\Gamma_e)$ with elements (t):

$$T_2(\Gamma_e) = \left[\frac{1}{2\pi T} \sum_{j=1}^{T} \frac{s_{u\varepsilon}(\omega_j; \Gamma_e)}{\det[s(\omega_j; \Gamma_e)]} e^{-i\omega_j t} \right]_{t}.$$

Define the modified Whittle approximation for $\rho \neq 1$ as follows:

$$\mathcal{L}_{W,m}(X, Y; \Gamma_e, \beta_x, \rho, \mu) = \mathcal{L}_W(X, Y^W(Y, X, \Gamma_e); \Gamma_{x,y}(\Gamma_e, \beta_x, \rho), \mu),$$

where series $Y^W(Y, X, \Gamma_e)$ that replaces Y in the Whittle approximation is constructed as $Y^W(Y, X, \Gamma_e) = Y + \delta_y(\Gamma_e)(x_0 - x_T)$ with $\delta_y(\Gamma_e) = \mathrm{Re}\left(\tilde{A}_{yy}^{-1}(\Gamma_e) T_2(\Gamma_e)\right)$. Define the corresponding conditional density for $\rho \neq 1$ by

$$\mathcal{L}_{W,m}(Y|X; \Gamma_e, \beta_x, \rho, \mu) = \frac{\mathcal{L}_{W,m}(X, Y; \Gamma_e, \beta_x, \rho, \mu)}{\int_Y \mathcal{L}_{W,m}(X, Y; \Gamma_e, \beta_x, \rho, \mu) dY}$$

and extend this density to $\rho = 1$ by taking the limit $\rho \to 1$. Then, the conditional density of Y given X that follows from the Gaussian likelihood,

[6] The norm $||\Gamma_{e,k}||$ is Euclidean.

https://doi.org/10.1017/S0266466620000079 Published online by Cambridge University Press

$\mathcal{L}_G(Y|X;\Gamma_e,\beta_x,\rho,\mu)$, is asymptotically equivalent to $\mathcal{L}_{W,m}(Y|X;\Gamma_e,\beta_x,\rho,\mu)$. In particular,

$$\mathcal{L}_G(Y|X;\Gamma_e,\beta_x,\rho,\mu) = \mathcal{L}_{W,m}(Y|X;\Gamma_e,\beta_x,\rho,\mu) + O_p(1).$$

The approximation error is $O_p(1)$ uniformly over $\rho \in \mathbb{R}$.

The intuition behind this result is as follows. The Whittle approximation to the conditional log-density of Y has the same quadratic form as the Gaussian distribution. We can determine the details of this approximation that correspond to the conditional variance and the conditional mean of Y given X. It appears that both the true conditional variance and its approximated version do not depend on ρ. Therefore, their convergence does not depend on the stationarity of x_t. We can apply the result of Dzhaparidze (1986) (Chap. 1.2, Lem. 1) to claim the asymptotic equivalence of the conditional variances.

The conditional expectation of Y given X takes the form $\beta_x X + \mathbb{E}(\varepsilon|u;\Gamma_e)$, where ε and u are column vectors composed of shocks ε_t and u_t, $t = 1,\dots,T$, respectively, and the conditional expectation depends on the covariance structure of shocks Γ_e. Under the Whittle approximation, the conditional expectation of Y is $\beta_x X + \mathbb{E}(\varepsilon|u;\Gamma_{e,W}(\Gamma_e))$, where expectation $\mathbb{E}(\varepsilon|u;\Gamma_{e,W}(\Gamma_e))$ is calculated using the wrong covariance structure of the shocks $\Gamma_{e,W}(\Gamma_e) \neq \Gamma_e$. For example, the covariance structure $\Gamma_{e,W}(\Gamma_e)$ is such that the relationship between ε_T and u_1 is the same as the relationship between ε_{T-1} and u_T. This causes an end-of-sample effect that is negligible provided that $|\rho| << 1$, but is significant if $\rho = \rho_T = 1 + c/T$. Consequently, Lemma 1 suggests a correction. For illustration, if the shocks are i.i.d., then series Y^W that replaces Y differs from Y only by the last observation, which is $y_T^W = y_T + \frac{s_{u\varepsilon}(0;\Gamma_e)}{s_{uu}(0;\Gamma_e)}(x_0 - x_T)$. After this end-of-sample correction in the Whittle approximation, the approximation error is not dependent on ρ, which ensures uniform convergence of the modified Whittle approximation to the Gaussian likelihood estimation. More details on the proof are given in the Appendix.

An alternative way to interpret Lemma 1 follows from the representation of the previously introduced Whittle approximation $\mathcal{L}_W(\mathcal{Z}_T;\Gamma,\mu)$ as the likelihood of Fourier transformations $d(\omega_j;\mathcal{Z}_T)$; see, for instance, Krafty and Collinge (2013). In particular, likelihood $\mathcal{L}_W(\mathcal{Z}_T;\Gamma,\mu)$ of the stationary series $\mathcal{Z}_T$ can be represented as a product of the marginal likelihoods of Fourier transformations $d(\omega_j;\mathcal{Z}_t)$, $j = 0,\dots,\lfloor T/2\rfloor$, as follows:

$$\ln \mathcal{L}_W(\mathcal{Z}_T;\Gamma,\mu) = j(T) + \sum_{j=0}^{\lfloor T/2\rfloor} \ln \ell_j(d(\omega_j;\mathcal{Z}_T);\Gamma,\mu),$$

where $j(T)$ is a normalizing constant. Therefore, in the stationary case, the asymptotic equivalence between the Gaussian likelihood and the Whittle approximation implies asymptotic independence of the Fourier transformations $d(\omega_j;\mathcal{Z}_t)$ across frequencies $\omega_j = (2\pi j)/T$, $j = 0,\dots,\lfloor T/2\rfloor$, for example, Hannan (1973b) and Brillinger (1975). As noted in Phillips (2007), this (independence) property

https://doi.org/10.1017/S0266466620000079 Published online by Cambridge University Press

does not hold in the unit-root case; therefore, the above decomposition does not apply. However, Corollary 1 states that a related decomposition of the modified Whittle approximation $\mathcal{L}_{W,m}(Y|X;\Gamma_e,\beta_x,\rho,\mu)$ holds for local-to-unity sequences, including the unit-root case.

Corollary 1. *Consider the Whittle approximation* $\mathcal{L}_W(X,Y;\Gamma_{x,y}(\Gamma_e,\beta_x,\rho),\mu)$ *and its decomposition into the likelihoods of Fourier transformations:* $\ln\mathcal{L}_W\left(X,Y;\Gamma_{x,y}(\Gamma_e,\beta_x,\rho),\mu\right)=j(T)+\sum_{j=0}^{\lfloor T/2\rfloor}\ln\ell_j\left(d(\omega_j;X,Y);\Gamma_{x,y}(\Gamma_e,\beta_x,\rho),\mu\right)$. *Then, the modified Whittle approximation in Lemma 1 has the following representation:*

$$\mathcal{L}_{W,m}(X,Y;\Gamma_e,\beta_x,\rho,\mu)=j(T)+\sum_{j=0}^{\lfloor T/2\rfloor}\ln\ell_j\left(d_{W,m}(\omega_j;X,Y,\Gamma_e);\Gamma_{x,y}(\Gamma_e,\beta_x,\rho),\mu\right),$$

$$\text{where } d_{W,m}(\omega_j;X,Y,\Gamma_e):=d(\omega_j;X,Y)+\begin{pmatrix}0\\ s_{\varepsilon u}(\omega_j;\Gamma_e)\left[s_{uu}(\omega_j;\Gamma_e)\right]^{-1}(2\pi T)^{-1/2}(x_0-x_T)\end{pmatrix}.$$

The proof follows from Lemmas 1 and 2 in the Appendix. An important observation from Corollary 1 is that Fourier transformations of Y remain asymptotically independent across frequencies ω_j conditional on all observations of X. The difference from the stationary case is that each density $\ell_j\left(d_m(\omega_j;X,Y,\Gamma_e);\mu,\Gamma_{x,y}(\Gamma_e,\beta_x,\rho)\right)$ is now a function of not only the Fourier transformations of Y and X at frequency ω_j but also of the difference x_0-x_T.

For our further analysis, it is convenient to obtain an explicit formula for the conditional part of the modified Whittle approximation $\mathcal{L}_{W,m}(Y|X;\Gamma_e,\beta_x,\rho,\mu)$ as a likelihood of the Fourier transformations of Y conditional on X, which is our next step undertaken in Proposition 1. For convenience, we denote a k-dimensional complex normal variable by $N^C(a,\Sigma)$, where a is a complex-valued column vector with k elements and Σ is a $k\times k$ complex-valued Hermitian positive definite matrix. The log density of $N^C(a,\Sigma)$ at point $z\in\mathcal{C}^k$ is $-k\ln(\pi)-\ln(\det\Sigma)-(z-a)^*\Sigma^{-1}(z-a)$.

Proposition 1. *Define the modified conditional Whittle approximation* $\mathcal{L}_{W,m}(Y|X;\Gamma_e,\beta_x,\rho,\mu)$ *as in Lemma 1. This approximation takes the form:*

$$\ln\mathcal{L}_{W,m}(Y|X;\Gamma_e,\beta_x,\rho,\mu)=j_{Y|X}(T)+\sum_{j=0}^{\lfloor T/2\rfloor}\ln\ell_{W,m}(d(\omega_j;Y)|X;\Gamma_e,\beta_x,\rho,\mu),$$

where $j_{Y|X}(T)$ *is a normalizing constant and densities* $\ell_{W,m}(d(\omega_j;Y)|X;\Gamma_e,\beta_x,\rho,\mu)$ *correspond to normal variables.*

In particular, for $j=1,\ldots,\lfloor(T-1)/2\rfloor$, *these distributions correspond to complex normals*

$$N^C\left(\frac{s_{yx}(\omega_j;\Gamma_{x,y}(\Gamma_e,\beta_x,\rho))}{s_{xx}(\omega_j;\Gamma_{x,y}(\Gamma_e,\beta_x,\rho))}d(\omega_j;\{x_{t-1}\}_{t=1}^T)-\frac{s_{\varepsilon u}(\omega_j;\Gamma_e)}{s_{uu}(\omega_j;\Gamma_e)}\frac{x_0-x_T}{\sqrt{2\pi T}},\frac{\det[s(\omega_j;\Gamma_e)]}{s_{uu}(\omega_j;\Gamma_e)}\right).$$

https://doi.org/10.1017/S0266466620000079 Published online by Cambridge University Press

If T is even and $j = T/2$, then the corresponding distribution is normal

$$N\left(\frac{s_{yx}(\omega_j;\Gamma_{x,y}(\Gamma_e,\beta_x,\rho))}{s_{xx}(\omega_j;\Gamma_{x,y}(\Gamma_e,\beta_x,\rho))}d(\omega_j;\{x_{t-1}\}_{t=1}^T) - \frac{s_{\varepsilon u}(\omega_j;\Gamma_e)}{s_{uu}(\omega_j;\Gamma_e)}\frac{x_0 - x_T}{\sqrt{2\pi T}}, \frac{\det[s(\omega_j;\Gamma_e)]}{s_{uu}(\omega_j;\Gamma_e)}\right).$$

For $j = 0$, the distribution is likewise normal with mean

$$\sqrt{\frac{T}{2\pi}}\mu_y + \left[\beta_x + \frac{s_{\varepsilon u}(0;\Gamma_e)}{s_{uu}(0;\Gamma_e)}(1-\rho)\right]\left(d(0;\{x_{t-1}\}_{t=1}^T) - \sqrt{\frac{T}{2\pi}}\mu_x\right) - \frac{s_{\varepsilon u}(0;\Gamma_e)}{s_{uu}(0;\Gamma_e)}\frac{x_0 - x_T}{\sqrt{2\pi T}}$$

and variance $\det[s(0;\Gamma_e)]/s_{uu}(0;\Gamma_e)$.

Note that the Fourier transformation of Y at zero frequency $d(0;Y)$ is $\sqrt{T/(2\pi)}\bar{y}_T$, where $\bar{y}_T$ is a sample average. Therefore, the last statement characterizes the approximate distribution of $\bar{y}_T$ conditional on X.

In our further analysis, we will use the explicit formula for the conditional likelihood $\ln \mathcal{L}_{W,m}$ derived in Proposition 1 and a marginal likelihood of X to arrive at a small set of the sufficient statistics for parameters β_x and ρ. By the sufficiency argument, tests with optimal properties can be searched within a class of tests based on these sufficient statistics.

It is also worth noting how the invariance principle employed in prior literature to eliminate the shift parameter μ_y applies to the Whittle approximation. We note that the null hypothesis $H_0 : \beta_x = 0$ and its alternative are invariant with respect to the shifts in Y. In other words, the shift in coordinates—a level with respect to which Y is measured—does not affect our testing problem; see Lehmann and Romano (2005). Following Jansson and Moreira (2006), we can restrict the set of possible tests to the tests that are invariant under transformations of the type "Y plus shift." The minimal statistic[7] that is invariant under this class of transformations is a set $\{d(\omega_1;Y), d(\omega_2;Y), \ldots, d(\omega_{\lfloor T/2\rfloor};Y)\}$. The conditional asymptotic likelihood of this minimal statistic is exactly the likelihood described in Proposition 1 without the term with $j = 0$, which corresponds to the density of the sample average.

Remarks. (i) Our analysis starts from an approximation of the true likelihood as $T \to \infty$. Therefore, we cannot address the optimality of the tests in finite samples even under the strong assumption of Gaussianity. Instead, we consider asymptotic inference in the Gaussian case for $\rho = \rho_T = 1 + c/T$, which is the second econometric framework analyzed by Jansson and Moreira (2006). (ii) The marginal distribution of X is not approximated. However, the methods presented in this paper can be readily extended to the approximated probability density of X. Given our choice of the exact marginal distribution of X in the next section, our

[7] In Jansson and Moreira (2006), the minimal statistic is $(y_2 - y_1, y_3 - y_2, \ldots, y_T - y_{T-1})$. Formally, our minimal statistic has less elements. However, its elements, except for the last one if T is even, are complex numbers.

https://doi.org/10.1017/S0266466620000079 Published online by Cambridge University Press

results can be directly compared with the studies that develop tests under the i.i.d. assumption for u_t shocks, such that these tests remain valid in the presence of serial correlation.

3. LIMITING EXPERIMENT AND NEARLY OPTIMAL TEST IN THE GAUSSIAN CASE

In this section, we continue to consider tests of $H_0 : \beta_x = 0$ under the Gaussianity assumption. We also impose a restriction on the x_t dynamics to be removed in Section 4. The entire set of assumptions is now as follows:

Assumption A'. Gaussianity and weak dependence: zero-mean process $e_t = (u_t, \varepsilon_t)$ is a stationary Gaussian process with maximal correlations[8] $r_{e,k}^{max}$ satisfying the condition $\sum_{k=1}^{\infty} \left(r_{e,k}^{max}\right)^{1-\epsilon} < \infty$ for some $0 < \epsilon < 1$. The spectrum $s(\omega; \Gamma_e)$ is positive definite with leading principal minors bounded away from zero for all $-\pi \geq \omega \geq \pi$.

The above assumption combines three conditions: the Gaussianity, the conditions from [p.1023] Phillips (1988) for the functional central limit theorem (FCLT), and the conditions from Dzhaparidze (1986) for the asymptotic equivalence of the Whittle approximation for stationary processes. In particular, this assumption implies that the strong mixing coefficients α_k of process e_t also satisfy the condition $\sum_{k=0}^{\infty} \alpha_k^{1-\epsilon} < \infty$ by the Kolmogorov–Rozanov theorem (see e.g. Rozanov, 1967, Thm. 10.2). Therefore, the process e_t satisfies the strong mixing conditions from Phillips (1988).

Assumption B'. AR(1) dynamics of x_t: u_t is i.i.d.

The test derived under the assumption above remains valid for more general dynamics of u_t in Section 4. Alternatively, x_t can be explicitly modeled as an AR(p) process; see Mikusheva (2012). This article focuses on the conditional part of the likelihood and maintains the simplest assumption about x_t. A similar procedure can be worked out for the alternative specification, as in Mikusheva (2012).

Assumption C'. Starting value: x_0 is a random variable with distribution that does not depend on β_x and ρ. Future shocks e_t, $t = 1, \ldots, T$ are independent of x_0 and $Ex_0^2 < \infty$.

Assumption D'. Long-run variance: the long-run variance ψ of e_t, which exists and is nondegenerate by Assumption A, is known.

Denote the natural partition of matrix ψ by ψ_{uu}, $\psi_{u\varepsilon} = \psi_{\varepsilon u}$, and $\psi_{\varepsilon\varepsilon}$. Also define the long-run correlation $r := \psi_{\varepsilon u}\psi_{uu}^{-1/2}\psi_{\varepsilon\varepsilon}^{-1/2} \in (-1, 1)$. The long-run variance of ε conditional on u is then denoted by $\psi_{\varepsilon\varepsilon|u} := \psi_{\varepsilon\varepsilon}(1 - r^2) \neq 0$.

[8] Maximal correlation $r_{e,k}^{max}$ is defined as sup corr(η_t, η_{t+k}) where the supremum is taken over all random variables η_t and η_{t+k} that are measurable with respect to σ-algebras generated by observations $(e_t, e_{t-1}, \ldots)$ and $(e_{t+k}, e_{t+k+1}, \ldots)$, respectively.

https://doi.org/10.1017/S0266466620000079 Published online by Cambridge University Press

By the above assumptions, the marginal likelihood of X equals

$$\ln \mathcal{L}_G(X;\mu_x,\rho,\psi_{uu}) = -\frac{T-1}{2}\ln(2\pi) - \frac{T-1}{2}\ln\psi_{uu} - \frac{1}{2}\sum_{t=1}^{T-1}\frac{(x_t-\mu_x-\rho(x_{t-1}-\mu_x)^2}{\psi_{uu}}$$

ignoring the density of x_0, which is not informative about β_x and ρ by our assumptions. The parameter ψ_{uu} is the long-run variance of u_t and coincides with the variance of u_t by Assumption B.

We next obtain an asymptotically equivalent joint likelihood of (X,Y) by combining the approximate conditional density of Y derived in the previous section with the marginal distribution of X,

$$\ln \mathcal{L}_{W|G}(Y,X;\Gamma_e,\beta_x,\rho,\mu) = j_{Y|X}(T) + \sum_{j=0}^{\lfloor T/2\rfloor}\ln \ell_{W,m}(d(\omega_j;Y)|X;\Gamma_e,\beta_x,\rho,\mu) + \ln \mathcal{L}_G(X;\mu_x,\rho,\psi_{uu}).$$

We restrict our attention to the tests that are invariant under transformations of the type "Y plus shift," that is, invariant to the shift in coordinates. As discussed in Section 2.1, this is achieved by removing the term corresponding to $j=0$. Finally, we consider semiparametric tests that depend on a fixed number of low frequencies $q < (T-1)/2$, thus, restricting our attention to the long-run frequency range.

In effect, we consider efficient tests based on observations X and $d(\omega_j;Y)$, $j=1,\dots,q$, where $d(\omega_j;Y)$ are asymptotically independent conditional on X. For this set of observations, we now follow the steps of Elliott et al. (2015) to derive an α-level nearly optimal test that yields the WAP that is asymptotically ε-close to the maximum WAP for the given prespecified weights.

For the asymptotic results we assume that ρ is local to unity and express the original testing problem in the neighborhood of $\beta_x=\beta_0$ and $\rho=1$ in terms of new parameters b and c, where $\beta_x = \beta_0 + \frac{1}{T}\frac{\psi_{\varepsilon\varepsilon|u}^{1/2}}{\psi_{uu}^{1/2}}b$ and $\rho = 1+c/T$. We show the following result.

THEOREM 1. *Consider the testing problem of the null hypothesis $H_0: b=0 \notin B, c\in\Theta_c$ against the alternative $H_a: b\in B, c\in\Theta_c$ based on observations X and Fourier transformations $d(\omega_j;Y)$, $j=1,\dots,q$, $q<(T-1)/2$. Under Assumptions A'–D', the limiting experiment is characterized by the following likelihood:*

$$f(\mathcal{R}|b,c) \propto \exp\left\{b\mathcal{R}_\beta + c\mathcal{R}_\rho - \frac{1}{2}\left(b - c\frac{r}{\sqrt{1-r^2}}\right)^2\mathcal{R}_{\beta\beta} - \frac{1}{2}c^2\mathcal{R}_{\rho\rho}\right\}, \quad (3)$$

https://doi.org/10.1017/S0266466620000079 Published online by Cambridge University Press

where the coefficient of proportionality does not depend on b and c, and the four sufficient statistics $\mathcal{R} = (\mathcal{R}_\beta, \mathcal{R}_\rho, \mathcal{R}_{\beta\beta}, \mathcal{R}_{\rho\rho})$ *are defined as follows:*

$$\mathcal{R}_\beta = 2\,\mathrm{Re}\left\{\sum_{j=1}^{q}\left(\int_0^1 J_c(\tau)e^{-2\pi ij\tau}d\tau\right)\left(\int_0^1 e^{-2\pi ij\tau}dW^z(\tau) + \left(b - \frac{cr}{\sqrt{1-r^2}}\right)\int_0^1 J_c(\tau)e^{-2\pi ij\tau}d\tau\right)^*\right\},$$

$$\mathcal{R}_\rho = \int_0^1 J_c(\tau)dJ_c(\tau) - r(1-r^2)^{-1/2}\mathcal{R}_\beta,$$

$$\mathcal{R}_{\beta\beta} = 2\sum_{j=1}^{q}\left|\int_{\tau=0}^1 e^{-2\pi ij\tau}J_c(\tau)d\tau\right|^2,$$

$$\mathcal{R}_{\rho\rho} = \int_0^1 [J_c(\tau)]^2 d\tau.$$

Process $J_c(\tau)$ *is the Ornstein–Uhlenbeck (O–U) process and* $W^z(\tau)$ *is a standard Brownian motion independent of* $J_c(\tau)$.

By definition, $J_c(\tau)$ has the dynamics $dJ_c(\tau) = cJ_c(\tau)d\tau + dW^u(\tau)$ starting from $J_c(0) = 0$, where $W^u(\tau)$ is a standard Brownian motion independent of $W^z(\tau)$.

Let us restate our objective in terms of the limiting problem characterized by the asymptotic likelihood in Theorem 1. Our goal is to find a test of the null hypothesis $H_0 : b = 0, c \in \Theta_c$ against the alternative $H_a : b \in \mathrm{B}, c \in \Theta_c$ that is based on sufficient statistics $\mathcal{R}$ whose distribution is given in (3) and (nearly) maximizes the WAP for the given weighting measure $F(b,c)$ over $(b,c) \in \mathrm{B} \times \Theta_c$.

At this point, it will be convenient to introduce additional notations for the tests and their properties. Let $\mathcal{D}$ denote a decision rule that governs when to reject H_0 given an observation of $\mathcal{R}$, that is, $\mathcal{D} : \mathcal{R} \to [0,1]$. This decision rule in effect defines a hypothesis test. The size of this test, given the parameter values $(b,c) = (0,c) \in H_0$, is $\alpha(\mathcal{D},c) = \mathbb{E}_{0,c}(\mathcal{D})$, where $\mathbb{E}_{b,c}(\mathcal{D}) = \int \mathcal{D}f(\mathcal{R}|b,c)d\mathcal{R}$. Because the null hypothesis is composite, the size $\alpha(\mathcal{D},c)$ is a function of c. Therefore, we also introduce $\alpha(\mathcal{D})$ that denotes the uniform size of the test d, such that $\alpha(\mathcal{D}) = \sup_{c\in\Theta_c}\alpha(\mathcal{D},c)$. We consider the tests that control the uniform size, that is, the tests with $\alpha(\mathcal{D}) \le \alpha$, where α is a preselected nominal significance level, see Lehmann and Stein (1948). Only such tests are considered valid. With this notation, we can define the *WAP* as follows: $WAP(\mathcal{D}) = \int \mathbb{E}_{b,c}(\mathcal{D})dF(b,c)$. An optimal test $\mathcal{D}^*$ maximizes $WAP(\mathcal{D})$ subject to restriction $\alpha(\mathcal{D}) \le \alpha$. The resulting maximally attainable power is $WAP^* = WAP(\mathcal{D}^*)$. A nearly optimal test $\mathcal{D}^{*,\epsilon}$ is defined as a valid test with $WAP(\mathcal{D}^{*,\epsilon}) \ge WAP^* - \epsilon$, where ϵ is a small positive number.

The search for nearly optimal tests can be simplified given the following two considerations: first, the objective of maximizing WAP in effect replaces the

https://doi.org/10.1017/S0266466620000079 Published online by Cambridge University Press

composite alternative hypothesis $H_a : b \in \mathrm{B}, c \in \Theta_c$ by a simple hypothesis[9] $H_{a,F}$: "the density of $\mathcal{R}$ is $\int_{(b,c)\in \mathrm{B}\times\Theta_c} f(\mathcal{R}|b,c)dF(b,c)$."

Second, using the result of Lem. 1 by Elliott et al. (2015), we can construct a set of power bounds from the powers of tests for simple null hypotheses. In particular, consider an arbitrary probability measure $\Lambda(c)$ on Θ_c. This distribution defines weights across possible parameter values $(b,c) = (0,c)$. We define the likelihood ratio statistic corresponding to the simple null $H_{0,\Lambda}$: "the density of $\mathcal{R}$ is $\int_{c\in\Theta_c} f(\mathcal{R}|0,c)d\Lambda(c)$" and the simple alternative hypothesis $H_{a,F}$: "the density of $\mathcal{R}$ is $\int_{(b,c)\in \mathrm{B}\times\Theta_c} f(\mathcal{R}|b,c)dF(b,c)$," as follows:

$$LR(\mathcal{R},\Lambda) = \frac{\int_{(b,c)\in \mathrm{B}\times\Theta_c} f(\mathcal{R}|b,c)dF(b,c)}{\int_{c\in\Theta_c} f(\mathcal{R}|b=0,c)d\Lambda(c)}.$$

Define a test with rejection rule $\mathcal{D}(\mathcal{R};\Lambda,K^0_\Lambda) = \mathbb{I}\left(\{LR(\mathcal{R},\Lambda) > K^0_\Lambda\right\}$, where K^0_Λ is a critical value chosen so that the size of the test under $H_{0,\Lambda}$ is exactly α, that is, $\alpha(\mathcal{D}(\mathcal{R};\Lambda,K^0_\Lambda),\Lambda) = \int \mathbb{E}_{0,c}(\mathcal{D})\Lambda(c)dc$. By Neyman–Pearson lemma, this test maximizes the WAP across tests subject to the restriction $\alpha(\mathcal{D},\Lambda) \le \alpha$. However, this test is generally invalid, in the sense that its uniform size $\alpha(\mathcal{D})$ can be larger than α. Nevertheless, it provides an upper bound on WAP^*. As shown in Lem. 1 by Elliott et al. (2015), $WAP^* \le WAP(\mathcal{D}(\mathcal{R};\Lambda,K^0_\Lambda))$. Therefore, if we find a valid test such that $WAP(\mathcal{D}) + \epsilon \ge WAP(\mathcal{D}(\mathcal{R};\Lambda,K^0_\Lambda))$, then it would satisfy the definition of the nearly optimal test.

Additionally, consider the second rejection rule $\mathcal{D}(\mathcal{R};\Lambda,K^1_\Lambda) = \mathbb{I}(LR(\mathcal{R},\Lambda) > K^1_\Lambda)$, where constant K^1_Λ is larger than K^0_Λ. The critical value K^1_Λ is chosen to make $\alpha(\mathcal{D}(\mathcal{R};\Lambda,K^1_\Lambda)) = \alpha$. In practice, we can find $K^1_\Lambda > K^0_\Lambda$ by verifying that $\alpha(\mathcal{D}(\mathcal{R};\Lambda,K^1_\Lambda),c_\iota) \le \alpha$ over a large grid of points $c_\iota \in \Theta_c$. This test is valid by construction and, therefore, provides a lower bound on WAP^*, that is,

$$WAP\left(\mathcal{D}(\mathcal{R};\Lambda,K^1_\Lambda)\right) \le WAP^* \le WAP\left(\mathcal{D}(\mathcal{R};\Lambda,K^0_\Lambda)\right). \quad \textbf{(4)}$$

It should be stressed that the above equality holds for any probability density Λ on Θ_c. However, some distributions Λ will result in uninformative bounds, while others may be very informative. If we can find $\Lambda^{*,\epsilon}$ such that the distance between the bounds is less than a small positive ϵ, then $\mathcal{D}(\mathcal{R};\Lambda^{*,\epsilon},K^1_{\Lambda^{*,\epsilon}})$ is the desired nearly optimal test. Thus, the problem of finding $\mathcal{D}^{*,\epsilon}$ turns into the problem of finding an informative $\Lambda^{*,\epsilon}$.

Note that the ideas relayed above are similar to the results obtained by [Thm. 3.8.1] Lehmann and Romano (2005). Although the latter depend on the existence of the least favorable distribution Λ^*—a distribution such that the test with the decision rule $\mathcal{D}(\mathcal{R},\Lambda^*,K^0_{\Lambda^*})$ also yields a valid test for the hypothesis H_0. In particular, Thm. 3.8.1 by Lehmann and Romano (2005) asserts that this test $\mathrm{D}(\mathcal{R},\Lambda^*,K^0_{\Lambda^*})$ yields the maximum WAP among all valid tests for

[9] Because the WAP is calculated by integral $\int_{(b,c)}\left(\int_{\mathcal{R}} \mathcal{D}(\mathcal{R})f(\mathcal{R}|b,c)d\mathcal{R}\right)dF(b,c)$, where $\mathcal{D}(\mathcal{R})$ is a rejection rule for some α-level test, the same WAP can be rewritten as $\int_{\mathcal{R}} \mathcal{D}(\mathcal{R})\left(\int_{(b,c)} f(\mathcal{R}|b,c)dF(b,c)\right)d\mathcal{R}$, which is the probability of rejecting the null hypothesis when the density of $\mathcal{R}$ is $\int_{(b,c)} f(\mathcal{R}|b,c)dF(b,c)$.

https://doi.org/10.1017/S0266466620000079 Published online by Cambridge University Press

the null hypothesis H_0 against the alternative $H_{a,F}$. In contrast, the result (4) holds for any probability measure Λ. In particular, for any candidate $\Lambda^{*,\epsilon}$, we know that the test $\mathcal{D}(\mathcal{R};\Lambda^{*,\epsilon},K^1_{\Lambda^{*,\epsilon}})$ is valid by construction and can verify the distance from the power bound by calculating $\epsilon = WAP(\mathcal{D}(\mathcal{R};\Lambda,K^0_{\Lambda^{*,\epsilon}})) - WAP(\mathcal{D}(\mathcal{R};\Lambda^{*,\epsilon},K^1_{\Lambda^{*,\epsilon}}))$[10]. The distribution $\Lambda^{*,\epsilon}$ is referred to as ϵ-approximate least favorable distribution (ϵ-ALFD); see Müller and Watson (2013) and Andrews, Moreira, and Stock (2008).

The crucial computational step of forming the nearly optimal test is to numerically determine $\Lambda^{*,\epsilon}$ and the corresponding critical value $K^1_{\Lambda^{*,\epsilon}}$. At this point, it is convenient to assume that Θ_c is a bounded set, an assumption that we reconsider in Section 5. For example, suppose $\Theta_c = [\underline{c},\bar{c}]$ with $\underline{c} < 0$ and $\bar{c} > 0$. If Θ_c is bounded, then Elliott et al. (2015) suggest searching for $\Lambda^{*,\epsilon}$ among discrete distributions. Intuitively, if Θ_c were bounded and discrete, then it would be possible to numerically determine the least favorable distribution on Θ_c. Suppose, $\Theta_{c,I}$ is a fine grid on Θ_c that expands as $I \to \infty$. If rejection probabilities are continuous in c, then $\sup_{c\in\Theta_{c,I}}\alpha(\mathcal{D},c)$ can become arbitrary close to $\sup_{c\in\Theta_c}\alpha(\mathcal{D},c)$ as $I \to \infty$.

Consider distributions Λ represented by sets of probabilities $(\lambda_1,\ldots,\lambda_I) \in [0,1]^I$ for values $\Theta_{c,I} = \{c_1,\ldots,c_I\} \subset \Theta_c$. Let $\alpha_\iota(\Lambda)$ be the actual significance level of the test with the decision rule $\mathcal{D}(\mathcal{R},\Lambda,K^0_\Lambda)$ if $c = c_\iota$, that is, $\alpha_\iota(\Lambda) = \alpha(\mathcal{D}(\mathcal{R},\Lambda,K^0_\Lambda),c_\iota)$. The insight of Elliott et al. (2015) is that a candidate $\Lambda^{*,\epsilon} = \Lambda^*[\Theta_{c,I}]$ can be searched within this class of discrete distributions by solving a nonlinear complementarity problem defined by the following conditions:

$$\lambda_\iota \geq 0,\ \alpha_\iota(\Lambda) \leq \alpha,\ \lambda_\iota(\alpha - \alpha_\iota(\Lambda)) = 0, \iota = 1,\ldots,I.$$

That is, distribution $\Lambda^*[\Theta_{c,I}]$ puts a nonzero weight on $c_\iota \in \Theta_c$ if the corresponding rejection probability is $\alpha_\iota = \alpha$. Furthermore, Elliott et al. (2015) note that $\Lambda^*[\Theta_{c,I}]$ is a fixed point of the map $G(\Lambda)$, for which the ιth element is

$$[G(\Lambda)]_\iota = \frac{\lambda_\iota + \max(\alpha_\iota(\Lambda) - \alpha, 0)}{\sum_{m=1}^M(\lambda_m + \max(\alpha_m(\Lambda) - \alpha, 0))}.$$

Define $\mu_\iota = \ln(K^0_\Lambda\lambda_\iota)$. Note that $\alpha_\iota(\Lambda)$ depends on $(\lambda_1,\ldots,\lambda_I)$ and K^0_Λ only through $(\mu_1,\ldots,\mu_I)$. It is convenient to optimize over $\mu_\iota \in \mathbb{R}$ rather than $\lambda_\iota \in [0,1]$ with the solution found by iterations: $\mu_\iota^{(j+1)} = \mu_\iota^{(j)} + \kappa_\mu(\alpha_\iota(\Lambda^{(j)}) - \alpha)$, where κ_μ is the sole tuning parameter equal to two for Monte-Carlo simulations in this article. Suppose, this numerical algorithm yields distribution $\Lambda^*[\Theta_{c,I}]$. For this distribution, we can verify condition $WAP(\mathcal{D}(\mathcal{R};\Lambda,K^0_\Lambda)) - WAP(\mathcal{D}(\mathcal{R};\Lambda,K^1_\Lambda)) < \epsilon$ with $\Lambda = \Lambda^*[\Theta_{c,I}]$. If this condition holds, then $\Lambda^{*,\epsilon} = \Lambda^*[\Theta_{c,I}]$. If not, then the entire search procedure is repeated with a finer grid of points c. This new grid must include points that violate the condition $\alpha(\mathcal{D}(\mathcal{R},\Lambda,K^0_\Lambda),c) \leq \alpha$ with $\Lambda = \Lambda^*[\Theta_{c,I}]$.

[10]This is an important point because numerical procedures to find $\Lambda()$ may not converge. However, if these numerical mechanisms yield Λ such that $WAP(\mathcal{D}(\mathcal{R};\Lambda,K^0_\Lambda)) - WAP(\mathcal{D}(\mathcal{R};\Lambda,K^1_\Lambda)) < \epsilon$, then this Λ yields a nearly optimal test $\mathcal{D}(\mathcal{R};\Lambda,K^1_\Lambda)$. Note that this statement does not imply that Λ is a good approximation to the unknown least favorable distribution if it exists.

https://doi.org/10.1017/S0266466620000079 Published online by Cambridge University Press

To summarize, the algorithm that determines $\Lambda^{*,\epsilon}$ and the corresponding critical value $K^1_{\Lambda^{*,\epsilon}}$ is implemented as follows: first, we choose an initial grid $\Theta_{c,I}$. Second, we find a candidate $\Lambda^*[\Theta_{c,I}]$ by mean-value iterations. Third, we find K^0 such that $\sup_{c\in\Theta_{c,I}}\alpha(\mathcal{D}(\mathcal{R};\Lambda^*[\Theta_{c,I}],K^0),c)=\alpha$. Fourth, we find the corresponding power bound $\bar{\pi}=WAP(\mathcal{D}(\mathcal{R};\Lambda^*[\Theta_{c,I}],K^0))$. Fifth, we find K^1 such that $\sup_{c\in\Theta_c}\alpha(\mathcal{D}(\mathcal{R};\Lambda^*[\Theta_{c,I}],K^1),c)=\alpha$ and calculate the corresponding $\underline{\pi}=WAP(\mathcal{D}(\mathcal{R};\Lambda^*[\Theta_{c,I}],K^1))$. In practice, this supremum is determined on a finer grid $\Theta_{c,I'}$. In parallel, we calculate $\alpha(\mathcal{D}(\mathcal{R};\Lambda^*[\Theta_{c,I}],K^0),c)$ for all $c\in\Theta_{c,I'}$ Finally, we verify that $\bar{\pi}-\underline{\pi}<\epsilon$. If not, then we repeat the same steps with a new grid that includes those points of $\Theta_{c,I}$ that have nonzero probability in $\Lambda^*[\Theta_{c,I}]$ and those points from $\Theta_{c,I'}$ for which the size $\alpha(\mathcal{D}(\mathcal{R};\Lambda^*[\Theta_{c,I}],K^0),c)$ is larger than α by more than a numerical error.

Note that by choosing a particular density $F(b,c)$ for the WAP calculations, we implicitly take a stand on which points are important. There are two considerations that can guide the choice of $F(b,c)$. First, the grid of c can be chosen to be uniform in the natural parameterization of this parameter. For example, Elliott et al. (2015) suggest a grid $c_\iota=-(0.25m)^2$ for integers $m=0,1,2,3,\ldots$. Note that this grid includes the important unit root case. In this article, we marginally extend the domain of F into the positive range to include mildly explosive processes, that is, $c_\iota=\pm(0.25m)^2$, where m are integers such that $c_\iota\in(-200,3)$. The choice of the lower bound $c=-200$ yields ρ of only $1/2$ at quarterly frequencies for a large time span of 100 years.

The second consideration is to choose $F(b,c)$ that yields tests close to the power curve generated by point optimal tests, see King (1987). King (1987) suggests to choose $F(b,c)$ so that it yields the same rejection probabilities across c. Elliott et al. (2015) consider the probability mass F in the WAP that puts equal weight on values (b,c) in the discrete set (b_ι,c_ι) with $b_\iota=1.645\sqrt{\frac{-2c_\iota+6}{1-r^2}}$. This choice of b_ι yields approximately equal powers across $c_\iota<0$ for the test of Elliott et al. (2015), whose test can be also obtained in the limit as $q\to\infty$, see Section 7. Likewise, this choice of b_ι yields approximately equal test powers across $c_\iota<0$ for the test with a small number of frequencies q.

Note that the empirical implementation of the test relies on a table of $\lambda_\iota K^1_{\Lambda^{*,\epsilon}},\iota=1,\ldots,I$ for any choice of r, q, and α. This table can be precalculated as is usually conducted for the critical values of the statistical tests. To implement the test, we require finite-sample analogs of $\mathcal{R}$, say R. The finite-sample analog of the rejection rule $\mathcal{D}(\mathcal{R};\Lambda^{*,\epsilon},K^1_{\Lambda^{*,\epsilon}})$ becomes

$$\mathrm{D}(\mathrm{R};\Lambda^{*,\epsilon},K^1_{\Lambda^{*,\epsilon}})=\left\{\frac{\sum_{\iota=1}^I f(\mathrm{R}|b_\iota,c_\iota)}{\sum_{\iota=1}^I f(\mathrm{R}|0,c_\iota)\lambda_\iota K^1_{\Lambda^{*,\epsilon}}}>1\right\}.$$

Statistics R can be defined as in the proof of Theorem 1:

$$R_\beta=\frac{4\pi}{T}\frac{1}{\sqrt{\psi_{\varepsilon\varepsilon|u}\psi_{uu}}}\mathrm{Re}\sum_{j=1}^q\left\{d(\omega_j;\{x_{t-1}\}_{t=1}^T)\left[d(\omega_j;Y)-\beta_0 d(\omega_j;\{x_{t-1}\}_{t=1}^T)\right.\right.$$

https://doi.org/10.1017/S0266466620000079 Published online by Cambridge University Press

$$- \frac{r\psi_{\varepsilon\varepsilon}^{1/2}}{\psi_{uu}^{1/2}}(d(\omega;\{x_t\}_{t=1}^{T}) - d(\omega_j;\{x_{t-1}\}_{t=1}^{T}))\Big]^*\Big\},$$

$$R_\rho = \psi_{uu}^{-1}T^{-1}\sum_{t=1}^{T-1}(x_{t-1}-\mu_x)(x_t - x_{t-1}) - \frac{r}{\sqrt{1-r^2}}R_\beta,$$

$$R_{\beta\beta} = \frac{4\pi}{T^2\psi_{uu}}\sum_{j=1}^{q} d(\omega_j;\{x_{t-1}\}_{t=1}^{T})d(\omega_j;\{x_{t-1}\}_{t=1}^{T})^*,$$

$$R_{\rho\rho} = \frac{1}{T^2\psi_{uu}}\sum_{t=1}^{T}(x_{t-1}-\mu_x)^2. \tag{5}$$

Remark. (i) The convergence of the fixed-point iterations for μ_ι can be assessed in a standard manner by examining values of probabilities $\lambda_\iota^{(j)}$ across iterations $j = 1, 2, \ldots$. For our choice of the weighting measure $F(b, c)$, these sequences appear to converge. However, note that the success of the method can be assessed directly by comparing the power of the test $\mathcal{D}(\mathcal{R}; \Lambda^{*,\epsilon}, K^0_{\Lambda^{*,\epsilon}})$, which is an α-size test under the simple hypothesis $H_{0,\Lambda^{*,\epsilon}}$, to the power of the test $\mathrm{D}(\mathcal{R}; \Lambda^{*,\epsilon}, K^1_{\Lambda^{*,\epsilon}})$, which has size α under H_0. For our choice of the weighting measure $F(b, c)$, the resulting power loss of 0.5% is obtained with $\Lambda^{*,\epsilon}$ defined on the domain of $F(b, c)$, that is, $\Theta_{c,I} = \{c_\iota\}$. (ii) An alternative algorithm to compute $\Lambda^{*,\epsilon}$ can be devised as a solution to a dual linear program; see Moreira, Mourão, and Moreira (2014) and Chiburis (2009). The details of this algorithm are presented in the Appendix. For the same choice of the weighting measure $F(b, c)$, as previously described, and for the same class of discrete distributions, among which $\Lambda^{*,\epsilon}$ is searched, we attain ϵ of 2%. Because our results appear better with the computational method of Elliott et al. (2015), we proceed with the latter algorithm. (iii) We emphasize the two important steps of the algorithm. The first is the size control, which is verified on the fine grid of c. The second step is calculation of ϵ, which is the exact power loss compared with the asymptotic power bound. Although it is not guaranteed that the general algorithm suggested by Elliott et al. (2015) always results in a small ϵ for any choice of $F(b, c)$, this algorithm produces good results in our case with $\epsilon = 0.5\%$. (iv) The algorithm described in this section applies to other choices of $F(b, c)$. As an example, we examined the case in which $F(b, c)$ assigns all of the probability mass to one point (b_F, c_F) and considered c_F between zero and -20. The resulting ϵ varied between 0.25% and 1.0% for the same class of discrete distributions of $\Lambda^{*,\epsilon}$, as described in this section, without new points added to the domain of $\Lambda^{*,\epsilon}$.

4. TEST IN THE GENERAL NON-GAUSSIAN CASE

We now relax the assumptions from the previous section. In particular, we replace the Gaussianity Assumption A' by the following more general condition.

https://doi.org/10.1017/S0266466620000079 Published online by Cambridge University Press

Assumption A. Mixing condition: for some $0 < \epsilon < 1$ and $\delta > 0$, $E||e_t||^{\frac{2}{\epsilon}+\delta} < \infty$, and strong mixing coefficients α_i are such that $\sum_{i=1}^{\infty} \alpha_i^{1-\epsilon} < \infty$.

We also remove the AR(1) assumption for x_t. The claim is that we can asymptotically achieve the same power as the asymptotic nearly optimal test in the Gaussian case by employing statistics R′ that weakly converge to the statistics $\mathcal{R}$. The next theorem shows that we can construct such statistics converging to $\mathcal{R}$ under the more general assumption stated above.

THEOREM 2. *Let $\hat{\psi}_{uu}$, $\hat{\psi}_{\varepsilon\varepsilon}$, and $\hat{r}$ be consistent estimates of the long-run variances $\psi_{uu} > 0$, $\psi_{\varepsilon\varepsilon} > 0$, and the long-run correlation $|r| < 1$, respectively. Let $\hat{\sigma}_{uu}$ be a consistent estimate of the variance of u_t. Suppose, x_0 is a random variable with $Ex_0^2 < \infty$ and dynamics represented by model (1). Under Assumption A, the following finite sample statistics R′ weakly converge to the limiting statistics $\mathcal{R}$ defined in Theorem 1.*

$$R'_\beta = \left(\hat{\psi}_{uu}\hat{\psi}_{\varepsilon\varepsilon}(1-\hat{r}^2)\right)^{-1/2} \frac{4\pi}{T} \mathrm{Re}\left[\sum_{j=1}^{q} d(\omega_j; \{x_{t-1}\}_{t=1}^T)(d(\omega_j; Y) - \beta_0 d(\omega_j; \{x_{t-1}\}_{t=1}^T) - \hat{r}\sqrt{\hat{\psi}_{\varepsilon\varepsilon}/\hat{\psi}_{uu}}(d(\omega_j; \{x_t\}_{t=1}^T) - d(\omega_j; \{x_{t-1}\}_{t=1}^T)))^*\right],$$

$$R'_\rho = \frac{1}{2T}(\hat{\psi}_{uu}^{-1}x_T^2 - \hat{\sigma}_{uu}^{-1}\sum_{t=1}^{T}(x_t - x_{t-1})^2) - \hat{r}(1-\hat{r}^2)^{-1/2}R'_\beta,$$

$$R'_{\beta\beta} = \hat{\psi}_{uu}^{-1}T^{-2}4\pi\sum_{j=1}^{q}|d(\omega_j; \{x_{t-1}\}_{t=1}^T)|^2,$$

$$R'_{\rho\rho} = \hat{\psi}_{uu}^{-1}T^{-2}\sum_{t=1}^{T}x_{t-1}^2.$$

The algorithm of the test now becomes as follows. First, we obtain the tables of $\lambda_\iota K'_\alpha$ for a given choice of the significance level α, the number of frequencies q, and a consistent estimate of r. Second, we calculate the statistics R′ as in Theorem 2. We reject H_0 if the average of the likelihood $f(\mathrm{R}'|b_\iota, c_\iota)$ across (c_ι, b_ι), $\iota = 1, \ldots, I$, is larger than $1/I\sum_{\iota=1}^{I} f(\mathrm{R}'|0, c_\iota)\lambda_\iota K_{\alpha, H_0}$.

5. UNBOUNDED Θ_C

As is usually the case with asymptotic results based on the local-to-unity approximations, there is a concern about the validity of the presented test for moderate values of $|\rho| < 1$. In particular, tests that behave differently if $c \to -\infty$ in the limiting problem and for a fixed $|\rho| < 1$ with $T \to \infty$ were shown to exhibit an incorrect size in the stationary case; see Phillips (2014). A separate concern is the performance of the test for low negative values of c. So far, we abstracted

https://doi.org/10.1017/S0266466620000079 Published online by Cambridge University Press

from problems associated with low negative values of c altogether by restricting our analysis to the case of the bounded set $\Theta_c = [\underline{c}, \overline{c}]$, to which c belongs. In particular, we used this assumption in constructing the ε-ALFD distribution in Section 3. Unfortunately, the likelihood ratio statistic $LR(\mathcal{R}, \Lambda)$ calculated with fixed $\Lambda = \Lambda^{*,\epsilon}$ does not converge to a standard likelihood ratio for $c \to -\infty$. However, Elliott et al. (2015) offer a solution to the aforementioned problems by incorporating a switch between a nearly optimal test, which is derived for local-to-unity sequences of ρ, and a test with a required significance level for sequences $c_T \to -\infty$. Herein, we apply this method to our testing problem.

Consider $\Theta_c = (-\infty, \overline{c}]$. Let R_T be a vector of finite-sample statistics that are asymptotically sufficient under local-to-unity asymptotics. For concreteness, we can consider R_T as the same as the statistics R' in Theorem 2. The asymptotic distribution of R_T depends on the limits of sequences $b_T = (\beta_T - \beta_0)T$ and $c_T = (\rho_T - 1)T$. If b_T and c_T are constant, then the weak limit of R_T is $\mathcal{R}$ defined in Theorem 1.

The suggested test with a switch involves two parts. The first part of the test takes the following form:

$$\mathrm{D}_n(\mathrm{R}_T) = \mathbb{I}\left\{ \frac{\int_{(b,c)\in \mathrm{B}\times\Theta_c} f(\mathrm{R}_T|b,c)dF(b,c)}{\int_{c\in\Theta_c} f(\mathrm{R}_T|b_0,c)d\Lambda(c)^{*,\epsilon}} > K_{\alpha,n} \right\},$$

where $K_{\alpha,n}$ is a constant that depends on the significance level α, and $\Lambda^{*,\epsilon}$ is the ε-ALFD distribution determined numerically as in Section 3 while assuming that c belongs to a bounded interval $[\underline{c}, \overline{c}]$ in Θ_c. As previously discussed, functions $f(.|b,c)$ are densities of the limiting statistics $\mathcal{R}$ given parameters b and c. Thus, $\mathrm{D}_n(\mathrm{R}_T) = \mathcal{D}(\mathrm{R}_T; \Lambda^{*,\epsilon}, K_{\alpha,n})$, where the decision rule $\mathcal{D}$ is defined in Section 3. By the continuous mapping theorem (CMT), the limit of this test is $\mathcal{D}(\mathcal{R}, \Lambda^{*,\epsilon}, K_{\alpha,n})$ if parameters b_T and c_T are constant.

In the following, we switch to another test in a, so called, near stationary region of c_T. This switch depends on a statistic $C(\mathrm{R}_T)$ with the following property: for all sequences $c_T \to -\infty$ and all sequences of b_T in $\mathbb{R}$, probability $\Pr(C(\mathrm{R}_T) < K_c)$ converges to 1. If $C(\mathrm{R}_T) < K_c$, the test switches to another statistic $t(\mathrm{R}_T)$, whose weak limit $t(\mathrm{R}_T) \Rightarrow \xi(b)$ remains the same for all sequences $c_T \to -\infty$ when $b_T = b$. The rejection rule after the switch becomes $\mathrm{D}_s(\mathrm{R}_T) = \mathrm{I}\{t(\mathrm{R}_T) > K_{\alpha,s}\}$, where $K_{\alpha,s}$ is another constant that, together with $K_{\alpha,n}$, ensures the asymptotic α-level of the test for all finite c in Θ_c.

Let this finite-sample switch test be denoted by $\mathrm{D}_{ns}(\mathrm{R}_T)$. Suppose that $C(\mathrm{R}_T)$ and $t(\mathrm{R}_T)$ are almost surely continuous in R_T. Then, the limit of $\mathrm{D}_{ns}(\mathrm{R}_T)$ for fixed $b_T = b$ and $c_T = c$ is the test defined on the space of $\mathcal{R}$ that switches from the rejection rule $\mathcal{D}(\mathcal{R}; \Lambda^{*,\epsilon}, K_{\alpha,n})$ to $\mathcal{D}_s(\mathcal{R}) = \{t(\mathcal{R}) > K_{\alpha,s}\}$ if $C(\mathcal{R}) < K_c$. Similarly, we can derive the limit of $\mathrm{D}_{ns}(\mathrm{R}_T)$ in the near stationary region, that is, for sequences $c_T \to -\infty$. In the limit, this test rejects if $\xi(b) > K_{\alpha,s}$.

The values of $K_{\alpha,n}$, $K_{\alpha,s}$, and K_c are adjusted to obtain an α-level test in the limiting problem for local-to-unity sequences of ρ_T. As previously discussed, the

https://doi.org/10.1017/S0266466620000079 Published online by Cambridge University Press

upper bound in inequality (4) can be used to evaluate the test power loss relative to the power bound. Furthermore, the next result shows that the test with the same choice of $K_{\alpha,n}$, $K_{\alpha,s}$, and K_c uniformly controls for the correct significance level over the unbounded Θ_c.

We use the following notation in the next theorem: $\mathrm{R}_T(c_T, b_T)$ denotes finite-sample statistics whose distribution depends on parameters c_T and b_T, and $\mathcal{R}(c,b)$ are asymptotic statistics whose distribution depends on parameters c and b.

THEOREM 3. *Let $b_T = (\beta_T - \beta_0)T$ and $c_T = (\rho_T - 1)T$. Define finite sample statistics $R_T(c_T, b_T) \in \Theta_R \subset \mathbb{R}^q$ that converge weakly to $\mathcal{R}(c,b) \in \Theta_R$ for any sequence c_T that has a finite limit c in $\Theta_c = (-\infty, \overline{c}]$ if $b = b_T = 0$. Define continuous functions $C(R) : \Theta_R \to \mathbb{R}$ and $t(R) : \Theta_R \to \mathbb{R}$ such that for all sequences c_T converging to $-\infty$ and $b_T = 0$, probability $Pr(C(R_T(c_T, b_T) < K_c)$ converges to 1 and $t(R_T(c_T, b_T))$ converges to a random variable $\xi(0)$. Select $K_{\alpha,s}$ such that $Pr(\xi(0) > K_{\alpha,s}) \leq \alpha$. Select $K_{\alpha,n}$ such that $D_{ns}(\mathcal{R}(c,b))$ is an α-level test of $H_0 : b = 0$, that is, $sup_{c\in\Theta_c} \limsup_{T\to\infty} E\,(D_{ns}(R_T(c_T = c, b_T = 0))) \leq \alpha$. Then, $\limsup_{T\to\infty} sup_{c\in\Theta_c} E\,(D_{ns}(R_T(c_T = c, b_T = 0))) \leq \alpha$.*

It is left to show that the finite-sample statistics R' introduced in Section 4 indeed converge to $\mathcal{R}$ for any sequence c_T that has its finite limit in Θ_c. The proof is based on Thm. 3.1 by Hansen (1992), requires a stronger assumption on mixing coefficients, and is given in the Appendix.

Finally, we discuss the choices of $C(R)$ and $t(R)$ for the switch rule and consider two options. The first option is a "safety switch." It serves to control rejection probabilities in the near stationary region. The second option is to switch to a standard test. This method also increases the test power in the near stationary region. Although the second option seems more appealing, it requires additional assumptions on the dynamics of u_t. The remaining part of this section provides the details for these two methods.

"Safety Switch." We note that the power of the long-run nearly optimal test naturally becomes low for low values of c. Indeed, one can show that, for a fixed distribution $\Lambda^{*,\epsilon}$, the probability to reject using the likelihood-ratio test $\mathcal{D}(\mathcal{R}(c,b), \Lambda^{*,\epsilon}, K^1_{\Lambda^{*,\epsilon}})$ converges to zero as c decreases to $-\infty$ for a fixed b. Likewise in small samples, with ρ being small, little information is contained in the first q Author Fourier transformations[11] of Y. Therefore, our suggestion is to use $\mathrm{D}_s = \alpha$. This choice of D_s does not require any additional assumptions and provides a robust method to control the size of the test in the near stationary region. This switch can be triggered by the switch statistic

$$C(\mathrm{R}') = \frac{\mathrm{R}'_\rho + \frac{r}{\sqrt{1-r^2}}\mathrm{R}'_\beta}{\mathrm{R}'_{\rho\rho}},$$

[11] In contrast, in large samples, a low c_T can occur with $\rho_T = 1 + c_T/T \approx 1$ as long as T is sufficiently large, in which case one should consider an econometric framework with $q/T + 1/q \to 0$ as described in Section 7.

https://doi.org/10.1017/S0266466620000079 Published online by Cambridge University Press

which is equal to $\frac{T}{2}\left(\frac{x_T^2}{\sum_{t=1}^T x_{t-1}^2} - \frac{\hat{\psi}_{uu}}{\hat{\sigma}_{uu}}\frac{\sum_{t=1}^T (x_t - x_{t-1})^2}{\sum_{t=1}^T x_{t-1}^2}\right)$ and converges to $-\infty$ for all $c_T \to -\infty$.

"Standard Maximum Likelihood (MLE) t-test." Elliott et al. (2015) suggest that the alternative test be based on statistic $t(\mathcal{R})$, which in our case is

$$\frac{\frac{\mathcal{R}_\beta}{\mathcal{R}_{\beta\beta}} + \frac{r}{\sqrt{1-r^2}}C(\mathcal{R})}{\sqrt{\mathcal{R}_{\beta\beta}^{-1} + \frac{r}{\sqrt{1-r^2}}\mathcal{R}_{\rho\rho}^{-1}}},$$

and the switch statistic $C(\mathcal{R})$. These statistics are motivated as follows. Statistic $C(\mathcal{R})$ is the maximum likelihood (ML) estimator of c. That is, the test switches if the ML estimate of c falls below a threshold K_c. Statistic $t(\mathcal{R})$ is the t-statistic based on the ML estimate of b.

The corresponding finite-sample statistic $C(\mathrm{R}') = \frac{T}{2}\left(\frac{x_T^2}{\sum_{t=1}^T x_{t-1}^2} - \frac{\hat{\psi}_{uu}}{\hat{\sigma}_{uu}}\frac{\sum_{t=1}^T (x_t - x_{t-1})^2}{\sum_{t=1}^T x_{t-1}^2}\right)$ converges to $-\infty$ for all $c_T \to -\infty$. Therefore, condition $\Pr(C(\mathrm{R}') < K_c) \to 1$ is satisfied. However, the limiting distribution of both $t(\mathrm{R})$ and $t(\mathrm{R}')$ depends on the model for the serial correlation and conditional heteroskedasticity of u_t; see Andrews and Guggenberger (2012). Therefore, additional assumptions are required. For example, assuming that u_t is a homoskedastic martingale difference sequence, we obtain that $t(\mathrm{R}) \Rightarrow N(0,1)$ for all $c_T \to -\infty$, see Theorem 5 in the Appendix.

If μ_x is unknown, its value can be replaced with the sample average $\bar{x}_T = \frac{1}{T}\sum_{t=1}^T x_{t-1}$ in the statistics R_ρ and $R_{\rho\rho}$. For example, consider $R_\beta^\mu = R_\beta$, $R_{\beta\beta}^\mu = R_{\beta\beta}$,

$$R_\rho^\mu = \frac{1}{T\psi_{uu}}\sum_{t=1}^{T-1}(x_{t-1} - \bar{x}_T)(x_t - x_{t-1}) - \frac{r}{\sqrt{1-r^2}}R_\beta^\mu \quad \text{and}$$

$$R_{\rho\rho}^\mu = \frac{1}{T^2\psi_{uu}}\sum_{t=1}^{T}(x_{t-1} - \bar{x}_T)^2.$$

In this case, $t(\mathrm{R}^\mu) \Rightarrow N(0,1)$ in the near stationary region. In the local-to-unity case, $t(\mathrm{R}^\mu) \Rightarrow t(\mathcal{R}^\mu)$, where $\mathcal{R}_\beta^\mu = \mathcal{R}_\beta$, $\mathcal{R}_{\beta\beta}^\mu = \mathcal{R}_{\beta\beta}$,

$$\mathcal{R}_\rho^\mu = \int_0^1 J_c^\mu(\tau)dJ_c(\tau) - r(1-r^2)^{-1/2}\mathcal{R}_\beta^\mu,$$

$$\mathcal{R}_{\rho\rho}^\mu = \int_0^1 \left(J_c^\mu(\tau)\right)^2 d\tau$$

and $J_c^\mu(\tau) = J_c(\tau) - \int_0^1 J_c(r)dr$ is the demeaned O–U process. It should be noted that the critical values for this switch test depend on the joint distribution of $\mathcal{R}$ in the nearly optimal test and $\mathcal{R}^\mu$ in the t-test.

https://doi.org/10.1017/S0266466620000079 Published online by Cambridge University Press

Finally, if u_t is serially correlated but its correlations are known or can be consistently estimated, then one can introduce modified statistics in the switch rule. Suppose, $x_t = \rho x_{t-1} + \Psi(L)\tilde{u}_t$, where $\tilde{u}_t$ is a homoskedastic martingale difference sequence and $\Psi(L)$ is invertible, that is, $\tilde{u}_t = \Psi(L)^{-1}u_t$. Consider $\tilde{x}_t = \Psi(L)^{-1}x_t$, which is a rotated version of x_t. The corresponding statistics are $\tilde{R}_\beta = R_\beta$, $\tilde{R}_{\beta\beta} = R_{\beta\beta}$,

$$\tilde{R}_\rho = \frac{1}{T\psi_{\tilde{u}\tilde{u}}}\sum_{t=1}^{T-1}(\tilde{x}_{t-1} - \bar{\tilde{x}}_T)(\tilde{x}_t - \tilde{x}_{t-1}) - \frac{r}{\sqrt{1-r^2}}\tilde{R}_\beta, \quad \text{and}$$

$$\tilde{R}_{\rho\rho} = \frac{1}{T^2\psi_{\tilde{u}\tilde{u}}}\sum_{t=1}^{T}(\tilde{x}_{t-1} - \bar{\tilde{x}}_T)^2,$$

where $\psi_{\tilde{u}\tilde{u}}$ is a long-run variance of $\tilde{u}_t$. For all $c_T \to -\infty$, $t(\tilde{R}) \Rightarrow N(0,1)$ following the same steps as in the proof of Theorem 5. Note that the critical values will depend on the joint distribution of $\mathcal{R}$ and the limits of $\tilde{R}_\rho$ and $\tilde{R}_{\rho\rho}$ in the local-to-unity case. Therefore, the critical values of this test will depend on $\Psi(L)$.

If T and q are small, Monte Carlo simulations demonstrate only negligible power differences between the tests based on the safety switch and the t-test. Herein, we present the results based on the MLE t-test for i.i.d. u_t. In our Monte Carlo simulations, we set $K_c = -130$, $\alpha = 5\%$, $K_{\alpha,s} = 1.645$ and adjust $K_{\alpha,n}$ up from $K^1_{\Lambda^{*,\epsilon}}$ to obtain an asymptotic α-level test for a grid[12] of $c \in \Theta_c$. We also reoptimize $\Lambda^{*,\epsilon}$ to remove the probability mass from low values of $c \in [\underline{c}, \overline{c}]$,

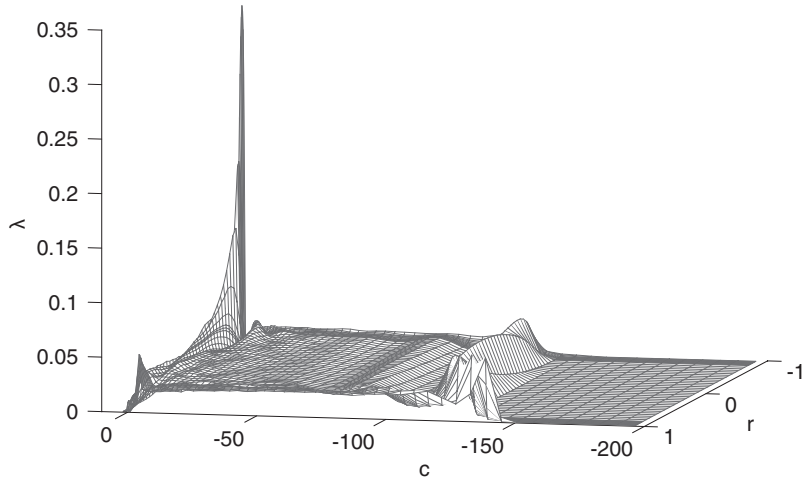


FIGURE 1. Approximate least favorable distribution $q = 2$.

[12] Although Θ_c is unbounded, controlling for the correct asymptotic test level is rather simple because the probability to reject H_0 quickly converges to α for $c < K_c$ under the simple switch test $D_s = \alpha$. Similarly, for the switch test

https://doi.org/10.1017/S0266466620000079 Published online by Cambridge University Press

for which the event $C(\mathcal{R}(c,b)) \geq K_c$ is unlikely. The resulting $\Lambda^{*,\epsilon}$ is plotted in Figure 1 for $q = 2$. The tables for other choices of q are available on the author's web page.

The algorithm of the test now becomes as follows. First, we obtain the tables of $\lambda_\iota K_{\alpha,n}$ for a given choice of the significance level α, the number of frequencies q, and a consistent estimate of r. Second, we calculate the statistics R′ as in Theorem 2. We reject H_0 if the average of the likelihood $f(\mathrm{R}'|b_\iota, c_\iota)$ across (c_ι, b_ι), $\iota = 1, \ldots, I$, is larger than $1/I \sum_{\iota=1}^{I} f(\mathrm{R}'|0, c_\iota)\lambda_\iota K_{\alpha,n}$ and estimated $C(\mathrm{R}')$ is larger than K_c, or if the t-test $t(\mathrm{R}) > K_{\alpha,s}$ and $C(\mathrm{R}') \leq K_c$.

6. MONTE CARLO SIMULATIONS

In this section, we test the performance of the long-run nearly optimal test presented in this article through Monte Carlo simulations. We assume that time is measured in annual units and consider three sample sizes: $T = 50$, $T = 100$, and $T = 1{,}000$, where the last represents the case with large T. We consider three possible persistence values for small samples: $\rho = 0.7$, which yields a half-life period of 2 years; $\rho = 0.9$, which yields a half-life period of 6 years; and $\rho = 1.0$, which corresponds to the unit-root case. For the large sample, the parameter $\rho_T = 1 + c/T$, and c is -30, -10, or zero, which correspond to ρ equal to 0.7, 0.9, or 1.0, respectively, if $T = 100$.

We consider two cases. In the first case, there is no short-run endogeneity and $E(\varepsilon_t|u_{t-1}, \ldots) = 0$. In this case, there is no reason to restrict the number of frequencies. However, we report the results in this case to evaluate the efficiency loss resulting from the data reduction. The benchmark is the nearly optimal test of Elliott et al. (2015) (denoted as EMW) that is based on the entire Gaussian likelihood with a switch to the Maximum Likelihood (MLE) t-test (see Appendix E.2). The simulated model is

$$y_t = \mu_y + \beta_x x_{t-1} + r u_t + \sqrt{1 - r^2} z_t,$$
$$x_t = \rho x_{t-1} + u_t,$$

where (u_t, z_t) are uncorrelated i.i.d. shocks. The slope β_x takes values in $\{0,\ 0.1,\ 0.5\}$. As it will follow from the simulation results, the value $\beta_x = 0.5$ corresponds to strong predictability and the value $\beta_x = 0.1$ is an intermediate case. We test the hypothesis $\beta_x = 0$ against the one-sided alternative $\beta_x > 0$ at the $\alpha = 5\%$ significance level.

The results of the simulations for the large sample are reported in Table 1. The actual significance levels for all choices of q, T, ρ, and r are lower than or approximately 5%. The EMW test yields the highest test power and serves as a benchmark because it uses all of the information from the Gaussian likelihood. Compared to this benchmark, the long-run nearly optimal test with $q = 10$ is close

based on the MLE t-statistic, the probability to reject H_0 for $c < K_c$ quickly becomes the probability of the event $\xi(0) > K_{\alpha,s}$, where $\xi(0)$ is a standard normal variable.

https://doi.org/10.1017/S0266466620000079 Published online by Cambridge University Press

in performance for $c = 0$ and $c = -10$. For the nearly optimal test with fewer frequencies $q < 10$, we see the gradual reduction in test powers as q decreases. This power loss becomes more apparent as ρ decreases. The steepest loss in accuracy occurs between $q = 4$ and $q = 2$. Tables 2 and 3 repeat the same analysis for small samples with $T = 100$ and $T = 50$. The long-run variance ψ is estimated from the fitted vector autoregression (VAR) on the OLS estimated $\hat{e}_t$. We obtain similar results as in the case with $T = 1,000$.

In the second case, the expectation $E(\varepsilon_t|u_{t-1}, u_{t-2}, \ldots)$ is not zero due to the omission of a relevant predictor. The omitted predictor $x_{2,t}$ is correlated with the included x_t, but it is substantially less persistent. This simulation design mimics the situation in which the relationship between y_t and x_t changes with the time horizon. Simulations come from the following model:

$$\begin{aligned} y_t &= \beta_x x_{t-1} + \gamma x_{2,t-1} + \varepsilon_t^*, \\ x_t &= \rho x_{t-1} + u_t, \\ x_{2,t} &= \rho_2 x_{2,t-1} + u_{2,t}, \end{aligned} \tag{6}$$

where (ε_t^*, u_t) are i.i.d. with the cross-correlation r. Series $x_{2,t}$ is unobserved. The shocks u_t and $u_{2,t}$ are perfectly correlated. Table 4 considers the case with $\text{corr}(u_t, u_{2,t}) = -1$ and Table 5 considers the case with $\text{corr}(u_t, u_{2,t}) = 1$. The persistence $\rho_2 = 0.2$ is chosen to be sufficiently small such that $x_{2,t}$ is close to an i.i.d. process at lower frequencies. The severity of the bias $\gamma = 0.1$ is chosen sufficiently large to affect the performance of the tests.

As follows from Tables 4 and 5, some size correction is already achieved at $q = 10$ frequencies. The smaller the q, the closer the rejection rates to the nominal significance level. The most accurate results are achieved with the larger sample size $T = 100$.

We complete our main Monte Carlo results by considering adjustments to the EMW test that correct the statistics in the basis of this method for short-run endogeneity. Because such adjustments have not been previously considered, we suggest corrections that are similar to those appearing in related papers by Campbell and Yogo (2006) and Magdalinos and Phillips (2009). In particular, these corrections are based on consistent estimates of the one-sided long-run variance $\sum_{h=1}^{\infty} Ee_t e_{t-h}'$. This matrix is estimated using the VAR for estimated errors $\hat{e}_t$ with the number of lags chosen by the Bayesian information (BIC) criterion. We find that the methods used to calculate $\hat{e}_t$ affect the performance of the test and, thus, report results from several such methods. The first method, as is usually suggested, estimates $\hat{e}_t$ from the OLS regression of $(x_t, y_t)'$ on x_{t-1} and a constant. This estimation method yields consistent estimates of $\hat{e}_t$ for local-to-unity sequences of ρ_T. In Tables 4 and 5, the results corresponding to this case are reported in column EMW^A. As it follows, although this method yields rejection rates of approximately 5% for the model with a negative short-run bias, it remains oversized by a large margin in the positive bias case. The second method estimates $\hat{\varepsilon}_t$ by regressing y_t on x_{t-1} and lags of $\hat{u}_t$. The results are reported in column EMW^B of Tables 4 and 5. This method does not solve the problem of high rejection rates under H_0

https://doi.org/10.1017/S0266466620000079 Published online by Cambridge University Press

TABLE 1. Large-sample simulations. Exogenous x_{t-1}

c	r	b	$q=2$	$q=4$	$q=6$	$q=8$	$q=10$	EMW
0	−0.5	50	99.58%	99.97%	99.99%	99.99%	100.00%	100.00%
0	−0.5	10	84.00%	89.83%	91.72%	92.59%	92.99%	93.97%
0	−0.5	0	4.89%	5.43%	5.33%	5.42%	5.43%	5.10%
0	0	50	99.43%	99.97%	100.00%	100.00%	100.00%	100.00%
0	0	10	77.85%	84.19%	85.81%	86.82%	87.17%	89.05%
0	0	0	4.95%	5.17%	5.18%	5.06%	4.91%	4.81%
0	0.5	50	99.58%	99.97%	99.99%	99.99%	100.00%	100.00%
0	0.5	10	84.00%	83.73%	85.21%	85.92%	92.99%	87.57%
0	0.5	0	4.89%	5.18%	5.15%	5.02%	5.43%	4.59%
−10	−0.5	50	97.79%	99.88%	99.99%	100.0%	100.00%	100.00%
−10	−0.5	10	41.19%	52.29%	55.82%	57.71%	59.23%	62.45%
−10	−0.5	0	4.77%	4.83%	4.63%	4.81%	4.78%	4.88%
−10	0	50	96.01%	99.73%	99.97%	100.0%	100.00%	100.00%
−10	0	10	37.72%	47.08%	51.49%	53.65%	54.57%	58.05%
−10	0	0	4.87%	5.07%	4.95%	5.05%	4.89%	4.59%
−10	0.5	50	97.19%	99.74%	99.91%	99.97%	99.97%	100.00%
−10	0.5	10	41.04%	49.29%	52.89%	54.54%	55.50%	59.33%
−10	0.5	0	5.10%	5.35%	5.18%	5.23%	5.07%	5.30%
−30	−0.5	50	81.83%	97.17%	99.47%	99.88%	99.95%	100.00%
−30	−0.5	10	15.14%	19.67%	22.40%	25.37%	26.38%	33.45%
−30	−0.5	0	4.70%	4.62%	4.72%	4.77%	4.65%	4.90%
−30	0	50	75.20%	93.79%	97.85%	99.08%	99.44%	100.00%
−30	0	10	14.96%	19.98%	22.61%	25.44%	26.60%	33.84%
−30	0	0	4.76%	4.73%	4.45%	5.00%	4.57%	4.62%
−30	0.5	50	79.57%	94.45%	97.83%	98.85%	99.22%	99.93%
−30	0.5	10	17.57%	23.17%	26.23%	27.74%	29.42%	35.31%
−30	0.5	0	5.07%	5.33%	4.94%	4.97%	4.95%	5.07%

Note: The table reports rejection rates for the hypothesis $H_0 : \beta_x = 0$ for model (1) if innovations (ε_t, u_t) are i.i.d. and the number of observations is $T = 1,000$. We compare the long-run nearly optimal test based on the lowest q frequencies of Fourier transformations of y_t to the nearly optimal test that is based on the overall Gaussian likelihood (EMW). The persistence parameter of x_t is $\rho_T = 1 + c/T$. The predictability parameter is $\beta_x = b/T$. The cross-correlation between ε_t and u_t is r.

given the positive omitted variable bias, and results in an oversized test for the negative omitted variable bias case. The third method estimates $\hat{e}_t$ on the basis of the interpretation of ρ as the highest root of the characteristic polynomial of the vector-process (y_t, x_t). Residuals $\hat{u}_t$ and $\hat{\varepsilon}$ are obtained from estimates of ρ and β_x that minimize the highest root of the characteristic polynomials for $x_t - \rho x_{t-1}$ and $y_t - \beta_x x_{t-1}$, respectively. These results are reported in column EMW^C. It appears that among the considered methods, this correction is the only one that produces the test with a correct level (below 5%). However, this test appears undersized for

https://doi.org/10.1017/S0266466620000079 Published online by Cambridge University Press

TABLE 2. Small-sample simulations with $T = 100$. Exogenous x_{t-1}

ρ	r	β_x	$q=2$	$q=4$	$q=6$	$q=8$	$q=10$	*EMW*
1.0	− 0.5	0.5	96.08%	99.11%	99.66%	99.79%	99.87%	99.98%
1.0	− 0.5	0.1	79.71%	86.65%	88.49%	89.67%	90.33%	90.94%
1.0	− 0.5	0	4.76%	4.81%	4.93%	4.99%	5.06%	4.47%
1.0	0	0.5	96.28%	99.43%	99.73%	99.87%	99.92%	100.00%
1.0	0	0.1	74.63%	81.30%	83.47%	84.39%	84.88%	86.24%
1.0	0	0	4.60%	4.85%	5.00%	4.80%	4.79%	4.81%
1.0	0.5	0.5	96.65%	99.54%	99.83%	99.91%	99.93%	99.98%
1.0	0.5	0.1	76.73%	82.05%	83.76%	84.54%	84.82%	86.55%
1.0	0.5	0	4.57%	4.56%	4.81%	4.86%	5.11%	4.77%
0.9	− 0.5	0.5	95.60%	99.71%	99.99%	99.96%	99.97%	99.98%
0.9	− 0.5	0.1	41.79%	53.00%	56.32%	58.91%	60.28%	62.83%
0.9	− 0.5	0	4.57%	4.68%	4.66%	4.75%	4.55%	4.44%
0.9	0	0.5	93.90%	99.52%	99.89%	99.96%	99.98%	100.00%
0.9	0	0.1	38.43%	47.98%	52.25%	54.37%	55.62%	59.22%
0.9	0	0	4.48%	4.32%	4.64%	4.52%	4.46%	4.32%
0.9	0.5	0.5	95.60%	99.61%	99.92%	99.93%	99.96%	99.99%
0.9	0.5	0.1	41.79%	49.63%	53.39%	54.91%	55.54%	60.53%
0.9	0.5	0	4.57%	4.18%	4.53%	4.55%	4.45%	5.03%
0.7	− 0.5	0.5	76.33%	95.86%	99.19%	99.66%	99.76%	99.94%
0.7	− 0.5	0.1	15.10%	19.73%	23.35%	26.41%	28.13%	34.14%
0.7	− 0.5	0	4.33%	4.13%	4.50%	4.38%	4.26%	4.29%
0.7	0	0.5	70.81%	92.56%	97.99%	99.03%	99.17%	99.56%
0.7	0	0.1	14.49%	19.66%	22.59%	25.18%	26.84%	35.39%
0.7	0	0	4.42%	4.40%	4.21%	4.09%	4.14%	4.52%
0.7	0.5	0.5	74.29%	93.88%	97.98%	99.06%	99.45%	99.97%
0.7	0.5	0.1	17.14%	22.35%	26.09%	27.15%	28.50%	36.98%
0.7	0.5	0	4.28%	4.45%	4.53%	4.43%	4.34%	4.91%

Note: The table reports rejection rates for the hypothesis $H_0 : \beta_x = 0$ for model (1) if innovations (ε_t, u_t) are i.i.d. and the number of observations is $T = 100$. We compare the long-run nearly optimal test based on the lowest q frequencies of Fourier transformations of y_t to the nearly optimal test that is based on the overall Gaussian likelihood (EMW). The persistence parameter of x_t is ρ. The predictability parameter is β_x. The cross-correlation between ε_t and u_t is r.

low values of ρ. We refer to $\hat{e}_t$ constructed by the latter method as $\hat{e}_t^C$ in the next subsection.

6.1. Additional Monte-Carlo Experiments

This section provides additional Monte Carlo evidence that includes methods suggested by Magdalinos and Phillips (2009), Campbell and Yogo (2006), Jansson

https://doi.org/10.1017/S0266466620000079 Published online by Cambridge University Press

TABLE 3. Small-sample simulations with $T = 50$. Exogenous x_{t-1}

ρ	r	β_x	$q = 2$	$q = 4$	$q = 6$	$q = 8$	$q = 10$	*EMW*
1.0	− 0.5	0.5	89.12%	94.71%	96.19%	96.75%	97.15%	97.38%
1.0	− 0.5	0.1	42.68%	48.51%	50.11%	51.33%	51.98%	50.58%
1.0	− 0.5	0	4.84%	4.96%	4.96%	5.08%	5.04%	4.25%
1.0	0	0.5	88.86%	94.45%	95.76%	96.40%	96.61%	97.23%
1.0	0	0.1	39.31%	44.86%	46.19%	46.92%	47.32%	47.42%
1.0	0	0	4.45%	4.83%	4.86%	4.91%	4.80%	4.86%
1.0	0.5	0.5	89.12%	95.48%	96.67%	97.18%	97.33%	97.95%
1.0	0.5	0.1	42.68%	46.02%	47.84%	48.44%	49.17%	49.65%
1.0	0.5	0	4.84%	4.55%	4.70%	4.55%	4.83%	4.70%
0.9	− 0.5	0.5	89.61%	96.23%	97.38%	97.78%	97.91%	98.36%
0.9	− 0.5	0.1	21.92%	26.06%	28.21%	29.01%	29.48%	29.74%
0.9	− 0.5	0	4.11%	4.16%	4.28%	4.23%	4.41%	4.17%
0.9	0	0.5	86.76%	94.51%	96.47%	97.19%	97.41%	98.01%
0.9	0	0.1	23.76%	27.77%	29.08%	29.48%	30.06%	30.50%
0.9	0	0	4.18%	4.27%	4.39%	4.46%	4.55%	4.42%
0.9	0.5	0.5	87.85%	94.94%	96.65%	97.16%	97.38%	98.14%
0.9	0.5	0.1	25.35%	28.59%	30.37%	31.32%	31.27%	33.91%
0.9	0.5	0	3.95%	4.08%	4.19%	4.25%	4.39%	4.69%
0.7	− 0.5	0.5	74.54%	91.74%	95.63%	96.74%	97.24%	97.97%
0.7	− 0.5	0.1	13.39%	16.08%	17.65%	19.03%	19.81%	19.31%
0.7	− 0.5	0	4.61%	4.56%	4.82%	5.04%	5.40%	3.92%
0.7	0	0.5	68.53%	86.73%	92.08%	94.22%	95.15%	97.20%
0.7	0	0.1	12.55%	15.26%	17.07%	18.02%	18.75%	20.64%
0.7	0	0	4.32%	4.19%	3.98%	4.31%	4.26%	4.28%
0.7	0.5	0.5	70.17%	86.37%	91.19%	92.92%	93.82%	96.80%
0.7	0.5	0.1	13.65%	16.47%	18.04%	18.51%	18.76%	22.32%
0.7	0.5	0	4.01%	3.91%	3.84%	4.05%	4.28%	4.65%

Note: The table reports rejection rates for the hypothesis $H_0 : \beta_x = 0$ for model (1) if innovations (ε_t, u_t) are i.i.d. and the number of observations is $T = 50$. We compare the long-run nearly optimal test based on the lowest q frequencies of Fourier transformations of y_t to the nearly optimal test that is based on the overall Gaussian likelihood (EMW). The persistence parameter of x_t is ρ. The predictability parameter is β_x. The cross-correlation between ε_t and u_t is r.

and Moreira (2006), and Müller and Watson (2013). We consider the case with $T = 50$ annual observations and predictable ε_t, as in the endogenous case in the previous subsection. In the next few paragraphs, we provide a short description for the considered methods as applied to the problem in this paper.

The *IVX approach by Magdalinos and Phillips (2009)* is an instrumental variables method in which auxiliary instruments are created from existing regressors. Because of the manner in which the instrumental variables are constructed, their

https://doi.org/10.1017/S0266466620000079 Published online by Cambridge University Press

TABLE 4. Small-sample simulations. Negative omitted variable bias Actual significance level.

T	ρ	r	$q=2$	$q=4$	$q=6$	$q=8$	$q=10$	EMW	EMW^A	EMW^B	EMW^C
100	1.0	−0.5	4.58%	4.73%	4.78%	4.63%	4.28%	2.93%	5.34%	5.75%	2.75%
100	1.0	0	4.95%	5.23%	4.97%	4.90%	4.71%	3.54%	5.80%	6.17%	4.15%
100	1.0	0.5	5.04%	5.06%	4.76%	4.80%	4.73%	3.15%	5.35%	5.67%	5.02%
50	1.0	−0.5	4.67%	4.90%	4.78%	4.82%	4.74%	5.02%	6.23%	7.49%	1.96%
50	1.0	0	5.44%	5.71%	5.56%	5.53%	5.40%	5.06%	7.47%	8.70%	4.66%
50	1.0	0.5	5.55%	5.60%	5.37%	5.16%	5.06%	4.21%	6.74%	7.78%	6.13%
100	0.9	−0.5	4.18%	3.75%	3.43%	3.12%	2.67%	1.56%	3.64%	5.59%	0.84%
100	0.9	0	4.08%	3.76%	3.64%	3.26%	2.94%	1.46%	4.30%	6.55%	1.89%
100	0.9	0.5	4.12%	3.49%	3.30%	3.09%	2.94%	1.47%	4.41%	6.51%	2.97%
50	0.9	−0.5	3.86%	3.88%	3.66%	3.43%	3.35%	3.22%	4.32%	7.82%	0.54%
50	0.9	0	4.53%	4.31%	3.75%	3.57%	3.45%	2.96%	5.69%	9.11%	2.13%
50	0.9	0.5	4.69%	4.16%	3.60%	3.63%	3.23%	2.52%	5.85%	8.67%	4.10%
100	0.7	−0.5	3.64%	3.06%	2.44%	2.09%	1.76%	0.88%	1.74%	9.28%	0.07%
100	0.7	0	3.76%	3.54%	2.63%	2.08%	1.71%	0.78%	2.29%	11.00%	0.40%
100	0.7	0.5	4.05%	3.28%	2.33%	1.88%	1.55%	0.68%	2.51%	10.06%	0.75%
50	0.7	−0.5	3.30%	2.73%	2.08%	2.22%	2.09%	2.63%	2.73%	11.47%	0.13%
50	0.7	0	3.73%	3.11%	2.60%	2.51%	2.28%	2.05%	3.90%	12.87%	0.58%
50	0.7	0.5	3.76%	3.14%	2.76%	2.57%	2.38%	1.45%	3.84%	12.59%	1.70%

Note: The table reports actual significance levels of the tests for the hypothesis $H_0 : \beta_x = 0$ in model (1), where $x_{2,t}$ is the omitted factor. We compare the long-run nearly optimal test based on the lowest q frequencies of Fourier transformations of y_t to the nearly optimal test that is based on the overall Gaussian likelihood (EMW). The methods EMW^A, EMW^B, and EMW^C are designed to correct the EMW method for the omitted variable bias based on the consistent estimates of the one-sided long-run variance $\sum_{h=1}^{\infty} E_t e_t e_{t-h}'$. In all cases, this sum is calculated based on the VAR estimates for $\hat{e}_t$. For EMW^A, $\hat{e}_t$ is estimated from ordinary least squares (OLS) regression of (x_t, y_t) on x_{t-1} and constant. For EMW^B, $\hat{\varepsilon}_t$ is estimated from the regression of y_t on x_{t-1}, lags of $\hat{u}_t$ and constant. For EMW^C, $\hat{\varepsilon}_t$ is estimated as demeaned series of $y_t - \hat{\beta}_x x_{t-1}$, where $\hat{\beta}_x$ minimizes the persistence of $\hat{\varepsilon}_t$. The persistence parameter of x_t is ρ. The cross-correlation between innovations ε_t^* and u_t is r, and between u_t and $u_{2,t}$ is negative one. The persistence of the omitted factor $x_{2,t}$ is $\rho_2 = 0.2$ and the corresponding slope is $\gamma = 0.1$.

degree of persistence is explicitly controlled. On one hand, this method is robust to the exact rate of convergence of ρ_T to 1 and yields mixed normal distributions for t-statistics for stationary, unit root, nearly integrated and mildly integrated regressors. On the other hand, this method results in a slower convergence rate. The method of Magdalinos and Phillips (2009) has a built-in procedure to correct for short-run endogeneity in ε_t, which is based on the kernel estimate of the one-sided long-run variance matrix $\sum_{h=1}^{\infty} E e_t e_{t-h}'$. Further, Kostakis, Magdalinos and Stamatogiannis (2015) present an improved version of the IVX method, but consider only the case with unpredictable ε_t. In our Monte Carlo simulations, we use the code provided by Michalis P. Stamatogiannis, which we modify to

https://doi.org/10.1017/S0266466620000079 Published online by Cambridge University Press

TABLE 5. Small-sample simulations, positive omitted variable bias, and actual significance level

T	ρ	r	$q=2$	$q=4$	$q=6$	$q=8$	$q=10$	EMW	EMW^A	EMW^B	EMW^C
100	1.0	−0.5	4.99%	5.65%	6.20%	6.51%	6.73%	13.55%	7.36%	7.10%	3.61%
100	1.0	0	5.22%	5.81%	5.93%	6.37%	6.54%	11.34%	7.45%	7.12%	5.30%
100	1.0	0.5	5.35%	5.54%	5.72%	5.83%	6.07%	9.48%	6.35%	6.17%	5.72%
50	1.0	−0.5	5.70%	6.98%	8.19%	9.17%	10.24%	17.49%	9.87%	9.52%	2.97%
50	1.0	0	6.11%	6.90%	7.53%	8.34%	9.04%	13.64%	9.89%	9.71%	5.86%
50	1.0	0.5	6.16%	6.82%	7.13%	7.62%	8.03%	10.33%	8.26%	7.93%	6.94%
100	0.9	−0.5	4.69%	4.90%	5.02%	5.46%	5.73%	18.72%	9.98%	9.19%	1.84%
100	0.9	0	4.76%	5.06%	5.12%	5.46%	5.78%	15.33%	10.50%	9.75%	3.49%
100	0.9	0.5	5.04%	5.01%	5.31%	5.49%	5.69%	13.30%	10.04%	9.29%	4.55%
50	0.9	−0.5	4.76%	5.86%	6.92%	8.32%	9.74%	19.10%	11.41%	10.87%	1.19%
50	0.9	0	5.53%	6.03%	6.81%	7.50%	8.42%	15.49%	12.54%	12.12%	3.51%
50	0.9	0.5	5.93%	6.32%	6.80%	7.44%	8.01%	12.08%	11.36%	10.51%	5.63%
100	0.7	−0.5	4.35%	4.24%	3.77%	3.91%	4.00%	25.60%	16.70%	16.40%	0.24%
100	0.7	0	4.66%	4.87%	4.33%	4.34%	4.56%	21.70%	17.91%	17.74%	1.29%
100	0.7	0.5	4.79%	4.80%	4.18%	4.43%	4.74%	18.36%	17.41%	16.46%	1.60%
50	0.7	−0.5	3.88%	4.11%	5.04%	6.06%	7.96%	21.70%	14.10%	16.73$	0.32%
50	0.7	0	4.78%	4.96%	5.77%	6.53%	7.66%	18.19%	16.43%	18.68%	1.39%
50	0.7	0.5	5.43%	5.71%	6.64%	7.55%	8.27%	14.42%	16.47%	17.38%	2.81%

Note: The table reports actual significance levels of the tests for the hypothesis $H_0 : \beta_x = 0$ in model (1), where $x_{2,t}$ is the omitted factor. We compare the long-run nearly optimal test based on the lowest q frequencies of Fourier transformations of y_t to the nearly optimal test that is based on the overall Gaussian likelihood (EMW). The methods EMW^A, EMW^B, and EMW^C are designed to correct the EMW method for the omitted variable bias based on the consistent estimates of the one-sided long-run variance $\sum_{h=1}^{\infty} E_t e_t e_{t-h}'$. In all cases, this sum is calculated based on the VAR estimates for $\hat{e}_t$. For EMW^A, $\hat{e}_t$ is estimated from OLS regression of (x_t, y_t) on x_{t-1} and constant. For EMW^B, $\hat{\varepsilon}_t$ is estimated from the regression of y_t on x_{t-1}, lags of $\hat{u}_t$ and constant. For EMW^C, $\hat{\varepsilon}_t$ is estimated as demeaned series of $y_t - \hat{\beta}_x x_{t-1}$, where $\hat{\beta}_x$ minimizes the persistence of $\hat{\varepsilon}_t$. The persistence parameter of x_t is ρ. The cross-correlation between innovations ε_t^* and u_t is r, and between u_t and $u_{2,t}$ is one. The persistence of the omitted factor $x_{2,t}$ is $\rho_2 = 0.2$ and the corresponding slope is $\gamma = 0.1$.

perform the test of Magdalinos and Phillips (2009). Note that condition $\mu_y = 0$, which is required by Magdalinos and Phillips (2009), holds in our Monte Carlo design. The important tuning parameters of the IVX method are (a) the preselected persistence of the auxiliary instruments and (b) the estimation method for the long-run variance matrices. For (a), we chose the instrument persistence parameter as in Kostakis et al. (2015). We find that other reasonable values of this parameter do not improve the results. For (ii), we considered two choices: one with the long-run one-sided variances estimated by kernel methods from OLS-estimated $\hat{e}_t$, as in Kostakis et al. (2015), and the other one with long-run variances calculated from the VAR model estimated on series $\hat{e}_t^C$. The actual rejection rates under H_0 for both of these methods are reported in Table 6 in columns $IVX(orig)$ and

https://doi.org/10.1017/S0266466620000079 Published online by Cambridge University Press

TABLE 6. Additional small-sample simulations. Actual significance level under omitted variable bias.

ρ	r	Q-test	Q-test (adj)	IVX (orig)	IVX (adj)	Opt. Inf. (orig)	Y-only (orig)	OLS
Negative bias $r_b = -1$								
1.0	− 0.5	4.01%	3.29%	39.50%	3.59%	7.03%	5.44%	4.16%
1.0	0	4.18%	4.71%	6.73%	4.56%	18.24%	5.44%	3.78%
1.0	0.5	3.51%	4.71%	0.83%	5.06%	6.03%	5.44%	1.71%
0.9	− 0.5	2.23%	0.86%	68.87%	1.48%	4.50%	5.44%	1.24%
0.9	0	2.33%	2.55%	13.86%	2.63%	8.07%	5.44%	2.18%
0.9	0.5	2.55%	2.99%	0.74%	3.41%	3.60%	5.44%	2.00%
0.7	− 0.5	1.37%	0.26%	61.71%	0.91%	3.80%	5.44%	0.87%
0.7	0	1.51%	0.90%	4.98%	1.49%	5.80%	5.44%	1.37%
0.7	0.5	2.23%	1.63%	0.06%	2.88%	2.80%	5.44%	1.53%
Positive bias $r_b = 1$								
1.0	− 0.5	9.09%	3.96%	47.37%	3.85%	9.24%	5.44%	8.04%
1.0	0	8.08%	5.83%	9.47%	5.10%	24.91%	5.44%	7.34%
1.0	0.5	6.67%	5.78%	1.53%	5.96%	10.33%	5.44%	3.87%
0.9	− 0.5	11.39%	1.51%	51.76%	2.07%	6.70%	5.44%	7.28%
0.9	0	11.42%	3.90%	6.06%	3.59%	15.30%	5.44%	10.73%
0.9	0.5	10.99%	4.95%	0.19%	4.89%	10.40%	5.44%	9.44%
0.7	− 0.5	13.43%	0.69%	85.26%	1.84%	6.10%	5.44%	9.77%
0.7	0	14.71%	1.73%	19.72%	2.20%	15.63%	5.44%	13.94%
0.7	0.5	15.98%	3.34%	0.01%	3.98%	10.01%	5.44%	14.63%

Note: The table reports actual significance levels of the tests for the hypothesis $H_0 : \beta_x = 0$ in model (1) with $T = 50$, where $x_{2,t}$ is the omitted factor. We compare the Q-test of Campbell and Yogo (2006), the IVX method of Magdalinos and Phillips (2009), Optimal Inference test by Jansson and Moreira (2006), Y-only test by Müller and Watson (2013), and the test based on the OLS in which y_t is regressed on x_{t-1} and the lags of $\Delta x_t = x_t - x_{t-1}$. The Q-test uses adjusted Bonferroni bounds based on the inversion of Mikusheva (2009) test. The adjusted version of the Q-test corrects for the short-run endogeneity as suggested in this paper. The original IVX test corrects for the short-run endogeneity using kernel estimator of the one-sided long-run variance matrices as in Magdalinos and Phillips (2009). The adjusted IVX test corrects for the short-run endogeneity as suggested in this paper. The OLS method uses adjusted Bonferroni bounds based on the inversion of Mikusheva (2009) test. The persistence parameter of x_t is ρ. The cross-correlation between innovations ε_t^* and u_t is r, and between u_t and $u_{2,t}$ is r_b. The persistence of the omitted factor $x_{2,t}$ is $\rho_2 = 0.2$ and the corresponding slope is $\gamma = 0.1$.

IVX(*adj*), respectively. As it follows from the table, although both methods are designed to adjust for short-run endogeneity, only the second method yields a 5% level test. Similar to the EMW^C, the latter method produces a somewhat undersized test.

The *optimal inference method by Jansson and Moreira (2006)* is optimal within a class of conditionally unbiased tests that are also invariant to parallel shifts in Y, assuming that e_t is i.i.d. Gaussian. This optimality preserves asymptotically in the non-Gaussian case and in the case of serially correlated u_t after proper

https://doi.org/10.1017/S0266466620000079 Published online by Cambridge University Press

adjustments. Because this method is optimal only in the class of conditionally unbiased tests, it can be outperformed by conditionally biased tests derived under the same assumptions, for instance, by tests of Campbell and Yogo (2006) and Elliott et al. (2015). For our Monte Carlo simulations, we use the code provided by Michael Jansson. The results are shown in Table 6. As with the unadjusted EMW method, the optimal inference produces a test with the wrong actual size in the presence of short-run endogeneity. However, note that this method is based on the same asymptotic statistics as the test by Elliott et al. (2015). Therefore, to correct for short-run endogeneity, we can use the same corrected finite-sample statistics as for the EMW method. Efforts were made to adapt the optimal inference method to the assumptions of this article. Although such adaptation is possible, given the limitations of the computational power and the increased instability of the code when used in conjunction with corrected statistics, we decided not to report these simulation results and recommend such adaptation for further research.

Q-test of Campbell and Yogo (2006). Campbell and Yogo (2006) develop an optimal test when ρ is known and e_t is Gaussian i.i.d. within a class of conditional tests that condition on the ancillary statistic $\sum_{t=1}^{T} x_{t-1}^2$. The dependence on ρ is then removed using adjusted Bonferroni bounds. Following the suggestions of Phillips (2014), we construct Bonferroni bounds on ρ by inverting the centered OLS t-statistic for ρ, as in Mikusheva (2007), assuming i.i.d. u_t, which is true in our simulations. This approach to constructing confidence intervals for ρ works uniformly well across $\rho \in \mathbb{R}$. Yet, the resulting test may still remain oversized for very low values of c given the use of adjustments in Bonferroni bounds that depend on the local-to-unity asymptotics. The Campbell and Yogo Q-test (2006) involves the same asymptotic statistics as the method of Elliott et al. (2015). Therefore, to correct for short-run endogeneity, we use the same corrected finite-sample statistics as for the EMW test. We update the code by Motohiro Yogo, provided on this web site, to use different confidence intervals for ρ. The results for the Campbell and Yogo (2006) Q-test are reported in Table 6 with and without corrections. As in the case with the EMW test, we achieve a correct test level after adjustments. Similar to the EMW and IVX approaches, the test seems undersized after corrections.

The *Y-only optimal test of Müller and Watson (2013)* nearly achieves a power bound within a class of tests based on q cosine transformations of Y and X series, robust to the trend model of X, and that maximize the test power against the unit-root model with cointegration. This class of tests is immune to short-run endogeneity and treats all of the asymptotic distribution parameters of cosine transformations as nuisance parameters. Conveniently, the Y-only test does not depend on the estimate of the long-run correlation coefficient r. We use the code provided by Mark Watson on his web site, with $q = 14$ as in Müller and Watson (2013). Other choices of q do not improve the performance of the test in our simulations, with higher q delivering oversized tests. As follows, from the Monte Carlo results in Table 6, this method yields the test with approximately correct size.

OLS regression with additional regressors. Finally, we consider a simple method that is a natural solution to the short-run endogeneity problem. For this method,

https://doi.org/10.1017/S0266466620000079 Published online by Cambridge University Press

we base our test on the t-statistics in the regression of y_t on x_{t-1} and lags of $\Delta x_t = x_t - x_{t-1}$, where the number of lags is chosen by the BIC. The resulting t-statistics have nonstandard distributions for persistent x_{t-1}, which depend on a nuisance parameter c; see Cavanagh et al. (1995). To remove this nuisance parameter from confidence intervals, we use the adjusted Bonferroni method based on the inversion of the centered t-statistic by Mikusheva (2007) that takes into account that u_t is i.i.d. The resulting actual rejection rates under the H_0 are reported in Table 6. It follows that this test is undersized in the negative bias case and oversized in the positive bias case. We conclude that this method does not provide a satisfactory solution to the short-run endogeneity problem. See also Ing (2003) for a related result in the univariate case.

We complete our Monte Carlo exercise by comparing powers of the tests with approximately correct levels for $T = 50$. In particular, we consider the long-run nearly optimal test with $q = 2$, the adjusted Q-test, the adjusted EMW test (EMW^C), and the Y-only test. The results are given in Table 7.

It is important to note that only the long-run nearly optimal test and the EMW test have the same optimality criterion, which is the WAP with the same weights. We start by comparing these two tests and find that their performance is approximately the same for $\rho = 1$. However, at $\rho = 0.9$, the long-run nearly optimal test outperforms the adjusted EMW test. The difference in performance increases further for $\rho = 0.7$. We conclude that corrections for short-run endogeneity result in nontrivial errors in small samples, even when the predictor variable is persistent.

Note that the Y-only test is the most robust among the considered set. Correspondingly, it yields the lowest rejection rates under alternatives across most combinations of ρ and r. The IVX method does not optimize any WAP a priori. However, it yields the highest rejection rates under alternative hypotheses for $\rho = 1$. At the same time, this test underperforms the long-run nearly optimal test for $\rho = 0.9$ and $\rho = 0.7$. Surprisingly, the adjusted Q-test often outperforms the adjusted EMW test, although its unadjusted version is asymptotically less efficient. The adjusted Q-test also outperforms the long-run nearly optimal tests for $\rho = 1.0$, performs on par for $\rho = 0.9$, and somewhat underperforms for $\rho = 0.7$ for most parameter combinations.

To summarize, the tests developed for the i.i.d. case can be potentially corrected to account for short-run endogeneity. Such corrections can be made based on estimates of one-sided long-run variances of $\hat{e}_t$. However, these corrections result in nontrivial errors in small samples. Depending on the estimation method for $\hat{e}_t$, the corrected tests can be oversized or undersized. For example, when estimates $\hat{e}_t^c$ are used, the resulting tests are undersized. As a result, they are often outperformed by the long-run test based on a small number of frequencies.

7. REMARK ON CASE WITH $Q \to \infty$

Suppose the number of frequencies q is not fixed but is allowed to increase at a slow rate so that $1/q + q/T \to 0$. Note that the limiting behavior of the

https://doi.org/10.1017/S0266466620000079 Published online by Cambridge University Press

TABLE 7. Additional small-sample simulations. Actual test powers under omitted variable bias.

ρ	r	β_x	$q = 2$	EMW^C	Q-test (adj)	IVX (adj)	Y-only (orig)
Negative bias $r_b = -1$							
1.0	− 0.5	0.5	92.52%	90.12%	95.91%	91.83%	75.37%
1.0	− 0.5	0.1	41.09%	34.49%	46.29%	67.78%	23.21%
1.0	0	0.5	93.59%	94.45%	97.25%	93.93%	78.18%
1.0	0	0.1	42.38%	42.16%	47.04%	65.10%	24.80%
1.0	0.5	0.5	94.94%	96.27%	97.49%	91.83%	80.93%
1.0	0.5	0.1	48.10%	51.67%	50.25%	62.29%	26.86%
0.9	− 0.5	0.5	78.08%	62.86%	81.95%	64.23%	35.73%
0.9	− 0.5	0.1	17.98%	6.63%	10.38%	12.13%	3.92%
0.9	0	0.5	80.55%	78.99%	87.84%	74.61%	45.26%
0.9	0	0.1	20.10%	14.38%	17.29%	18.97%	8.28%
0.9	0.5	0.5	84.58%	85.36%	89.43%	76.49%	53.47%
0.9	0.5	0.1	22.96%	22.41%	21.79%	21.89%	14.06%
0.7	− 0.5	0.5	37.62%	13.78%	30.13%	14.18%	6.56%
0.7	− 0.5	0.1	7.09%	0.93%	0.78%	2.07%	3.82%
0.7	0	0.5	40.76%	30.03%	44.64%	24.99%	14.22%
0.7	0	0.1	8.29%	2.45%	3.04%	3.69%	5.69%
0.7	0.5	0.5	58.34%	48.25%	56.94%	37.02%	20.02%
0.7	0.5	0.1	9.26%	5.00%	5.08%	5.52%	8.21%
Positive bias $r_b = 1$							
1.0	− 0.5	0.5	92.01%	87.15%	94.23%	87.81%	75.45%
1.0	− 0.5	0.1	46.09%	39.43%	49.63%	67.77%	24.47%
1.0	0	0.5	93.45%	94.09%	96.96%	92.66%	76.56%
1.0	0	0.1	44.55%	45.39%	48.95%	62.71%	24.19%
1.0	0.5	0.5	95.31%	96.91%	97.77%	94.47%	77.91%
1.0	0.5	0.1	47.80%	51.96%	59.96%	58.14%	23.83%
0.9	− 0.5	0.5	75.17%	54.02%	74.38%	52.71%	38.19%
0.9	− 0.5	0.1	21.83%	10.04%	14.98%	16.12%	5.14%
0.9	0	0.5	79.63%	76.88%	86.60%	70.26%	44.86%
0.9	0	0.1	22.46%	17.97%	22.31%	20.42%	10.30%
0.9	0.5	0.5	86.42%	88.64%	91.66%	80.20%	50.33%
0.9	0.5	0.1	24.28%	25.56%	26.53%	22.07%	15.20%
0.7	− 0.5	0.5	33.74%	9.19%	22.96%	9.46%	8.59%
0.7	− 0.5	0.1	8.76%	2.64%	2.33%	3.62%	3.75%

https://doi.org/10.1017/S0266466620000079 Published online by Cambridge University Press

TABLE 7. (Continued).

ρ	r	β_x	$q=2$	EMW^C	Q-test (adj)	IVX (adj)	Y-only (orig)
0.7	0	0.5	40.62%	29.48%	44.77%	23.22%	14.68%
0.7	0	0.1	10.02%	4.47%	6.05%	5.35%	6.37%
0.7	0.5	0.5	58.57%	61.91%	58.08%	48.23%	19.06%
0.7	0.5	0.1	11.92%	8.54%	9.94%	7.59%	9.03%

Note: The table reports rejection rates of the tests for the hypothesis $H_0 : \beta_x = 0$ in model (1) with $T = 50$, where $x_{2,t}$ is the omitted factor. We compare the long-run nearly optimal test based on the lowest q frequencies of Fourier transformation of y_t to the nearly optimal test based on the overall Gaussian likelihood (EMW), the Q-test of Campbell and Yogo (2006), the IVX method of Magdalinos and Phillips (2009) and Y-only test by Müller and Watson (2013). The EMW method, Q-test, and the IVX method are adjusted for the omitted variable bias as suggested in this article. The Q-test uses adjusted Bonferroni bounds based on the inversion of Mikusheva (2009) test. The persistence parameter of x_t is ρ. The predictability parameter is β_x. The cross-correlation between innovations ε_t^* and u_t is r, and between u_t and $u_{2,t}$ is r_b. The persistence of the omitted factor $x_{2,t}$ is $\rho_2 = 0.2$ and the corresponding slope is $\gamma = 0.1$.

finite-sample statistics R'_ρ and $\mathrm{R}'_{\rho\rho}$ is not affected by this change as they do not depend on q. The statistics R'_β and $\mathrm{R}'_{\beta\beta}$ depend on q through the averaged periodograms $F_T(1,q;\mathcal{Z}_T) := 2\pi T^{-1}\mathrm{Re}\left[\sum_{j=1}^{q} I_T(\omega_j;\mathcal{Z}_T)\right]$, where $I_T(\omega_j;\mathcal{Z}_T)$ are periodograms of relevant variables $\mathcal{Z}_T$. Following Marinucci and Robinson (2003), define averaged periodograms more generally as $F_T(m,n;\mathcal{Z}_T) := 2\pi T^{-1}\mathrm{Re}\left[\sum_{j=m}^{n} I_T(\omega_j;\mathcal{Z}_T)\right]$.

Note that $F_T(1,T-1;\mathcal{Z}_T)$ is exactly the sample variance matrix of $\mathcal{Z}_T$. Also, $F_T(1,T-1;\mathcal{Z}_T) = F_T(1,q;\mathcal{Z}_T) + F_T(q+1,T-q-1;\mathcal{Z}_T) + F_T(T-q,T-1;\mathcal{Z}_T)$ and $F_T(1,q;\mathcal{Z}_T) = F_T(T-q,T-1;\mathcal{Z}_T)$. Therefore, $F_T(1,T-1;\mathcal{Z}_T) = 2F_T(1,q;\mathcal{Z}_T) + F_T(q+1,T-q-1;\mathcal{Z}_T)$. If the variation of the remainder $F_T(q+1,T-q-1;\mathcal{Z}_T)$ is negligible compared to the variation of the periodograms at low frequencies $F_T(1,q;\mathcal{Z}_T)$ for $1/q+q/T\to\infty$, then the limiting behavior of $2\,F_T(1,q;\mathcal{Z}_T)$ is the same as that of the variance matrix up to a nonrandom constant. Finally, the condition $E(\varepsilon_t|u_{t-1},\ldots)\neq 0$ affects only the remainder $F_T(q+1,T-q-1;\mathcal{Z}_T)$, where $Z_t=(x_{t-1},y_t)'$; see Marinucci and Robinson (2003). This result requires stronger assumptions, for example, an MA-representation for e_t driven by i.i.d. shocks that have finite fourth moments.

Consider the following finite-sample statistics that are based on sample variances:

$$\mathrm{Y}_\beta = \left(\psi_{uu}\psi_{\varepsilon\varepsilon}(1-r^2)\right)^{-1/2}\frac{1}{T}\sum_{t=1}^{T}\Big[(x_{t-1}-\bar{x}_T)(y_t-\beta_0 x_{t-1} - r\sqrt{\psi_{\varepsilon\varepsilon}/\psi_{uu}}(x_t-x_{t-1}))\Big],$$

$$\mathrm{Y}_{\beta\beta} = \psi_{uu}^{-1}T^{-2}\sum_{t=1}^{T}(x_{t-1}-\bar{x}_T)^2,$$

https://doi.org/10.1017/S0266466620000079 Published online by Cambridge University Press

where $\bar{x}_T = \frac{1}{T}\sum_{t=1}^{T} x_{t-1}$. Their limits for $\rho_T = 1 + c/T$ and m.d.s. ε_t are considered by Jansson and Moreira (2006). In particular, $(\mathrm{Y}_\beta, \mathrm{Y}_{\beta\beta}) \Rightarrow (\mathcal{Y}_\beta, \mathcal{Y}_{\beta\beta})$, where

$$\mathcal{Y}_\beta = \int_0^1 J_c^\mu(\tau) dW^z(\tau) + (b - \frac{r}{\sqrt{1-r^2}}c) \int_0^1 \left(J_c^\mu(\tau)\right)^2 d\tau,$$
$$\mathcal{Y}_{\beta\beta} = \int_0^1 \left(J_c^\mu(\tau)\right)^2 d\tau,$$

with the demeaned O–U process defined as $J_c^\mu(\tau) := J_c(\tau) - \int_0^1 J_c(\tau)d\tau$. As was discussed above, $\mathrm{Y}_\beta = 1/2\ \mathrm{R}'_\beta$ and $\mathrm{Y}_{\beta\beta} = 1/2\ \mathrm{R}'_{\beta\beta}$, where R'_β and $\mathrm{R}'_{\beta\beta}$ are calculated with $q = T - 1$ frequencies. Based on the asymptotic relation between $F_T(1, T-1; \mathcal{Z}_T)$ and $2\ F_T(1, q; \mathcal{Z}_T)$, we conclude that R' has the same limit as $\mathrm{Y} := (\mathrm{Y}_\beta, \mathrm{R}'_\rho, \mathrm{Y}_{\beta\beta}, \mathrm{R}'_{\rho\rho})$ under condition $1/q + q/T \to \infty$. That is R' weakly converges to $(\mathcal{Y}_\beta, \mathcal{R}_\rho, \mathcal{Y}_{\beta\beta}, \mathcal{R}_{\rho\rho})$. Note that $\mathcal{Y}$ appear as the asymptotic limit of sufficient statistics in the optimal tests based on the complete Gaussian likelihood; see Jansson and Moreira (2006) and Elliott et al. (2015). The claim for the optimality now follows directly from the results of Elliott et al. (2015). The condition $E(\varepsilon_t | u_{t-1}, \ldots) = 0$ is not required.

8. APPLICATION STEPS

Let us summarize the nearly optimal test as a series of steps. The input to the test is the hypothesized null value β_0, the series of $y_1, \ldots, y_T$ and $x_0, \ldots, x_T$, and the frequency q at which long-run predictability is to be tested. For example, for the data span of $T = 50$ years and the forecasting time horizon of 10 years, the cut-off frequency is $q = 5$.

The first step of the algorithm is to consistently estimate the long-run covariance matrix $\hat{\psi}$ between the series of $y_t - \beta_0 x_{t-1}$ and $x_t - \rho x_{t-1}$. This estimate is obtained from the vector autoregressive model for (y_t, x_t). Following notations, define elements of $\hat{\psi}$ matrix:

$$\psi = \begin{pmatrix} \psi_{\varepsilon\varepsilon} & \psi_{\varepsilon u} \\ \psi_{u\varepsilon} & \psi_{uu} \end{pmatrix},$$

and the derived parameters $\hat{r} = \hat{\psi}_{\varepsilon u}(\hat{\psi}_{\varepsilon\varepsilon}\hat{\psi}_{uu})^{-1/2}$ and $\hat{\psi}_{\varepsilon\varepsilon|u} = (1 - \hat{r}^2)\hat{\psi}_{\varepsilon\varepsilon}$. Based on q , $\hat{r}$, and significance level α, the test routine computes constant thresholds $K_{\alpha,s}$, K_c, and $K_{\alpha,n}$, and the probability weights λ_ι for the least favorable distribution of ρ.

The second step is to determine whether to apply the "standard t-test" or the likelihood ratio test. To this end, we assess the persistence in x_t by computing statistic

https://doi.org/10.1017/S0266466620000079 Published online by Cambridge University Press

$$C(R^{\mu}) = T\frac{\sum_{t=1}^{T}(x_{t-1}-\bar{x}_T)(x_t-x_{t-1})}{\sum_{t=1}^{T}(x_{t-1}-\bar{x}_T)^2}$$

and verifying condition $C(R^{\mu}) \le K_c$. If this condition is satisfied, then the test switches to the t-statistic. For example, in the i.i.d. case with

$$t(\mathcal{R}^{\mu}) = \frac{\frac{\mathcal{R}_{\beta}}{\mathcal{R}_{\beta\beta}} + \frac{r}{\sqrt{1-r^2}}C(\mathcal{R}^{\mu})}{\sqrt{\mathcal{R}_{\beta\beta}^{-1} + \frac{r}{\sqrt{1-r^2}}\left(\mathcal{R}_{\rho\rho}^{\mu}\right)^{-1}}},$$

the rejection rule becomes $|t(\mathcal{R}^{\mu})| > K_{\alpha,s}$. Conversely, when $C(R^{\mu}) > K_c$, we proceed to construct the likelihood ratio statistic. First, we form Fourier transformations of the series at natural frequencies $\omega_j = (2\pi j)/T, j = 1, \ldots, q$:

$$d(\omega_j; Y) = \sum_{t=1}^{T} y_t \exp\{-i\omega_j t\},$$

$$d(\omega_j; \{x_t\}_{t=1}^{T}) = \sum_{t=1}^{T} x_t \exp\{-i\omega_j t\},$$

$$d(\omega_j; \{x_t\}_{t=1}^{T}) = \sum_{t=1}^{T} x_{t-1} \exp\{-i\omega_j t\},$$

and compute sufficient statistics as in Theorem 2:

$$\mathrm{R}'_{\beta} = \left(\hat{\psi}_{uu}\hat{\psi}_{\varepsilon\varepsilon}(1-\hat{r}^2)\right)^{-1/2}\frac{4\pi}{T}\,\mathrm{Re}$$
$$\left[\sum_{j=1}^{q} d(\omega_j; \{x_{t-1}\}_{t=1}^{T})(d(\omega_j; Y) - \beta_0 d(\omega_j; \{x_{t-1}\}_{t=1}^{T})\right.$$
$$\left. - \hat{r}\sqrt{\hat{\psi}_{\varepsilon\varepsilon}/\hat{\psi}_{uu}}(d(\omega_j; \{x_t\}_{t=1}^{T}) - d(\omega_j; \{x_{t-1}\}_{t=1}^{T})))^*\right],$$

$$\mathrm{R}'_{\rho} = \frac{1}{2T}(\hat{\psi}_{uu}^{-1}x_T^2 - \hat{\sigma}_{uu}^{-1}\sum_{t=1}^{T}(x_t - x_{t-1})^2) - \hat{r}(1-\hat{r}^2)^{-1/2}\mathrm{R}'_{\beta},$$

$$\mathrm{R}'_{\beta\beta} = \hat{\psi}_{uu}^{-1}T^{-2}4\pi\sum_{j=1}^{q}|d(\omega_j; \{x_{t-1}\}_{t=1}^{T})|^2,$$

$$\mathrm{R}'_{\rho\rho} = \hat{\psi}_{uu}^{-1}T^{-2}\sum_{t=1}^{T}x_{t-1}^2.$$

https://doi.org/10.1017/S0266466620000079 Published online by Cambridge University Press

The final step is to form the likelihood ratio:

$$\frac{\sum_{\iota=1}^{I} f(\mathrm{R}'|b_\iota, c_\iota)}{\sum_{\iota=1}^{I} f(\mathrm{R}'|0, c_\iota)\lambda_\iota},$$

where the density root is

$$f(\mathrm{R}'|b,c) = \exp\left\{ b\mathrm{R}'_\beta + c\mathrm{R}'_\rho - \frac{1}{2}\left(b - c\frac{r}{\sqrt{1-r^2}}\right)^2 \mathrm{R}'_{\beta\beta} - \frac{1}{2}c^2\mathrm{R}'_{\rho\rho} \right\}$$

and the sums are calculated across points $c_\iota = (0.25 * 6)^2, (0.25 * 5)^2, \ldots, -(-0.25 * 56)^2$ and $b_\iota = 1.645\sqrt{\frac{-2c_\iota + 6}{1-\hat{r}^2}}$. The hypothesis of no long-run predictability is rejected when the above statistic is larger than the threshold $K_{\alpha,n}$.

9. CONCLUSION

We consider the long-run predictability testing problem in which the predictor x_{t-1} is nearly integrated and the y_t is close to the unpredictable noise. We suggest a semiparametric test that efficiently uses the information from the lowest q frequencies in the modified Whittle approximation. This test is nearly optimal in the sense of Elliott et al. (2015).

Based on simulations, we find that the long-run test with the $q = 10$ lowest frequencies yields powers close to the nearly optimal test based on the entire Gaussian likelihood. At the same time, we reduce size distortions due to endogeneity with $q < 10$ in small samples.

The modified Whittle approximation is derived in Lemma 1 and Corollary 1 and offers an approximation to the conditional distribution of Y given X uniformly over the persistence parameter ρ.

REFERENCES

Andrews, D.W.K. & P. Guggenberger (2012) Asymptotics for LS, GLS, and feasible GLS statistics in an AR(1) model with conditional heteroskedasticity. *Journal of Econometrics* 169(2), 196–210.

Andrews, D.W.K., M. Moreira, & J.H. Stock (2008) Efficient two-sided nonsimilar invariant tests in IV regression with weak instruments. *Journal of Econometrics* 146, 241–254.

Bandi, F.M. & B. Perron (2008) Long-run risk-return trade-offs. *Journal of Econometrics* 143, 349–374.

Boudoukh, J., M. Richardson, & R. Whitelaw (2016) New evidence on the forward premium puzzle. *Journal of Financial and Quantitative Analysis* 51(3), 875–897.

Brillinger, D.R. (1975) *Time Series: Data Analysis and Theory*. Holt, Rinehart and Winston.

Campbell, J.Y. & M. Yogo (2006) Efficient tests of stock return predictability. *Journal of Financial Economics* 81(1), 27–60.

Cavanagh, C.L., G. Elliott, & J.H. Stock (1995) Inference in models with nearly integrated regressors. *Econometric Theory* 11(05), 1131–1147.

https://doi.org/10.1017/S0266466620000079 Published online by Cambridge University Press

Chiburis, R.C. (2009) *Approximately most powerful tests for moment inequalities*. Chapter of Ph.D. dissertation, Department of Economics, Princeton University.
Christensen, B.J. & M.O. Nielsen (2006) Asymptotic normality of narrow-band least squares in the stationary fractional cointegration model and volatility forecasting. *Journal of Econometrics* 133(1), 343–371.
Davidson, J. (2002) Establishing conditions for the functional central limit theorem in nonlinear and semiparametric time series. *Journal of Econometrics* 106(2), 243–269.
Corbae, D., S. Ouliaris, & P.C.B. Phillips (2002) Band spectral regression with trending data. *Econometrica* 70(3), 1067–1110.
Dahlhaus, R. (2000) A likelihood approximation for locally stationary processes. *Annals of Statistics* 28(6), 1762–1794.
Davis, P.J. (1979) *Circulant Matrices*. Wiley.
Dzhaparidze, K. (1986) *Parameter Estimation and Hypothesis Testing in Spectral Analysis of Stationary Time Series*. Springer.
Durlauf, S.N. & P.C.B. Phillips (1986) Multiple time series regression with integrated processes. *Review of Economic Studies* 53(4), 473–495.
Nikolay, G . (2009) A new look at the forward premium puzzle. *Journal of Financial Econometrics* 7(3), 312–338.
Elliott, G. & U.K. Müller (2014) Pre and post break parameter inference. *Journal of Econometrics* 180, 141–157.
Elliott, G., U. Müller, & M. Watson (2015) Nearly optimal tests when a nuisance parameter is present under the null hypothesis. *Econometrica* 83(2), 771–811.
Elliott, G., T.J. Rothenberg, & J.H. Stock (1996) Efficient tests for an autoregressive unit root. *Econometrica* 64, 813–836.
Elliott, G. & J.H. Stock (1994) Inference in time series regression when the order of integration of a regressor is unknown. *Econometric Theory* 10(3/4), 672–700.
Fisher, M.E. & J.J. Seater (1993) Long-run neutrality and superneutrality in an ARIMA framework. *The American Economic Review* 83(3), 402–415.
Goyal, A. & I. Welch (2008) A comprehensive look at the empirical performance of equity premium prediction. *The Review of Financial Studies* 21(4), 1455–1508.
Granger, C.W.J. (1969) Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* 37(3), 424–438.
Granger, C.W.J. & M. Hatanaka (1964) *Spectral Analysis of Economic Time Series*. Princeton University Press.
Hamilton, J.D. (1994) *Time Series Analysis*. Princeton University Press.
Hansen, B.E. (1999) The grid bootstrap and the autoregressive model. *The Review of Economics and Statistics* 81(4), 594–607.
Hannan, E.J. (1973a) The asymptotic theory of linear time-series models. *Journal of Applied Probability* 10(1), 130–145.
Hannan, E.J. (1973b) Central limit theorems for time series regression. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 26, 157–170.
Hansen, B.E. (1992) Convergence to stochastic integrals for dependent heterogenous processes. *Econometric Theory* 8, 489–500.
Hjalmarsson, E. (2011) New methods for inference in long-horizon regressions, *Journal of Financial and Quantitative Analysis* 46(3), 815–839.
Hjalmarsson, E. (2007) Fully modified estimation with nearly integrated regressors. *Finance Research Letters* 4(2), 92–94.
Hwang, J. & Y. Sun (2018) Simple, robust, and accurate f and t tests in cointegrated systems. *Econometric Theory* 34(5), 949–984.
Hurvich, C.M., E. Moulines, & P. Soulier (2005) Estimating long memory in volatility. *Econometrica* 73(4), 1283–1328.

https://doi.org/10.1017/S0266466620000079 Published online by Cambridge University Press

Ibragimov, I.A. & Y.V. Linnik (1971) *Independent and Stationary Sequences of Random*. Londo Mathematical Society.

Ibragimov, I.A. & Y.A. Rozanov (1978) *Gaussian Random Processes*. Springer-Verlag.

Ing, C.-K. (2003) Multistep prediction in autoregressive processes. *Econometric Theory* 19(2), 254–279.

Jansson, M. & M. Moreira (2006) Optimal inference in regression models with nearly integrated regressors. *Econometrica* 74, 681–714.

Jeganathan, P. (1995) Some aspects of asymptotic theory with applications to time series models. *Econometric Theory* 11, 818–887.

Hamilton, J.H. (1994) *The Time Series Analysis*. Princeton University Press.

Herrndorf, N. (1984) A functional central limit theorem for weakly dependent sequences of random variables. *The Annals of Probability* 12(1), 141–153.

Hong, H. & J.C. Stein (1999) A unified theory of underreaction, momentum trading, and overreaction in asset markets. *The Journal of Finance* 54(6), 2143–2184.

Kapetanios, G. & F. Papailias (2011) *Block Bootstrap and Long Memory*. Working papers 679, School of Economics and Finance, Queen Mary University of London.

Kim, Y.M. & D.J. Nordman (2012) Properties of a block bootstrap under long-range dependence. *Sankhya A* 73(1), 79–109.

Kostakis, A., T. Magdalinos, & M.P. Stamatogiannis (2015) Robust econometric inference for stock return predictability. *Review of Financial Studies* 28(5), 1506–1553.

Krafty, R.T. & W.O. Collinge (2013) Penalized multivariate whittle likelihood for power spectrum estimation. *Biometrika* 100(2), 447–458.

Lehmann, E.L. & J.P. Romano (2005) *Testing Statistical Hypotheses*. Springer.

Lobato, I.N. (1997) Consistency of the averaged cross-periodogram in long memory series. *Journal of Time Series Analysis* 18, 137–155.

Magdalinos, T. & P.C.B. Phillips (2009) *Econometric Inference in the Vicinity of Unity*. CoFie Working Paper, Singapore Management University.

Marinucci, D. & P.M. Robinson (2003) Semiparametric frequency domain analysis of fractional cointegration. In P.M. Robinson (ed.), *Time Series with Long Memory*, pp. 334–373. Oxford University Press.

Marmol, F. & C. Velasco (2004) Consistent testing of cointegrating relationships. *Econometrica* 72, 1809–1844.

McLeish, D.L. (1974) Dependent central limit theorems and invariance principles. *The Annals of Probability* 2(4), 620–628.

Mikusheva, A. (2007) Uniform inference in autoregressive models. *Econometrica* 75(5), 1411–1452.

Mikusheva, A. (2012) One-dimensional inference in autoregressive models with the potential presence of a unit root. *Econometrica* 80(1), 173–212.

Müller, U. & M.W. Watson (2013) Low-frequency robust cointegration testing. *Journal of Econometrics* 174, 66–81.

Mishkin, F.S. (1990) What does the term structure of interest rates tell us about future inflation? *Journal of Monetary Economics* 25, 77–95.

Müller, U.K. & M.W. Watson (2008) Testing models of low-frequency variability. *Econometrica* 76, 979–1016.

Moreira, M.J., R. Mourão, & H. Moreira (2014) A Critical Value Function Approach, with an Application to Persistent Time-Series. Centre for Microdata Methods and Practice Working Paper, The Institute for Fiscal Studies, UCL.

Nielsen, M.O. & P. Frederiksen (2011) Fully modified narrow-band least squares estimation of weak fractional cointegration. *Econometrics Journal* 14, 77–120.

Park, J.Y. (2006) A bootstrap theory for weakly integrated processes. *Journal of Econometrics* 133(2), 639–672.

https://doi.org/10.1017/S0266466620000079 Published online by Cambridge University Press

Phillips, P.C.B. & V. Solo (1992) Asymptotics for linear processes. *Annals of Statistics* 20(2), 971–1001.
Phillips, P.C.B. (1987) Towards a unified asymptotic theory for autoregression. *Biometrika* 74(3), 535–547.
Phillips, P.C.B. (1987) Asymptotic expansions in nonstationary vector autoregressions. *Econometric Theory* 3(1), 45–68.
Phillips, P.C.B. (1988) Regression theory for near-integrated time series. *Econometrica* 56(5), 1021–1043.
Phillips, P.C.B. (1991) Spectral regression for cointegrated time series. In W.A. Barnett, J. Powell, G. Tauchen (eds.), *Nonparametric and Semiparametric Methods in Economics and Statistics*, pp. 413–435. Cambridge University Press.
Phillips, P.C.B. (2007) Unit root log periodogram regression. *Journal of Econometrics* 138, 104–124.
Phillips, P.C.B. (2014a) On confidence intervals for autoregressive roots and predictive regression. *Econometrica* 82(3), 1177–1195.
Phillips, P.C.B. (2014b) Optimal estimation of cointegrated systems with irrelevant instruments. *Journal of Econometrics* 178, 210–224.
Phillips, P.C.B. & B.E. Hansen (1990) Statistical inference in instrumental variables regression with I(1). *The Review of Economic Studies* 57(1), 99–125.
Politis, D.N., J.P. Romano, & M. Wolf (1999) *Subsampling*. Springer.
Poskit, D.S. (2007) Properties of the sieve boostrap for fractionally integrated and non-invertible processes. *Journal of Time Series Analysis* 29(2), 224–250.
Richardson, M. & J. Stock (1989) Drawing inferences from statistics based on multi-year asset. *Journal of Financial Economics* 25(2), 323–348.
Ramey, V.A. & S. Zubairy (2018) Government spending multipliers in good times and in bad: Evidence from U.S. historical data. *Journal of Political Economy* 126(2), 850–901.
Robinson, P.M. (1994) Semiparametric analysis of long-memory time series. *Annals of Statistics* 22, 515–539.
Robinson, P.M. (2003) *Time Series with Long Memory*. Oxford University Press.
Sizova, N. (2013) Long-horizon return regressions with historical volatility. *Journal of Business & Economic Statistics* 31(4), 546–559.
Rozanov, Y.A. (1967) Stationary random processes. In G.M. Jenkins, E. Parzen (eds.), *Holden-Day Series in Time Series Analysis*. Holden-Day.
Sizova, N. (2014) A frequency-domain alternative to long-horizon regressions with application to return predictability. *Journal of Empirical Finance* 28, 261–272.
Stambaugh, R.F. (1999) Predictive regressions. *Journal of Financial Economics* 54, 375–421.
Stock, J.H. (1994) *Unit Roots Structural Breaks and Trends*. Elsevier Science B.V.
Stock, J. & M. Watson (1996) *Confidence Sets in Regressions with Highly Serially Correlated Regressors*. Working Paper, Harvard University.
Valkanov, R. (2003) Long-horizon regressions: Theoretical results and applications. *Journal of Financial Economics* 68(2), 201–232.
Ventosa-Santaulària, D. (2009) Spurious regression. *Journal of Probability and Statistics*.
Wolf, M. (2000) Stock returns and dividend yields revisited: A new way to look at an old problem. *Journal of Business & Economic Statistics* 18(1), 18–30.
Wooldridge, J.M. & H. White (1988) Some invariance principles and central limit theorems for dependent heterogeneous processes. *Econometric Theory* 4(2), 210–230.
Velasco, C. (1999) Gaussian semiparametric estimation of nonstationary time series. *Journal of Time Series Analysis* 20, 87–127.
Whittle, P. (1953) The analysis of multiple stationary time series. *Journal of the Royal Statistical Society Series B (Methodological)* 15(1), 125–139.
Whittle, P. (1951) *Hypothesis Testing in Times Series Analysis*. Hafner.

https://doi.org/10.1017/S0266466620000079 Published online by Cambridge University Press

APPENDIX

A. PROOFS OF LEMMA 1 AND LEMMA 2

Proof of lemma 1. Without loss of generality, let $\mu_x = 0$ and $\mu_y = 0$. Conditional on $X = (x_0, \ldots, x_{T-1})'$, the distribution of the vector $Y = (y_1, \ldots, y_T)'$ is normal with mean $\mathbb{E}(Y|X) = \Pi X = \beta_x X + \mathbb{E}(\varepsilon|u)$, where Π is a $T \times T$ matrix, $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_T')$ and $u = (u_1, \ldots, u_{T-1})'$ are column vectors of shocks. The true conditional variance of Y is the matrix $B = \mathrm{Var}(\varepsilon|u)$.

For the Whittle approximation, the sum $\sum_{j=1}^{T} \mathrm{tr}\left[s(\omega_j; \Gamma_{x,y})^{-1} \mathcal{I}(\omega_j; X, Y)\right]$ can be rewritten as a quadratic term $(X'Y')\tilde{\Omega}^{-1}(X'Y')'$, where matrix $\tilde{\Omega}$ replaces the variance of the Gaussian vector $(X'Y')'$ under the Whittle likelihood. Therefore, matrix $\tilde{\Omega}$ holds information about the conditional moments of Y under the Whittle likelihood.

Denote the natural partition of matrices $s(\omega; \Gamma_{x,y})$ by $s_{xx}(\omega; \Gamma_{x,y})$, $s_{xy}(\omega; \Gamma_{x,y})$, $s_{yx}(\omega; \Gamma_{x,y})$, and $s_{yy}(\omega; \Gamma_{x,y})$. Note that matrix $\tilde{\Omega}^{-1}$ can be comformably partitioned into four blocks, $T \times T$ each: $\tilde{A}_{xx}$, $\tilde{A}_{x,y}$, $\tilde{A}_{y,x}$, and $\tilde{A}_{y,y}$, where the elements of the blocks are as follows:

$$\tilde{A}_{xx}(t,s) = (2\pi T)^{-1} \sum_{j=1}^{T} s_{yy}(\omega_j; \Gamma_{x,y}) \left[\det\{s(\omega_j; \Gamma_{x,y})\}\right]^{-1} \exp\{-i\omega_j(t-s)\},$$

$$\tilde{A}_{yy}(t,s) = (2\pi T)^{-1} \sum_{j=1}^{T} s_{xx}(\omega_j; \Gamma_{x,y}) \left[\det\{s(\omega_j; \Gamma_{x,y})\}\right]^{-1} \exp\{-i\omega_j(t-s)\},$$

$$\tilde{A}_{xy}(t,s) = -(2\pi T)^{-1} \mathrm{Re}\left(\sum_{j=1}^{T} s_{yx}(\omega_j; \Gamma_{x,y}) \left[\det\{s(\omega_j; \Gamma_{x,y})\}\right]^{-1} \exp\{-i\omega_j(t-s)\} \right)$$

and $\tilde{A}_{yx}(t,s) = \tilde{A}_{xy}(s,t)$. The above matrices follow directly from the identity $\sum_{j=1}^{T} \mathrm{tr}\left[s(\omega_j; \Gamma_{x,y})^{-1} \mathcal{I}(\omega_j; X, Y)\right] = (X'Y')\tilde{\Omega}^{-1}(X'Y')'$, where $s(\omega; \Gamma_{x,y})$ is given by equation (2).

Let $\tilde{\Omega}_{xx}$, $\tilde{\Omega}_{xy}$, $\tilde{\Omega}_{yx}$, and $\tilde{\Omega}_{yy}$ be the corresponding partition of the matrix $\tilde{\Omega}$. The conditional expectation of Y under the Whittle likelihood is $\tilde{\mathbb{E}}(Y|X) = \tilde{\Pi}X$, where $\tilde{\Pi} = \tilde{\Omega}_{yx}\tilde{\Omega}_{xx}^{-1} = -\tilde{A}_{yy}^{-1}\tilde{A}_{yx}$. Furthermore, $\widetilde{\mathrm{Var}}(Y|X) := \tilde{B} = \tilde{\Omega}_{yy} - \tilde{\Omega}_{yx}\tilde{\Omega}_{xx}^{-1}\tilde{\Omega}_{xy} = \tilde{A}_{yy}^{-1}$.

The difference between the Gaussian and Whittle log likelihoods will depend on the following term:

$$\begin{aligned}
&(Y - \tilde{\Pi}X)'\tilde{B}^{-1}(Y - \tilde{\Pi}X) - (Y - \Pi X)B^{-1}(Y - \Pi X) \\
&\quad = 2X'(\Pi - \tilde{\Pi})'\tilde{B}^{-1}(Y - \Pi X) + X'(\Pi - \tilde{\Pi})'\tilde{B}^{-1}(\Pi - \tilde{\Pi})X \\
&\qquad - (Y - \Pi X)'(B^{-1} - \tilde{B}^{-1})(Y - \Pi X). \qquad \textbf{(A.7)}
\end{aligned}$$

First consider the last element that comes from the approximation error in conditional variances. From equation (2), it follows that $s_{xx}(\omega_j; \Gamma_{x,y}) \left[\det\{s(\omega_j; \Gamma_{x,y})\}\right]^{-1} = s_{uu}(\omega_j; \Gamma_e) \left[\det(s(\omega_j; \Gamma_e))\right]^{-1}$. Therefore,

$$\tilde{B}^{-1} = \tilde{A}_{yy} = \left[\sum_{j=1}^{T} \frac{s_{uu}(\omega_j; \Gamma_e)}{\det(s(\omega_j; \Gamma_e))} e^{-i\omega_j(t-s)} \frac{1}{2\pi T} \right]_{t,s},$$

https://doi.org/10.1017/S0266466620000079 Published online by Cambridge University Press

does not depend on ρ. Similarly, $B = \mathrm{Var}(\varepsilon|u)$ and $Y - \mathbb{E}(Y|X) = \varepsilon - \mathbb{E}(\varepsilon|u)$ do not depend on ρ. As follows from Dzhaparidze (1986), under assumptions of Lemma 1, $(Y-\mathbb{E}(Y|X))(B^{-1}-\tilde{B}^{-1})(Y-\mathbb{E}(Y|X)) \sim O_p(1)$ for a fixed $\rho < 1$. Therefore, this relation also holds for any $\rho \in \mathbb{R}$ as well as sequences ρ_T. We conclude that the difference in conditional variances does not result in an asymptotically significant deviation of the Whittle approximation from the true likelihood.

We next consider the other elements in the difference (A.7) that stem from the errors in the conditional mean. From equation (2), it follows that

$$s_{xy}(\omega_j;\Gamma_{x,y})\left[\det\{s(\omega_j;\Gamma_{x,y})\}\right]^{-1} = \beta_x s_{uu}(\omega_j;\Gamma_e)\left[\det(s(\omega_j;\Gamma_e))\right]^{-1} + (e^{-i\omega_j}-\rho)s_{u\varepsilon}(\omega_j;\Gamma_e)\left[\det(s(\omega_j;\Gamma_e))\right]^{-1}.$$

Therefore,

$$-\tilde{A}_{yx} = \beta_x\tilde{A}_{yy} + \mathrm{Re}\left[\frac{1}{2\pi T}\sum_{j=1}^{T}\frac{s_{u\varepsilon}(\omega_j;\Gamma_e)}{\det(s(\omega_j;\Gamma_e))}(e^{-i\omega_j(t-s+1)} - \rho e^{-i\omega_j(t-s)})\right]_{t,s}$$

and

$$\tilde{\mathbb{E}}(Y|X) = \tilde{\Pi}X = -\tilde{A}_{yy}^{-1}\tilde{A}_{yx}X = \beta_x X + \tilde{A}_{yy}^{-1}\mathrm{Re}(T_1)U + \tilde{A}_{yy}^{-1}\mathrm{Re}(T_2)(x_0 - x_T),$$

where $U = (u_1,\dots,u_T)'$, T_1 is a matrix $T \times T$ with the elements $T_1(t,s) = \frac{1}{2\pi T}\sum_{j=1}^{T}\frac{s_{u\varepsilon}(\omega_j;\Gamma_e)}{\det(s(\omega_j;\Gamma_e))}e^{-i\omega_j(t-s)}$, and T_2 is a $T \times 1$ vector with the elements $T_2(t) = \frac{1}{2\pi T}\sum_{j=1}^{T}\frac{s_{u\varepsilon}(\omega_j;\Gamma_e)}{\det(s(\omega_j;\Gamma_e))}e^{-i\omega_j t}$. Therefore,

$$\tilde{\mathbb{E}}(Y|X) = \tilde{\Pi}X = \beta_x X - \tilde{A}_{\varepsilon\varepsilon}^{-1}\tilde{A}_{\varepsilon u}U + \tilde{A}_{yy}^{-1}\mathrm{Re}(T_2)(x_0 - x_T), \tag{A.8}$$

where $T \times T$ matrices $\tilde{A}_{\varepsilon\varepsilon}(t,s) = (2\pi T)^{-1}\sum_{j=1}^{T}s_{uu}(\omega_j;\Gamma_e)\left[\det\{s(\omega_j;\Gamma_e)\}\right]^{-1}\exp\{-i\omega_j(t-s)\}$ and $\tilde{A}_{\varepsilon u}(t,s) = -(2\pi T)^{-1}\mathrm{Re}\left[\sum_{j=1}^{T}s_{u\varepsilon}(\omega_j;\Gamma_e)\left[\det\{s(\omega_j;\Gamma_e)\}\right]^{-1}\exp\{-i\omega_j(t-s)\}\right]$ are defined by analogy to $\tilde{A}_{yy}$ and $\tilde{A}_{yx}$. Therefore, $\tilde{\mathbb{E}}(\varepsilon|U) := -\tilde{A}_{\varepsilon\varepsilon}^{-1}\tilde{A}_{\varepsilon u}U$ is exactly the conditional mean of ε given U implied by the Whittle approximation to the joint Gaussian likelihood of (ε, U). As follows from Dzhaparidze (1986), $\left(\tilde{\mathbb{E}}(\varepsilon|U) - \mathbb{E}(\varepsilon|U)\right)'\mathrm{Var}(E|U)^{-1}\left(\tilde{\mathbb{E}}(\varepsilon|U) - \mathbb{E}(\varepsilon|U)\right) = O_p(1)$. Note that this error does not depend on ρ. Finally, we substitute the conditional expectation in (A.8) into the formula for the approximation error (A.7) and obtain that the difference between the (conditional parts of) Gaussian and Whittle likelihoods is $O_p(1) + (x_0 - x_T)^2\mathrm{Re}(T_2')\tilde{A}_{yy}^{-1}\mathrm{Re}(T_2) + 2(x_0 - x_T)\mathrm{Re}(T_2')(Y - \Pi X)$.

Consider a local-to-unity sequence of $\rho_T = 1 + c/T$. For such sequences, as follows from result (1) in Lemma 3, the term $(x_0 - x_T)^2$ is of the stochastic order $O_p(T)$. We conclude that the Whittle approximation (in its regular form) and the Gaussian likelihood are not equivalent for $\rho = 1$ and are not uniformly equivalent for $|\rho| < 1$. However, we can construct an equivalent modification to the Whittle approximation if we replace series Y in the Whittle approximation by the transformed series $Y^W = Y + \delta_y(x_0 - x_T)$, where δ_y is a $T \times 1$ vector equal to $\mathrm{Re}\left[\tilde{A}_{yy}^{-1}T_2\right]$.

https://doi.org/10.1017/S0266466620000079 Published online by Cambridge University Press

The resulting approximation error takes the form:

$$
\begin{aligned}
&(Y^W-\tilde{\Pi}X)'\tilde{B}^{-1}(Y^W-\tilde{\Pi}X)-(Y-\Pi X)B^{-1}(Y-\Pi X)\\
&\quad=2(\mathbb{E}(\varepsilon|u)-\tilde{\mathbb{E}}(\varepsilon|U))'\tilde{B}^{-1}(\varepsilon-\mathbb{E}(\varepsilon|u))+(\mathbb{E}(\varepsilon|u)-\tilde{\mathbb{E}}(\varepsilon|U))'\tilde{B}^{-1}(\mathbb{E}(\varepsilon|u)-\tilde{\mathbb{E}}(\varepsilon|U))\\
&\qquad-(\mathbb{E}(\varepsilon|u)-\tilde{\mathbb{E}}(\varepsilon|U))'(B^{-1}-\tilde{B}^{-1})(\mathbb{E}(\varepsilon|u)-\tilde{\mathbb{E}}(\varepsilon|U)).
\end{aligned}
$$

The above error does not depend on ρ. Therefore, the modified Whittle approximation is uniform in ρ. □

LEMMA 2. *Suppose,* $\delta_y=\mathrm{Re}\left(\tilde{A}_{yy}^{-1}T_2\right)$, *where matrix* $\tilde{A}_{yy}$ *and vector* T_2 *are defined in Lemma 1. Then,* $\sum_{t=1}^{T}e^{i\omega_{j_0}t}\delta_y(t)=s_{u\varepsilon}(\omega_{j_0};\Gamma_e)\left(s_{uu}(\omega_{j_0};\Gamma_e)\right)^{-1}$ *for* $j_0=1,\ldots,T$.

Proof. Define complex vector $T_y=\tilde{A}_{yy}^{-1}T_2$. The real part of $\sum_{t=1}^{T}e^{i\omega_{j_0}t}\delta_y(t)$ equals $\frac{1}{2}\mathrm{Re}(\sum_{t=1}^{T}(e^{i\omega_{j_0}t}+e^{-i\omega_{j_0}t})T_y(t))$. The imaginary part of $\sum_{t=1}^{T}e^{i\omega_{j_0}t}\delta_y(t)$ equals $\frac{1}{2}\mathrm{Im}(\sum_{t=1}^{T}(e^{-i\omega_{j_0}t}-e^{i\omega_{j_0}t})T_y(t))$. Therefore, consider

$$
\sum_{t=1}^{T}e^{i\omega_{j_0}t}T_y(t)=\sum_{j=1}^{T}\frac{s_{u\varepsilon}(\omega_j;\Gamma_e)}{\det(s(\omega_j;\Gamma_e))}\frac{\sum_{t=1}^{T}\sum_{s=1}^{T}\tilde{A}_{yy}^{-1}(t,s)e^{-i(\omega_j s-\omega_{j_0}t)}}{2\pi T}.
$$

The result for $\sum_{t=1}^{T}e^{-i\omega_{j_0}t}T_y(t)$ is derived analogously. Note that the matrix $\tilde{A}_{yy}^{-1}$ is circulant. That is, its elements allow for the following representation: $\tilde{A}_{yy}^{-1}(t,s)=\tilde{\gamma}(s-t)$, and for any $k\in[1-T,-1]$, $\tilde{\gamma}(k)=\tilde{\gamma}(T+k)$. This property of the matrix $\tilde{A}_{yy}^{-1}$ follows from the same property of $\tilde{A}_{yy}$, which is verified directly from the definition of $\tilde{A}_{yy}$.

Because matrix $\tilde{A}_{yy}^{-1}$ is circulant, all of the sums $\sum_{t=1}^{T}\sum_{s=1}^{T}\tilde{A}_{yy}^{-1}(t,s)e^{-i(\omega_j s-\omega_{j_0}t)}$ are zeroes as long as $\omega_j\neq\omega_{j_0}+2\pi n$, $n=0,\pm1,\pm2,\ldots$. Note that

$$
\sum_{t=1}^{T}\sum_{s=1}^{T}\tilde{\gamma}(s-t)e^{-i(\omega_j s-\omega_{j_0}t)}=\sum_{t=1}^{T}\sum_{k=t-1}^{t-T}\tilde{\gamma}(k)e^{-i(\omega_j-\omega_{j_0})t+i\omega_j k}=
$$

$$
\underbrace{\tilde{\gamma}(0)\sum_{t=1}^{T}e^{-i(\omega_j-\omega_{j_0})t}}_{0}+\sum_{k=1}^{T-1}\tilde{\gamma}(k)e^{i\omega_j k}\sum_{t=k+1}^{T}e^{-i(\omega_j-\omega_{j_0})t}+\sum_{k=1-T}^{T-1}\tilde{\gamma}(k)e^{i\omega_j k}\sum_{t=1}^{T+k}e^{-i(\omega_j-\omega_{j_0})t}=
$$

$$
=\frac{e^{-i(\omega_j-\omega_{j_0})}}{1-e^{-i(\omega_j-\omega_{j_0})}}\left(\sum_{k=1}^{T-1}\tilde{\gamma}(k)e^{i\omega_j k}(e^{-i(\omega_j-\omega_{j_0})k}-1)+\sum_{k=1-T}^{-1}\tilde{\gamma}(k)e^{i\omega_j k}(1-e^{-i(\omega_j-\omega_{j_0})k})\right).
$$

After replacing $\tilde{\gamma}(k)$ by $\tilde{\gamma}(k+T)$ in the second sum, we obtain the result $\sum_{t=1}^{T}\sum_{s=1}^{T}\tilde{\gamma}(s-t)e^{-i(\omega s-\omega_0 t)}=0$. Therefore,

$$
\sum_{t=1}^{T}e^{i\omega_{j_0}t}T_y(t)=\frac{s_{u\varepsilon}(\omega_{j_0};\Gamma_e)}{\det(s(\omega_{j_0};\Gamma_e))}\frac{\sum_{t=1}^{T}\sum_{s=1}^{T}\tilde{A}_{yy}^{-1}(t,s)e^{-i\omega_{j_0}(s-t)}}{2\pi T}.
$$

https://doi.org/10.1017/S0266466620000079 Published online by Cambridge University Press

Note also that

$$\sum_{t=1}^{T}\sum_{s=1}^{T}\tilde{A}_{yy}^{-1}(t,s)e^{-i\omega_{j_0}(s-t)} = \sum_{t=1}^{T}\sum_{s=1}^{T}\tilde{\gamma}(s-t)e^{-i\omega_{j_0}(s-t)} = T\sum_{k=0}^{T-1}\tilde{\gamma}(k)e^{-i\omega_{j_0}k}.$$

The latter sum $\lambda_{j_0}(\tilde{A}_{yy}^{-1}) := \sum_{k=0}^{T-1}\tilde{\gamma}(k)e^{-i\omega_{j_0}k}$ is the eigenvalue of $\tilde{A}_{yy}^{-1}(t,s)$ that corresponds to the eigenvector $(1, e^{-i\omega_{j_0}}, \ldots, e^{-i\omega_{j_0}(T-1)})'$ (see Davis, 1979). Because $\tilde{A}_{yy}$ is also circulant, the same result holds:

$$\sum_{t=1}^{T}\sum_{s=1}^{T}\tilde{A}_{yy}(t,s)e^{-i\omega_{j_0}(s-t)} = T\lambda_{j_0}(\tilde{A}_{yy}).$$

From the known relation between the eigenvalues of a matrix and its inverse, $\lambda_{j_0}(\tilde{A}_{yy}) = \lambda_{j_0}(\tilde{A}_{yy}^{-1})^{-1}$, we obtain

$$\sum_{t=1}^{T}e^{i\omega_{j_0}t}T_y(t) = \frac{s_{u\varepsilon}(\omega_{j_0};\Gamma_e)}{\det(s(\omega_{j_0};\Gamma_e))}\frac{T/2\pi}{\sum_{t=1}^{T}\sum_{s=1}^{T}\tilde{A}_{yy}(t,s)e^{-i\omega_{j_0}(s-t)}}.$$

Lastly, substituting the definition of $\tilde{A}_{yy}(t,s)$, we obtain

$$\sum_{t=1}^{T}\sum_{s=1}^{T}\tilde{A}_{yy}(t,s)e^{-i\omega_{j_0}(t-s)} = \sum_{t=1}^{T}\sum_{s=1}^{T}\tilde{A}_{yy}(t,s)e^{-i\omega_{j_0}(s-t)} = \frac{T}{2\pi}\frac{s_{uu}(\omega_{j_0};\Gamma_e)}{\det(s(\omega_{j_0};\Gamma_e))},$$

and $\sum_{t=1}^{T}e^{i\omega_{j_0}t}T_y(t) = \frac{s_{u\varepsilon}(\omega_{j_0};\Gamma_e)}{s_{uu}(\omega_{j_0};\Gamma_e)}$. The same result holds for $\sum_{t=1}^{T}e^{-i\omega_{j_0}t}T_y(t)$. Therefore, $\sum_{t=1}^{T}e^{i\omega_{j_0}t}\delta_y(t) = \frac{s_{u\varepsilon}(\omega_{j_0};\Gamma_e)}{s_{uu}(\omega_{j_0};\Gamma_e)}$. □

B. PROOF OF PROPOSITION 1

Proof of Proposition 1. First, we review the decomposition of the Whittle likelihood into the joint probability densities of Fourier transformations. Note that for $j = 1, \ldots, T-1$, the sum $\sum_{t=1}^{T}\exp\{-i\omega_j t\}$ is zero. Therefore, for any vector process $\mathcal{Z}_T = \{Z_t, t = 1, \ldots, T\}$, $d_\mu(\omega_j;\mathcal{Z}_T) = d(\omega_j;\mathcal{Z}_T)$, where $d(\omega_j;\mathcal{Z}_T)$ is the Fourier transformation of $\mathcal{Z}_T$ and $d_\mu(\omega_j;\mathcal{Z}_T)$ is the Fourier transformation of the recentered $\mathcal{Z}_T - \mu$ series. Also note that $d(\omega_{T-j};\mathcal{Z}_T)$ is a complex conjugate of $d(\omega_j;\mathcal{Z}_T)$. Therefore, if T is even, then $d(\omega_{T/2};\mathcal{Z}_T)$ is real-valued. Finally, $d_\mu(\omega_T;\mathcal{Z}_T) = d_\mu(\omega_0;\mathcal{Z}_T) = d(\omega_0;\mathcal{Z}_T) - (2\pi)^{-1/2}T^{1/2}\mu$.

We represent the formula of the log Whittle likelihood as a sum of log densities of $d(\omega_j;\mathcal{Z}_T)$ for $j = 0, \ldots, \lfloor T/2\rfloor$. We start by combining the terms with $d(\omega_j;\mathcal{Z}_T)$ and $d(\omega_{T-j};\mathcal{Z}_T)$ for $j = 1, \ldots, \lfloor (T-1)/2\rfloor$ into one element

$$\begin{aligned}\ln l_j(d(\omega_j;\mathcal{Z}_T)) &= -2\ln\pi - \frac{1}{2}\ln\left[\det|s(\omega_j;\Gamma)|\right] - \frac{1}{2}\ln\left[\det|s(\omega_{T-j};\Gamma)|\right]\\ &\quad - \frac{1}{2}\mathrm{tr}\left[s(\omega_j;\Gamma)^{-1}d(\omega_j;\mathcal{Z}_T)d^*(\omega_j;\mathcal{Z}_T) + s(\omega_{T-j};\Gamma)^{-1}d(\omega_{T-j};\mathcal{Z}_T)d^*(\omega_{T-j};\mathcal{Z}_T)\right]\\ &= -2\ln\pi - \ln\left[\det|s(\omega_j;\mathcal{Z}_T)|\right] - \mathrm{tr}\left[s(\omega_j;\Gamma)^{-1}d(\omega_j;\mathcal{Z}_T)d^*(\omega_j;\mathcal{Z}_T)\right],\end{aligned}$$

https://doi.org/10.1017/S0266466620000079 Published online by Cambridge University Press

which corresponds to the density of a complex normal variable $N^C(0, s(\omega_j;\Gamma))$ (e.g., Brillinger 1975, p. 89). If T is even, then for the median frequency with $j = T/2$, the observation $d(\omega_{T/2};\mathcal{Z}_T)$ and spectrum $s(\omega_{T/2};\Gamma)$ are real-valued. The corresponding term in the Whittle approximation

$$\ln l_j(d(\omega_{T/2};\mathcal{Z}_T)) = -\ln 2\pi - 1/2\ln\left[\det|s(\omega_{T/2};\Gamma)|\right] - \frac{1}{2}\left[d'(\omega_{T/2};\mathcal{Z}_T)s(\omega_{T/2};\Gamma)^{-1}d(\omega_{T/2};\mathcal{Z}_T)\right]$$

is the density of a normal variable $N(0, s(\omega_{T/2};\mathcal{Z}_T))$. Similarly, for the zero-frequency $\omega_0 = 0$, the observation $d(\omega_0;\mathcal{Z}_T) = d(0;\mathcal{Z}_T)$ and spectrum $s(\omega_0;\Gamma) = s(0;\Gamma)$ are real-valued. The corresponding term

$$\ln l_j(d(\omega_0;\mathcal{Z}_T)) = -\ln 2\pi - 1/2\ln[\det|s(0;\Gamma)|] - \frac{1}{2}\left[(d(0;\mathcal{Z}_T) - \mu(2\pi)^{-1/2}T^{1/2})'s(0;\Gamma)^{-1} \times (d(0;\mathcal{Z}_T) - \mu(2\pi)^{-1/2}T^{1/2})\right]$$

is a log pdf of a normal variable $N(\mu(2\pi)^{-1/2}T^{1/2}, s(0;\Gamma))$.

We consider $Z_t = (x_{t-1}, y_t')$ with covariance structure $\Gamma = \Gamma_{x,y}(\Gamma_e, \beta_x, \rho)$ and $\mu = (\mu_x, \mu_y')$. The conditional distributions of $d(\omega_j; Y)$ that follow from the Whittle approximation are obtained by dividing the joint densities into the conditional and marginal parts. We use the identities $\det(s(\omega_j;\Gamma_{x,y}))\left[s_{xx}(\omega_j;\Gamma_{x,y})\right]^{-1} = \det(s(\omega_j;\Gamma_e))\left[s_{uu}(\omega_j;\Gamma_e)\right]^{-1}$ and $s_{yx}(\omega_j;\Gamma_{x,y})\left[s_{xx}(\omega_j;\Gamma_{x,y})\right]^{-1} = \beta_x + s_{\varepsilon u}(\omega_j;\Gamma_e)(\exp\{i\omega_j\} - \rho)$ to obtain the final formulas. In particular, under the Whittle approximation (in its regular form), $d(\omega_j; Y)$, $0 < j < T/2$ are complex-normal:

$$N^C\left(\left[\beta_x + s_{\varepsilon u}(\omega_j;\Gamma_e)(\exp\{i\omega_j\} - \rho)\right]d(\omega_j; X), 2\frac{\det[s(\omega_j;\Gamma_e)]}{s_{uu}(\omega_j;\Gamma_e)}\right).$$

For the modified Whittle approximation, the above formula applies to the conditional densities $d(\omega_j; Y^W)$. We note that $d(\omega_j; Y) = d(\omega_j; Y^W) - (2\pi T)^{-1/2}s_{\varepsilon u}(\omega_j;\Gamma_e)s_{uu}(\omega_j;\Gamma_e)^{-1}(x_0 - x_T)$ as follows from Lemma 2. Therefore, the adjusted Whittle approximation corresponds to the conditional distributions of $d_y(\omega_j)$ that are complex-normal:

$$N^C\left(\left[\beta_x + s_{\varepsilon u}(\omega_j;\Gamma_e)(\exp\{i\omega_j\} - \rho)\right]d(\omega_j; X) - (2\pi T)^{-1/2}\frac{s_{\varepsilon u}(\omega_j;\Gamma_e)}{s_{uu}(\omega_j;\Gamma_e)}(x_0 - x_T), 2\frac{\det[s(\omega_j;\Gamma_e)]}{s_{uu}(\omega_j;\Gamma_e)}\right).$$

The cases with $j = T/2$ and $j = 0$ are obtained analogously. □

C. PROOF OF THEOREMS 1 and 2

The proof of Theorem 1 relies on the results by Phillips (1987, 1988) and the related statements listed below. All of the following weak convergence results hold jointly. Any set of conditions that insure the validity of the FCLT can replace the assumptions in Lemma 3. For instance, these results hold under Assumption A and its restriction to the Gaussian

https://doi.org/10.1017/S0266466620000079 Published online by Cambridge University Press

case—Assumption A'. We use brackets $\lfloor\ \rfloor$ to denote rounding toward minus infinity. Stronger results for Fourier transformations are considered by Phillips (2007).

LEMMA 3.

(1) Under Assumption A, Phillips (1987, Lem. 1, and 1988, Lem. 3.1) proved that the process $\frac{1}{\sqrt{\psi_{uu}}}\frac{x_{\lfloor\tau T\rfloor}}{\sqrt{T}}$, $\tau\in[0,1]$ *weakly converges to the O–U process* $J_c(\tau)$ *starting from* $J_c(0)=0$. *The O–U process follows the dynamics* $dJ_c(\tau)=cd\tau+dW^u(\tau)$, *where* $W^u(\tau)$ *is a standard Brownian motion. By the FCLT,* $\frac{1}{\sqrt{\psi_{uu}}}\frac{\sum_{t=1}^{\lfloor\tau T\rfloor}u_t}{\sqrt{T}}\Rightarrow W^u(\tau)$ *and* $\frac{1}{\sqrt{\psi_{\varepsilon\varepsilon}}}\frac{\sum_{t=1}^{\lfloor\tau T\rfloor}\varepsilon_t}{\sqrt{T}}\Rightarrow W^{\varepsilon}(\tau)$, *where the vector process* $(W^u(\tau),W^{\varepsilon}(\tau))$ *is a two-dimensional Brownian motion with correlation* $r=\psi_{\varepsilon u}(\psi_{\varepsilon\varepsilon}\psi_{uu})^{-1/2}$ *and unit marginal variances. Note that the same result holds for the demeaned process* $\frac{1}{\sqrt{\psi_{uu}}}\frac{(x_{\lfloor\tau T\rfloor}-\mu_x)}{\sqrt{T}}\Rightarrow J_c(\tau)$ *as* $\mu_x/\sqrt{T}\to 0$.

(2) From CMT and result (1), it follows that the Fourier transformation of $X=(x_0,\ldots,x_{T-1})$ *calculated at frequency* $\omega_{j_0}=2\pi j_0/T$, *that is,* $d_x(\omega_{j_0})=\frac{1}{\sqrt{2\pi T}}\sum_{t=1}^{T}x_{t-1}e^{-i\omega_{j_0}t}$, *has the following limit,*

$$\frac{d_x(\omega_{j_0})}{T}=\frac{1}{\sqrt{2\pi}}\int_{\tau=0}^{1}\frac{x_{\lfloor\tau T\rfloor}}{\sqrt{T}}e^{-i\frac{2\pi j_0(\lfloor\tau T\rfloor+1)}{T}}d\tau\Rightarrow\sqrt{\frac{\psi_{uu}}{2\pi}}\int_{\tau=0}^{1}J_c(\tau)e^{-i2\pi j_0\tau}d\tau,$$

where the convergence holds jointly across $j_0=1,\ldots,q$.

(3) Define $S_{\varepsilon,t}=\varepsilon_t+\cdots+\varepsilon_1$. *Then, the Fourier transformation of* ε_t *has the following representation:* $\frac{\sum_{t=1}^{T}\varepsilon_t e^{-i\omega_{j_0}t}}{\sqrt{2\pi T}}=\frac{T(e^{i\omega_{j_0}}-1)}{\sqrt{2\pi}}\sum_{t=1}^{T}\frac{S_{\varepsilon,t-1}}{\sqrt{T}}e^{-2\pi j_0 i\frac{t}{T}}\frac{1}{T}+\frac{1}{\sqrt{2\pi}}\frac{S_{\varepsilon,T}}{\sqrt{T}}$. *Because the FCLT holds for the partial sums* $S_{\varepsilon,t}$ *and* $\lim_{T\to\infty}T(e^{i\omega_{j_0}}-1)=i2\pi j_0$, *by applying the CMT we obtain* $\frac{\sum_{t=1}^{T}\varepsilon_t e^{-i\omega_{j_0}t}}{\sqrt{2\pi T}}\Rightarrow j_0 i\sqrt{2\pi\psi_{\varepsilon\varepsilon}}\int_{\tau=0}^{1}W^{\varepsilon}(\tau)e^{-2\pi j_0 i\tau}d\tau+\sqrt{\frac{\psi_{\varepsilon\varepsilon}}{2\pi}}W^{\varepsilon}(1)$, *and, therefore,* $\frac{\sum_{t=1}^{T}\varepsilon_t e^{-i\omega_{j_0}t}}{\sqrt{2\pi T}}\Rightarrow\sqrt{\frac{\psi_{\varepsilon\varepsilon}}{2\pi}}\int_{\tau=0}^{1}e^{-2\pi j_0 i\tau}dW^{\varepsilon}(\tau)$, *where the convergence holds jointly across* $d_{\varepsilon}(\omega_{j_0})$, $j_0=1,\ldots,q$.

(4) For any two processes, $(w_{1,t},w_{2,t})$, *which satisfy the conditions for the bivariate FCLT (e.g., Assumption A) partial sums* $\frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor\tau T\rfloor}w_{1,t}$ *and* $\frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor\tau T\rfloor}w_{2,t}$ *converge jointly with the Fourier transformations* $\frac{\sum_{t=1}^{\lfloor\tau T\rfloor}w_{1,t}e^{-i\omega_{j_0}t}}{\sqrt{2\pi T}}$ *and* $\frac{\sum_{t=1}^{\lfloor\tau T\rfloor}w_{2,t}e^{-i\omega_{j_0}t}}{2\pi\sqrt{T}}$ *to the processes* $W_{w_1}(\tau)$, $W_{w_2}(\tau)$, $\frac{1}{\sqrt{2\pi}}\int_{\tau=0}^{1}e^{-2\pi j_0 i\tau}dW_{w_1}(\tau)$, *and* $\frac{1}{\sqrt{2\pi}}\int_{\tau=0}^{1}e^{-2\pi j_0 i\tau}dW_{w_2}(\tau)$, *respectively, where* $W_w(\tau)=(W_{w_1}(\tau),W_{w_2}(\tau))'$ *is a Brownian motion with a variance equal to the long-run variance of* $(w_{1,t},w_{2,t})$. *The proof is obtained by using the representation for the Fourier transformations in result (3).*

Proof of Theorem 1. As follows from Lemma 1 and Proposition 1, the following is the asymptotically equivalent joint likelihood of X and $d(\omega_j;Y)$, $j=1,\ldots,q$:

$$\ln\mathcal{L}_{W|G}(X,\{d(\omega_j;Y)\}_{j=1}^{q})=$$

$$\sum_{j=1}^{q}\ln l_j(d(\omega_j)|X)-\frac{T-1}{2}\ln(2\pi)-\frac{T-1}{2}\ln\psi_{uu}-\frac{1}{2}\sum_{t=1}^{T-1}\frac{(x_t-\mu_x-\rho(x_{t-1}-\mu_x)^2)}{\psi_{uu}},$$

https://doi.org/10.1017/S0266466620000079 Published online by Cambridge University Press

where $\ln l_j(d(\omega_j;Y)|X) = -\ln 2\pi - \ln\left[\frac{\det[s(\omega_j;\Gamma_e)]}{s_{uu}(\omega_j;\Gamma_e)}\right] - \left(\frac{\det[s(\omega_j;\Gamma_e)]}{s_{uu}(\omega_j;\Gamma_e)}\right)^{-1}|d(\omega_j;Y) - \mu_j|^2$ and $\mu_j = \frac{s_{yx}(\omega_j;\Gamma_{x,y})}{s_{xx}(\omega_j;\Gamma_{x,y})}d(\omega_j;X) - \frac{s_{\varepsilon u}(\omega_j;\Gamma_e)}{s_{uu}(\omega_j;\Gamma_e)}\frac{x_0-x_T}{\sqrt{2\pi T}} = \left(\beta_x + \frac{s_{\varepsilon u}(\omega_j;\Gamma_e)(e^{i\omega}-\rho)}{s_{uu}(\omega_j;\Gamma_e)}\right)d(\omega_j;X) - \frac{s_{\varepsilon u}(\omega_j;\Gamma_e)}{s_{uu}(\omega_j;\Gamma_e)}\frac{x_0-x_T}{\sqrt{2\pi T}} = \beta_x d(\omega_j;X) + \frac{s_{\varepsilon u}(\omega_j;\Gamma_e)}{s_{uu}(\omega_j;\Gamma_e)}(d(\omega_j;X_+) - \rho\, d(\omega_j;X))$, with $X_+ = (x_1,\ldots,x_T')$.

We re-parameterize $\beta_x = \beta_0 + \frac{1}{T}\frac{\psi_{\varepsilon\varepsilon|u}^{1/2}}{\psi_{uu}^{1/2}}b$ and $\rho = 1 + c/T$, and we express the likelihood in terms of new parameters (b,c) as follows: $\ln\mathcal{L}_{W|G}(X,\{d(\omega_j;Y)\}_{j=1}^q|b,c) = \ln\mathcal{L}_{W|G}(X,\{d(\omega_j;Y)\}_{j=1}^q|0,0) + bR_\beta + cR_\rho - \frac{1}{2}(b - \frac{r}{\sqrt{1-r^2}}c)^2R_{\beta\beta} - \frac{1}{2}c^2R_{\rho\rho} + \epsilon_{\tilde{L}}$, where $\ln\mathcal{L}_{W|G}(X,\{d(\omega_j;Y)\}_{j=1}^q|0,0)$ is the likelihood calculated at $b=0$ and $c=0$ and sufficient statistics are

$$R_\beta = \frac{4\pi}{T}\frac{1}{\sqrt{\psi_{\varepsilon\varepsilon|u}\psi_{uu}}}\mathrm{Re}\sum_{j=1}^q\left[d(\omega_j;X)\left[d(\omega_j;Y) - \beta_0 d(\omega;X) - \frac{r\psi_{\varepsilon\varepsilon}^{1/2}}{\psi_{uu}^{1/2}}(d(\omega;X_+) - d(\omega_j;X))\right]^*\right],$$

$$R_\rho = \psi_{uu}^{-1}T^{-1}\sum_{t=1}^{T-1}(x_{t-1}-\mu_x)(x_t - x_{t-1}) - \frac{r}{\sqrt{1-r^2}}R_\beta,$$

$$R_{\beta\beta} = \frac{4\pi}{T^2\psi_{uu}}\sum_{j=1}^q d(\omega_j;X)d(\omega_j;X)^*,$$

$$R_{\rho\rho} = \frac{1}{T^2\psi_{uu}}\sum_{t=1}^T(x_{t-1}-\mu_x)^2.$$

The component $\epsilon_{\tilde{L}}$ includes errors resulting from the difference between $s(\omega_j;\Gamma_e)$, $j = 1,\ldots,q$, and $s(0;\Gamma_e) = \psi/(2\pi)$. In particular, $\epsilon_{\tilde{L}} = b\epsilon_{r_\beta} + c\epsilon_{r_\rho} - \epsilon_{r_{\beta\beta}}$, where

$$\epsilon_{r_\beta} = \frac{2}{T}\mathrm{Re}\sum_{j=1}^q\left[\left\{\frac{s_{uu}(\omega_j;\Gamma_e)}{\det[s(\omega_j;\Gamma_e)]} - \frac{s_{uu}(0;\Gamma_e)}{\det[s(0;\Gamma_e)]}\right\}\sqrt{\frac{\psi_{\varepsilon\varepsilon|u}}{\psi_{uu}}}d(\omega_j;X)\left[d(\omega_j;Y) - \beta_0 d(\omega_j;X)\right]^*\right]$$
$$-\frac{2}{T}\mathrm{Re}\sum_{j=1}^q\left[\left\{\frac{s_{u\varepsilon}(\omega_j;\Gamma_e)}{\det[s(\omega_j;\Gamma_e)]} - \frac{s_{u\varepsilon}(0;\Gamma_e)}{\det[s(0;\Gamma_e)]}\right\}\sqrt{\frac{\psi_{\varepsilon\varepsilon|u}}{\psi_{uu}}}d(\omega_j;X)(d^*(\omega_j;X_+) - d^*(\omega_j;X))\right]$$

$$\epsilon_{r_\rho} = -\frac{2}{T}\mathrm{Re}\sum_{j=1}^q\left[\left\{\frac{s_{\varepsilon u}(\omega_j;\Gamma_e)}{\det[s(\omega_j;\Gamma_e)]} - \frac{s_{\varepsilon u}(0;\Gamma_e)}{\det[s(0;\Gamma_e)]}\right\}d(\omega_j;X)\left[d(\omega_j;Y) - \beta_0 d(\omega_j;X)\right]^*\right]$$
$$+\frac{2}{T}\mathrm{Re}\sum_{j=1}^q\left[\left\{\frac{|s_{\varepsilon u}(\omega_j;\Gamma_e)|^2}{\det[s(\omega_j;\Gamma_e)]s_{uu}(\omega_j;\Gamma_e)} - \frac{|s_{\varepsilon u}(0;\Gamma_e)|^2}{\det[s(0;\Gamma_e)]s_{uu}(0;\Gamma_e)}\right\}\right.$$
$$\left.d(\omega_j;X)(d^*(\omega_j;X_+) - d^*(\omega_j;X))\right]$$

https://doi.org/10.1017/S0266466620000079 Published online by Cambridge University Press

$$\epsilon_{r_{\beta\beta}} = \frac{2}{T^2}\sum_{j=1}^{q}\frac{\psi_{\varepsilon\varepsilon|u}}{\psi_{uu}}|d(\omega_j;X)|^2 \times\left(\frac{s_{uu}(\omega_j;\Gamma_e)}{\det[s(\omega_j;\Gamma_e)]}\left|b-c\frac{s_{\varepsilon u}(\omega_j;\Gamma_e)}{s_{uu}(\omega_j;\Gamma_e)}\sqrt{\frac{\psi_{uu}}{\psi_{\varepsilon\varepsilon|u}}}\right|^2 -\frac{s_{uu}(0;\Gamma_e)}{\det[s(0;\Gamma_e)]}\left|b-c\frac{s_{\varepsilon u}(0;\Gamma_e)}{s_{uu}(0;\Gamma_e)}\sqrt{\frac{\psi_{uu}}{\psi_{\varepsilon\varepsilon|u}}}\right|^2\right)$$

Note that under Assumption A', because strong mixing coefficients satisfy the same condition as the maximal correlation for stationary Gaussian processes, the spectrum $s(\omega;\Gamma_e)$ is continuous at zero and converges to $s(0;\Gamma_e)$. In fact by Thm. 8 (p. 181) of Ibragimov and Rozanov (1978), the spectrum has at least one derivative and, therefore, $s(2\pi j/T;\Gamma_e) - s(0;\Gamma_e)$ converges to zero at rate T^{-1} for a fixed j. Therefore, the error $\epsilon_{\tilde{L}}$ does not appear in the limiting distribution of the likelihood process $\ln\mathcal{L}_{W|G}(X,\{d(\omega_j,Y)\}_{j=1}^{q}|b,c)$ unless the principal part $bR_\beta + cR_\rho - \frac{1}{2}(b-\frac{r}{1-r^2}c)^2R_{\beta\beta} - \frac{1}{2}c^2R_{\rho\rho}$ is of order $1/T$ or smaller relative to its elements separately. We now show that this is not the case.

Using the CMT and the results listed in Lemma 3, we find the weak limits of the sufficient statistics $R = (R_\beta, R_c, R_{\beta\beta}, R'_{cc})$. First,

$$R_\beta = (\psi_{uu}\psi_{\varepsilon\varepsilon|u})^{-1/2}\frac{4\pi}{T}\,\mathrm{Re}\left[\sum_{j=1}^{q}d(\omega_j;X)\left(b\sqrt{\frac{\psi_{\varepsilon\varepsilon|u}}{\psi_{uu}}}\frac{d(\omega_j;X)}{T}+\sqrt{\psi_{\varepsilon\varepsilon|u}}d(\omega_j;z) -c\frac{r\psi_{\varepsilon\varepsilon}^{1/2}}{\psi_{uu}^{1/2}}\frac{d(\omega_j;X)}{T}\right)^*\right],$$

where $d(\omega_j;z)$ is the Fourier transformation of $z_t := (\psi_{\varepsilon\varepsilon|u})^{-1/2}(\varepsilon_t - \frac{r\psi_{\varepsilon\varepsilon}^{1/2}}{\psi_{uu}^{1/2}}u_t)$. Note that the long-run variance of z_t is 1 and, therefore, as follows from result (3) in Lemma 3, its Fourier transformation has the limit $d(\omega_j;z) \Rightarrow \frac{1}{\sqrt{2\pi}}\sum_{\tau=0}^{1}e^{-2\pi ij\tau}dW^z(\tau)$, where $W^z(\tau)$ is a Brownian process. As follows from result (2) in Lemma 3, the Fourier transformation of x_t has the limit $\frac{d(\omega_j;X)}{T} \Rightarrow \sqrt{\frac{\psi_{uu}}{2\pi}}\sum_{\tau=0}^{1}e^{-2\pi ij\tau}J_c(\tau)d\tau$. Furthermore, because the long-run covariance between z_t and u_t is zero, then from results (1) and (4) in Lemma 3, it follows that $J_c(\tau)$ and $W^z(\tau)$ are independent. Thus, we obtain the limit $R_\beta \Rightarrow \mathcal{R}_\beta$, where

$$\mathcal{R}_\beta = 2\,\mathrm{Re}\left[\sum_{j=1}^{q}\left(\int_0^1 J_c(\tau)e^{-2\pi ij\tau}d\tau\right)\left(\int_0^1 e^{-2\pi ij\tau}dW^z(\tau) +(b-\frac{cr}{\sqrt{1-r^2}})\int_0^1 J_c(\tau)e^{-2\pi ij\tau}d\tau\right)^*\right].$$

Second, combining the latter result and the result by Jansson and Moreira (2006, Lem. 3), we obtain

$$R_\rho \Rightarrow \mathcal{R}_\rho = \int_0^1 J_c(\tau)dJ_c(\tau) - r(1-r^2)^{-1/2}\mathcal{R}_\beta.$$

https://doi.org/10.1017/S0266466620000079 Published online by Cambridge University Press

Finally, the limits $R_{\rho\rho} \Rightarrow \mathcal{R}_{\rho\rho} = \int_0^1 (J_c(\tau))^2 d\tau$ and $R_{\beta\beta} \Rightarrow \mathcal{R}_{\beta\beta} = 2\sum_{j=1}^{q} |\sum_{\tau=0}^{1} e^{-2\pi ij\tau} J_c(\tau) d\tau|^2$ follow from results (1) and (2) in Lemma 3, respectively.

Therefore the likelihood ratio process $\ln \mathcal{L}_{W|G}(X, \{d(\omega_j; Y)\}_{j=1}^{q} | b, c) - \ln \mathcal{L}_{W|G}(X, \{d(\omega_j; Y)\}_{j=1}^{q} | b = 0, c = 0)$ will have $b\mathcal{R}_\beta + c\mathcal{R}_\rho - \frac{1}{2}\left(b - c\frac{r}{\sqrt{1-r^2}}\right)^2 \mathcal{R}_{\beta\beta} - \frac{1}{2}c^2\mathcal{R}_{\rho\rho}$ as its limiting distribution. By arguments of Elliott et al. (2015), this asymptotic ratio characterizes the asymptotic behavior of all possible tests for the null hypothesis $H_0 : b = 0, c \in \Theta_c$ against the alternative $H_a : b \in \mathrm{B}, c \in \Theta_c$ under the restrictions imposed in this article. □

Proof of Theorem 2. Note that the convergence results in Theorem 1 rely mostly on Lemma 3, which can be shown both under Assumptions A' and A. The only exception is the convergence of R_ρ that relies on the i.i.d. condition; see Jansson and Moreira (2006). However, note that $\int_0^1 J_c(\tau) dJ_c(\tau) = \frac{1}{2}(J_c(1)^2 - 1)$, and, therefore, the limit $\mathcal{R}_\rho$ can be represented as follows

$$\mathcal{R}_\rho = \frac{1}{2}(J_c(1)^2 - 1) - r(1-r^2)^{-1/2}\mathcal{R}_\beta.$$

By result (1) of Lemma 3, x_T^2/T converges to $\psi_{uu} J_c(1)^2$. Also, $\frac{1}{T}\sum_{t=1}^{T}(x_t - x_{t-1})^2$ converges to σ_{uu} in probability. Therefore, R'_ρ converges to $\mathcal{R}_\rho$ if ψ is known.

Finally, the estimated elements of the long-run variance $\hat{\psi}$ introduce errors that are small in probability. For example, consider $\mathrm{R}'_\beta = \sqrt{\frac{\psi_{uu}\psi_{\epsilon\epsilon}(1-r^2)}{\hat{\psi}_{uu}\hat{\psi}_{\epsilon\epsilon}(1-\hat{r}^2)}}(\mathrm{R}_\beta + \epsilon_{r'_\beta})$ that includes the error

$$\epsilon_{r'_\beta} = 4\pi \frac{1}{\sqrt{\psi_{\varepsilon\varepsilon|u}\psi_{uu}}}\left(\frac{r\psi_{\varepsilon\varepsilon}^{1/2}}{\psi_{uu}^{1/2}} - \frac{\hat{r}\hat{\psi}_{\varepsilon\varepsilon}^{1/2}}{\hat{\psi}_{uu}^{1/2}}\right) \mathrm{Re}\sum_{j=1}^{q}\left(\frac{d(\omega_j; X)}{T}(d(\omega_j; X_+) - d(\omega_j; X))\right).$$

Note that $T^{-1}d(\omega_j; X)(d(\omega_j; X_+) - d(\omega_j; X)) = T^{-1}d(\omega_j; X)(d(\omega_j; U) + c\, d(\omega_j; X)/T)$.The term $d(\omega_j; U)$ is the Fourier transformation of u_t. Therefore, as follows from Lemma 3 result (3), $d(\omega_j; U)$ is bounded in probability. The same holds for the term $d(\omega_j; X)/T$ by result (2) in Lemma 3. Therefore, $T^{-1}d(\omega_j; T)(d(\omega_j; X_+) - d(\omega_j; X)) \sim O_p(1)$. By assumptions, $\frac{\hat{r}\hat{\psi}_{\varepsilon\varepsilon}^{1/2}}{\hat{\psi}_{uu}^{1/2}}$ is a consistent estimate of $\frac{r\psi_{\varepsilon\varepsilon}^{1/2}}{\psi_{uu}^{1/2}}$. Therefore, the error $\epsilon_{r'_\beta}$ converges to zero. Similarly, $\hat{\psi}_{uu}\hat{\psi}_{\epsilon\epsilon}(1-\hat{r}^2)$ is a consistent estimate of $\psi_{uu}\psi_{\epsilon\epsilon}(1-r^2)$. Therefore, the statistic R'_β has the same limit as R_β. □

D. PROOF OF THEOREMS 3–5

Proof. The proof follows from (ii.b) result in Prop. 5 by Elliott and Müller (2014) for the tests in the models with structural breaks. For any small positive number $0 < \Delta_\alpha < 1 - \alpha$, consider a sequence $(T_k^*, c_{T_k^*})$ such that $T_k^* \to \infty$ as $k \to \infty$ and

$$\limsup_{T\to\infty} \sup_{c\in\Theta_c} \mathbb{E}[D_{ns}(\mathrm{R}_T(c_T = c, b_T = 0))] \le \lim_{k\to\infty} \mathbb{E}\left[D_{ns}(\mathrm{R}_{T_k^*}(c_{T_k^*}, b_{T_k^*} = 0))\right] + \Delta_\alpha.$$

https://doi.org/10.1017/S0266466620000079 Published online by Cambridge University Press

Because Θ_c is closed, the sequence $c_{T_k^*}$ has a subsequence c_{T_k} with $T_k \to \infty$ as $k \to \infty$, such that it is either (a) converging to a finite $c^* \in \Theta_c$ or (b) it is converging to $-\infty$. In both cases it holds that $\lim_{k\to\infty} \mathbb{E}\left[D_{ns}(\mathrm{R}_{T_k*}(c_{T_k^*}, b_{T_k^*} = 0))\right] = \lim_{k\to\infty} \mathbb{E}\left[D_{ns}(\mathrm{R}_{T_k}(c_{T_k}, b_{T_k} = 0))\right]$.

In case (a), decompose $\lim_{k\to\infty} \mathbb{E}\left[D_{ns}(\mathrm{R}_{T_k}(c_{T_k}, b_{T_k} = 0))\right]$ into the sum:

$$\lim_{k\to\infty} \mathbb{E}\left[D_{ns}(\mathrm{R}_{T_k}(c^*, 0))\right] + \left(\lim_{k\to\infty} \mathbb{E}\left[D_{ns}(\mathrm{R}_{T_k}(c_{T_k}, 0))\right] - \lim_{k\to\infty} \mathbb{E}\left[D_{ns}(\mathrm{R}_{T_k}(c^*, 0))\right]\right).$$

For the first summand,

$$\lim_{k\to\infty} \mathbb{E}\left[D_{ns}(\mathrm{R}_{T_k})(c^*, 0)\right] \le \sup_{c\in\Theta_c} \limsup_{T\to\infty} \mathbb{E}[D_{ns}(\mathrm{R}_T(c_T = c, b_T = 0))] \le \alpha.$$

The second summand is zero by the CMT, because R_{T_k} weakly converges to $\mathcal{R}$ along any converging sequence $c_{T_k} \to c^* \in \Theta_c$.

For case (b), note that $D_{ns}(\mathrm{R}_{T_k}) = D_n(\mathrm{R}_{T_k})\mathbb{I}(C(\mathrm{R}_{T_k}) \ge K_c) + D_s(\mathrm{R}_{T_k})\mathbb{I}(C(\mathrm{R}_{T_k}) < K_c)$. Therefore,

$$D_{ns}(\mathrm{R}_{T_k}) \le \mathbb{I}(C(\mathrm{R}_{T_k}) \ge K_c) + D_s(\mathrm{R}_{T_k}).$$

Therefore,

$$\lim_{k\to\infty} \mathbb{E}\left[D_{ns}(\mathrm{R}_{T_k}(c_{T_k}, 0))\right] \le \lim_{k\to\infty} \mathbb{E}\left[D_s(\mathrm{R}_{T_k}(c_{T_k}, 0))\right] + \lim_{k\to\infty} \Pr(C(\mathrm{R}_{T_k}(c_{T_k}, 0)) < K_c).$$

The second element $\Pr(C(\mathrm{R}_{T_k}(c_{T_k}, 0)) < K_c)$ converges to zero and the first element converges to $\Pr(\xi(0) > K_{\alpha,s}) \le \alpha$ along any sequence $c_{T_k} \to -\infty$. We conclude that in both cases, $\limsup_{T\to\infty} \sup_{c\in\Theta_c} \mathbb{E}[D_{ns}(\mathrm{R}_T(c_T = c, b_T = 0))] \le \alpha + \Delta_\alpha$. Because this inequality holds for any small $\Delta_\alpha > 0$, it is also true that $\limsup_{T\to\infty} \sup_{c\in\Theta_c} \mathbb{E}[D_{ns}(\mathrm{R}_T(c_T = c, b_T = 0))] \le \alpha$. □

Finally, we need to prove that R′ statistics suggested in this paper converge to $\mathcal{R}$ for any converging sequence c_T in Θ_c and $b_T = 0$. This result follows from Lem. 4 in Andrews and Guggenberger (2012) and is based on Thm. 3.1 by Hansen (1992), which requires stronger assumptions on mixing coefficients of innovations and their moments. As before, $\rho_T = 1 + c_T/T$, that is, parameters explicitly depend on the sample size T.

THEOREM 4. *Suppose, x_0 is a random variable with $Ex_0^2 < \infty$ and, for some $1/2 < \epsilon < 1$ and $\delta > 0$, $E||e_t||^{\frac{1}{\epsilon-\frac{1}{2}}+\delta} < \infty$, and strong mixing coefficients α_i are such that $\sum_{i=1}^{\infty} \alpha_i^{1-\epsilon} < \infty$. Consider a converging sequence $c_T \to c^* \in \Theta_c$. Then $R'(c_T, b_T = 0)$ weakly converge to the limiting statistics $\mathcal{R}(c^*, b = 0)$ as defined in Theorem 1.*

Proof. We first show that results of Lemma 3 hold along the sequences $c_T \to c^* \in \Theta_c$. Because result (3) does not depend on c_T and results (2) and (4) follow from CMT , we only need to prove result (1) of Lemma 3. In particular, we should show that the process $\frac{1}{\sqrt{\psi_{uu}}} \frac{x_{\lfloor \tau T \rfloor}}{\sqrt{T}}$, $\tau \in [0, 1]$ weakly converges to the O–U process $J_{c^*}(\tau)$ along the sequences $c_T \to c^*$. We note that

$$\frac{1}{\sqrt{T}} x_{\lfloor \tau T \rfloor} = \frac{\mu_x}{\sqrt{T}} + \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor \tau T \rfloor} (1 + \frac{c_T}{T})^{\lfloor \tau T \rfloor - t} u_t + (1 + \frac{c_T}{T})^{\lfloor \tau T \rfloor} \frac{x_0 - \mu_x}{\sqrt{T}}.$$

https://doi.org/10.1017/S0266466620000079 Published online by Cambridge University Press

Given the condition on the initial value x_0, we can write:

$$\frac{1}{\sqrt{T}}x_{\lfloor \tau T \rfloor} = \frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor \tau T \rfloor}(1+\frac{c_T}{T})^{\lfloor \tau T \rfloor - t}u_t + o_p(1). \tag{D.9}$$

We decompose $u_t = u_{\epsilon,t} + u_{z,t-1} - u_{z,t}$, where $u_{\epsilon,t} := \sum_{i=0}^{\infty}(E_t u_{t+i} - E_{t-1}u_{t+i})$ and $u_{z,t} := \sum_{i=1}^{\infty} E_t u_{t+i}$, see [Thm. 4.4] Hansen (1992). Note that $E_{t-1}u_{\epsilon,t} = 0$ and consider

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor \tau T \rfloor}\left(1+\frac{c_T}{T}\right)^{\lfloor \tau T \rfloor - t}u_{\epsilon,t}.$$

Define process $U_T(r) = \left(1+\frac{c_T}{T}\right)^{-\lfloor rT \rfloor}$, which converges to $U(r) = \exp(-rc^*)$ as $T \to \infty$. Also define a martingale process $V_T(r) = \frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor rT \rfloor} u_{\epsilon,t}$. By Thm. 3.1 of Hansen (1992), $\int_{r=0}^{1} U_T(r)dV_T(r)$ weakly converges to $\int_{r=0}^{1}\exp\{(\tau - r)c^*\}dW(r) := J_{c^*}(\tau)$.

The remaining part in expression (D.9) is the sum

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor \tau T \rfloor}(1+\frac{c_T}{T})^{\lfloor \tau T \rfloor - t}(u_{z,t-1} - u_{z,t}).$$

Up to an $o_p(1)$ term, this sum can be rewritten as $\frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor \tau T \rfloor} u_{z,t}(1+\frac{c_T}{T})^{\lfloor \tau T \rfloor - t}\left((1+\frac{c_T}{T})^{-1} - 1\right)$ and its absolute value can be bounded as follows:

$$\frac{1}{\sqrt{T}}\left|\sum_{t=1}^{\lfloor \tau T \rfloor} u_{z,t}(1+\frac{c_T}{T})^{\lfloor \tau T \rfloor - t}\left((1+\frac{c_T}{T})^{-1} - 1\right)\right|$$
$$\leq \frac{1}{\sqrt{T}}\sup_{t\leq T}|u_{z,t}|\sum_{t=1}^{\lfloor \tau T \rfloor}\left|(1+\frac{c_T}{T})^{\lfloor \tau T \rfloor - t}\left((1+\frac{c_T}{T})^{-1} - 1\right)\right|.$$

From the condition on mixing coefficients of u_t it follows that $\frac{1}{\sqrt{T}}\sup_{t\leq T}|u_{z,t}| \to 0$, see Thm. 3.1 by Hansen (1992). Furthermore, $\sum_{t=1}^{\lfloor \tau T \rfloor}\left|(1+\frac{c_T}{T})^{\lfloor \tau T \rfloor - \lfloor rT \rfloor}\left((1+\frac{c_T}{T})^{-1} - 1\right)\right|$ converges to $|c^*|\int_{r=0}^{\tau}\exp\{(\tau - r)c^*\}dr < \infty$. Therefore, $\frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor \tau T \rfloor}(1+\frac{c_T}{T})^{\lfloor \tau T \rfloor - t}(u_{z,t-1} - u_{z,t})$ converges to zero as $T \to \infty$ and $\frac{1}{\sqrt{T}}x_{\lfloor \tau T \rfloor}$ converges to the O–U process $J_c^*(\tau)$.

Therefore, the results in Lemma 3 hold also for converging sequences c_T. Finally, convergence of $\mathrm{R}'(c_T, 0)$ to $\mathcal{R}(c^*, 0)$ follows from Lemma 3 for sequences, following the proof steps in Theorem 2. □

THEOREM 5. *Consider model (1) with stationary and ergodic u_t that is a martingale difference sequence with finite $\mathbb{E}(u_t|u_{t-1}, u_{t-2}, \ldots) = \psi_{uu} < \infty$. Suppose, the initial observation x_0 is a random variable that does not depend on T, has finite variance, and is independent of future realizations of $e_t = (\varepsilon_t, u_t)$. Additionally, let the mixing condition*

https://doi.org/10.1017/S0266466620000079 Published online by Cambridge University Press

in Assumption A hold for e_t. Then the t-statistic

$$t(R) = \frac{\frac{R_\beta}{R_{\beta\beta}} + \frac{r}{\sqrt{1-r^2}} C(R)}{\sqrt{R^{-1}_{\beta\beta} + \frac{r}{\sqrt{1-r^2}} R^{-1}_{\rho\rho}}}, \quad with \quad C(R) = \frac{R_\rho + \frac{r}{\sqrt{1-r^2}} R_\beta}{R_{\rho\rho}}$$

calculated from R defined in (5) converges to $N(0,1)$ for all sequences $c_T = (\rho_T - 1)T \to -\infty$.

Proof. Note that the distribution of process $e_t = (\varepsilon_t, u_t)$ does not depend on c. Therefore, the limiting distribution of $d(\omega_j; \{e_t\}_{t=1}^T), \omega_j = 2\pi j/T, j = 1, \ldots, q$, conditional on $u_1, \ldots, u_T$ is the same for any sequence c_T. By Thm. 4.4.1 in Brillinger (1975), in the limit, $d(\omega_j; \{\varepsilon_t\}_{t=1}^T) - r\sqrt{\frac{\psi_{\varepsilon\varepsilon}}{\psi_{uu}}} d(\omega_j; \{u_t\}_{t=1}^T)$, $j = 1, \ldots, q$ is $\sqrt{\frac{\psi_{\varepsilon\varepsilon|u}}{2\pi}} z_j$, where z_j, $j = 1, \ldots, q$ are independent complex random variables with vector $\{\mathrm{Re}(z_j), \mathrm{Im}(z_j)\}$ jointly distributed as $N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}\right)$. Because, x_0 is independent of future realizations of e_t, the distribution of $d(\omega_j; \{\varepsilon_t\}_{t=1}^T)$, $j = 1, \ldots, q$, conditional on $x_0, x_1, \ldots, x_T$ is the same as its distribution conditional on $u_1, u_2, \ldots, u_T$.

Also, note that $d(\omega_j; \{u_t\}_{t=1}^T) = d(\omega_j; \{x_{t-1}\}_{t=1}^T)(e^{i\omega_j} - \rho_T) + \frac{x_T - x_0}{\sqrt{2\pi T}}$. Because $\mathrm{var}\left(T^{-1/2}(x_T - x_0)\right) = \psi_{uu}(-c_T(2 + c_T/T))^{-1} \to 0$ for all $c_T \to -\infty$, we observe that Fourier transformations $d(\omega_j; \{u_t\}_{t=1}^T)$ are asymptotically equivalent to $d(\omega_j; \{x_{t-1}\}_{t=1}^T)(e^{i\omega_j} - \rho_T)$. That is, $d(\omega_j; \{u_t\}_{t=1}^T) - d(\omega_j; \{x_{t-1}\}_{t=1}^T)(e^{i\omega_j} - \rho_T) \to_p 0$.

Similarly, the difference $d(\omega_j; \{x_t\}_{t-1}^T) - d(\omega_j; \{x_{t-1}\}_{t=1}^T) = \frac{x_T - x_0}{\sqrt{2\pi T}} + (e^{i\omega_j} - 1)$ $d(\omega_j; \{x_{t-1}\}_{t=1}^T)$ converges to zero for all sequences $c_T \to \infty$ because $||\frac{e^{i\omega_j} - 1}{e^{i\omega_j} - \rho_T}||^2 \to 0$ for $j = 1, \ldots, q$ and $c_T \to -\infty$. Therefore, $T\mathrm{R}_\beta(1 - \rho_T)$ for $b = 0$ is asymptotically equivalent to

$$4\pi \frac{r}{\sqrt{1-r^2}} \frac{\sum_{j=1}^q ||d(\omega_j; \{u_t\}_{t=1}^T)||^2}{\psi_{uu}} + \sqrt{8\pi} \frac{\mathrm{Re}\left(\sum_{j=1}^q d(\omega_j; \{u_t\}_{t=1}^T) z_j^*\right)}{\sqrt{\psi_{uu}}}$$

for any $c_T \to -\infty$. From definition of z_j^*, we obtain that the conditional distribution of $\mathrm{Re}\left(d(\omega_j; \{u_t\}_t^T) z_j^*\right)$ for $j = 1, \ldots, q$ is that of independent normal variables with means zero and variances equal to $\frac{1}{2}||d(\omega_j; \{u_t\}_{t=1}^T)||^2$. Therefore, $\mathrm{Re}\left(\sum_{j=1}^q d(\omega_j; \{u_t\}_{t=1}^T) z_j^*\right) = z\sqrt{\frac{1}{2}\sum_{j=1}^q ||d(\omega_j; \{u_t\}_{t=1}^T)||^2}$, where z is a standard normal variable that is independent of X.

Similarly, $T^2\mathcal{R}_{\beta\beta}(1 - \rho_T)^2$ is asymptotically equivalent to

$$4\pi \frac{\sum_{j=1}^q ||d(\omega_j; \{u_t\}_{t=1}^T)||^2}{\psi_{uu}}.$$

Statistic

$$\frac{1}{T(1-\rho_T)} C(\mathrm{R}) = (1 - \rho_T)^{-1} \left(\frac{\sum_{t=1}^{T-1} (x_{t-1} - \mu_x)(x_t - \mu_x)}{\sum_{t=1}^T (x_{t-1} - \mu_x)^2} - 1 \right)$$

https://doi.org/10.1017/S0266466620000079 Published online by Cambridge University Press

converges to -1 in probability for all $c_T \to -\infty$, see Giraitis and Phillips (2006). Therefore, $\frac{1}{(1-\rho_T)T}\left(\frac{R_\beta}{R_{\beta\beta}} - \frac{r}{\sqrt{1-r^2}}C(\mathrm{R})\right)$ is asymptotically equivalent to

$$z\sqrt{\frac{\psi_{uu}}{4\pi}}\sqrt{\frac{1}{\sum_{j=1}^{q}||d(\omega_j;\{u_t\}_{t=1}^{T})||^2}}.$$

Finally, $T^2R_{\beta\beta}(1-\rho_T)^2 = c_T^2 R_{\beta\beta}$ is asymptotically equivalent to

$$\frac{4\pi}{\psi_{uu}}\sum_{j=1}^{q}||d(\omega_j;\{u_t\}_{t=1}^{T})||^2,$$

and $T(1-\rho_T^2)R_{\rho\rho} \to 1$ by Lem. 2 of Giraitis and Phillips (2006). Therefore, $R_{\rho\rho}/R_{\beta\beta} = O_p(c_T)$ and $\mathrm{R}_{\beta\beta}^{-1} + \frac{r}{\sqrt{1-r^2}}\mathrm{R}_{\rho\rho}^{-1}$ is asymptotically equivalent to $\mathrm{R}_{\beta\beta}^{-1}$ for all $c_T \to -\infty$.

From the above results, it follows that statistic $t(\mathrm{R})$ converges to the limit of

$$\frac{\sqrt{\frac{\psi_{uu}}{4\pi}}\sqrt{\frac{1}{\sum_{j=1}^{q}||d(\omega_j;\{u_t\}_{t=1}^{T})||^2}}z}{\sqrt{\frac{\psi_{uu}}{4\pi}\left(\sum_{j=1}^{q}||d(\omega_j;\{u_t\}_{t=1}^{T})||^2\right)^{-1}}}$$

for all $c_T \to -\infty$. The latter is exactly equal to the standard normal z. □

E. ADDITIONAL METHODS

E.1. Nearly Optimal Test Based on the Entire Gaussian Likelihood

In Section 6, we compare the long-run nearly optimal test presented in this article to the nearly optimal test based on the entire Gaussian likelihood. The latter has the same structure as the test presented in this article, but employs a different set of asymptotic statistics. The asymptotic inference problem (see Jansson and Moreira 2006 and Elliott, Müller and Watson, 2015) is based on the likelihood:

$$f(\mathcal{Y}|b,c) \propto \exp\left\{b\mathcal{Y}_\beta + c\mathcal{Y}_\rho - \frac{1}{2}\left(b - c\frac{r}{\sqrt{1-r^2}}\right)^2\mathcal{Y}_{\beta\beta} - \frac{1}{2}c^2\mathcal{Y}_{\rho\rho}\right\},$$

with the four asymptotic statistics $\mathcal{Y} = (\mathcal{Y}_\beta, \mathcal{Y}_\rho, \mathcal{Y}_{\beta\beta}, \mathcal{Y}_{\rho\rho})$ defined as follows:

$$\mathcal{Y}_\beta = \int_0^1 J_c^\mu(\tau)dW^z(\tau) + \left(b - \frac{cr}{\sqrt{1-r^2}}\right)\int_0^1 \left(J_c^\mu(\tau)\right)^2 d\tau,$$

$$\mathcal{Y}_\rho = \int_0^1 J_c(\tau)dJ_c(\tau) - r(1-r^2)^{-1/2}\mathcal{Y}_\beta,$$

$$\mathcal{Y}_{\beta\beta} = \int_0^1 \left(J_c^\mu(\tau)\right)^2 d\tau,$$

$$\mathcal{Y}_{\rho\rho} = \int_0^1 \left(J_c(\tau)\right)^2 d\tau,$$

https://doi.org/10.1017/S0266466620000079 Published online by Cambridge University Press

where $J_c^{\mu}(\tau) := J_c(\tau) - \int_0^1 J_c(r)dr$ is the demeaned O–U process. The EMW test switches between the likelihood ratio rejection rule

$$\mathcal{D}_n(\mathcal{Y}) = \left\{ \frac{\int_{(b,c)\in \mathrm{B}\times\Theta_c} f(\mathcal{Y}|b,c)dF(b,c)}{\int_{c\in\Theta_c} f(\mathcal{Y}|b_0,c)d\Lambda(c)^{*,\epsilon}} > K_{\alpha,n} \right\}$$

and the t-test

$$\mathcal{D}_s(\mathcal{Y}) = \left\{ \frac{\mathcal{Y}_\beta \frac{}{} + \frac{r}{\sqrt{1-r^2}} C(\mathcal{Y})}{\sqrt{\mathcal{Y}_{\beta\beta}^{-1} + \frac{r}{\sqrt{1-r^2}}\mathcal{Y}_{\rho\rho}^{-1}}} > K_{\alpha,s} \right\},$$

if $C(\mathcal{Y}) = \frac{\mathcal{Y}_\rho + \frac{r}{\sqrt{1-r^2}}\mathcal{Y}_\beta}{\mathcal{Y}_{\rho\rho}}$ is less than threshold K_c. As previously defined, $F(b,c)$ is the weighting measure in the WAP.

For the i.i.d. Monte Carlo experiments in Section 6, this test is implemented by replacing $\mathcal{Y}$ with the small-sample counterparts, as in Jansson and Moreira (2006). In particular,

$$Y_\beta = \frac{1}{T}\frac{1}{\sqrt{\psi_{\varepsilon\varepsilon|u}\psi_{uu}}}\sum_{t=1}^{T}\left[(x_{t-1}-\bar{x})\left[y_t - \beta_0 x_{t-1} - \frac{r\psi_{\varepsilon\varepsilon}^{1/2}}{\psi_{uu}^{1/2}}(x_t - x_{t-1})\right]\right],$$

$$Y_\rho = \psi_{uu}^{-1}T^{-1}\sum_{t=1}^{T-1} x_{t-1}(x_t - x_{t-1}) - \frac{r}{\sqrt{1-r^2}}Y_\beta,$$

$$Y_{\beta\beta} = \frac{1}{T^2\psi_{uu}}\sum_{t=1}^{T}(x_{t-1}-\bar{x})^2,$$

$$Y_{\rho\rho} = \frac{1}{T^2\psi_{uu}}\sum_{t=1}^{T}(x_{t-1})^2.$$

For the Monte Carlo experiments in the endogenous case, the statistic Y_β is adjusted. In particular,

$$Y'_\beta = Y_\beta - \frac{\sum_{j=1}^{\infty} E\left(u_t\varepsilon_{t+j}\right) - \frac{r\psi_{\varepsilon\varepsilon}^{1/2}}{\psi_{uu}^{1/2}}\sum_{j=1}^{\infty} E\left(u_t u_{t+j}\right)}{\sqrt{\psi_{uu}\psi_{\varepsilon\varepsilon|u}}}.$$

This correction follows from the results of Hansen (1992). Other statistics can be also adjusted if u_t is serially correlated, as in Jansson and Moreira (2006).

E.2. Linear Programming Approach to Construction of ϵ-ALFD

Note that the fixed-point iterations used in this paper to compute $\Lambda^{*,\epsilon}$ may not always converge. However, as discussed in the main text, we can always directly assess the quality of the resulting test by comparing its power to the power bound, while controlling for the correct asymptotic test level. There is also an alternative approach to computing $\Lambda^{*,\epsilon}$ based on the linear programming method, which we shortly describe here.

Note that the nearly optimal test presented in this article belongs to the class of tests considered by Moreira et al. (2014), with the test statistic being the likelihood ratio.

https://doi.org/10.1017/S0266466620000079 Published online by Cambridge University Press

Therefore, the critical value function approach of Moreira et al. (2014) can be applied as follows. Define the probability distribution $h(\mathcal{R}) = \int_{(b,c)\in\mathbb{B}\times\Omega_c} f(\mathcal{R}|b,c)dF(b,c)$, where $F(b,c)$ defines weights in the WAP. Also define distribution $h_0(\mathcal{R}) = \int_{(c)\times\Omega_c} f(\mathcal{R}|b = 0,c)dF'(c)$, where $F'(c)$ is an arbitrary probability measure on Ω_c. Consider a likelihood-ratio statistic $\psi(\mathcal{R}) = h(\mathcal{R})/h_0(\mathcal{R})$. That is, $\psi(\mathcal{R})$ is the likelihood ratio between the distribution of $\mathcal{R}$ under the alternative hypothesis that corresponds to the weights in the WAP and an auxiliary reference distribution under H_0. We reject the H_0 hypothesis when this likelihood ratio is high. The critical value for this test $K_\alpha(\mathcal{R})$ can depend on the observed $\mathcal{R}$ and is chosen to obtain the correct size of the test while maximizing the power to reject H_0.

This test is equivalent to the nearly optimal test presented in this article for a discrete Ω_c under the following choice of the critical value function:

$$K_\alpha(\mathcal{R}) = \sum_{\iota=1}^{I} K_{\alpha,H_0}\lambda_\iota \frac{f(\mathcal{R}|b=0,c_\iota)}{h_0(\mathcal{R})}.$$

Next, suppose that $\mathcal{R}_j, j = 1,\ldots,J$ is a sample of size J simulated from distribution $h_0(\mathcal{R})$. Then, the approximate optimal weights $K_{\alpha,H_{0,\Lambda}}\lambda_\iota$ can be obtained in the limit with $J\to\infty$ as the Lagrange multipliers in the following linear program:

$$\max_{\phi_J\in[0,1]^J} \frac{1}{J}\left(\frac{h(\mathcal{R}_1)}{h_0(\mathcal{R}_1)}\phi_{J,1} + \cdots + \frac{h(\mathcal{R}_J)}{h_0(\mathcal{R}_1)}\phi_{J,J}\right),$$

$$s.t.\ \frac{1}{J}\left(\frac{f(\mathcal{R}_1|b=0,c=c_\iota)}{h_0(\mathcal{R}_1)}\phi_{J,1} + \cdots + \frac{f(\mathcal{R}_J|b=0,c=c_\iota)}{h_0(\mathcal{R}_1)}\phi_{J,J}\right) \le \alpha,\ \iota = 1,\ldots,I.$$

To implement this method, we use the C++ package GLPK (version 4.45). Following Moreira et al. (2014), we select $h_0(\mathcal{R})$ that puts equal weights on points $c_\iota, \iota = 1,\ldots,I$ and implement this method with $J = 10,000$. As described in the main text, the best attained ϵ (power loss) is 2%, which is higher than with the computational method of Elliott et al. (2015). We obtain the same result with the linear programming method of [Alg. A.1]ch09.

https://doi.org/10.1017/S0266466620000079 Published online by Cambridge University Press