Econometric Theory, **36**, 2020, 907–960.
doi:10.1017/S0266466619000367

A MAX-CORRELATION WHITE NOISE TEST FOR WEAKLY DEPENDENT TIME SERIES*

JONATHAN B. HILL
University of North Carolina
KAIJI MOTEGI
Kobe University

This article presents a bootstrapped p-value white noise test based on the maximum correlation, for a time series that may be weakly dependent under the null hypothesis. The time series may be prefiltered residuals. The test statistic is a normalized weighted maximum sample correlation coefficient $\max_{1\leq h\leq \mathcal{L}_n}\sqrt{n}|\hat{\omega}_n(h)\hat{\rho}_n(h)|$, where $\hat{\omega}_n(h)$ are weights and the maximum lag $\mathcal{L}_n$ increases at a rate slower than the sample size n. We only require uncorrelatedness under the null hypothesis, along with a moment contraction dependence property that includes mixing and nonmixing sequences. We show Shao's (2011, *Annals of Statistics* 35, 1773–1801) dependent wild bootstrap is valid for a much larger class of processes than originally considered. It is also valid for residuals from a general class of parametric models as long as the bootstrap is applied to a first-order expansion of the sample correlation. We prove the bootstrap is asymptotically valid without exploiting extreme value theory (standard in the literature) or recent Gaussian approximation theory. Finally, we extend Escanciano and Lobato's (2009, *Journal of Econometrics* 151, 140–149) automatic maximum lag selection to our setting with an unbounded lag set that ensures a consistent white noise test, and find it works extremely well in controlled experiments.

1. INTRODUCTION

We present a bootstrap white noise test based on the maximum (in absolute value) autocorrelation. The data may be observed or filtered residuals. A new asymptotic theory approach is used relative to the literature, one that sidesteps

*We thank three referees, Co-Editor Michael Jansson, and Editor Peter C.B. Phillips for helpful comments and suggestions that led to significant improvements to our manuscript. We also thank Eric Ghysels, Shigeyuki Hamori, Peter R. Hansen, Yoshihiko Nishiyama, Kenichiro Tamaki, Kozo Ueda, and Zheng Zhang, seminar participants at the Kyoto Institute of Economic Research, UNC Chapel Hill, the University of Essex, and Kobe University, and conference participants at the 10th Spring Meeting of JSS, 2016 SWET, 2016 AMES, 2016 JJSM, 2016 NBER-NSF Time Series Conference, the 15th International Conference of WEAI, and SETA 2019 for helpful comments. The second author is grateful for financial support from JSPS KAKENHI (Grant Number 16K17104), Kikawada Foundation, Mitsubishi UFJ Trust Scholarship Foundation, Nomura Foundation, and Suntory Foundation. Address correspondence to Jonathan B. Hill, Department of Economics, University of North Carolina at Chapel Hill, Chapel Hill, NC. USA; e-mail: jbhill@email.unc.edu; web: https://jbhill.web.unc.edu.

© The Author(s) 2020. Published by Cambridge University Press.

https://doi.org/10.1017/S0266466619000367 Published online by Cambridge University Press

deriving the asymptotic distribution of a max-correlation statistic, or working with tools specific to Gaussian approximations and couplings. We operate solely on the bootstrapped p-value. Convergence in finite dimensional distributions of the sample correlation is combined with new theory for handling convergence of arbitrary arrays. The latter is applicable for dealing with the maximum of an increasing sequence of correlations, in particular, when residuals based on a plug-in estimator are used.

The class of time series models considered here is

$$y_t = f(x_{t-1}, \phi_0) + u_t \text{ and } u_t = \epsilon_t \sigma_t(\theta_0), \tag{1}$$

where $\phi \in \mathbb{R}^{k_\phi}$, $k_\phi \geq 0$, and $f(x, \phi)$ is a level response function. The error ϵ_t satisfies $E[\epsilon_t] = 0$, $E[\epsilon_t^2] < \infty$, and the regressors are $x_t \in \mathbb{R}^{k_x}$, $k_x \geq 0$. We assume $\{x_t, y_t\}$ are strictly stationary in order to focus ideas. Volatility $\sigma_t^2(\theta_0)$ is a process measurable with respect to $\mathcal{F}_{t-1} \equiv \sigma(y_\tau, x_\tau : \tau \leq t-1)$, where θ_0 is decomposed as $[\phi_0', \delta_0'] \in \mathbb{R}^{k_\theta}$, $\delta_0 \in \mathbb{R}^{k_\delta}$ are volatility-specific parameters, and $(k_\theta, k_\delta) \geq 0$. The dimensions of ϕ_0 and δ_0 (hence θ_0) may be zero, depending on the model desired and the interpretation of the test variable ϵ_t. Thus, $k_\phi = 0$ implies a volatility model $y_t = \epsilon_t \sigma_t(\theta_0)$, if $k_\delta = 0$ then $y_t = f(x_{t-1}, \phi_0) + \epsilon_t$, and $y_t = \epsilon_t$ when $k_\theta = 0$ (i.e., a filter is not used). We want to test if $\{\epsilon_t\}$ is a white noise process:

$$H_0 : E[\epsilon_t \epsilon_{t-h}] = 0 \ \forall h \in \mathbb{N} \text{ against } H_1 : E[\epsilon_t \epsilon_{t-h}] \neq 0 \text{ for some } h \in \mathbb{N}.$$

Notice ϵ_t need not have a zero conditional mean: we do not require, for example, $E[\epsilon_t | x_{t-1}] = 0$ a.s. This implies that we do not require $\sigma_t^2(\theta_0)$ to be a conditional variance. Together, (1) allows for model mis-specification. Nevertheless, (1) is assumed correct in some sense, whether H_0 is true or not, in view of $E[\epsilon_t] = 0$ and possibly other moment conditions used to identify θ_0. Thus, θ_0 should be thought of as a pseudo-true value that can be identified, often by unconditional moment conditions (Kullback and Leibler, 1951; Sawa, 1978). Complete assumptions are given in Section 2: see especially Assumption 2.

Unless $y_t = \epsilon_t$ such that y_t is known to have a zero mean, let $\hat{\theta}_n = [\hat{\phi}_n', \hat{\delta}_n']$ estimate θ_0 where n is the sample size, and define the residual, and its sample serial covariance and correlation at lag $h \geq 1$

$$\epsilon_t(\hat{\theta}_n) \equiv \frac{u_t(\hat{\phi}_n)}{\sigma_t(\hat{\theta}_n)} \equiv \frac{y_t - f(x_{t-1}, \hat{\phi}_n)}{\sigma_t(\hat{\theta}_n)} \text{ and}$$

$$\hat{\gamma}_n(h) \equiv \frac{1}{n} \sum_{t=1+h}^{n} \epsilon_t(\hat{\theta}_n) \epsilon_{t-h}(\hat{\theta}_n) \text{ and } \hat{\rho}_n(h) \equiv \frac{\hat{\gamma}_n(h)}{\hat{\gamma}_n(0)}.$$

In the pure volatility model, set $f(x_{t-1}, \hat{\phi}_n) = 0$, and in the level model, set $\sigma_t(\hat{\theta}_n) = 1$.

Our primary test statistic is the normalized weighted sample maximum correlation,

https://doi.org/10.1017/S0266466619000367 Published online by Cambridge University Press

$$\hat{\mathcal{T}}_n \equiv \sqrt{n} \max_{1 \le h \le \mathcal{L}_n} \left| \hat{\omega}_n(h) \hat{\rho}_n(h) \right|,$$

where $\hat{\omega}_n(h) > 0$ are possibly stochastic weights with $\hat{\omega}_n(h) \xrightarrow{p} \omega(h) > 0$, where $\omega(h)$ are nonstochastic. The weights allow for (i) control for variable dispersion across lags that affect empirical power, or (ii) a decrease in accuracy in probability when n is small and h is large. In the former case, $\hat{\omega}_n(h)$ may be an inverted standard deviation estimator. In the latter case, we might use $\hat{\omega}_n(h) = (n-2)/(n-h)$ as in Ljung and Box (1978). Despite the generality afforded by weights, we find using $\hat{\omega}_n(h) = 1$ results in accurate size and comparably high power in Monte Carlo simulations. Indeed, using an inverted standard deviation $\hat{\omega}_n(h)$ does not improve test performance in our experiments due to estimation error associated with $\hat{\omega}_n(h)$.

The number of lags $\mathcal{L}_n$ can converge to a finite positive integer: the theory follows trivially from the proofs of our main results. In that case, our test would not be a formal test of the white noise hypothesis. We want $\mathcal{L}_n \to \infty$ as $n \to \infty$ in order to ensure a white noise test, and that $\mathcal{L}_n = o(n)$ to ensure $\hat{\gamma}_n(h) = E[\epsilon_t \epsilon_{t-h}] + O_p(1/\sqrt{n})$ for each $h \in \{1, ..., \mathcal{L}_n\}$ and therefore yield a consistent test. The limit theory in that case requires more than convergence in finite dimensional distributions based on classic arguments (e.g., Hoffmann-Jrgensen, 1984, 1991), which is one of the major challenges we address in this article.

Interest in the maximum of an increasing sequence of deviated covariances $\sqrt{n}\max_{1 \le h \le \mathcal{L}_n} |\hat{\gamma}_n(h) - \gamma(h)|$ dates in some form to Berman (1964) and Hannan (1974). See also Xiao and Wu (2014) and their references. In this literature, the test variable is observed, and the exact asymptotic distribution form of a suitably normalized $\sqrt{n}\max_{1 \le h \le \mathcal{L}_n} |\hat{\gamma}_n(h) - \gamma(h)|$ is sought. Xiao and Wu (2014) impose a moment contraction property on y_t, and $\mathcal{L}_n = O(n^{\upsilon})$ for some $\upsilon \in (0, 1)$ that is smaller with greater allowed dependence. They show $a_n\{\sqrt{n}\max_{1 \le h \le \mathcal{L}_n} |\hat{\gamma}_n(h) - \gamma(h)|/(\sum_{h=0}^{\infty} \gamma(h)^2)^{1/2} - b_n\} \xrightarrow{d} \exp\{-\exp\{-x\}\}$, a Gumbel distribution, with normalizing sequences a_n, $b_n \sim (2\ln(n))^{1/2}$. See, also, Jirak (2011). Xiao and Wu (2014) do not prove their blocks-of-blocks bootstrap is valid under their assumptions, and only observed data are allowed. The moment contraction property is also more restrictive than the Near Epoch Dependence (NED) property used here (see the supplemental material, Appendix B).

Chernozhukov, Chetverikov, and Kato (2013, 2015, 2017) significantly improve on results in the literature on Gaussian approximations and couplings, cf. Yurinskii (1977), Dudley and Philipp (1983), Portnoy (1986), and Le Cam (1988). They allow for arbitrary dependence across the sequence of sample means of independent data, and the sequence length may grow at a rate of order $e^{Kn^{\varsigma}}$ for some K, $\varsigma > 0$. Sample autocorrelations, however, only exist for lags $\{0, \ldots, n-1\}$, and are Fisher consistent for the population autocorrelations for lags h up to order $o(n)$. The independence assumption, however, is not feasible for a white noise test since $\epsilon_t \epsilon_{t-h}$ is at best a martingale difference, and may be generally dependent under either hypothesis. Furthermore, a Gaussian approximation theory cannot handle

https://doi.org/10.1017/S0266466619000367 Published online by Cambridge University Press

the maximum distance between $\hat{\rho}_n(h)$ based on residuals $\epsilon_t(\hat{\theta}_n)\epsilon_{t-h}(\hat{\theta}_n)$, and its version based on $\epsilon_t\epsilon_{t-h}$ (and other components due to the plug-in estimator $\hat{\theta}_n$) because $\epsilon_t\epsilon_{t-h}$ is typically not Gaussian even if ϵ_t is.[1] Chernozhukov, Chetverikov, and Kato (2014, Appendix B), cf. Chernozhukov, Chetverikov, and Kato (2018, Supplemental Appendix), allow for *almost surely* bounded β-mixing data, but the above problem involving filtered data is not resolved, boundedness rules out many time series of practical interest, and our NED environment eclipses a mixing environment (see Section 2.1 below and see, e.g., Davidson, 1994, Chap. 17).

Zhang and Wu (2017) extend results in Chernozhukov, Chetverikov, and Kato (2013) to a large class of dependent processes (see also Zhang and Cheng, 2014, for an extension to geometrically dependent data in a bootstrap setting). Their framework is the functional dependence or moment contraction notions popularized in, for example, Wu (2005). The possibility of filtered data is ignored, which requires a non-Gaussian approximation theory. Furthermore, it is not obvious which processes satisfy the conditions of their main Thm. 3.2 (e.g., nonlinear Autoregressive Moving Average-Generalized Autoregressive Conditional Heteroscedasticity (ARMA-GARCH), stochastic volatility).

Compared with the above literature, we use a different asymptotic theory approach. We sidestep extreme value theoretic methods by exploiting convergence of $\{\sqrt{n}(\hat{\gamma}_n(h) - \gamma(h)) : 1 \leq h \leq \mathcal{L}\}$ to a Gaussian process, for each finite $\mathcal{L} \in \mathbb{N}$. Since that is not sufficient for weak convergence in the classic sense of Hoffmann-Jørgensen (1984, 1991), we develop new theory for double array convergence, which is associated with arguments dating to Ramsey (1930). This allows us to prove that under H_0 the maximum distance over $1 \leq h \leq \mathcal{L}_n$ between $\sqrt{n}\hat{\rho}_n(h)$ and its bootstrapped version converges to zero for some sequence of positive integers $\{\mathcal{L}_n\}$, with $\mathcal{L}_n \to \infty$ and $\mathcal{L}_n = o(n)$, without using extreme value theoretic arguments or Gaussian approximation theory. Under additional technical conditions presented in the supplemental material (Appendix G), we show $\mathcal{L}_n = O(n^c/\ln(n))$ must also hold, for some $c \in (0, 1)$ that depends on the rate of convergence of the weights $\hat{\omega}_n(h) \xrightarrow{p} \omega(h)$, and an asymptotic approximation expansion for the plug-in $\hat{\theta}_n$. Under standard regularity conditions $c = 1/2$, hence $\mathcal{L}_n = O(\sqrt{n}/\ln(n))$. These are our primary contributions. As in Chernozhukov, Chetverikov, and Kato (2013), we do not require $\sqrt{n}\max_{1\leq h\leq \mathcal{L}_n}|\hat{\omega}_n(h)\hat{\rho}_n(h)|$ to converge in law under H_0 since the bootstrap is asymptotically valid irrespective of the asymptotic properties of $\sqrt{n}\max_{1\leq h\leq \mathcal{L}_n}|\hat{\omega}_n(h)\hat{\rho}_n(h)|$.

[1] When filtered data are used, we must prove in Lemma 2.1 that $\max_{1\leq h\leq \mathcal{L}_n}|1/\sqrt{n}\sum_{t=1}^n \epsilon_t(\hat{\theta}_n)\epsilon_{t-h}(\hat{\theta}_n) - 1/\sqrt{n}\sum_{t=1}^n z_t(h)| \xrightarrow{p} 0$ for some sequence $\{\mathcal{L}_n\}$, $\mathcal{L}_n \to \infty$, and some process $\{z_t(h)\}$ that is a function of $\epsilon_t\epsilon_{t-h}$ and components of $\hat{\theta}_n$. We then prove in Lemma 2.2 that $\max_{1\leq h\leq \mathcal{L}_n}|1/\sqrt{n}\sum_{t=1}^n z_t(h)| \xrightarrow{d} \max_{1\leq h\leq\infty}|\mathcal{Z}(h)|$ for some Gaussian process $\{\mathcal{Z}(h)\}$. Under suitable memory and heterogeneity restrictions, the Gaussian approximation theory of Zhang and Wu (2017), cf. Chernozhukov, Chetverikov, and Kato (2013), can handle $\max_{1\leq h\leq \mathcal{L}_n}|1/\sqrt{n}\sum_{t=1}^n z_t(h)| \xrightarrow{d} \max_{1\leq h\leq\infty}|\mathcal{Z}(h)|$ since $\{\mathcal{Z}(h)\}$ is Gaussian. But their theory cannot determine $\max_{1\leq h\leq \mathcal{L}_n}|1/\sqrt{n}\sum_{t=1}^n \epsilon_t(\hat{\theta}_n)\epsilon_{t-h}(\hat{\theta}_n) - 1/\sqrt{n}\sum_{t=1}^n z_t(h)| \xrightarrow{p} 0$ because that would require $1/\sqrt{n}\sum_{t=1}^n z_t(h)$ itself to be Gaussian for each n. The latter generally does not hold because $\epsilon_t\epsilon_{t-h}$ is not Gaussian even if ϵ_t is.

https://doi.org/10.1017/S0266466619000367 Published online by Cambridge University Press

Our asymptotic theory covers a class of continuous transforms of $[\sqrt{n}\hat{\omega}_n(h)\ \hat{\rho}_n(h)]_{h=1}^{\mathcal{L}_n}$, including the maximum, but also a weighted average $n\sum_{h=1}^{\mathcal{L}_n}\hat{\omega}_n^2(h)\ \hat{\rho}_n^2(h)$, and therefore portmanteau statistics (cf. Ljung and Box, 1978; Hong, 1996, 2001). Hong (1996, 2001) presents spectral density methods for testing for uncorrelatedness, and the proposed test statistic is simply a normalized portmanteau. The latter is shown to be asymptotically normal under regularity conditions that ensure $\sqrt{n}\hat{\rho}_n^2(h)$ is asymptotically independent across h under H_0. The approach taken here alleviates the necessity for the normalized $n\sum_{h=1}^{\mathcal{L}_n}\hat{\omega}_n^2(h)\hat{\rho}_n^2(h)$ to converge in law under H_0, hence we do not require asymptotic independence. Furthermore, as opposed to Hong (1996, 2001), our test statistic achieves the parametric rate of convergence because we do not use self-normalization. See Remark 17 in Section 2.

We perform a bootstrap p-value test using Shao's (2011) dependent wild bootstrap, and prove its validity. In order to control for the use of filtered sampling errors, the bootstrap is applied to a first-order expansion of the sample covariance. Delgado and Velasco (2011) take a different approach by using orthogonally transformed jointly standardized correlations in order to control for residuals and dependence. They assume a fixed maximum lag $\mathcal{L}$, however, due to joint standardization.

Finally, in order to resolve the choice of $\{\mathcal{L}_n\}$ in practice, we extend Escanciano and Lobato's (2009) automatic maximum lag selection method to our setting. They develop a Q-test with a bounded maximum lag that is selected based on the magnitude of the maximum correlation. We allow for selection from an increasing set of integers, and provide a new asymptotic theory for the automatic maximum lag.

General dependence under the null is allowed in different ways in Hong (1996), Romano and Thombs (1996), Shao (2011), and Guay, Guerre, and Lazarová (2013), among others. Our NED setting is similar to that of Lobato (2001) and Nankervis and Savin (2010, 2012), but the former works with observed data and requires a fixed maximum lag, and we allow for a substantially larger class of filters and parametric estimators than the latter. NED encompasses mixing and nonmixing processes, hence our setting is more general than Zhu's (2015) for his block-wise random weighting bootstrap.

Shao (2011), Guay, Guerre, and Lazarová (2013), and Xiao and Wu (2014) use a moment contraction property from Wu (2005) and Wu and Min (2005) with (potentially far) greater moment conditions than imposed here (e.g., Shao, 2011; Guay, Guerre, and Lazarová, 2013). Shao (2011) requires a complicated eighth order cumulant condition that is only known to hold under geometric memory, and residuals are not treated. Xiao and Wu (2014) only require slightly more than a fourth moment, as we do, but do not allow for residuals. We show in the supplemental material (Appendix B) that our NED setting is more general than the moment contraction properties employed in Shao (2011) and Guay, Guerre, and Lazarová (2013), and allows for slower memory decay than Xiao and Wu (2014).

https://doi.org/10.1017/S0266466619000367 Published online by Cambridge University Press

Test statistics that combine serial correlations have a vast history dating to Box and Pierce's (1970) Q-test. Many generalizations exist, including letting the maximum lag increase (Hong, 1996, 2001); bootstrapping or re-scaling for size correction under weak dependence (Romano and Thombs, 1996; Lobato, 2001; Horowitz et al., 2006; Kuan and Lee, 2006; Zhu, 2015); using a Lagrange Multiplier type statistic to account for weak dependence (e.g., Andrews and Ploberger, 1996; Lobato, Nankervis, and Savin, 2002); exploiting an expansion and orthogonal projection to produce pivotal statistics (Lobato, 2001; Kuan and Lee, 2006; Delgado and Velasco, 2011); and using endogenous maximum lag selection (Escanciano and Lobato, 2009; Guay, Guerre, and Lazarová, 2013). Recently, Dalla, Giraitis, and Phillips (2019) improve correlogram and portmanteau statistics to uncover hidden dependencies and nonstationarities specifically associated with heteroskedasticity.

A related class of estimators exploits the periodogram, an increasing sum of sample correlations, dating to Grenander and Rosenblatt (1952) (e.g., Hong, 1996; Deo, 2000; Delgado, Hidalgo, and Velasco, 2005; Shao, 2011; Zhu and Li, 2015). Hong (1996) standardizes a periodogram resulting in less than $\sqrt{n}$-local power, while Cramér–von Mises and Kolmogorov–Smirnov transforms in Deo (2000), Delgado, Hidalgo, and Velasco (2005), and Shao (2011) result in $\sqrt{n}$-local power. Guay, Guerre, and Lazarová (2013) show that Hong's (1996) standardized portmanteau test (but not a Cramér–von Mises test) can detect local-to-null correlation values at a rate faster than $\sqrt{n}$ provided an adaptive increasing maximum lag is used. Finally, a weighted sum of correlations also arises in Andrews and Ploberger's (1996) sup-LM test (cf. Nankervis and Savin, 2010).

A simulation study shows our proposed max-correlation test with Shao's (2011) dependent wild bootstrap and automatic lag (denoted $\hat{\mathcal{T}}^{dw}(\mathcal{L}_n^*)$) dominates a variety of other tests. In this article, we compare $\hat{\mathcal{T}}^{dw}(\mathcal{L}_n^*)$ and Shao's (2011) dependent wild bootstrap spectral Cramér–von Mises test, which is proposed for observed data. In the supplemental material (Appendix H), we consider other tests, including Hong's (1996) test based on a standardized periodogram, a CvM test with Zhu and Li's (2015) block-wise random weighting bootstrap, and Andrews and Ploberger's (1996) sup-LM test with the dependent wild bootstrap. Overall the CvM test is one of the strongest competitors of our test. First, generally $\hat{\mathcal{T}}^{dw}(\mathcal{L}_n^*)$ achieves sharp size. Second, $\hat{\mathcal{T}}^{dw}(\mathcal{L}_n^*)$, the sup-LM and the CvM tests lead to roughly comparable power when there exist autocorrelations at small lags. Third, $\hat{\mathcal{T}}^{dw}(\mathcal{L}_n^*)$ has high power while others have nearly trivial power when there exist autocorrelations at remote lags. Thus, of the tests under study, $\hat{\mathcal{T}}^{dw}(\mathcal{L}_n^*)$ is the only white noise test that accomplishes both sharp size in general and high power across cases. The sharp performance of $\hat{\mathcal{T}}^{dw}(\mathcal{L}_n^*)$ stems from the fact that the automatic lag selection mechanism trims redundant lags under H_0, and hones in on the most informative lag under H_1.

The remainder of the article is as follows. Section 2 contains the assumptions and main results. Automatic lag selection is developed in Section 3, and a Monte

https://doi.org/10.1017/S0266466619000367 Published online by Cambridge University Press

Carlo study follows in Section 4. Concluding remarks are left for Section 5. Proofs are gathered in Appendix A and the supplemental material (Appendix F).

Throughout, $|\cdot|$ is the l_1-matrix norm; $||\cdot||$ is the l_2 -matrix norm; $||\cdot||_p$ is the L_p-norm. $I(\cdot)$ is the indicator function: $I(A) = 1$ if A is true, else $I(A) = 0$. $\mathcal{F}_t \equiv \sigma(y_\tau, x_\tau : \tau \leq t)$. All random variables lie in a complete probability measure space $(\Omega, \mathcal{P}, \mathcal{F})$, hence $\sigma(\cup_{t\in\mathbb{Z}}\mathcal{F}_t)\subseteq\mathcal{F}$. We drop the (pseudo) true value θ_0 from function arguments when there is no confusion.

2. MAX-CORRELATION TEST

We first lay out the assumptions and derive some fundamental properties of the correlation maximum. We then derive the main results.

2.1. Assumptions and Asymptotic Expansion

An expansion of $\epsilon_t(\hat{\theta}_n)$ around θ_0 is required in order to ensure the bootstrapped statistic captures the influence of the estimator $\hat{\theta}_n$ on $\sqrt{n}\hat{\rho}_n(h)$. This is accomplished under various regularity assumptions. Let $\{\upsilon_t\}$ be a stationary α-mixing process with σ-fields $\mathfrak{V}_s^t \equiv \sigma(\upsilon_\tau : s \leq \tau \leq t)$ and $\mathfrak{V}_t \equiv \mathfrak{V}_{-\infty}^t$, and coefficients $\alpha_m^{(\upsilon)} = \sup_{\mathcal{A}\subset\mathfrak{V}_t^\infty, \mathcal{B}\subset\mathfrak{V}_{-\infty}^{t-m}}|P(\mathcal{A}\cap\mathcal{B}) - P(\mathcal{A})P(\mathcal{B})| \to 0$ as $m \to \infty$. We say L_q-bounded $\{\epsilon_t\}$ is stationary L_q-NED with size $\lambda > 0$ on a mixing base $\{\upsilon_t\}$ when $||\epsilon_t - E[\epsilon_t|\mathfrak{V}_{t-m}^{t+m}]||_q = O(m^{-\lambda-\iota})$ for tiny $\iota > 0$.[2] If $\epsilon_t = \upsilon_t$ then $||\epsilon_t - E[\epsilon_t|\mathfrak{V}_{t-m}^{t+m}]||_q = 0$, hence NED includes mixing sequences, but it also includes nonmixing sequences since it covers infinite lag functions of mixing sequences that need not be mixing. NED is related to McLeish's (1975) mixingale property. See Davidson (1994, Chap. 17) for historical references and deep results.

Assumption 1 (data generating process)**.**

a. $\{x_t, y_t\}$ are stationary, ergodic, and $L_{2+\delta}$-bounded for tiny $\delta > 0$.
b. ϵ_t is stationary, ergodic, $E[\epsilon_t] = 0$, L_r-bounded, $r > 4$, and L_4-NED with size 1/2 on stationary α-mixing $\{\upsilon_t\}$ with coefficients $\alpha_h^{(\upsilon)} = O(h^{-r/(r-4)-\iota})$ for tiny $\iota > 0$.
c. The weights satisfy $\hat{\omega}_n(h) > 0$ *a.s.* and $\hat{\omega}_n(h) = \omega(h) + O_p(1/n^\kappa)$ for some $\kappa > 0$ and nonrandom $\omega(h) \in (0, \infty)$, for each h.

Remark 1. The assumption $E[\epsilon_t] = 0$ is typically imposed in practice with inclusion of a regression model constant term. It is important that the necessary steps for ensuring $E[\epsilon_t] = 0$ are taken, since otherwise a white noise test may reject due merely to $E[\epsilon_t] \neq 0$.

Remark 2. Ergodicity is not required in principle, but imposed to allow easily for laws of large numbers on functions of $f(x_t, \phi)$ and $\sigma_t^2(\theta)$ and their derivatives. Indeed, NED does not necessarily carry over to arbitrary measurable transforms of

[2]This definition of size is slightly different from the conventional one, for example, Davidson (1994, p. 262). We use de Jong's (1997, Def. 1) definition because we use his central limit theorem for NED arrays.

https://doi.org/10.1017/S0266466619000367 Published online by Cambridge University Press

an NED process. α-Mixing, for example, implies ergodicity, it extends to measurable transforms, and is a subclass of NED. Lobato, Nankervis, and Savin (2002) impose a similar NED property. Nankervis and Savin (2010), who generalize the white noise test of Andrews and Ploberger (1996), allow for NED observed y_t, but mistakenly assume y_t is only L_2-NED.[3]

Remark 3. The requirement $\hat{\omega}_n(h) = \omega(h) + O_p(1/n^\kappa)$ will hold under suitable moment conditions, depending on how $\hat{\omega}_n(h)$ is constructed. If $\hat{\omega}_n(h)$ is a standard deviation for the sample correlation, for example, $\kappa = 1/2$ can hold under the existence of higher moments and a broad memory property like α-mixing, even for some kernel estimators (e.g., Andrews, 1991).

If $y_t = \epsilon_t$ is known then a filter is not required and Assumption 1 suffices for our main results. In this case, if y_t is i.i.d. under H_0, then it only needs to be L_2-bounded.

The next assumption is required if a filter is used. Let $\mathbf{0}_l$ be an l-dimensional zero vector. Define

$$G_t(\phi) \equiv \left[\frac{\partial}{\partial \phi'} f(x_{t-1}, \phi), 0'_{k_\delta}\right]' \in \mathbb{R}^{k_\theta} \text{ and } s_t(\theta) \equiv \frac{1}{2}\frac{\partial}{\partial \theta} \ln \sigma_t^2(\theta) \tag{2}$$

$$\mathcal{D}(h) \equiv E\left[\left(\epsilon_t s_t + \frac{G_t}{\sigma_t}\right)\epsilon_{t-h}\right] + E\left[\epsilon_t\left(\epsilon_{t-h}s_{t-h} + \frac{G_{t-h}}{\sigma_{t-h}}\right)\right] \in \mathbb{R}^{k_\theta}.$$

We do not require a filter for the above entities to make sense. If $y_t = \epsilon_t$, for example, then $G_t(\phi)$, $s_t(\theta)$ and therefore $\mathcal{D}(h)$ are each just zero.

We require notation that makes use of estimating equations $m_t \in \mathbb{R}^{k_m}$ and a matrix $\mathcal{A} \in \mathbb{R}^{k_\theta \times k_m}$ defined under Assumption 2.c. Define

$$r_t(h) \equiv \frac{\epsilon_t\epsilon_{t-h} - E[\epsilon_t\epsilon_{t-h}] - \mathcal{D}(h)'\mathcal{A}m_t}{E[\epsilon_t^2]} \text{ and } \rho(h) \equiv \frac{E[\epsilon_t\epsilon_{t-h}]}{E[\epsilon_t^2]} \tag{3}$$

$$z_t(h) \equiv r_t(h) - \rho(h)r_t(0) = \frac{\epsilon_t\epsilon_{t-h} - \rho(h)\epsilon_t^2 - (\mathcal{D}(h) - \rho(h)\mathcal{D}(0))'\mathcal{A}m_t}{E[\epsilon_t^2]}.$$

The process that arises in the key approximation is

$$\mathcal{Z}_n(h) \equiv \frac{1}{\sqrt{n}}\sum_{t=1+h}^{n} z_t(h). \tag{4}$$

Assumption 2 (Plug-in: response and identification)**.**

a. *Level response.* $f : \mathbb{R}^{k_x} \times \Phi \to \mathbb{R}$, where Φ is a compact subset of $\mathbb{R}^{k_\phi}$, $k_\phi \geq 0$; $f(x, \phi)$ is Borel measurable for each ϕ, and for each x three times continuously differentiable, where $(\partial/\partial\phi)^j f(x, \phi)$ is Borel measurable for each ϕ and $j = 1, 2, 3$; $E[\sup_{\phi\in\mathcal{N}_{\phi_0}} |(\partial/\partial\phi)^j f(x_t, \phi)|^6] < \infty$ for $j = 0, 1, 2, 3$ and some compact set with positive measure $\mathcal{N}_{\phi_0} \subseteq \Phi$ containing ϕ_0.

[3] A Gaussian central limit theorem requires the *product*, in our case $\epsilon_t\epsilon_{t-h}$, to be L_2-NED, which holds when ϵ_t is L_p-bounded, $p > 4$, and L_4-NED (Davidson, 1994, Thm. 17.9).

https://doi.org/10.1017/S0266466619000367 Published online by Cambridge University Press

b. *Volatility.* $\sigma_t^2 : \Theta \to [0,\infty)$ where $\Theta = \Phi \times \Delta \in \mathbb{R}^{k_\theta}$, and Δ is a compact subset of $\mathbb{R}^{k_\delta}$, $k_\delta \geq 0$; $\sigma_t^2(\theta)$ is $\mathcal{F}_{t-1}$-measurable, continuous, and three times continuously differentiable, where $(\partial/\partial\theta)^j \ln \sigma_t^2(\theta)$ is Borel measurable for each θ and $j = 1, 2, 3$; $\inf_{\theta\in\Theta}|\sigma_t^2(\theta)| \geq \iota > 0$ *a.s.* and $E[\sup_{\theta\in\mathcal{N}_{\theta_0}}|(\partial/\partial\theta)^j \ln \sigma_t^2(\theta)|^4] < \infty$ for $j = 0, 1, 2, 3$ and some compact subset $\mathcal{N}_{\theta_0} \subseteq \Theta$ containing θ_0.

c. *Estimator.* $\hat{\theta}_n \in \Theta$ for each n, and for a unique interior point $\theta_0 \in \Theta$, we have $\sqrt{n}(\hat{\theta}_n - \theta_0) = \mathcal{A}n^{-1/2}\sum_{t=1}^n m_t(\theta_0) + \mathcal{R}_m(n)$, where the $k_m \times 1$ stochastic remainder $\mathcal{R}_m(n) = O_p(n^{-\zeta})$ for some $\zeta > 0$, with $\mathcal{F}_t$ - measurable estimating equations $m_t = [m_{i,t}]_{i=1}^{k_m} : \Theta \to \mathbb{R}^{k_m}$ for $k_m \geq k_\theta$, and nonstochastic $\mathcal{A} \in \mathbb{R}^{k_\theta \times k_m}$. Moreover, zero-mean $m_t(\theta_0)$ is stationary, ergodic, $L_{r/2}$-bounded, and L_2-NED with size 1/2 on $\{\upsilon_t\}$, where $r > 4$ and $\{\upsilon_t\}$ appear in Assumption 1.b.

d. *Finite dimensional variance.* Let $\mathcal{L} \in \mathbb{N}$ be arbitrary, and let $\lambda \equiv [\lambda_h]_{h=1}^{\mathcal{L}} \in \mathbb{R}^{\mathcal{L}}$. Then $\liminf_{n\to\infty}\inf_{\lambda'\lambda=1} E[(\sum_{h=1}^{\mathcal{L}} \lambda_h \mathcal{Z}_n(h))^2] > 0$.

Remark 4. Smoothness (a) and (b) ensures a stochastic equicontinuity property for uniform laws of large numbers. Nondifferentiability can be allowed provided certain other smoothness conditions involving, for example, bracketing numbers apply (see, e.g., Pakes and Pollard, 1989; Arcones and Yu, 1994).

Remark 5. $E[\sup_{\phi\in\mathcal{N}_{\phi_0}}|(\partial/\partial\phi)^j f(x_t,\phi)|^4] < \infty$ and $E[\sup_{\theta\in\mathcal{N}_{\theta_0}}|(\partial/\partial\theta)^j \ln \sigma_t^2(\theta)|^4] < \infty$ are used to prove a uniform law of large numbers, where the former can imply higher moment bounds than in Assumption 1, depending on the response f. Fourth moments are required due to a required residual cross-product expansion. $E[\sup_{\theta\in\mathcal{N}_{\theta_0}}|(\partial/\partial\theta)^j \ln \sigma_t^2(\theta)|^4] < \infty$ holds for many linear and nonlinear volatility models, for example, GARCH, Quadratic GARCH, GJR-GARCH (Francq and Zakoïan, 2004, 2010). The sixth moment bound $E[\sup_{\phi\in\mathcal{N}_{\phi_0}}|(\partial/\partial\phi)^j f(x_t,\phi)|^6] < \infty$ is used to determine the rate of convergence of the correlation expansion approximation, which itself is used to bound the rate of increase of $\mathcal{L}_n$ in Lemma 2.1.

Remark 6. $\hat{\theta}_n$ under (c) asymptotically is a linear function of some zero-mean $\mathcal{F}_t$-measurable process $m_t(\theta_0)$. This includes M-estimators, Generalized Method of Moments (GMM), and (Generalized) Empirical Likelihood with smooth or nonsmooth estimating equations, and estimators with nonsmooth criteria and asymptotic expansions like Least Absolute Deviations and quantile regression. Typically, $m_t(\theta_0)$ is a function of u_t or ϵ_t and the gradients $(\partial/\partial\phi)f(x_t, \phi_0)$ and/or $(\partial/\partial\theta)\sigma_t^2(\theta_0)$, in which case $E[m_t] = 0$ represents an orthogonality condition that identifies θ_0, even if ϵ_t is not white noise. The assumption that m_t is NED in (c), in conjunction with Assumption 1, implies linear combinations of $\epsilon_t\epsilon_{t-h}$ and m_t are NED (Davidson, 1994, Thm. 17.8), which promotes Gaussian finite dimensional asymptotics for the residuals cross-product.

https://doi.org/10.1017/S0266466619000367 Published online by Cambridge University Press

Remark 7. The approximation error in (c) $\sqrt{n}(\hat{\theta}_n - \theta_0) = \mathcal{A}n^{-1/2}\sum_{t=1}^{n} m_t(\theta_0) + O_p(n^{-\zeta})$ is of order $n^{-\zeta}$ for some $\zeta > 0$. Generally $\zeta = 1/2$ under suitable regularity conditions. This allows us to describe the order of convergence for the remainder term in an asymptotic expansion of the sample correlation, which we require when deriving an upper bound on $\mathcal{L}_n$. All other technical arguments only require $\sqrt{n}(\hat{\theta}_n - \theta_0) = \mathcal{A}n^{-1/2}\sum_{t=1}^{n} m_t(\theta_0) + o_p(1)$.

Remark 8. (d) is a standard nondegeneracy assumption for finite dimensional asymptotics.

The theory developed in this article extends to a class of measurable functions of $[\sqrt{n}\hat{\rho}_n(h)]_{h=1}^{\mathcal{L}_n}$. Specifically

$$\vartheta : \mathbb{R}^{\mathcal{L}} \to [0,\infty) \text{ for arbitrary } \mathcal{L} \in \mathbb{N}, \tag{5}$$

which satisfies the following: $\vartheta(x)$ is continuous; lower bound $\vartheta(a) = 0$ *if and only if* $a = 0$; upper bound $\vartheta(a) \leq K\mathcal{L}\mathcal{M}$ for some $K > 0$ and any $a = [a_h]_{h=1}^{\mathcal{L}}$ such that $|a_h| \leq \mathcal{M}$ for each h; divergence $\vartheta(a) \to \infty$ as $||a|| \to \infty$; monotonicity $\vartheta(a_{\mathcal{L}_1}) \leq \vartheta([a'_{\mathcal{L}_1}, c'_{\mathcal{L}_2-\mathcal{L}_1}]')$ where $(a_{\mathcal{L}}, c_{\mathcal{L}}) \in \mathbb{R}^{\mathcal{L}}, \forall \mathcal{L}_2 \geq \mathcal{L}_1$ and any $c_{\mathcal{L}_2-\mathcal{L}_1} \in \mathbb{R}^{\mathcal{L}_2-\mathcal{L}_1}$; and the triangle inequality $\vartheta(a+b) \leq \vartheta(a) + \vartheta(b)\ \forall a, b \in \mathbb{R}^{\mathcal{L}_n}$. Examples include the maximum $\vartheta(a) = \max_{1\leq h\leq \mathcal{L}}|a_h|$, and sums $\vartheta(a) = \sum_{h=1}^{\mathcal{L}}|a_h|$ and $\vartheta(a) = \sum_{h=1}^{\mathcal{L}} a_h^2$, where $a = [a_h]_{h=1}^{\mathcal{L}}$. The lower bound $\vartheta(a) = 0$ *if and only if* $a = 0$ ensures we omit cases where test power is not asymptotically one. As one example, when $\tilde{\vartheta}(a) = \sum_{h=1}^{\mathcal{L}} a_h$ the statistic $\tilde{\vartheta}\left([\sqrt{n}\hat{\omega}_n(h)\hat{\rho}_n(h)]_{h=1}^{\mathcal{L}_n}\right)$ need not diverge under the alternative because $\tilde{\vartheta}(a) = 0$ is possible for $a \neq 0$.

We do not show that ϑ depends on $\mathcal{L}$ to reduce notation. The general test statistic is therefore

$$\hat{\mathcal{T}}_n \equiv \vartheta\left(\left[\sqrt{n}\hat{\omega}_n(h)\hat{\rho}_n(h)\right]_{h=1}^{\mathcal{L}_n}\right).$$

Both $\max_{1\leq h\leq \mathcal{L}_n}|\sqrt{n}\hat{\omega}_n(h)\hat{\rho}_n(h)|$ and a weighted portmanteau $n\sum_{h=1}^{\mathcal{L}_n}\hat{\omega}_n^2(h)\hat{\rho}_n^2(h)$ are covered. Note that the normalization $\vartheta([\sqrt{n}\hat{\omega}_n(h)\hat{\rho}_n(h)]_{h=1}^{\mathcal{L}_n}) = (2\mathcal{L}_n)^{-1/2}\sum_{h=1}^{\mathcal{L}_n}\hat{\omega}_n(h)\{n\hat{\rho}_n^2(h) - 1\}$ and similar normalized spectral density estimators used in Hong (1996, eqn. (3)) and Hong (2001) are not covered here because it violates positivity $\vartheta : \mathbb{R}^{\mathcal{L}} \to [0,\infty)$, lower bound $\vartheta(a) = 0$ *if and only if* a=0, and monotonicity. The fix $\vartheta([\sqrt{n}\hat{\omega}_n(h)\hat{\rho}_n(h)]_{hP=1}^{\mathcal{L}_n}) = (2\mathcal{L}_n)^{-1/2}|\sum_{h=1}^{\mathcal{L}_n}\hat{\omega}_n(h)\{n\hat{\rho}_n^2(h) - 1\}|$ still violates $\vartheta(a) = 0$ *if and only if* $a = 0$, and monotonicity.

The following result establishes a key (non-Gaussian) approximation theory for an increasing sequence of serial correlations. See Appendix A for all proofs. Recall $\kappa > 0$ in $\hat{\omega}_n(h) = \omega(h) + O_p(1/n^{\kappa})$ and $\zeta > 0$ in $\sqrt{n}(\hat{\theta}_n - \theta_0) = \mathcal{A}n^{-1/2}\sum_{t=1}^{n} m_t(\theta_0) + O_p(n^{-\zeta})$, cf. Assumptions 1c and 2c. These will determine an upper bound on $\mathcal{L}_n \to \infty$.

LEMMA 2.1. *Let Assumptions 1 and 2 hold. Then*

$$\begin{aligned}\tilde{\mathcal{X}}_n(h) &\equiv \left|\sqrt{n}\hat{\omega}_n(h)\left\{\hat{\rho}_n(h) - \rho(h)\right\} - \omega(h)\frac{1}{\sqrt{n}}\sum_{t=1+h}^{n}\{r_t(h) - \rho(h)r_t(0)\}\right| \\ &= O_p\left(1/n^{\min\{\zeta,\kappa,1/2\}}\right).\end{aligned} \tag{6}$$

https://doi.org/10.1017/S0266466619000367 Published online by Cambridge University Press

Moreover, for some nonunique monotonic sequence of positive integers $\{\mathcal{L}_n\}$, $\mathcal{L}_n \to \infty$ *and* $\mathcal{L}_n = o(n)$, *we have:* $|\vartheta(\sqrt{n}[\hat{\omega}_n(h)\{\hat{\rho}_n(h) - \rho(h)\}]_{h=1}^{\mathcal{L}_n}) - \vartheta([\omega(h)\mathcal{Z}_n(h)]_{h=1}^{\mathcal{L}_n})| \leq \vartheta([\sqrt{n}\hat{\omega}_n(h)\{\hat{\rho}_n(h) - \rho(h)\} - \omega(h)\mathcal{Z}_n(h)]_{h=1}^{\mathcal{L}_n}) \xrightarrow{p} 0$. *Therefore, under the null hypothesis*

$$\left| \vartheta\left(\left[\sqrt{n}\hat{\omega}_n(h)\hat{\rho}_n(h)\right]_{h=1}^{\mathcal{L}_n}\right) - \vartheta\left(\left[\omega(h)\frac{1}{\sqrt{n}}\sum_{t=1+h}^{n}\left\{\frac{\epsilon_t\epsilon_{t-h} - \mathcal{D}(h)'\mathcal{A}m_t}{E[\epsilon_t^2]}\right\}\right]_{h=1}^{\mathcal{L}_n}\right)\right| \xrightarrow{p} 0. \tag{7}$$

Finally, if $\vartheta(\cdot)$ *is the maximum transform, and* $(n^{\min\{\zeta,\kappa,1/2\}}/\ln(n))\tilde{\mathcal{X}}_n(h)$ *for all* h *is uniformly integrable, then* $\mathcal{L}_n = O(n^{\min\{\zeta,\kappa,1/2\}}/\ln(n))$ *must be satisfied.*

Remark 9. The sequence $\{\mathcal{L}_n\}$ is not unique because for any other $\{\mathring{\mathcal{L}}_n\}$, $\mathring{\mathcal{L}}_n \to \infty$ and $\lim_{n\to\infty}\{\mathring{\mathcal{L}}_n/\mathcal{L}_n\} \leq 1$, monotonicity $\vartheta(a_k) \leq \vartheta([a_k', c_{l-k}']') \ \forall a_k \in \mathbb{R}^k$ and $\forall c_{l-k} \in \mathbb{R}^{l-k}$ implies as $n \to \infty$:

$$\vartheta\left(\left[\sqrt{n}\hat{\omega}_n(h)\{\hat{\rho}_n(h) - \rho(h)\} - \omega(h)\mathcal{Z}_n(h)\right]_{h=1}^{\mathring{\mathcal{L}}_n}\right) \leq \vartheta\left(\left[\sqrt{n}\hat{\omega}_n(h)\{\hat{\rho}_n(h) - \rho(h)\} - \omega(h)\mathcal{Z}_n(h)\right]_{h=1}^{\mathcal{L}_n}\right) \xrightarrow{p} 0, \tag{8}$$

hence $|\vartheta(\sqrt{n}[\hat{\omega}_n(h)\{\hat{\rho}_n(h) - \rho(h)\}]_{h=1}^{\mathring{\mathcal{L}}_n}) - \vartheta([\omega(h)\mathcal{Z}_n(h)]_{h=1}^{\mathring{\mathcal{L}}_n})| \xrightarrow{p} 0$. Indeed, by an identical argument, trivially (8) applies for *any* positive integer sequence $\{\mathring{\mathcal{L}}_n\}$ that satisfies $\limsup_{n\to\infty}\{\mathring{\mathcal{L}}_n/\mathcal{L}_n\} < 1$, covering the case $\mathring{\mathcal{L}}_n \to (0, \infty)$. All subsequent results therefore extend to this general case, but we do not highlight it because it does not promote a consistent test.

Remark 10. An upper bound on $\mathcal{L}_n$ requires the mapping ϑ to be specified, so we work with the maximum. By Lemma A.2 in Appendix A, if $(n^{\min\{\zeta,\kappa,1/2\}}/\ln(n))\tilde{\mathcal{X}}_n(h)$ is uniformly integrable, then from standard arguments $\mathcal{L}_n = O(n^{\min\{\zeta,\kappa,1/2\}}/\ln(n))$ must hold. We do not tackle the case where uniform integrability fails to hold. Additional technical conditions laid out in Appendix G yield uniform integrability. In particular, we require $||1/\hat{\gamma}_n(0)||_p = O(1)$ for some $p > 1$, which generally cannot be easily verified. Furthermore, $E[\epsilon_t^6] < \infty$, the plug-in $\sqrt{n}||\hat{\theta}_n - \theta_0||_4 = O(1)$ and plug-in remainder $||n^{\lambda}\mathcal{R}_m(n)||_q = O(1)$ for some $\lambda > 0$ and $q > 2$, and test statistic weight $n^{\min\{\kappa,\zeta,1/2\}}||\hat{\omega}_n(h) - \omega(h)||_r = O(1)$ for some $r > 2$. Conditions like $\sqrt{n}||\hat{\theta}_n - \theta_0||_4 = O(1)$ generally require moment conditions higher than $E[\epsilon_t^6] < \infty$ (Appendix G: Ex. 1). These are relatively mild conditions and hold for most of the data generating processes under the simulation

https://doi.org/10.1017/S0266466619000367 Published online by Cambridge University Press

study in Section 4.[4] In those cases, $\zeta = 1/2$ and $\hat{\omega}_n(h) = \omega(h) = 1$ so that $\kappa = \infty$, hence $\mathcal{L}_n = O(\sqrt{n}/\ln(n))$.

The proof of Lemma 2.1 relies on a new twofold argument. It is new because it cannot rely on Gaussian approximation theory for high dimensional processes. First, we prove $\mathcal{A}_{\mathcal{L},n} \equiv \vartheta([\sqrt{n}\hat{\omega}_n(h)\{\hat{\rho}_n(h) - \rho(h)\} - \omega(h)\mathcal{Z}_n(h)]_{h=1}^{\mathcal{L}}) \xrightarrow{p} 0$ for each $\mathcal{L} \in \mathbb{N}$. Using standard weak convergence theory, this does not suffice to show $\mathcal{A}_{\mathcal{L}_n,n} \xrightarrow{p} 0$ for some $\mathcal{L}_n \to \infty$. This follows because weak convergence, in the broad sense of Hoffmann-Jørgensen (1984, 1991), to a Gaussian limit with a version that has uniformly bounded and uniformly continuous sample paths, is equivalent to convergence in finite dimensional distributions, the existence of a pseudo metric d on N such that (N, d) is a totally bounded pseudo metric space, and a stochastic equicontinuity property based on d holds. If d is the Euclidean distance, for example, then (N, d) is not totally bounded because $\mathbb{N}$ is not compact. See Dudley (1978, 1984) and Pollard (1990, Chap. 9 and 10). We take an approach different from Hoffman–Jorgensen's (1984) notion of weak dependence. We prove that $\mathcal{A}_{\mathcal{L},n} \xrightarrow{p} 0$ for each $\mathcal{L} \in \mathbb{N}$ directly implies $\mathcal{A}_{\mathcal{L}_n,n} \xrightarrow{p} 0$ for some sequence of positive integers $\{\mathcal{L}_n\}$ that satisfies $\mathcal{L}_n \to \infty$ and $\mathcal{L}_n = o(n)$. See Lemmas A.1– A.3 in Appendix A. Thus, by sidestepping the Hoffmann-Jørgensen (1984, 1991) view of weak dependence, which requires more than convergence in finite dimensional distributions, we are able to show that such convergence suffices. Our approach has deep roots in Ramsey (1930) theory, based on its implications for monotone subsequences (e.g., Boehme and Rosenfeld, 1974; Thomason, 1988; Myers, 2002) as applied to Frechét spaces (Boehme and Rosenfeld, 1974).

The same array argument, coupled with extant central limit theory for NED arrays, yields the following fundamental Gaussian approximation result for the Lemma 2.1 approximation process $\{\mathcal{Z}_n(h) : 1 \leq h \leq \mathcal{L}_n\}$. Recall $\mathcal{Z}_n(h) \equiv 1/\sqrt{n}\sum_{t=1+h}^{n} z_t(h)$ where $z_t(h) \equiv r_t(h) - \rho(h)r_t(0)$ and $r_t(h) \equiv \{\epsilon_t\epsilon_{t-h} - E[\epsilon_t\epsilon_{t-h}] - \mathcal{D}(h)'\mathcal{A}m_t\}/E[\epsilon_t^2]$.

LEMMA 2.2. *Let Assumptions 1 and 2 hold. Let* $\{\mathcal{Z}(h) : h \in \mathbb{N}\}$ *be a zero-mean Gaussian process with variance* $\lim_{n\to\infty} 1/n\sum_{s,t=1}^{n} E[z_s(h)z_t(h)] < \infty$, *and covariance function* $E[\mathcal{Z}(h)\mathcal{Z}(\tilde{h})] = \lim_{n\to\infty} 1/n\sum_{s,t=1}^{n} E[z_s(h)z_t(\tilde{h})]$. *Then for some* $\{\mathcal{Z}(h) : h \in \mathbb{N}\}$ *and some nonunique monotonic sequence of positive integers* $\{\mathcal{L}_n\}$, $\mathcal{L}_n \to \infty$ *and* $\mathcal{L}_n = o(n)$: $\vartheta([\omega(h)\mathcal{Z}_n(h)]_{h=1}^{\mathcal{L}_n}) \xrightarrow{d} \vartheta([\omega(h)\mathcal{Z}(h)]_{h=1}^{\infty})$.

Remark 11. If an estimator $\hat{\theta}_n$ is not required then $\mathcal{D}(h) = 0$ and the covariance function $E[\mathcal{Z}(h)\mathcal{Z}(\tilde{h})]$ reduces accordingly. If additionally ϵ_t is i.i.d. under the null then $E[\mathcal{Z}(h)\mathcal{Z}(\tilde{h})] = E[\epsilon_t^2\epsilon_{t-h}^2]/(E[\epsilon_t^2])^2$, which equals 1 if $h \neq 0$, and otherwise $E[\epsilon_t^4]/(E[\epsilon_t^2])^2$. If $\hat{\theta}_n$ is not required, then we can bypass our array convergence

[4] Some processes in the simulation study evidently fail to have higher moments, but are used to demonstrate the sensitivity of the proposed test to moment condition failure. See Section 4.

https://doi.org/10.1017/S0266466619000367 Published online by Cambridge University Press

argument and use the Gaussian approximation argument in Zhang and Wu (2017), under their moment contraction assumptions.

Remark 12. An upper bound on the rate $\mathcal{L}_n \to \infty$ can be provided in the maximum case under various dependence settings. For example, Chernozhukov, Chetverikov, and Kato (2014, Appendix B) impose boundedness and a β-mixing property, and Zhang and Cheng (2014), and Zhang and Wu (2017) work with a functional dependence property. Under their conditions a limit theory that supports our Lemma 2.1 expansion is evidently possible, while the bound on $\mathcal{L}_n \to \infty$ follows from our Lemma A.2 and results in Appendix G. In that case, their Thm. 3.2 will apply, hence $\mathcal{L}_n(\ln(\mathcal{L}_n))^{3q/2} = o(n^{q/2-1+\iota})$ for some $\iota > 0$, provided $E|z_t(h)|^q < \infty$ for all h and some $q \geq 4$.[5] The latter moment bound generally requires ϵ_t to be L_8-bounded. Put $q = 8$ to yield $\mathcal{L}_n(\ln(\mathcal{L}_n))^{12} = o\left(n^{3+\iota}\right)$. Hence $\mathcal{L}_n \to \infty$ as fast as $Kn^{3-\iota}$ for tiny $\iota > 0$ is allowed. Since we require $\mathcal{L}_n = o(n)$ for sample covariance consistency, the binding upper bound on $\mathcal{L}_n \to \infty$ comes from Lemma 2.1, for example, $\mathcal{L}_n = O(\sqrt{n}/\ln(n))$ under standard regularity conditions and $\hat{\omega}_n(h) = \omega(h) = 1$. We leave for future study a Gaussian approximation theory for high dimensional, heterogeneous, and possibly nonstationary NED processes.

Combine Lemmas 2.1 and 2.2 and invoke the triangle inequality to yield the following main result.

THEOREM 2.3. *Under Assumptions 1 and 2,* $\vartheta([\sqrt{n}\hat{\omega}_n(h)\{\hat{\rho}_n(h) - \rho(h)\}]_{h=1}^{\mathcal{L}_n}) \xrightarrow{d} \vartheta([\omega(h)\mathcal{Z}(h)]_{h=1}^{\infty})$ *for some monotonic sequence of positive integers* $\{\mathcal{L}_n\}$ *that is not unique,* $\mathcal{L}_n \to \infty$ *and* $\mathcal{L}_n = o(n)$*, where* $\{\mathcal{Z}(h) : h \in \mathbb{N}\}$ *is a zero-mean Gaussian process with variance* $\lim_{n\to\infty} n^{-1}\sum_{s,t=1}^{n} E[z_s(h)z_t(h)] < \infty$*, and covariance function* $\lim_{n\to\infty} n^{-1}\sum_{s,t=1}^{n} E[z_s(h)z_t(\tilde{h})]$*. Therefore, under the null hypothesis* $\vartheta([\sqrt{n}\hat{\omega}_n(h)\hat{\rho}_n(h)]_{h=1}^{\mathcal{L}_n}) \xrightarrow{d} \vartheta([\omega(h)\mathcal{Z}(h)]_{h=1}^{\infty})$*, where* $\{\mathcal{Z}(h) : h \in \mathbb{N}\}$ *is a zero-mean Gaussian process with variance* $\lim_{n\to\infty} n^{-1}\sum_{s,t=1}^{n} E[r_s(h)r_t(h)] < \infty$ *and* $r_t(h) \equiv \{\epsilon_t\epsilon_{t-h} - \mathcal{D}(h)'\mathcal{A}m_t\}/E[\epsilon_t^2]$*. Moreover, if* $\vartheta(\cdot)$ *is the maximum transform and* $(n^{\min\{\zeta,\kappa,1/2\}}/\ln(n))\tilde{\mathcal{X}}_n(h)$ *for all h is uniformly integrable, where* $\tilde{\mathcal{X}}_n(h)$ *is defined in (6), then* $\mathcal{L}_n = O(n^{\min\{\zeta,\kappa,1/2\}}/\ln(n))$ *must be satisfied.*

We now have a fundamental result for the maximum weighted autocorrelation under white noise.

COROLLARY 2.4. *Under Assumptions 1 and 2,* $\max_{1\leq h\leq \mathcal{L}_n}|\sqrt{n}\hat{\omega}_n(h)\{\hat{\rho}_n(h) - \rho(h)\}| \xrightarrow{d} \max_{1\leq h\leq\infty}|\omega(h)\mathcal{Z}(h)|$ *for some monotonic sequence of positive integers* $\{\mathcal{L}_n\}$ *that is not unique,* $\mathcal{L}_n \to \infty$ *and* $\mathcal{L}_n = o(n)$*, where* $\{\mathcal{Z}(h) : h \in \mathbb{N}\}$ *is defined in Theorem 2.3. Therefore, under the white noise null*

[5] See Zhang and Wu (2017, Thm. 3.2, p. 1900). They yield the optimal bound $\mathcal{L}_n(\ln(\mathcal{L}_n))^{3q/2} = o(n^{q/2-1})$, up to a multiplicative logarithmic term that is trumped by n^{ι} for any tiny $\iota > 0$.

https://doi.org/10.1017/S0266466619000367 Published online by Cambridge University Press

hypothesis $\max_{1\le h\le \mathcal{L}_n}|\sqrt{n}\hat{\omega}_n(h)\hat{\rho}_n(h)|\xrightarrow{d}\max_{1\le h\le\infty}|\omega(h)\mathcal{Z}(h)|$. *Furthermore, if* $(n^{\min\{\zeta,\kappa,1/2\}}/\ln(n))\tilde{\mathcal{X}}_n(h)$ *for all h is uniformly integrable, where* $\tilde{\mathcal{X}}_n(h)$ *is defined in (6), then* $\mathcal{L}_n = O(n^{\min\{\zeta,\kappa,1/2\}}/\ln(n))$ *must be satisfied.*

Remark 13. The conclusions of Theorem 2.3 and Corollary 2.4 do not require $\vartheta\left([\sqrt{n}\hat{\omega}_n(h)\hat{\rho}_n(h)]_{h=1}^{\mathcal{L}_n}\right)$ to have a well-defined limit law under the null. This is decidedly different from the max-correlation literature in which $\lim_{n\to\infty}\max_{1\le h\le\mathcal{L}_n}|\omega(h)\mathcal{Z}(h)|$ is characterized under suitable conditions that ensure asymptotic independence $E[\mathcal{Z}(i)\mathcal{Z}(j)]\to 0$ as $|i-j|\to 0$. See, for example, Leadbetter, Lindgren, and Rootzén (1983, Chap. 6) and Hüsler (1986). We do not require asymptotic independence, nor therefore convergence in law.

2.2. Bootstrapped *p*-Value Test

We work with Shao's (2011) dependent wild bootstrap. Recall $m_t(\theta)$ are the estimating equations for $\hat{\theta}_n$, let $\widehat{\mathcal{A}}_n$ be a consistent estimator of $\mathcal{A}$ in Assumption 2.c, and define

$$\hat{\mathcal{D}}_n(h)\equiv\frac{1}{n}\sum_{t=h+1}^{n}\left\{\left(\epsilon_t(\hat{\theta}_n)s_t(\hat{\theta}_n)+\frac{G_t(\hat{\theta}_n)}{\sigma_t(\hat{\theta}_n)}\right)\epsilon_{t-h}(\hat{\theta}_n)+\epsilon_t(\hat{\theta}_n)\right.$$
$$\left.\times\left(\epsilon_{t-h}(\hat{\theta}_n)s_{t-h}(\hat{\theta}_n)+\frac{G_{t-h}(\hat{\theta}_n)}{\sigma_{t-h}(\hat{\theta}_n)}\right)\right\}. \tag{9}$$

We now operate on an approximation of $\epsilon_t(\hat{\theta}_n)\epsilon_{t-h}(\hat{\theta}_n)$ expanded around θ_0 under H_0, cf. Lemma 2.1

$$\widehat{\mathcal{E}}_{n,t,h}(\hat{\theta}_n)\equiv\epsilon_t(\hat{\theta}_n)\epsilon_{t-h}(\hat{\theta}_n)-\hat{\mathcal{D}}_n(h)'\widehat{\mathcal{A}}_n m_t(\hat{\theta}_n).$$

In practice, $G_t(\theta)$ and $\sigma_t(\theta)$ are typically unobserved and must be iteratively approximated based on initial conditions. Examples include linear and nonlinear AR-GARCH models. In such cases, $\hat{\mathcal{D}}_n(h)$ is infeasible. Meitz and Saikkonen (2011), among others, lay out sufficient conditions for the Quasi-Maximum Likelihood (QML) estimator for a large class of Autoregressive-Generalized Autoregressive Conditional Heteroscedasticiy (AR-GARCH) models to be consistent and asymptotically normal, including smoothness conditions similar to Assumption 2 that include Lipschitz properties imposed on $f(x_t,\phi)$ and $\sigma_t(\theta)$. In their setting, initial conditions vanish geometrically fast and therefore do not play a role in asymptotics both for the QML estimator, and for sample statistics like a feasible version of $\hat{\mathcal{D}}_n(h)$. See their Assumptions DGP, E, and C1–C3.

2.3. Dependent Wild Bootstrap

The wild bootstrap is proposed for i.i.d. and m.d.s. sequences (Wu, 1986; Liu, 1988; Hansen, 1996). Shao (2010, 2011) generalizes the idea to allow for

https://doi.org/10.1017/S0266466619000367 Published online by Cambridge University Press

dependent sequences. Shao (2010) allows for general dependence by using block-wise i.i.d. random draws as weights, with a covariance function that equals a kernel function. His requirements rule out a truncated kernel, but allow a Bartlett kernel among others. We follow Shao (2011) whose draws effectively have a truncated kernel covariance function.

The algorithm is as follows. Set a block size b_n such that $1 \leq b_n < n$, $b_n \to \infty$ and $b_n/n \to 0$. Denote the blocks by $\mathcal{B}_s = \{(s-1)b_n + 1, \ldots, sb_n\}$ with $s = 1, \ldots, n/b_n$. Assume for simplicity that the number of blocks n/b_n is an integer. Generate i.i.d. random numbers $\{\xi_1, \ldots, \xi_{n/b_n}\}$ with $E[\xi_i] = 0$, $E[\xi_i^2] = 1$, and $E[\xi_i^4] < \infty$. Define an auxiliary variable $\varphi_t = \xi_s$ if $t \in \mathcal{B}_s$. Compute $\hat{\mathcal{T}}_n^{(dw)} \equiv \vartheta([\sqrt{n}\hat{\rho}_n^{(dw)}(h)]_{h=1}^{\mathcal{L}_n})$ from

$$\hat{\rho}_n^{(dw)}(h) \equiv \frac{1}{1/n\sum_{t=1}^n \epsilon_t^2(\hat{\theta}_n)} \frac{1}{n} \sum_{t=1+h}^n \varphi_t \left\{ \widehat{\mathcal{E}}_{n,t,h}(\hat{\theta}_n) - \frac{1}{n} \sum_{s=1+h}^n \widehat{\mathcal{E}}_{n,s,h}(\hat{\theta}_n) \right\}. \tag{10}$$

Repeat M times, resulting in bootstrapped statistics $\{\hat{\mathcal{T}}_{n,i}^{(dw)}\}_{i=1}^M$, and an approximate p-value $\hat{p}_{n,M}^{(dw)} \equiv 1/M\sum_{i=1}^M I(\hat{\mathcal{T}}_{n,i}^{(dw)} \geq \hat{\mathcal{T}}_n)$. The test proposed rejects the null at nominal size α when $\hat{p}_{n,M}^{(dw)} < \alpha$. The wild bootstrap has block size $b_n = 1$ and no re-centering with $1/n\sum_{s=1+h}^n \widehat{\mathcal{E}}_{n,s,h}(\hat{\theta}_n)$.

We use a sample version of the first-order expansion variable $\epsilon_t\epsilon_{t-h} - \mathcal{D}(h)'\mathcal{A}m_t$ from Lemma 2.1. It is *incorrect* to use just $\epsilon_t(\hat{\theta}_n)\epsilon_{t-h}(\hat{\theta}_n)$, as with

$$\hat{\rho}_n^{(dw)}(h) \equiv \frac{1}{1/n\sum_{t=1}^n \epsilon_t^2(\hat{\theta}_n)} \frac{1}{n} \sum_{t=1+h}^n \varphi_t \left\{ \epsilon_t(\hat{\theta}_n)\epsilon_{t-h}(\hat{\theta}_n) - \frac{1}{n} \sum_{s=1+h}^n \epsilon_t(\hat{\theta}_n)\epsilon_{t-h}(\hat{\theta}_n) \right\}. \tag{11}$$

This follows since φ_t is mean zero and independent of the data, hence $1/n\sum_{t=1+h}^n \varphi_t\epsilon_t(\hat{\theta}_n)\epsilon_{t-h}(\hat{\theta}_n) = 1/n\sum_{t=1+h}^n \varphi_t\epsilon_t\epsilon_{t-h} + o_p(1/\sqrt{n})$, yet $1/n\sum_{s=1+h}^n \epsilon_t(\hat{\theta}_n)\epsilon_{t-h}(\hat{\theta}_n) = E[\epsilon_t\epsilon_{t-h}] + O_p(1/\sqrt{n})$ by standard first-order arguments and $E[m_t] = 0$. Hence, $\sqrt{n}\hat{\rho}_n^{(dw)}(h)$ from (11) is equivalent to $1/\sqrt{n}\sum_{t=1+h}^n \varphi_t\epsilon_t\epsilon_{t-h}/E[\epsilon_t^2]$ asymptotically with probability approaching one, which under the null has the same asymptotic properties as $1/\sqrt{n}\sum_{t=1+h}^n \epsilon_t\epsilon_{t-h}/E[\epsilon_t^2]$. The latter is not equivalent to the Lemma 2.1 first-order expansion process $\{\mathcal{Z}_n(h)\}$ because asymptotic information from the estimator $\hat{\theta}_n$ has been scrubbed out by the bootstrap variable φ_t. The bootstrapped $\hat{\rho}_n^{(dw)}(h)$ in (10), however, contains the required information.

Shao (2011) imposes Wu's (2005) moment contraction property with an eighth moment, which we denote MC_8 (for details, see Appendix B). He then applies a Hilbert space approach for weak convergence of a spectral density process $\{\hat{S}_n(\lambda) : \lambda \in [0, \pi]\}$ to yield convergence for $\int_0^\pi \hat{S}_n^2(\lambda)d\lambda$.[6] Only observed data are considered. There are several reasons why a different approach is required here. First, $\hat{S}_n(\lambda)$ is a sum of all $\{\hat{\gamma}_n(h) : 1 \leq h \leq n-1\}$, and Shao (2011, proof of Thm.

[6] See, for example, Politis and Romano (1994) for applications of weak convergence in a Hilbert space to bootstrapped statistics.

https://doi.org/10.1017/S0266466619000367 Published online by Cambridge University Press

3.1) uses a variance of conditional variance bound for probability convergence based on Chebyshev's inequality. This requires $E[\epsilon_t^8] < \infty$ and a complicated eighth order joint cumulant series bound which is only known to hold when ϵ_t is *geometric* MC_8 (see Shao and Wu, 2007). Second, we only need convergence in distribution of $\sqrt{n}\hat{\gamma}_n(h)$, coupled with a new array convergence result, which are easier to handle than weak convergence of $\{\hat{S}_n(\lambda) : \lambda \in [0, \pi]\}$ on a Hilbert space. Third, in the Hilbert space approach, the supremum is not a continuous mapping from the space of square integrable (with respect to Lebesgue measure) functions on $[0, \pi]$. It is therefore not clear how, or if, Shao's (2011, Thm. 3.1) proof applies to our statistic.

In order to prove that the bootstrapped $\hat{\rho}_n^{(dw)}(h)$ has the same finite dimensional limit distributions as $\hat{\rho}_n(h)$ under the null, it is helpful to have the equations $m_t(\theta)$ in the Assumption 2.c expansion $\sqrt{n}(\hat{\theta}_n - \theta_0) = \mathcal{A}n^{-1/2}\sum_{t=1}^n m_t(\theta_0) + o_p(1)$ to be a smooth parametric function for a required uniform law of large numbers. As with response smoothness under Assumption 2.a,b, more general smoothness properties are achievable at the expense of more intense notation.[7]

Assumption 2.c′. $\hat{\theta}_n \in \Theta$ for each n, and for a unique interior point $\theta_0 \in \Theta$ we have $\sqrt{n}(\hat{\theta}_n - \theta_0) = \mathcal{A}n^{-1/2}\sum_{t=1}^n m_t(\theta_0) + \mathcal{R}_m(n)$ where the $k_m \times 1$ stochastic remainder $\mathcal{R}_m(n) = O_p(n^{-\zeta})$ for some $\zeta > 0$, with $\mathcal{F}_t$-measurable estimating equations $m_t = [m_{i,t}]_{i=1}^{k_m} : \Theta \to \mathbb{R}^{k_m}$ for $k_m \geq k_\theta$; and nonstochastic $\mathcal{A} \in \mathbb{R}^{k_\theta \times k_m}$. $m_t(\theta)$ is twice continuously differentiable, $(\partial/\partial\theta)^j m_t(\theta)$ is Borel measurable for each θ and $j = 1, 2$, and $E[\sup_{\theta\in\Theta}|(\partial/\partial\theta)^i m_{j,t}(\theta)|] < \infty$ for each $i = 0, 1, 2$ and $j = 1,..., k_m$. Moreover, zero-mean m_t is stationary, ergodic, $L_{r/2}$-bounded, and L_2-NED with size 1/2 on $\{\upsilon_t\}$, where $r > 4$ and $\{\upsilon_t\}$ appear in Assumption 1.b. □

The bootstrapped p-value leads to a valid and consistent test. Note $\kappa > 0$ in $\hat{\omega}_n(h) = \omega(h) + O_p(1/n^\kappa)$ and $\zeta > 0$ in $\sqrt{n}(\hat{\theta}_n - \theta_0) = \mathcal{A}n^{-1/2}\sum_{t=1}^n m_t(\theta_0) + O_p(n^{-\zeta})$, cf. Assumptions 1.c and 2.c′.

THEOREM 2.5. *Let Assumptions 1, 2.a,b,c′,d hold, and let the number of bootstrap samples $M = M_n \to \infty$. There exists a nonunique monotonic sequence of maximum lags $\{\mathcal{L}_n\}$, $\mathcal{L}_n \to \infty$ and $\mathcal{L}_n = o(n)$, such that under H_0, $P(\hat{p}_{n,M}^{(dw)} < \alpha) \to \alpha$, and if H_0 is false then $P(\hat{p}_{n,M}^{(dw)} < \alpha) \to 1$. Furthermore, $\mathcal{L}_n = O(n^{\min\{\zeta,\kappa,1/2\}}/\ln(n))$ must be satisfied.*

Remark 14. A similar theory applies to an approximate p-value computed by wild bootstrap where φ_t is i.i.d. $N(0, 1)$, provided ϵ_t forms an m.d.s. under the null.

Remark 15. The test operates on $\sqrt{n}\hat{\rho}_n(h)$ and $\sqrt{n}\hat{\rho}_n^{(dw)}(h)$ and therefore achieves the parametric rate of local asymptotic power against the sequence of

[7] Nonsmoothness can be allowed provided certain bracketing or other smoothness properties are applied like a Lipschitz condition or the Vapnick–Chervonenkis class, which ensure a required stochastic equicontinuity condition. See, for example, Andrews (1987), Arcones and Yu (1994) and Gaenssler and Ziegler (1994).

https://doi.org/10.1017/S0266466619000367 Published online by Cambridge University Press

alternatives: $H_1^L : \rho(h) = r(h)/\sqrt{n}$ for each h where $r(h)$ are fixed constants, $|r(h)| \leq \sqrt{n}$. See Appendix D, especially Thm. D.1.

Remark 16. The bound $\mathcal{L}_n = O(n^{\min\{\zeta,\kappa,1/2\}}/\ln(n))$ generally must hold. A uniform integrability condition is not imposed here, as it is in Lemma 2.1, cf. Lemma A.2.a, since the proof operates on conditional probabilities. The latter are bounded and therefore uniformly integrable. Furthermore, the conditional probabilities imbed any given transform ϑ over lags $1, ..., \mathcal{L}$. The maximum transform requirement from Lemma A.2 for bounding $\mathcal{L}_n$ is merely applied to the conditional probabilities themselves over lags $\mathcal{L} \in \{1, ..., \mathcal{L}_n\}$. See the proof of Theorem 2.5, cf. Lemma A.7.b.

Remark 17. Hong's (1996, Sect. 2) encompassing class of statistics, which includes $(2\mathcal{L}_n)^{-1/2} \sum_{h=1}^{\mathcal{L}_n} \hat{\omega}_n(h)\{n\hat{\rho}_n^2(h) - 1\}$, does not achieve the parametric rate of convergence due to the normalizing term $\mathcal{L}_n^{-1/2}$. The rate logically is $n^{1/2}/\mathcal{L}_n^{1/4}$ hence Hong's (1996) class of statistics have nontrivial power against $n^{1/2}/\mathcal{L}_n^{1/4}$ -local alternatives. We by-pass self-normalization by working solely in a bootstrap framework based on finite dimensional asymptotics. As noted above, our transform class ϑ does not allow for self-normalized statistics. Moreover, we do not need to know the limit distribution of $\vartheta([\sqrt{n}\hat{\omega}_n(h)\hat{\rho}_n(h)]_{h=1}^{\mathcal{L}_n})$, nor even be guaranteed that it has one. Our approach eases the burden of self-normalization with an increasing maximum lag: we retain $\sqrt{n}$-asymptotics and therefore $\sqrt{n}$-local power, even for Hong's (1996) (non-normalized) $\sum_{h=1}^{\mathcal{L}_n} \hat{\omega}_n(h) n\hat{\rho}_n^2(h)$.

3. AUTOMATIC MAXIMUM LAG SELECTION

We approach lag selection from the perspective of the practitioner by providing a data-driven, or automatic, lag selection method. Our method closely follows Escanciano and Lobato (2009), whose work is motivated by the automatic Neyman test proposed in Inglot and Ledwina (2006). Let $\mathcal{L}_n^*$ denote the data-driven lag selected. Under H_0, Escanciano and Lobato's (2009) method leads to $P(\mathcal{L}_n^* = 1) \to 1$ because higher lags do not provide useful information and incur a high penalty for their use (see below for details). Contrary to their Q-test method, however, we allow $\mathcal{L}_n \to \infty$ and by using a bootstrap, we do not need to standardize the sample autocorrelations.

In order to ease notation, we only work with the max-correlation statistic and weight $\hat{\omega}_n(h) = 1$, but all subsequent results carry over to the general transform ϑ and general $\hat{\omega}_n(h) \xrightarrow{p} \omega(h) > 0$ with few additional proof steps. Hence $\kappa = \infty$ in Assumption 1.c.

The optimal lag $\mathcal{L}_n^*$ is chosen from a set $\{1, ..., \bar{\mathcal{L}}_n\}$ for some prechosen upper-bound $\bar{\mathcal{L}}_n \to \infty$. In the case of the maximum and $\hat{\omega}_n(h) = \omega(h) = 1$, we have from expansion Lemma 2.1 and dependent wild bootstrap Theorem 2.5 that $\bar{\mathcal{L}}_n = O(n^{\min\{\zeta,1/2\}}/\ln(n))$ must hold, where $\zeta > 0$ appears in the Assumption 2.c or 2.c′ plug-in expansion $\sqrt{n}(\hat{\theta}_n - \theta_0) = \mathcal{A}n^{-1/2} \sum_{t=1}^n m_t(\theta_0) + O_p(n^{-\zeta})$. Under standard regularity conditions many plug-in estimators will satisfy $\zeta = 1/2$, hence

https://doi.org/10.1017/S0266466619000367 Published online by Cambridge University Press

$\bar{\mathcal{L}}_n = O(\sqrt{n}/\ln(n))$. We use the integer part of $\delta\sqrt{n}/(\ln(n))$ for certain $\delta > 0$ in our simulation study below. We only consider sequences $\{\mathcal{L}_n\}$ that satisfy $\mathcal{L}_n/\bar{\mathcal{L}}_n \to [0, K]$ for any finite $K > 0$ and we assume the results of Section 2 hold for any such $\{\mathcal{L}_n\}$. We save notation by fixing $K = 1$.

We also need to allow for selection of *any* positive integer sequence $\{\mathcal{L}_n\}$ that satisfies $\mathcal{L}_n/\bar{\mathcal{L}}_n \to [0, 1]$, hence $\mathcal{L}_n \to (0, \infty]$ is assumed such that $\mathcal{L}_n \to \mathcal{L}$, a finite positive integer, is possible. This is required because Escanciano and Lobato's (2009) method leads to $P(\mathcal{L}_n^* = 1) \to 1$ under H_0. See Remark 9 for discussion of the validity of our main results when $\mathcal{L}_n \to (0, \infty)$.

Escanciano and Lobato (2009) work with a penalized Q-statistic, with a penalty that is an increasing function of the number of included lags. Similarly, define the *penalized max-correlation* test statistic

$$\hat{\mathcal{T}}_n^{\mathcal{P}}(\mathcal{L}) \equiv \hat{\mathcal{T}}_n(\mathcal{L}) - \mathcal{P}_n(\mathcal{L}) \text{ where } \hat{\mathcal{T}}_n(\mathcal{L}) \equiv \sqrt{n} \max_{1 \le h \le \mathcal{L}} \left|\hat{\rho}_n(h)\right| \tag{12}$$

with penalty function $\mathcal{P}_n(\cdot)$

$$\mathcal{P}_n(\mathcal{L}) = \begin{cases} \sqrt{\mathcal{L}\ln n} & \text{if } \hat{\mathcal{T}}_n(\mathcal{L}) \le \sqrt{q\ln n} \\ \sqrt{2\mathcal{L}} & \text{if } \hat{\mathcal{T}}_n(\mathcal{L}) > \sqrt{q\ln n}, \end{cases} \tag{13}$$

where q is a fixed positive constant. A small value of q leads to the AIC penalty $\sqrt{2\mathcal{L}}$ being chosen with high probability, while a large q promotes selection of the BIC penalty. Escanciano and Lobato (2009) use $q = 2.4$, a choice motivated by their own simulation evidence, and evidence from Inglot and Ledwina (2006). Inglot and Ledwina (2006) develop an automatic Neyman test, and the portmanteau test explored in Escanciano and Lobato (2009) belongs to a class of smooth tests proposed in Neyman (1937). Hence, it is not surprising that their q values are similar. We find a slightly larger value $q = 3$ leads to strong results across null and alternative hypotheses for our test: see the discussion in Section 4.1, and see Figure 1.

The chosen maximum lag $\mathcal{L}_n^*$ for each n is

$$\mathcal{L}_n^* = \min\left\{\mathcal{L}_n : 1 \le \mathcal{L}_n \le \bar{\mathcal{L}}_n : \hat{\mathcal{T}}_n^{\mathcal{P}}(\mathcal{L}_n) \ge \hat{\mathcal{T}}_n^{\mathcal{P}}(l) \text{ for each } l = 1, ..., \bar{\mathcal{L}}_n\right\}. \tag{14}$$

We chose $\{\mathcal{L}_n\}$ from those integer sequences satisfying $\mathcal{L}_n \ge 1$ and $\mathcal{L}_n \le \bar{\mathcal{L}}_n$ to ensure $\mathcal{L}_n/\bar{\mathcal{L}}_n \to [0, 1]$ holds in practice, but in theory we may select *any* $\{\mathcal{L}_n\}$ such that $\mathcal{L}_n \ge 1$ and $\mathcal{L}_n/\bar{\mathcal{L}}_n \to [0, 1]$. Notice l may be a function of n, for example, $l = \bar{\mathcal{L}}_n - 1$. The penalties $(\sqrt{\mathcal{L}\ln n}, \sqrt{2\mathcal{L}})$ are related to Escanciano and Lobato's (2009, p. 144) penalties $(\mathcal{L}\ln n, 2\mathcal{L})$ for a fixed horizon Q-statistic. We need the square root because the max-correlation operates on $\sqrt{n}\hat{\rho}_n(h)$ rather than $n\hat{\rho}_n^2(h)$. Contrary to Escanciano and Lobato (2009), however, our test statistic *and* penalty are based on the max-correlation, we allow for diverging sequences $\{\mathcal{L}_n\}$, and we do not need to standardize the correlations because we use a bootstrap.[8]

[8]Escanciano and Lobato (2009, second remark following Thm. 2) claim that a diverging maximum lag is possible for their Q-test with an automatic lag, but an asymptotic theory is not presented. Furthermore, it is not obvious that

https://doi.org/10.1017/S0266466619000367 Published online by Cambridge University Press

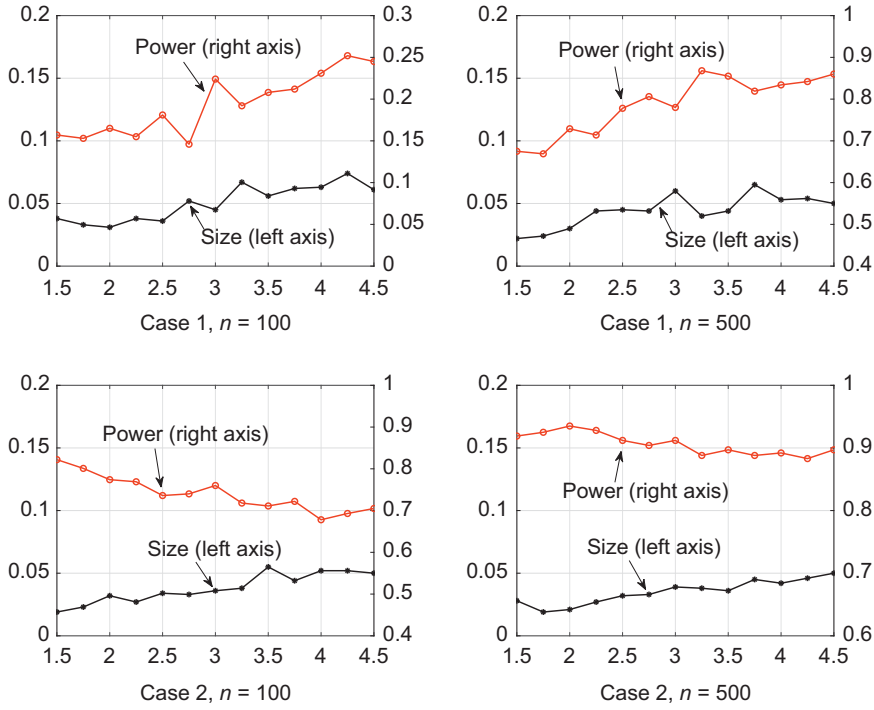


Fig. 1- Colour online, B/W in print

FIGURE 1. Empirical size and size-adjusted power of $\hat{\mathcal{T}}^{dw}(\mathcal{L}_n^*)$ with $\alpha = 0.05$. Note: We plot empirical size and size-adjusted power of the bootstrapped max-correlation test with automatic lag selection given nominal size 5%. In Case 1, the empirical size and empirical quantiles for size adjustment are computed under Scenario #1 (i.i.d. y_t and mean filter) with i.i.d. error; then the size-adjusted power is computed under Scenario #4 (AR(2) y_t and AR(1) filter) with i.i.d. error. In Case 2, the empirical size and empirical quantiles for size adjustment are computed under Scenario #5 (GARCH y_t and no filter) with i.i.d. error; then the size-adjusted power is computed under Scenario #5 with MA(2) error. The tuning parameter that affects the penalty term $\mathcal{P}_n(\mathcal{L})$ is $q \in \{1.50, 1.75, \ldots, 4.50\}$. The largest possible lag length is $\bar{\mathcal{L}}_n = [10\sqrt{n}/(\ln n)]$, which implies that $\bar{\mathcal{L}}_{100} = 21$ and $\bar{\mathcal{L}}_{500} = 35$. $J = 1{,}000$ Monte Carlo samples and $M = 500$ bootstrap samples are generated.

Define $\rho(\infty) \equiv \lim_{h\to\infty} \rho(h)$, and $h^* \equiv \min\{h : h = \arg\max_{1\le h\le\infty} |\rho(h)|\}$, the smallest lag at which the largest correlation in magnitude occurs.

THEOREM 3.1. *Let Assumptions 1 and 2 hold.* (*a*) *Under* H_0, *if* $(n^{\min\{\zeta,\kappa,1/2\}}/\ln(n))\tilde{\mathcal{X}}_n(h)$ *for all* h *is uniformly integrable, where* $\tilde{\mathcal{X}}_n(h)$ *is defined in* (6), *then* $\mathcal{L}_n = O(n^{\min\{\zeta,\kappa,1/2\}}/\ln(n))$ *must hold, and* $P(\mathcal{L}_n^* = 1) \to 1$ (*b*) *Under* H_1, $\mathcal{L}_n^* \xrightarrow{p} h^*$ *provided* $\bar{\mathcal{L}}_n = o(n/\ln(n))$.

their current fixed maximum lag proof can extend to the unbounded maximum lag case. By their eqn. (11), they need $\sum_{l=1}^{\bar{\mathcal{L}}_n} P(\mathcal{L}_n = l) \to 0$ *as* $n \to \infty$ hence $P(\mathcal{L}_n = l) \to 0$ fast enough when $\bar{\mathcal{L}}_n \to \infty$, which may not hold under their current assumptions.

https://doi.org/10.1017/S0266466619000367 Published online by Cambridge University Press

Remark 18. We only require the more lenient $\bar{\mathcal{L}}_n = o(n/\ln(n))$ for the proof of (b). The more stringent restriction $\bar{\mathcal{L}}_n = O(n^{\min\{\zeta, 1/2\}}/\ln(n))$ arises under (a) since there we must prove $P(\mathcal{P}_n(\mathcal{L}_n) = \sqrt{\mathcal{L}_n \ln(n)}) \to 1$ by using Lemma 2.1.

Remark 19. Under H_1, the optimal lag selected satisfies $\mathcal{L}_n^* \xrightarrow{p} h^*$. Notice h^* may be any value in $\mathbb{N}$ because we allow the maximum lag under consideration for finite samples to diverge $\bar{\mathcal{L}}_n \to \infty$. This ensures a consistent white noise test. The reason h^* is selected asymptotically is the penalized max-correlation favors choosing lags that are *at least* as large as the most informative lag(s), the lag(s) at which the max-correlation takes place. A nice advantage of the procedure is $\mathcal{L}_n^*$ converges to the *smallest* of such *most informative lags*, ensuring as $n \to \infty$ that the greatest number of data points possible are used for computing that correlation magnitude. A portmanteau statistic, however, sums over *all* squared correlations over a finite set of lags, hence its version is optimized at the largest fixed lag $\bar{h}$ under consideration, hence $P(\mathcal{L}_n^* = \bar{h}) \to 1$ (see the proof of Thm. 2 in Escanciano and Lobato, 2009).

Remark 20. The proof that $\mathcal{L}_n^*$ converges to 1 in probability under H_0 only exploits the null property $\sqrt{n}\hat{\rho}_n(h) = O_p(1)$. The latter *also* holds under the $\sqrt{n}$-local alternative $\rho(h) = r(h)/\sqrt{n}$, since $\sqrt{n}\hat{\rho}_n(h) = \sqrt{n}(\hat{\rho}_n(h) - \rho(h)) + r(h) = O_p(1)$. Thus, $P(\mathcal{L}_n^* = 1) \to 1$ under $H_1^L : \rho(h) = r(h)/\sqrt{n}$ as well. This means the max-correlation test with our proposed automatic lag selection will have trivial asymptotic local power against all directions from the null with $r(1) = 0$.[9]

Remark 21. In our proof, for example, under H_0, we show $\hat{\mathcal{T}}_n^{\mathcal{P}}(\mathcal{L}_n) \geq \hat{\mathcal{T}}_n^{\mathcal{P}}(l)$ for each $l = 1, \ldots, \bar{\mathcal{L}}_n$ *if and only if* $\mathcal{L}_n \to 1$, while by definition $\mathcal{L}_n^*$ is the least of all such sequences. We do this by inspecting an equivalent expression for $P(\hat{\mathcal{T}}_n^{\mathcal{P}}(\mathcal{L}_n) \geq \hat{\mathcal{T}}_n^{\mathcal{P}}(l))$ *for each* $1 \leq l \leq \bar{\mathcal{L}}_n$ (equivalence holds asymptotically with probability approaching one). Escanciano and Lobato (2009, proof of Thm. 1), by contrast, look at the joint probability that $\mathcal{L}_n^* \neq 1$, hence they must show $\sum_{l=2}^{\bar{\mathcal{L}}} P(\mathcal{L}_n^* = l) \to 0$ where $\bar{\mathcal{L}}$ is fixed and finite (see Escanciano and Lobato, 2009, eqn. (11)). The joint probability argument does not obviously transfer to the case where $\bar{\mathcal{L}} = \bar{\mathcal{L}}_n \to \infty$ since $\sum_{l=2}^{\bar{\mathcal{L}}_n} P(\mathcal{L}_n^* = l) \to 0$ requires $P(\mathcal{L}_n^* = l) \to 0$ sufficiently fast, which need not hold under their assumptions.

4. MONTE CARLO EXPERIMENTS

We now perform a Monte Carlo experiment to gauge the merits of the max-correlation test and automatic lag (labeled $\hat{\mathcal{T}}^{dw}(\mathcal{L}_n^*)$). A main competitor studied here is a Shao's (2011) dependent wild bootstrap spectral Cramér-von Mises test (CvM^{dw}). See Section 4.1 for the simulation design and Section 4.2 for results. In the supplemental material (Appendix H), we study other tests, including the

[9] Simply consider $r(h) = 0 \; \forall h \neq 2$ and $r(2) \neq 0$. Since $\mathcal{L}_n^* \xrightarrow{p} 1$, the above local drift cannot be detected asymptotically (with probability greater than the size of the test). We thank a referee for pointing this out.

https://doi.org/10.1017/S0266466619000367 Published online by Cambridge University Press

max-correlation with a prechosen nonrandom lag $\mathcal{L}_n$, the Ljung–Box test, Hong's (1996) test based on a standardized periodogram, a CvM test with Zhu and Li's (2015) block-wise random weighting bootstrap, and Andrews and Ploberger's (1996) sup-LM test with the dependent wild bootstrap. CvM^{dw} is one of the strongest competitors in terms of empirical size and power.

4.1. Simulation Design

We consider a variety of data generating processes, filters, and estimation methods. We first construct an error term e_t that drives an observed variable y_t. Let ν_t be i.i.d. $N(0, 1)$. We consider i.i.d. $e_t = \nu_t$; GARCH(1,1) $e_t = \nu_t w_t$ with random volatility process $w_1^2 = 1$ and $w_t^2 = 1 + 0.2e_{t-1}^2 + 0.5w_{t-1}^2$ for $t \geq 2$; MA(2) $e_t = \nu_t + 0.5\nu_{t-1} + 0.25\nu_{t-2}$ for $t \geq 3$, with initial values $e_1 = 0$ and $e_2 = \nu_2 + 0.5\nu_1$; and AR(1) $e_t = 0.7e_{t-1} + \nu_t$ for $t \geq 2$ with initial $e_1 = 0$. Each error process is strictly stationary and ergodic.[10] We use each of the four error terms in each of the following six scenarios.

Scenario #1: Simple $y_t = e_t$; mean filter $\epsilon_t = y_t - E[y_t]$; $\hat{\phi}_n = 1/n \sum_{t=1}^n y_t$.

Scenario #2: Bilinear $y_t = 0.5e_{t-1}y_{t-2} + e_t$; mean filter $\epsilon_t = y_t - E[y_t]$; $\hat{\phi}_n = 1/n \sum_{t=1}^n y_t$.

Scenario #3: AR(2) $y_t = 0.3y_{t-1} - 0.15y_{t-2} + e_t$; AR(2) filter $\epsilon_t = y_t - \phi_1 y_{t-1} - \phi_2 y_{t-2}$; least squares.

Scenario #4: AR(2) $y_t = 0.3y_{t-1} - 0.15y_{t-2} + e_t$; AR(1) filter $\epsilon_t = y_t - \phi_1 y_{t-1}$; least squares.

Scenario #5: GARCH(1,1) $y_t = \sigma_t e_t$, $\sigma_t^2 = 1 + 0.2y_{t-1}^2 + 0.5\sigma_{t-1}^2$; no filter.

Scenario #6: GARCH(1,1) $y_t = \sigma_t e_t$, $\sigma_t^2 = 1 + 0.2y_{t-1}^2 + 0.5\sigma_{t-1}^2$; GARCH(1,1) filter $\epsilon_t = y_t/\sigma_t$ with $\sigma_t^2 = \omega + \alpha y_{t-1}^2 + \beta\sigma_{t-1}^2$; quasi-maximum likelihood.[11]

In #5 and #6, e_t is standardized so that $E[e_t^2] = 1$.

The null is true for #1, #2, #3, #5, and #6 when the error e_t is i.i.d. or GARCH. For #4, the null is false for any error e_t because a misspecified AR(1) filter is used. This results in an AR(1) test variable ϵ_t, with geometrically decaying autocorrelations when e_t is i.i.d. or GARCH.

In #1–#4, y_t is stationary for each error. The GARCH(1,1) process in #5–#6 is strong when e_t is i.i.d., and semi-strong when e_t is GARCH(1,1) since it is an

[10] Ergodicity follows since each error process is stationary α-mixing. See, for example, Kolmogorov and Rozanov (1960) for processes with continuous bounded spectral densities (e.g., stationary Gaussian AR, Gaussian MA(2)); Nelson (1990) for GARCH process stationarity; and Carrasco and Chen (2002) for mixing properties of stationary GARCH processes.

[11] QML is performed using the iterated process $\tilde{\sigma}_1^2(\theta) = \omega$ and $\tilde{\sigma}_t^2(\theta) = \omega + \alpha y_{t-1}^2 + \beta\tilde{\sigma}_{t-1}^2(\theta)$ for $t = 2, \ldots, n$. We impose $(\omega, \alpha, \beta) > 0$ and $\alpha + \beta \leq 1$ during estimation.

https://doi.org/10.1017/S0266466619000367 Published online by Cambridge University Press

adapted m.d.s. (Drost and Nijman, 1993), hence in those cases y_t is stationary (Nelson, 1990; Lee and Hansen, 1994). If e_t is MA(2) or AR(1), then both $\{e_t, y_t\}$ are serially correlated. In the MA(2) error case, it can be verified that GARCH y_t is stationary due to the finite feedback structure. It is unknown whether GARCH y_t with an AR(1) error has a stationary solution (see, e.g., Drost and Nijman, 1993; Straumann and Mikosch, 2006).

All of our chosen tests require a finite fourth moment on the test variable ϵ_t, and in all cases $E[e_t^4] < \infty$. In #1–#4, $E[\epsilon_t^4] < \infty$ holds for each error type e_t. In Scenario #6, we test the standardized error $\epsilon_t = e_t = y_t/\sigma_t$ which has a finite fourth moment in all cases.

In Scenario #3, we do not include a constant term in the filter in order to reduce estimator dispersion, and because $E[y_t] = 0$ is known within this experiment. In practice, a constant term would be included to ensure $E[\epsilon_t] = 0$.

In Scenario #5, however, we test GARCH $\epsilon_t = y_t$ itself. $E[\epsilon_t^4] < \infty$ holds when e_t is i.i.d. or MA(2), but it is unknown in theory whether a fourth moment exists when e_t is GARCH(1,1) or AR(1).[12] Test results in the latter case should therefore be interpreted with some caution.

We also consider three additional scenarios in which remote autocorrelations are present. Only an i.i.d. error e_t is used for the following processes in order to focus on autocorrelation remoteness.

Scenario #7: Remote MA(6) $y_t = e_t + 0.25e_{t-6}$; mean filter $\epsilon_t = y_t - E[y_t]$; $\hat{\phi}_n = 1/n\sum_{t=1}^n y_t$.

Scenario #8: Remote MA(12) $y_t = e_t + 0.25e_{t-12}$; mean filter $\epsilon_t = y_t - E[y_t]$; $\hat{\phi}_n = 1/n\sum_{t=1}^n y_t$.

Scenario #9: Remote MA(24) $y_t = e_t + 0.25e_{t-24}$; mean filter $\epsilon_t = y_t - E[y_t]$; $\hat{\phi}_n = 1/n\sum_{t=1}^n y_t$.

In Remote MA(q), $\rho(h)\neq 0$ *if and only if* $h = q$. Hence, any test with a maximum lag less than q should fail to detect serial dependence.

We draw $J = 1,000$ Monte Carlo samples of size $n \in \{100, 250, 500, 1{,}000\}$. We draw $2n$ observations and retain the last n observations for analysis. The rejection frequency of any test corresponds to its empirical size when the tested variable ϵ_t is white noise, and empirical power when ϵ_t is correlated. In Table 1, we summarize the dependence property of ϵ_t under each scenario and error e_t.

Our proposed test is the max-correlation test with the dependent wild bootstrap and automatic lag, denoted $\hat{\mathcal{T}}^{dw}(\mathcal{L}_n^*)$. The test statistic is $\hat{\mathcal{T}}_n(\mathcal{L}_n^*) \equiv \sqrt{n}\max_{1\leq h\leq \mathcal{L}_n^*}$

[12] As an experiment not presented in this article, we simulated $J = 10{,}000$ sample paths $\{y_t\}_{t=1}^{250}$ from GARCH $\epsilon_t = y_t$ with GARCH(1,1) or AR(1) error e_t. We inspected the median over all J samples of the 4^{th} moment for subsamples $\{y_t\}_{t=1}^T$ with $T = 50, \ldots, 250$. Denote this statistic as $k(T)$. $k(T)$ grows exponentially in T, suggesting a fourth moment does not exist for either process.

https://doi.org/10.1017/S0266466619000367 Published online by Cambridge University Press

TABLE 1. Dependence of test variable ϵ_t under each scenario and error e_t

Scenario	#1	#2	#3	#4	#5	#6	#7, #8, #9
DGP	Simple	Bilinear	AR(2)	AR(2)	GARCH	GARCH	Remote MA
Filter	Mean	Mean	AR(2)	AR(1)	None	GARCH	Mean
i.i.d. e_t	**i.i.d.**	**wn**	**i.i.d.**	corr	**m.d.s.**	**i.i.d.**	Remote corr
GARCH e_t	**m.d.s.**	**wn**	**m.d.s.**	corr	**m.d.s.**	**m.d.s.**	Not considered
MA(2) e_t	corr	corr	corr	corr	corr	corr	Not considered
AR(1) e_t	corr	corr	corr	corr	corr	corr	Not considered

Note: wn = non-m.d.s. white noise; corr = autocorrelated; remote corr = autocorrelation is present at a remote lag. Bold text is used to highlight when the null is true.

$|\hat{\omega}_n(h)\hat{\rho}_n(h)|$ with weight $\hat{\omega}_n(h) = 1$.[13] We compute the bootstrapped statistic $\hat{\mathcal{T}}_{n,i}^{(dw)}(\mathcal{L}_{n,i}^*) \equiv \sqrt{n}\max_{1\le h\le \mathcal{L}_{n,i}^*}|\hat{\rho}_{n,i}^{(dw)}(h)|$ for each bootstrap sample $i \in \{1,\ldots,M\}$ with $M = 500$. $\hat{\rho}_{n,i}^{(dw)}(h)$ is computed via (10) based on the Lemma 2.1 correlation expansion, which correctly accounts for the first-order (asymptotic) impact of the *ith* sample's plug-in $\hat{\theta}_{n,i}$. Note that $\mathcal{L}_{n,i}^*$ is the automatic lag for the *ith* bootstrap sample specifically. The dependent wild bootstrap requires a choice of block size b_n. Shao (2011) uses $b_n = b\sqrt{n}$ with $b \in \{.5, 1, 2\}$, leading to qualitatively similar results. Therefore, we use the middle value $b = 1$.[14] The approximate p-value is computed as $\hat{p}_{n,M}^{(dw)} = 1/M\sum_{i=1}^{M} I(\hat{\mathcal{T}}_{n,i}^{(dw)}(\mathcal{L}_{n,i}^*) \ge \hat{\mathcal{T}}_n(\mathcal{L}_n^*))$.

The automatic lag selection requires a choice set $\{1,..,\bar{\mathcal{L}}_n\}$ with maximum possible lag length $\bar{\mathcal{L}}_n$, and the tuning parameter q (cf. (13) and (14)). Let $[z]$ denote the integer part of z. We set $\bar{\mathcal{L}}_n = [\delta\sqrt{n}/(\ln n)]$ with $\delta = 10$ so that $\bar{\mathcal{L}}_n \in \{21, 28, 35, 45\}$ for $n \in \{100, 250, 500, 1{,}000\}$, respectively. In the present simulation design, $\bar{\mathcal{L}}_n$ satisfies the Lemma 2.1 and Theorem 3.1 requirement $\bar{\mathcal{L}}_n = O(\sqrt{n}/\ln(n))$ for all processes except possibly when the test variable is GARCH y_t with AR(1) error e_t. Similar and larger values lead to qualitatively similar results.[15]

In order to choose a plausible value of q for the penalty function in (13), we perform a preliminary simulation study that computes empirical size and size-adjusted power for the max-correlation test with $\hat{\mathcal{T}}_n(\mathcal{L}_n^*)$ across $q \in \{1.50, 1.75, \ldots, 4.50\}$. We consider two cases in order to highlight empirical size and power properties. In Case 1, size is computed under Scenario #1 with an i.i.d. error; and size-adjusted

[13] Other plausible weights include an inverted standard deviation based on a Heteroscedasticity Autocorrelation Consistent (HAC) estimator, and/or the Ljung and Box (1978) weights. In this article, we demonstrate that the uniform weight leads to accurate size and high power. In simulations not reported here, we also find that an inverted standard deviation, either parametric (when known) or nonparametric, is suboptimal due to the added sampling error.

[14] We compared $b_n = b\sqrt{n}$ across $b \in \{0.5, 1, 2\}$ in unreported simulations and found there is little difference in test performance.

[15] In experiments not reported here, we also used $\bar{\mathcal{L}}_n = [\delta n/(\ln(n))^c]$ for various c and δ and found essentially the same results. Thus, the value $\bar{\mathcal{L}}_n = [\delta\sqrt{n}/(\ln n)]$ is not essential to test performance, but does satisfy Lemma 2.1 and Theorem 3.1 for most processes under study.

https://doi.org/10.1017/S0266466619000367 Published online by Cambridge University Press

power is computed under #4 with an i.i.d. error. In Case 2, size is computed under #5 with an i.i.d. error; and size-adjusted power is computed under #5 with MA(2) error. For each case, sample size is $n \in \{100, 500\}$; nominal size is $\alpha = 0.05$; $J = 1000$ Monte Carlo samples and $M = 500$ bootstrap samples are generated. See Figure 1 for results. Variation of empirical size and size-adjusted power for the test based on $\hat{\mathcal{T}}^{dw}(\mathcal{L}_n^*)$ across the values of q is fairly small in each experiment, implying that a choice of q should not have a critical impact on the test performance. For each case and sample size, we obtain relatively accurate size and high power around $q = 3$, hence $q = 3$ is used.

We also perform the dependent wild bootstrap Cramér–von Mises test in Shao (2011), CvM^{dw}. This test is based on the sample spectral distribution function $F_n(\lambda) \equiv \int_0^\lambda I_n(\omega)d\omega$ with periodogram $I_n(\omega) \equiv (2\pi)^{-1}\sum_{h=1-n}^{n-1}\hat{\gamma}_n(h)e^{-h\omega}$. Define

$$S_n(\lambda) \equiv \sqrt{n}(F_n(\lambda) - \hat{\gamma}_n(0)\psi_0(\lambda)) = \sum_{h=1}^{n-1}\sqrt{n}\hat{\gamma}_n(h)\psi_h(\lambda),$$

where $\psi_h(\lambda) = (h\pi)^{-1}\sin(h\lambda)$ if $h \neq 0$, else $\psi_h(\lambda) = \lambda(2\pi)^{-1}$. The CvM test statistic is $\mathcal{C}_n = \int_0^\pi S_n^2(\lambda)d\lambda$, which has a nonstandard limit distribution under the null.[16] We then use Shao's (2011, Sect. 3) dependent wild bootstrap based on the Lemma 2.1 correlation expansion to compute an approximate p-value. Note that all $\mathcal{L}_n = n - 1$ lags are used by construction. Shao (2011) does not consider the use of a filter, but we apply the test to all scenarios for the sake of comparison, and use the correlation expansion to control for a filter when used.

4.2. Simulation Results

4.2.1. *Automatic Lag.* We first check the performance of the automatic lag selection itself. Recall that by Theorem 3.1 $\mathcal{L}_n^* \xrightarrow{p} 1$ under H_0, and under H_1 $\mathcal{L}_n^* \xrightarrow{p} h^*$, the smallest lag at which the largest correlation occurs. Under Scenarios #1–#6 when the error e_t is i.i.d. or GARCH the null is false only for #4. In the latter case, the test variable ϵ_t is AR(1) hence $h^* = 1$.

In Table 2, we report the median of optimal lags $\{\mathcal{L}_n^{*(1)}, \ldots, \mathcal{L}_n^{*(J)}\}$ for each scenario, where $\mathcal{L}_n^{*(j)}$ is the j^{th} sample's optimal lag. We also report the smallest lag at which the largest correlation occurs, h^*. In most cases, we compute h^* analytically. In a few cases an analytic solution is not feasible, so we use a large sample simulation. We generate 50,000 samples of size $n = 50{,}000$, and the autocorrelations for ϵ_t for each sample. We then report the median computed h^* across all samples.

In #1–#6, when H_0 is true or autocorrelations exist at small lags, the median of $\mathcal{L}_n^{*(j)}$ is 1 or 2. This (nearly) matches the predictions of Theorem 3.1 and the reported h^* in most cases. In just two cases, (i) bilinear with GARCH error and (ii) GARCH with GARCH error and without a filter, the reported h^* is 4. This is higher

[16] In practice, we use a numerical integral based on the midpoint approximation with an increment of 0.01.

https://doi.org/10.1017/S0266466619000367 Published online by Cambridge University Press

TABLE 2. Median of automatically selected lags $\mathcal{L}_n^*$

e_t	i.i.d.	GARCH(1,1)	MA(2)	AR(1)
n	{100, 250, 500, 1000}	{100, 250, 500, 1000}	{100, 250, 500, 1000}	{100, 250, 500, 1000}
#1	{1, 1, 1, 1}	{1, 1, 1, 1}	{1, 1, 1, 1}	{1, 1, 1, 1}
	$H_0, h^* = 1$	$H_0, h^* = 1$	$H_1, h^* = 1$	$H_1, h^* = 1$
#2	{1, 1, 1, 1}	{1, 2, 2, 2}	{1, 1, 1, 1}	{1, 1, 1, 1}
	$H_0, h^* = 1$	$H_1, \hat{h}^* = 4$	$H_1, \hat{h}^* = 1$	$H_1, \hat{h}^* = 1$
#3	{1, 1, 1, 1}	{1, 1, 1, 1}	{1, 1, 1, 1}	{1, 1, 1, 2}
	$H_0, h^* = 1$	$H_0, h^* = 1$	$H_1, h^* = 1$	$H_1, h^* = 1$
#4	{2, 2, 2, 2}	{2, 2, 2, 2}	{1, 1, 2, 1}	{1, 1, 1, 1}
	$H_1, \hat{h}^* = 1$	$H_1, \hat{h}^* = 1$	$H_1, \hat{h}^* = 1$	$H_1, \hat{h}^* = 1$
#5	{1, 1, 1, 1}	{1, 1, 2, 2}	{1, 1, 1, 1}	{1, 1, 1, 1}
	$H_0, h^* = 1$	$H_1, \hat{h}^* = 4$	$H_1, \hat{h}^* = 1$	$H_1, \hat{h}^* = 1$
#6	{1, 1, 1, 1}	{1, 1, 1, 1}	{1, 1, 1, 1}	{1, 1, 1, 1}
	$H_0, h^* = 1$	$H_0, h^* = 1$	$H_1, h^* = 1$	$H_1, h^* = 1$
#7	{1, 1, 6, 6}	–	–	–
	$H_1, h^* = 6$	–	–	–
#8	{1, 1, 12, 12}	–	–	–
	$H_1, h^* = 12$	–	–	–
#9	{1, 1, 1, 24}	–	–	–
	$H_1, h^* = 24$	–	–	–

Note: #1: simple $y_t = e_t$ with a mean filter. #2: bilinear process with a mean filter. #3: AR(2) process with an AR(2) filter. #4: AR(2) process with an AR(1) filter. #5: GARCH(1,1) process without a filter. #6: GARCH(1,1) process with a GARCH filter. #7: Remote MA(6) process with a mean filter. #8: Remote MA(12) process with a mean filter. #9: Remote MA(24) process with a mean filter. The error term e_t is i.i.d., GARCH(1,1), MA(2), or AR(1) in Scenarios #1–#6, while it is i.i.d. in Scenarios #7–#9. This table reports the median of automatic lags for actual test statistics, $\mathcal{L}_n^*$, across $J = 1{,}000$ Monte Carlo samples. The largest possible lag length is $\bar{\mathcal{L}}_n = [10\sqrt{n}/(\ln n)]$. The tuning parameter that affects the penalty term $\mathcal{P}_n(\mathcal{L})$ is $q = 3$. H_0 implies the test variable $\{\epsilon_t\}$ is white noise, while H_1 implies $\{\epsilon_t\}$ is serially correlated. The smallest lag at which the largest correlation occurs, h^*, is reported if it can be computed analytically. Otherwise, we report a simulation-based $\hat{h}^*$ as follows. $J = 50{,}000$ Monte Carlo samples of size $n = 50{,}000$ are generated, and sample autocorrelations of $\{\epsilon_t\}$ at $h = 1, \ldots, 20$ are computed. Let $\hat{h}_j^*$ be the smallest lag at which the largest correlation occurs for the *jth* sample, then the reported $\hat{h}^*$ is the median of $\{\hat{h}_1^*, \ldots, \hat{h}_J^*\}$.

than the optimally selected lag (1 or 2). These are the only cases where the median of $\mathcal{L}_n^{*(j)}$ deviates by more than 0 or 1 from h^*. In both of these cases, the process is highly volatile, possibly causing the aberrant deviation of $\mathcal{L}_n^{*(j)}$ from h^*, and the low empirical size of the max-correlation test. As suggested in Section 4.1, either of these processes may fail the required moment conditions for the underlying theory surrounding $\mathcal{L}_n^*$.

https://doi.org/10.1017/S0266466619000367 Published online by Cambridge University Press

In #7–#9, where autocorrelations exist at remote lags, the median of $\mathcal{L}_n^{*(j)}$ pinpoints those lags given a large enough sample size. Under Remote MA(12), for example, the median is 1 for $n \leq 250$ but exactly 12 for $n \geq 500$.

4.2.2. *Empirical Size.* We now report rejection frequencies associated with nominal size $\alpha \in \{0.01, 0.05, 0.10\}$. See Table 3 for $\hat{\mathcal{T}}^{dw}(\mathcal{L}_n^*)$ under #1–#6; Table 4 for CvM^{dw} under #1–#6; and Table 5 for both tests under #7–#9.

We begin with Scenario #1 (simple), $n = 100$, and i.i.d. error. The empirical size with respect to nominal sizes $\alpha \in \{0.010, 0.050, 0.100\}$ is $\{0.017, 0.068, 0.128\}$ for $\hat{\mathcal{T}}^{dw}(\mathcal{L}_n^*)$ and $\{0.023, 0.081, 0.138\}$ for CvM^{dw}, hence $\hat{\mathcal{T}}^{dw}(\mathcal{L}_n^*)$ has reasonably sharp size that is sharper than CvM^{dw}. A similar implication holds for #2 (bilinear), $n = 100$, and i.i.d. error, where the empirical size is $\{0.008, 0.047, 0.090\}$ for $\hat{\mathcal{T}}^{dw}(\mathcal{L}_n^*)$ and $\{0.018, 0.076, 0.149\}$ for CvM^{dw}. In general, the empirical size of the test based on $\hat{\mathcal{T}}^{dw}(\mathcal{L}_n^*)$ is at least as good as (and often better than) size associated with CvM^{dw}.

The reason why $\hat{\mathcal{T}}^{dw}(\mathcal{L}_n^*)$ achieves fairly sharp size in most cases is that, as confirmed in Table 2, $\mathcal{L}_n^*$ is sufficiently close to 1 in most samples under H_0. That feature cuts redundant lags and improves the size of the test. In fact, we find in the supplemental material (Appendix H) that $\hat{\mathcal{T}}^{dw}(\mathcal{L}_n^*)$ achieves the sharpest size among a variety of tests.[17] CvM^{dw} uses all $\mathcal{L}_n = n - 1$ lags, but the greatest weight is assigned to small lags by construction. Hence CvM^{dw} leads to have fairly accurate size in most cases, although generally the max-correlation test dominates.

4.2.3. *Empirical Power.* In #1–#6, the relative performance of $\hat{\mathcal{T}}^{dw}(\mathcal{L}_n^*)$ and CvM^{dw} under H_1 varies across cases. The former is more powerful than the latter in some cases, but not in other cases. In general, there is not a drastic gap between the two tests. See # 2, $n = 1,000$, and AR(1) error, for example. The empirical power with respect to $\alpha \in \{0.010, 0.050, 0.100\}$ is $\{0.723, 0.823, 0.864\}$ for $\hat{\mathcal{T}}^{dw}(\mathcal{L}_n^*)$ and $\{0.474, 0.697, 0.810\}$ for CvM^{dw}. But in # 3, with $n = 1,000$, and an AR(1) error, power is $\{0.599, 0.847, 0.922\}$ for $\hat{\mathcal{T}}^{dw}(\mathcal{L}_n^*)$ and $\{0.688, 0.876, 0.923\}$ for CvM^{dw}.

In #7–#9, however, $\hat{\mathcal{T}}^{dw}(\mathcal{L}_n^*)$ dominates CvM^{dw} completely (see Table 5). $\hat{\mathcal{T}}^{dw}(\mathcal{L}_n^*)$ successfully detects remote autocorrelations given a large enough sample size, while CvM^{dw} fails to detect them for any n. The power of $\hat{\mathcal{T}}^{dw}(\mathcal{L}_n^*)$ under #8 (Remote MA(12)), for instance, is $\{0.013, 0.067, 0.117\}$ for $n = 100$, $\{0.024, 0.134, 0.244\}$ for $n = 250$, $\{0.371, 0.673, 0.770\}$ for $n = 500$, and $\{0.983, 0.997,$

[17] In Scenario #2 (bilinear) with a GARCH error, the max-correlation test is undersized, even in large samples $n = 1,000$ Generally, however, inverted standard errors do not improve the max-correlation performance due to standard error sampling error. The primary cause is the bilinear process combined with a GARCH error results in extreme volatility, which undermines the efficacy of the bootstrap. The test is even more undersized under Scenario #5 (GARCH) with a GARCH error. The CvM test is also undersized for Scenario #2 with a GARCH error. It is, however, less affected than the max-correlation test in Scenario #5 with a GARCH error. Weighting the correlations for a max-correlation test might alleviate the under-rejection, for example, using weights equal to the inverted standard errors. The least volatile correlations in this case are given the greatest weight.

https://doi.org/10.1017/S0266466619000367 Published online by Cambridge University Press

https://doi.org/10.1017/S0266466619000367 Published online by Cambridge University Press

TABLE 3. Rejection frequencies of max-correlation test with automatic lag $\mathcal{L}_n^*$ (Scenarios #1–#6)

	i.i.d. error: $e_t = v_t$					
	#1. Simple	#2. Bilin	#3. AR2/AR2	#4. AR2/AR1	#5. GARCH/wo	#6. GARCH/w
n	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%
100	0.017, 0.068, 0.128	0.008, 0.047, 0.090	0.002, 0.061, 0.129	0.034, 0.197, 0.327	0.006, 0.031, 0.068	0.026, 0.091, 0.140
250	0.011, 0.045, 0.087	0.012, 0.042, 0.087	0.005, 0.048, 0.093	0.178, 0.479, 0.634	0.005, 0.029, 0.063	0.019, 0.064, 0.125
500	0.007, 0.047, 0.096	0.010, 0.036, 0.077	0.004, 0.045, 0.094	0.462, 0.803, 0.894	0.005, 0.033, 0.083	0.016, 0.053, 0.096
1,000	0.012, 0.050, 0.109	0.008, 0.056, 0.103	0.006, 0.057, 0.097	0.929, 0.996, 0.998	0.008, 0.038, 0.077	0.011, 0.047, 0.101
	GARCH(1,1) error: $e_t = v_t w_t$ with $w_t^2 = 1 + 0.2e_{t-1}^2 + 0.5w_{t-1}^2$					
	#1. Simple	#2. Bilin	#3. AR2/AR2	#4. AR2/AR1	#5. GARCH/wo	#6. GARCH/w
n	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%
100	0.005, 0.026, 0.075	0.007, 0.021, 0.040	0.004, 0.054, 0.109	0.027, 0.150, 0.248	0.001, 0.003, 0.012	0.026, 0.090, 0.162
250	0.004, 0.031, 0.069	0.008, 0.023, 0.038	0.007, 0.040, 0.084	0.116, 0.312, 0.451	0.004, 0.010, 0.015	0.011, 0.063, 0.107
500	0.001, 0.031, 0.063	0.017, 0.029, 0.042	0.008, 0.032, 0.078	0.283, 0.588, 0.733	0.003, 0.005, 0.006	0.012, 0.046, 0.089
1,000	0.006, 0.032, 0.071	0.014, 0.026, 0.031	0.008, 0.033, 0.085	0.741, 0.925, 0.961	0.002, 0.002, 0.002	0.006, 0.051, 0.102

(Continued)

https://doi.org/10.1017/S0266466619000367 Published online by Cambridge University Press

TABLE 3. (Continued)

	MA(2) Error: $e_t = v_t + 0.5v_{t-1} + 0.25v_{t-2}$					
	#1. Simple	#2. Bilin	#3. AR2/AR2	#4. AR2/AR1	#5. GARCH/wo	#6. GARCH/w
n	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%
100	0.693, 0.901, 0.951	0.582, 0.769, 0.825	0.012, 0.068, 0.135	0.242, 0.601, 0.762	0.461, 0.707, 0.788	0.908, 0.966, 0.979
250	0.993, 0.998, 1.00	0.841, 0.935, 0.962	0.006, 0.060, 0.114	0.677, 0.927, 0.982	0.707, 0.834, 0.868	0.992, 0.993, 0.993
500	1.00, 1.00, 1.00	0.932, 0.965, 0.976	0.019, 0.086, 0.151	0.972, 0.999, 0.999	0.798, 0.874, 0.908	1.00, 1.00, 1.00
1,000	1.00, 1.00, 1.00	0.968, 0.982, 0.986	0.063, 0.184, 0.257	1.00, 1.00, 1.00	0.900, 0.934, 0.952	1.00, 1.00, 1.00
	AR(1) Error: $e_t = 0.7e_{t-1} + v_t$					
	#1. Simple	#2. Bilin	#3. AR2/AR2	#4. AR2/AR1	#5. GARCH/wo	#6. GARCH/w
n	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%
100	0.531, 0.752, 0.847	0.477, 0.637, 0.685	0.021, 0.128, 0.227	0.263, 0.636, 0.788	0.179, 0.345, 0.432	0.987, 0.991, 0.991
250	0.903, 0.979, 0.990	0.676, 0.774, 0.812	0.041, 0.217, 0.355	0.714, 0.954, 0.984	0.156, 0.271, 0.334	1.00, 1.00, 1.00
500	0.998, 1.00, 1.00	0.715, 0.819, 0.860	0.181, 0.512, 0.636	0.991, 1.00, 1.00	0.111, 0.176, 0.230	1.00, 1.00, 1.00
1,000	1.00, 1.00, 1.00	0.723, 0.823, 0.864	0.599, 0.847, 0.922	1.00, 1.00, 1.00	0.066, 0.126, 0.158	1.00, 1.00, 1.00

Note: #1: Simple $y_t = e_t$ with a mean filter. #2: Bilinear $y_t = 0.5e_{t-1}y_{t-2} + e_t$ with a mean filter. #3: AR(2) $y_t = 0.3y_{t-1} - 0.15y_{t-2} + e_t$ with an AR(2) filter. #4: AR(2) $y_t = 0.3y_{t-1} - 0.15y_{t-2} + e_t$ with an AR(1) filter. #5: GARCH(1,1) $y_t = \sigma_t e_t$, $\sigma_t^2 = 1 + 0.2y_{t-1}^2 + 0.5\sigma_{t-1}^2$ without (wo) a filter. #6: GARCH(1,1) $y_t = \sigma_t e_t$, $\sigma_t^2 = 1 + 0.2y_{t-1}^2 + 0.5\sigma_{t-1}^2$ with (w) a GARCH filter. For each scenario, $v_t \overset{i.i.d.}{\sim} N(0,1)$. The largest possible lag length is $\bar{\mathcal{L}}_n = [10\sqrt{n}/(\ln n)]$, and the tuning parameter that affects the penalty term $\mathcal{P}_n(\mathcal{L})$ is $q = 3$. This table reports rejection frequencies with respect to nominal size $\alpha \in \{0.01, 0.05, 0.10\}$ across $J = 1{,}000$ Monte Carlo samples, where the number of bootstrap samples is $M = 500$.

https://doi.org/10.1017/S0266466619000367 Published online by Cambridge University Press

TABLE 4. Rejection frequencies of Cramér–von Mises test CvM^{dw} (Scenarios #1–#6)

	i.i.d. error: $e_t = v_t$					
	#1. Simple	#2. Bilin	#3. AR2/AR2	#4. AR2/AR1	#5. GARCH/wo	#6. GARCH/w
n	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%
100	0.023, 0.081, 0.138	0.018, 0.076, 0.149	0.020, 0.086, 0.167	0.133, 0.338, 0.483	0.021, 0.077, 0.141	0.034, 0.087, 0.144
250	0.016, 0.072, 0.144	0.030, 0.085, 0.154	0.011, 0.065, 0.127	0.370, 0.615, 0.735	0.011, 0.058, 0.118	0.019, 0.065, 0.112
500	0.010, 0.051, 0.102	0.014, 0.072, 0.124	0.012, 0.059, 0.132	0.710, 0.882, 0.939	0.009, 0.053, 0.103	0.016, 0.072, 0.141
1,000	0.008, 0.060, 0.108	0.016, 0.063, 0.106	0.010, 0.049, 0.102	0.974, 0.991, 0.993	0.015, 0.058, 0.107	0.013, 0.057, 0.103
	GARCH(1,1) error: $e_t = v_t w_t$ with $w_t^2 = 1 + 0.2e_{t-1}^2 + 0.5w_{t-1}^2$					
	#1. Simple	#2. Bilin	#3. AR2/AR2	#4. AR2/AR1	#5. GARCH/wo	#6. GARCH/w
n	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%
100	0.017, 0.081, 0.149	0.002, 0.030, 0.070	0.026, 0.086, 0.168	0.118, 0.287, 0.430	0.006, 0.049, 0.103	0.036, 0.100, 0.168
250	0.013, 0.059, 0.108	0.029, 0.048, 0.083	0.012, 0.058, 0.127	0.242, 0.501, 0.648	0.009, 0.037, 0.080	0.020, 0.075, 0.132
500	0.015, 0.066, 0.115	0.026, 0.038, 0.075	0.011, 0.051, 0.104	0.550, 0.802, 0.881	0.013, 0.052, 0.111	0.026, 0.072, 0.143
1,000	0.010, 0.060, 0.116	0.004, 0.014, 0.028	0.008, 0.056, 0.105	0.880, 0.973, 0.993	0.006, 0.032, 0.065	0.049, 0.065, 0.073

(Continued)

https://doi.org/10.1017/S0266466619000367 Published online by Cambridge University Press

TABLE 4. (Continued)

	MA(2) Error: $e_t = v_t + 0.5v_{t-1} + 0.25v_{t-2}$					
	#1. Simple	#2. Bilin	#3. AR2/AR2	#4. AR2/AR1	#5. GARCH/wo	#6. GARCH/w
n	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%
100	0.898, 0984, 0.995	0.450, 0.743, 0.866	0.029, 0.113, 0.182	0.570, 0.805, 0.898	0.681, 0.908, 0.969	0.878, 0.927, 0.940
250	0.999, 1.00, 1.00	0.769, 0.924, 0.968	0.019, 0.086, 0.189	0.951, 0.996, 0.999	0.903, 0.979, 0.994	0.983, 0.989, 0.991
500	1.00, 1.00, 1.00	0.884, 0.966, 0.990	0.032, 0.144, 0.250	1.00, 1.00, 1.00	0.959, 0.994, 0.995	0.995, 0.998, 0.998
1,000	1.00, 1.00, 1.00	0.974, 0.994, 0.997	0.068, 0.295, 0.471	1.00, 1.00, 1.00	0.986, 0.997, 1.00	0.998, 0.998, 0.998
	AR(1) Error: $e_t = 0.7e_{t-1} + v_t$					
	#1. Simple	#2. Bilin	#3. AR2/AR2	#4. AR2/AR1	#5. GARCH/wo	#6. GARCH/w
n	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%
100	0.925, 0.996, 1.00	0.282, 0.567, 0.741	0.064, 0.193,.299	0.472, 0.741, 0.849	0.564, 0.818,0.923	0.958, 0.970, 0.973
250	0.999, 1.00, 1.00	0.341, 0.572, 0.718	0.136, 0.341,.465	0.935, 0.991, 0.999	0.680, 0.849, 0.912	0.984, 0.987, 0.988
500	1.00, 1.00, 1.00	0.393, 0.630, 0.781	0.325, 0.592,.700	0.999, 1.00, 1.00	0.700, 0.852, 0.918	0.999, 1.00, 1.00
1,000	1.00, 1.00, 1.00	0.474, 0.697, 0.810	0.688, 0.876,.923	1.00, 1.00, 1.00	0.750, 0.877, 0.929	0.998, 0.999, 0.999

Note: #1: Simple $y_t = e_t$ with a mean filter. #2: Bilinear $y_t = 0.5e_{t-1}y_{t-2} + e_t$ with a mean filter. #3: AR(2) $y_t = 0.3y_{t-1} - 0.15y_{t-2} + e_t$ with an AR(2) filter. #4: AR(2) $y_t = 0.3y_{t-1} - 0.15y_{t-2} + e_t$ with an AR(1) filter. #5: GARCH(1,1) $y_t = \sigma_t e_t$, $\sigma_t^2 = 1 + 0.2y_{t-1}^2 + 0.5\sigma_{t-1}^2$ without (wo) a filter. #6: GARCH(1,1) $y_t = \sigma_t e_t$, $\sigma_t^2 = 1 + 0.2y_{t-1}^2 + 0.5\sigma_{t-1}^2$ with (w) a GARCH filter. For each scenario, $v_t \overset{i.i.d.}{\sim} N(0,1)$. The dependent wild bootstrap with $M = 500$ samples is used to compute an approximate p-value of the Cramér–von Mises test. All $\mathcal{L}_n = n-1$ lags are used. This table reports rejection frequencies with respect to nominal size $\alpha \in \{0.01, 0.05, 0.10\}$ across $J = 1{,}000$ Monte Carlo samples.

TABLE 5. Rejection frequencies of $\hat{\mathcal{T}}^{dw}(\mathcal{L}_n^*)$ and CvM^{dw} (Scenarios #7–#9)

Max-correlation test with automatically selected lag $\hat{\mathcal{T}}^{dw}(\mathcal{L}_n^*)$										
		#7. MA(6)			#8. MA(12)			#9. MA(24)		
n	$\bar{\mathcal{L}}_n$	1%	5%	10%	1%	5%	10%	1%	5%	10%
100	21	0.016	0.084	0.139	0.013	0.067	0.117	0.017	0.065	0.118
250	28	0.155	0.289	0.352	0.024	0.134	0.244	0.012	0.042	0.088
500	35	0.710	0.812	0.826	0.371	0.673	0.770	0.024	0.097	0.192
1,000	45	0.999	1.00	1.00	0.983	0.997	0.997	0.578	0.833	0.918

Cramér–von Mises test CvM^{dw}										
		#7. MA(6)			#8. MA(12)			#9. MA(24)		
n	$\mathcal{L}_n$	1%	5%	10%	1%	5%	10%	1%	5%	10%
100	99	0.040	0.098	0.171	0.034	0.110	0.179	0.029	0.098	0.186
250	249	0.026	0.080	0.142	0.025	0.087	0.155	0.022	0.088	0.143
500	499	0.014	0.087	0.175	0.026	0.092	0.161	0.024	0.071	0.133
1,000	999	0.038	0.160	0.320	0.017	0.083	0.166	0.028	0.079	0.144

Note: Scenario #7: Remote MA(6) $y_t = e_t + 0.25e_{t-6}$ with a mean filter. Scenario #8: Remote MA(12) $y_t = e_t + 0.25e_{t-12}$ with a mean filter. Scenario #9: Remote MA(24) $y_t = e_t + 0.25e_{t-24}$ with a mean filter. For each scenario, $e_t \overset{i.i.d.}{\sim} N(0,1)$. For each test, the dependent wild bootstrap is used to compute an approximate p-value. For the max-correlation test, the largest possible lag length is $\bar{\mathcal{L}}_n = [10\sqrt{n}/(\ln n)]$, and the tuning parameter that affects the penalty term $\mathcal{P}_n(\mathcal{L})$ is $q = 3$. For the Cramér–von Mises test, all $\mathcal{L}_n = n - 1$ lags are used. We report rejection frequencies with respect to nominal size $\alpha \in \{0.01, 0.05, 0.10\}$ across $J = 1{,}000$ Monte Carlo samples.

0.997} for $n = 1{,}000$. Logically, power increases as n grows. The reason that $\hat{\mathcal{T}}^{dw}(\mathcal{L}_n^*)$ detects remote autocorrelations is confirmed in Table 2 (cf. Theorem 3.1.b): $\mathcal{L}_n^*$ converges to $h^* = 12$ when $n \geq 500$ under #8. The power of CvM^{dw}, by contrast, is {0.034, 0.110, 0.179} for $n = 100$, {0.025, 0.087, 0.155} for $n = 250$, {0.026, 0.092, 0.161} for $n = 500$, and {0.017, 0.083, 0.166} for $n = 1{,}000$. CvM^{dw} has (almost) no power against the remote autocorrelation even when $n = 1{,}000$. In fact, we find in Appendix H that $\hat{\mathcal{T}}^{dw}(\mathcal{L}_n^*)$ is the only test that has power against remote autocorrelations among a variety of tests, which have decent size.

The reason why CvM^{dw} fails to capture remote autocorrelations is that it incorporates *all* available sample correlations, while assigning the greatest weight to small lags. That feature delivers sharp size and high power against adjacent correlations like Scenarios #1–#6, but critically low power against remote correlations like Scenarios #7–#9.

The (nonweighted) max-correlation, by contrast, operates on the most informative serial correlation over a range of lags $\{1, ..., \mathcal{L}_n^*\}$. The optimal maximum lag selected $\mathcal{L}_n^*$ asymptotically hones in on the most informative lag range: the range that includes the smallest lag at which the greatest correlation in

https://doi.org/10.1017/S0266466619000367 Published online by Cambridge University Press

magnitude occurs. Thus, in large samples in particular, $\hat{\mathcal{T}}^{dw}(\mathcal{L}_n^*)$ delivers the single most informative serial correlation for test purposes, as opposed to a weighted sum of all, and therefore potentially less useful, correlations. That feature itself generally delivers accurate size (or under-rejections in some cases) and competitive power for Scenarios #1–#6, and dominant power against remote correlations.

In some cases against adjacent correlations, power is not dominant when a large prechosen nonrandom $\mathcal{L}_n$ is used (see Appendix H), but such a shortcoming is alleviated by using our proposed automatic lag $\mathcal{L}_n^*$. The combined max-correlation with automatic lag and bootstrapped p-value leads to a dominant test over all when size and power are considered, in comparison to a variety of tests.

5. CONCLUSION

We present a bootstrap max-correlation test of the white noise hypothesis for regression model residuals. The maximum correlation over an increasing lag length has a long history in the statistics literature, but only in terms of characterizing its limit distribution using extreme value theory and only for observed data. We apply a bootstrap method to a first-order correlation expansion in order to account for the impact of a plug-in $\hat{\theta}_n$ used to compute model residuals. We prove that Shao's (2011) dependent wild bootstrap yields a valid test in a more general environment than Shao (2011) or Xiao and Wu (2014) considered. Our approach does not require showing that the original and bootstrapped max-correlation test statistics have the same limit properties under the null, allowing us to bypass the extreme value theory approach altogether. We also extend Escanciano and Lobato's (2009) automatic lag selection to our setting with an (asymptotically) unbounded lag set. We prove that the automatic lag converges in probability to one under the null, and the smallest lag at which the largest correlation in magnitude occurs under the alternative. In both cases, the procedure hones in on the most informative lag, offering the greatest number of data points for analysis, for the given hypothesis.

Simulation experiments show that our test with the automatic lag generally outperforms a variety of other tests. It achieves sharper empirical size in most cases than other tests since the automatic lag $\mathcal{L}_n^*$ is sufficiently close to 1 under the null hypothesis. When there exist serial correlations at small lags, the max-correlation test and some strong competitors such as the Cramér–von Mises test with the dependent wild bootstrap lead to roughly comparable empirical power. When there exist correlations only at remote lags, the max-correlation test has (potentially) high power while the Cramér–von Mises test has nearly trivial power for any sample size due to its weighting structure. Other tests also have comparatively lower power. This striking difference stems from the fact that the automatic lag $\mathcal{L}_n^*$ pinpoints the relevant remote lag, while other tests by construction incorporate many lags into a test statistic (the CvM test gives the greatest weight to low lags, making it useless against remote lags).

https://doi.org/10.1017/S0266466619000367 Published online by Cambridge University Press

SUPPLEMENTARY MATERIAL

To view supplementary material for this article, please visit http://dx.doi.org/10.1017/S0266466619000367

REFERENCES

Andrews, D.W.K. (1987) Consistency in nonlinear econometric models: A generic uniform law of large numbers. *Econometrica* 55, 1465–1471.

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

Andrews, D.W.K. & W. Ploberger (1996) Testing for serial correlation against an ARMA(1,1) process. *Journal of the American Statistical Association* 91, 1331–1342.

Arcones, M.A. & B.Yu (1994) Central limit theorems for empirical and U-processes of stationary mixing sequences. *Journal of Theoretical Probability* 7, 47–71.

Berman, S.M. (1964) Limit theorems for the maximum term in stationary sequences. *Annals of Mathematical Statistics* 35, 502–516.

Billingsley, P. (1995) *Probability and Measure, 3rd Edition*. Wiley.

Billingsley, P. (1999) *Convergence of Probability Measures, 2nd Edition*. Wiley.

Boehme, T.K. & M. Rosenfeld (1974) An example of two compact Hausdorff Fréchet spaces whose product is not Fréchet. *Journal of the London Mathematical Society* 8, 339–344.

Box, G.E.P. & D.A. Pierce (1970) Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association* 65, 1509–1526.

Carrasco, M. & X. Chen (2002) Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory* 18, 17–39.

Chatterjee, S. (2005) An error bound in the Sudakov-Fernique inequality. Available at https://arxiv.org/abs/math/0510424.

Chernozhukov, V., D. Chetverikov, & K. Kato (2013) Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Annals of Statistics* 41, 2786–2819.

Chernozhukov, V., D. Chetverikov, & K. Kato (2014) Testing many moment inequalities. Available at arXiv:1312.7614.

Chernozhukov, V., D. Chetverikov, & K. Kato (2015) Comparison and anti-concentration bounds for maxima of Gaussian random vectors. *Probability Theory and Related Fields* 162, 47–70.

Chernozhukov, V., D. Chetverikov, & K. Kato (2017) Central limit theorems and bootstrap in high dimensions. *Annals of Probability* 45, 2309–2352.

Chernozhukov, V., D. Chetverikov, & K. Kato (2018) Inference on causal and structural parameters using many moment inequalities. *Review of Economic Studies* 86, 1867–1900.

Dalla, V., L. Giraitis, & P.C.B. Phillips (2019) Robust tests for white noise and cross-correlation. Cowles Foundation Discussion Paper CFDP 2194.

Davidson, J. (1994) *Stochastic Limit Theory*. Oxford University Press.

de Jong, R.M. (1997) Central limit theorems for dependent heterogeneous random variables. *Econometric Theory* 13, 353–367.

Delgado, M.A., J. Hidalgo, & C. Velasco (2005) Distribution free goodness-of-fit tests for linear processes. *Annals of Statistics* 33, 2568–2609.

Delgado, M.A. & C. Velasco (2011) An asymptotically pivotal transform of the residuals sample autocorrelations with application to model checking. *Journal of the American Statistical Association* 106, 946–958.

Deo, R.S. (2000) Spectral tests of the martingale hypothesis under conditional heteroscedasticity. *Journal of Econometrics* 99, 291–315.

Drost, F.C. & T.E. Nijman (1993) Temporal aggregation of GARCH processes. *Econometrica* 61, 909–927.

https://doi.org/10.1017/S0266466619000367 Published online by Cambridge University Press

Dudley, R.M. (1978) Central limit theorems for empirical measures. *Annals of Probability* 6, 899–929.
Dudley, R.M. (1984) A course on empirical processes. *Ecole d'Eté de Probabilités de Saint-Flour XII-1982. Lecture Notes in Math* 1097, 2–142.
Dudley, R.M. & W. Philipp (1983) "Invariance principles for sums of Banach space valued random elements and empirical processes. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 62, 509–552.
EscancianoJ.C. & I.N. Lobato (2009) An automatic portmanteau test for serial correlation. *Journal of Econometrics* 151, 140–149.
Francq, C. & J.-M. Zakoïan (2004) Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli* 10, 605–637.
Francq, C. and J.-M. Zakoïan (2010) *GARCH Models: Structure, Statistical Inference and Financial Applications*. Wiley, UK.
Gaenssler, P. & K. Ziegler (1994) A uniform law of large numbers for set-indexed processes with applications to empirical and partial-sum processes. In J. Hoffmann-Jørgensen, J. Kuelbs, and M.B. Marcus (eds.), *Probability in Banach Spaces*, vol. 9, pp. 385–400. Birkhauser.
Giné, E. & J. Zinn (1990) Bootstrapping general empirical measures. *Annals of Probability* 18, 851–869.
Grenander, U. & M. Rosenblatt (1952) On spectral analysis of stationary time series. *Proceedings of the National Academy of Sciences of the United States of America*, 38, 519–521.
Guay, A., E. Guerre, & S. Lazarová (2013) Robust adaptive rate-optimal testing for the white noise hypothesis. *Journal of Econometrics* 176, 134–145.
Hannan, E.J. (1974) The uniform convergence of autocovariances. *Annals of Statistics* 2, 803–806.
Hansen, B.E. (1996) Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica* 64, 413–430.
Hoffmann-Jørgensen, J. (1984) Convergence of Stochastic Processes on Polish Spaces. (mimeo).
Hoffmann-Jørgensen, J. (1991) Convergence of stochastic processes on Polish spaces. In *Various Publications Series 39*. Aarhus Universitet, Aarhus, Denmark (mimeo).
Hong, Y. (1996) Consistent testing for serial correlation of unknown form. *Econometrica* 64, 837–864.
Hong, Y. (2001) A test for volatility spillover with application to exchange rates. *Journal of Econometrics* 103, 183–224.
Horowitz, J.L., I.N. Lobato, J.C. Nankervis, and N.E. Savin (2006) Bootstrapping the Box-Pierce Q test: A robust test of uncorrelatedness. *Journal of Econometrics* 133, 841–862.
Hüsler, J. (1986) Extreme values of non-stationary random sequences. *Journal of Applied Probability* 23, 937–950.
Inglot, T. & T. Ledwina (2006) Towards data driven selection of a penalty function for data driven Neyman tests. *Linear Algebra and its Applications* 417, 124–133.
Jirak, M. (2011) On the maximum of covariance estimators. *Journal of Multivariate Analysis* 102, 1032–1046.
Kolmogorov, A.N. & Y.A. Rozanov (1960) On strong mixing conditions for stationarity Gaussian processes. *Theory of Probability and its Applications* 5, 204–208.
Kuan, C.-M. & W.-M. Lee (2006) Robust M tests without consistent estimation of the asymptotic covariance matrix. *Journal of the American Statistical Association* 101, 1264–1275.
Kullback, S. & R.A. Leibler (1951) On information and sufficiency. *Annals of Mathematical Statistics* 22, 79–86.
Le Cam, L. (1988) On the Prokhorov distance between the empirical process and the associated Gaussian bridge." Discussion Paper, Dept. of Statistics, University of California, Berkely, CA.
Leadbetter, M.R., G. Lindgren, & H. Rootzén (1983) *Extremes and Related Properties of Random Sequences and Processes*, Springer Series in Statistics. Springer-Verlag New York.
Lee, S.-W. & B.E. Hansen (1994) Asymptotic theory for the GARCH(1,1) quasi-maximum likelihood estimator. *Econometric Theory* 10, 29–52.

https://doi.org/10.1017/S0266466619000367 Published online by Cambridge University Press

Lipschitz, S. (1965) *General Topology*. McGraw Hill.
Liu, R.Y. (1988) Bootstrap procedures under some non-I.I.D. models. *Annals of Statistics* 16, 1696–1708.
Ljung, G.M. & G.E.P. Box (1978) On a measure of lack of fit in time series models. *Biometrika* 65, 297–303.
Lobato, I.N. (2001) Testing that a dependent process is uncorrelated. *Journal of the American Statistical Association* 96, 1066–1076.
Lobato, I.N., J.C. Nankervis, & N.E. Savin (2002) Testing for zero autocorrelation in the presence of statistical dependence. *Econometric Theory* 18, 730–743.
McLeish, D.L. (1975) A maximal inequality and dependent strong laws. *Annals of Probability* 3, 829–839.
Meitz, M. & P. Saikkonen (2011) Parameter estimation in nonlinear AR-GARCH models. *Econometric Theory* 27, 1236–1278.
Myers, J.S. (2002) The minimum number of monotone subsequences." *The Electronic Journal of Combinatorics* 9, R4.
Nankervis, J.C. & N.E. Savin (2010) Testing for serial correlation: Generalized Andrews-Ploberger tests. *Journal of Business & Economic Statistics* 28, 246–255.
Nankervis, J.C. & N.E. Savin (2012) Testing for uncorrelated errors in ARMA models: Non-standard Andrews-Ploberger tests. *Econometrics Journal* 15, 516–534.
Nelson, D.B. (1990) Stationarity and persistence in the GARCH(1,1) model." *Econometric Theory* 6, 318–334.
Neyman, J. (1937) "Smootht est for goodness of fit. *Scandinavian Aktuarietidskr* 20, 149–199.
Pakes, A. & D. Pollard (1989) Simulation and the asymptotics of optimization estimators. *Econometrica* 57, 1027–1057.
Politis, D.N. & J.P. Romano (1994) Limit theorems for weakly dependent Hilbert space valued random variables with application to the stationary bootstrap. *Statistica Sinica* 4, 461–476.
Pollard, D. (1984) *Convergence of Stochastic Processes*. Springer-Verlag New York.
Pollard, D. (1990) Empirical processes: Theory and applications. In *NSF-CBMS Regional Conference Series in Probability and Statistics*, p. 2.
Portnoy, S. (1986) On the central limit theorem in R^p when $p \to \infty$. *Probability Theory and Related Fields* 73, 571–583.
Ramsey, F.P. (1930) On a problem of formal logic. *Proceedings of the London Mathematical Society* 30, 264–286.
Romano, J.P. & L.A. Thombs (1996) Inference for Autocorrelations under Weak Assumptions. *Journal of the American Statistical Association* 91, 590–600.
Sawa, T. (1978) Information criteria for discriminating among alternative regression models. *Econometrica* 46, 1273–1291.
Shao, X. (2010) The dependent wild bootstrap. *Journal of the American Statistical Association* 105, 218–235.
Shao, X. (2011) A bootstrap-assisted spectral test of white noise under unknown dependence. *Journal of Econometrics* 162, 213–224.
Shao, X. & W.B. Wu (2007) Asymptotic spectral theory for nonlinear time series. *Annals of Statistics* 35, 1773–1801.
Straumann, D. & T. Mikosch (2006) Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: A stochastic recurrence equations approach. *Annals of Statistics* 34, 5.
Thomason, A. (1988) A disproof of a conjecture of Erdös in Ramsey theory." *Journal of the London Mathematical Society* 39, 246–255.
Wu, C.F.J. (1986) Jackknife, bootstrap and other resampling methods in regression analysis. *Annals of Statistics* 14, 1261–1295.
Wu, W.B. (2005) Nonlinear system theory: Another look at dependence. *Proceedings of the National Academy of Sciences of the United States of America* 102, 14150–14154.
Wu, W.B. & W. Min (2005) On linear processes with dependent innovations. *Stochastic Processes and their Applications* 115, 939–958.

https://doi.org/10.1017/S0266466619000367 Published online by Cambridge University Press

Xiao, H. & W.B. Wu (2014) Portmanteau test and simultaneous inference for serial covariances. *Statistica Sinica* 24, 577–599.

Yurinskii, V.V. (1977) On the error of the Gaussian approximation for convolutions. *Theory of Probability and its Applications* 22, 236–247.

Zhang, D. & W.B. Wu (2017) Gaussian approximation for high dimensional time series. *Annals of Statistics* 45, 1895–1919.

Zhang, X. & G. Cheng (2014) Bootstrapping high dimensional time series. Available at arXiv:1406.1037.

Zhu, K. (2015) Bootstrapping the Portmanteau tests in weak auto-regressive moving average models. *Journal of the Royal Statistical Society Series B* 78, 463–485.

Zhu, K. & W.K. Li (2015) A bootstrapped spectral test for adequacy in weak ARMA models. *Journal of Econometrics* 187, 113–130.

APPENDIX

A: Proofs

We assume all random variables exist on a complete measure space such that majorants and integrals over uncountable families of measurable functions are measurable, and probabilities where applicable are outer probability measures. See Pollard's (1984, Appendix C) *permissibility* criteria, and see Dudley's (1984, p. 101) *admissible Suslin* property.

We use the following variance bound for NED sequences repeatedly. If w_t is zero mean, L_p-bounded for some $p > 2$, and L_2-NED with size 1/2, on an α-mixing base with decay $O(h^{-p/(p-2)-\iota})$, then by Thm. 17.5 in Davidson (1994) and Thm. 1.6 in McLeish (1975):

$$E\left[\left(1/\sqrt{n}\sum_{t=1}^{n} w_t\right)^2\right] = O(1). \tag{A.1}$$

The following results are key steps toward sidestepping extreme value theory and Gaussian approximations when working with the maximum. The first result expands on a result in Boehme and Rosenfeld (1974, Lem. 1) for first countable topological spaces. The latter is intimately linked to array convergence implications of theory developed in Ramsey (1930), cf. Boehme and Rosenfeld (1974), Thomason (1988), and Myers (2002). Recall that any metric space is a first countable topological space (see, e.g., Lipschitz, 1965, p. 131).

LEMMA A.1. *Assume the array* $\{\mathcal{A}_{k,n} : 1 \leq k \leq \mathcal{I}_n\}_{n\geq 1}$ *lies in a first countable topological space, where* $\{\mathcal{I}_n\}_{n\geq 1}$ *is a sequence of positive integers,* $\mathcal{I}_n \to \infty$ *as* $n \to \infty$. *Let* $\lim_{n\to\infty} \mathcal{A}_{k,n} = a_k$ *for each fixed* k, *and* $\lim_{k\to\infty} a_k = a$. *Then* $\lim_{n\to\infty} \mathcal{A}_{\mathcal{L}_n,n} = a$ *for some nonunique monotonic sequence of positive integers* $\{\mathcal{L}_n\}$, *where* $\mathcal{L}_n \leq \mathcal{I}_n$, *and* $\mathcal{L}_n \to \infty$. *If* $\mathcal{I}_n = n$ *then* $\mathcal{L}_n \leq n$. *Moreover,* $\lim_{n\to\infty} \mathcal{A}_{\tilde{\mathcal{L}}_n,n} = a$ *for any other monotonic sequence of positive integers* $\{\tilde{\mathcal{L}}_n\}_{n=1}^{\infty}$, $\tilde{\mathcal{L}}_n \to \infty$ *that satisfies* $\tilde{\mathcal{L}}_n/\mathcal{L}_n \to 0$, *hence* $\mathcal{L}_n = o(n)$ *can always be assured.*

Proof. In view of $\lim_{n\to\infty} \mathcal{A}_{k,n} = a_k$ for each fixed k, and $\lim_{k\to\infty} a_k = a$, there always exists a monotonic sequence of positive integers $\{\mathcal{L}_n\}_{n=1}^{\infty}$, $\mathcal{L}_n \to \infty$ as $n \to \infty$, satisfying

$$\left|\mathcal{A}_{\mathcal{L}_n,n} - a_{\mathcal{L}_n}\right| \leq \frac{1}{\mathcal{L}_n}. \tag{A.2}$$

https://doi.org/10.1017/S0266466619000367 Published online by Cambridge University Press

Similarly, for any $\{\mathcal{L}_n\}_{n=1}^{\infty}$ that satisfies (A.2), any other monotonic sequence of positive integers $\{\tilde{\mathcal{L}}_n\}_{n=1}^{\infty}$, $\tilde{\mathcal{L}}_n \to \infty$ and $\lim_{n\to\infty} \tilde{\mathcal{L}}_n/\mathcal{L}_n \leq 1$ also satisfies (A.2). Hence $\{\mathcal{L}_n\}_{n=1}^{\infty}$ is not unique, and $\mathcal{L}_n = o(n)$ is always feasible.

Inequality (A.2) yields $a_{\mathcal{L}_n} - 1/\mathcal{L}_n \leq \mathcal{A}_{\mathcal{L}_n,n} \leq a_{\mathcal{L}_n} + 1/\mathcal{L}_n$ for each n. Then, by the definition of a limit

$$\lim_{n\to\infty} a_{\mathcal{L}_n} - \frac{1}{\mathcal{L}_n} \leq \lim_{n\to\infty} \mathcal{A}_{\mathcal{L}_n,n} \leq \lim_{n\to\infty} a_{\mathcal{L}_n} + \frac{1}{\mathcal{L}_n}.$$

The Sandwich theorem, $\mathcal{L}_n \to \infty$ and therefore $\lim_{n\to\infty} a_{\mathcal{L}_n} = a$, now yields:

$$\lim_{n\to\infty} \mathcal{A}_{\mathcal{L}_n,n} = \lim_{n\to\infty} a_{\mathcal{L}_n} = a.$$

See also Boehme and Rosenfeld (1974, Lem. 1). This proves $\lim_{n\to\infty} \mathcal{A}_{\mathcal{L}_n,n} = a$ for some nonunique monotonic sequence of positive integers $\{\mathcal{L}_n\}_{n=1}^{\infty}$, where $\mathcal{L}_n = o(n)$ is always feasible. By construction, $\mathcal{L}_n \leq \mathcal{I}_n$ must be satisfied, hence if $\mathcal{I}_n = n$ then $\mathcal{L}_n \leq n$. $\mathcal{QED}$. □

The next result uses Lemma A.1 as the basis for deriving *in probability* convergence of a function of an increasing set of random variables. This result forms the basis for the proof of the non-Gaussian correlation expansion Lemma 2.1.

Recall the continuous $\vartheta : \mathbb{R}^{\mathcal{L}_n} \to [0,\infty)$ satisfies: lower bound $\vartheta(a) = 0$ *if and only if* $a = 0$; upper bound $\vartheta(a) \leq K\mathcal{LM}$ for some $K > 0$ and any $a = [a_h]_{h=1}^{\mathcal{L}}$ such that $|a_h| \leq \mathcal{M}$ for each h; divergence $\vartheta(a) \to \infty$ as $||a|| \to \infty$; monotonicity $\vartheta(a_{\mathcal{L}_1}) \leq \vartheta([a'_{\mathcal{L}_1}, c'_{\mathcal{L}_2-\mathcal{L}_1}]')$ where $(a_{\mathcal{L}}, c_{\mathcal{L}}) \in \mathbb{R}^{\mathcal{L}}$, $\forall \mathcal{L}_2 \geq \mathcal{L}_1$ and any $c_{\mathcal{L}_2-\mathcal{L}_1} \in \mathbb{R}^{\mathcal{L}_2-\mathcal{L}_1}$; and the triangle inequality $\vartheta(a+b) \leq \vartheta(a) + \vartheta(b)$ $\forall a, b \in \mathbb{R}^{\mathcal{L}_n}$.

LEMMA A.2. *Let* $\{\mathcal{X}_n(i), \mathcal{Y}_n(i) : i \in \mathbb{N}\}_{n\geq 1}$ *be arrays of random variables.*

a. *If* $\mathcal{X}_n(i) \xrightarrow{p} 0$ *as* $n \to \infty$ *for each* $i \in \mathbb{N}$, *then* $\vartheta([\mathcal{X}_n(i)]_{i=1}^{\mathcal{L}_n}) \xrightarrow{p} 0$ *for some nonunique monotonic sequence of positive integers* $\{\mathcal{L}_n\}$ *with* $\mathcal{L}_n \to \infty$. *Moreover,* $\mathcal{L}_n = o(n)$ *can always be assured. In particular,* $\vartheta([\mathcal{X}_n(i)]_{i=1}^{\tilde{\mathcal{L}}_n}) \xrightarrow{p} 0$ *for any other monotonic sequence of positive integers* $\{\tilde{\mathcal{L}}_n\}$ *with* $\tilde{\mathcal{L}}_n \to \infty$ *and* $\lim_{n\to\infty} \tilde{\mathcal{L}}_n/\mathcal{L}_n \leq 1$.

   Furthermore, let ϑ *be the maximum transform:* $\vartheta([\mathcal{X}_n(i)]_{i=1}^{\mathcal{L}_n}) = \max_{1\leq i\leq \mathcal{L}_n} |\mathcal{X}_n(i)|$. *If for some nonstochastic* $g : \mathbb{N} \to [0,\infty)$ *with* $g(n) \to \infty$ *as* $n \to \infty$, $g(n)\mathcal{X}_n(i) \xrightarrow{p} 0$ *and* $g(n)\mathcal{X}_n(i)$ *is uniformly integrable for each* i, *then* $\mathcal{L}_n = O(g(n))$ *must hold.*

b. *If* $\mathcal{X}_n(i) - \mathcal{Y}_n(i) \xrightarrow{p} 0$ *as* $n \to \infty$ *for each* $i \in \mathbb{N}$, *then for some nonunique monotonic sequence of positive integers* $\{\mathcal{L}_n\}$ *with* $\mathcal{L}_n \to \infty$: $|\vartheta([\mathcal{X}_n(i)]_{i=1}^{\mathcal{L}_n}) - \vartheta([\mathcal{Y}_n(i)]_{i=1}^{\mathcal{L}_n})| \leq |\vartheta([\mathcal{X}_n(i) - \mathcal{Y}_n(i)]_{i=1}^{\mathcal{L}_n})| \xrightarrow{p} 0$. *Moreover,* $\mathcal{L}_n = o(n)$ *can always be assured. In particular, the claim holds for any other monotonic sequence of positive integers* $\{\tilde{\mathcal{L}}_n\}$ *with* $\tilde{\mathcal{L}}_n \to \infty$ *and* $\lim_{n\to\infty} \tilde{\mathcal{L}}_n/\mathcal{L}_n \leq 1$.

   Furthermore, let ϑ *be the maximum transform:* $\vartheta([\mathcal{X}_n(i)]_{i=1}^{\mathcal{L}_n}) = \max_{1\leq i\leq \mathcal{L}_n} |\mathcal{X}_n(i)|$. *If for some nonstochastic* $g : \mathbb{N} \to [0,\infty)$ *with* $g(n) \to \infty$ *as* $n \to \infty$, $g(n)(\mathcal{X}_n(i) - \mathcal{Y}_n(i)) \xrightarrow{p} 0$ *and* $g(n)(\mathcal{X}_n(i) - \mathcal{Y}_n(i))$ *is uniformly integrable for each* i, *then* $\mathcal{L}_n = O(g(n))$ *must hold.*

Remark A.1. $\mathcal{L}_n = o(n)$ is required for sample correlation consistency.

https://doi.org/10.1017/S0266466619000367 Published online by Cambridge University Press

Remark A.2. Due to difficulties associated with having a general plug-in and therefore non-Gaussian approximation theory without imposing dependence or heterogeneity restrictions, we only bound $\mathcal{L}_n$ for the maximum case using classic arguments. An improved upper bound on $\mathcal{L}_n \to \infty$ is possible with additional assumptions. The Gaussian approximation or extreme value theoretic approaches yield (possibly far) sharper bounds, but require information that generally does not hold when $X_n(i)$ is a filtered functional of a plug-in estimator as discussed in Section 1. Indeed, under Gaussianicity the Sudakov–Fernique inequality is available (see, e.g., Chatterjee, 2005; Chernozhukov, Chetverikov, and Kato, 2015). See also the tools developed in Zhang and Wu (2017, Sects. 6 and 7): in their stationary functional dependence setting, a sharper bound than (A.3), below, is available for a Gaussian approximation (see, e.g., Zhang and Wu, 2017, Thm. 6.2).

Remark A.3. In principle, we can let $\{g(n)\}_{n=1}^{\infty}$ be an array $\{g_n(i) : i = 1, 2, ...\}_{n=1}^{\infty}$ allowing for a different rate of convergence for each i, for example, $g_n(i)\mathcal{X}_n(i) \xrightarrow{p} 0$. We only treat a single sequence $\{g(n)\}_{n=1}^{\infty}$ for brevity, and given the environment in which we apply the theory.

Proof.

Claim (a). By assumption, each $\mathcal{X}_n(i) \xrightarrow{p} 0$, therefore, $\vartheta([\mathcal{X}_n(i)]_{i=1}^{k}) \xrightarrow{p} 0$ for each k. Define $\mathcal{A}_{k,n} \equiv 1 - \exp\{-\vartheta([\mathcal{X}_n(i)]_{i=1}^{k})\} \in [0, 1]$ *a.s.* $\forall(k, n)$ and $\mathcal{P}_{k,n} \equiv \int_0^{\infty} P(\mathcal{A}_{k,n} > \epsilon)d\epsilon$. Note that $\mathcal{P}_{k,n} \leq \mathcal{P}_{k+1,n}$ because $\mathcal{A}_{k,n} \leq \mathcal{A}_{k+1,n}$ by monotonicity of ϑ. Lebesgue's dominated convergence theorem, and $\mathcal{A}_{k,n} \xrightarrow{p} 0$, therefore yield for each k

$$\lim_{n\to\infty} \mathcal{P}_{k,n} = \lim_{n\to\infty} \int_0^{\infty} P\left(\mathcal{A}_{k,n} > \epsilon\right) d\epsilon = \lim_{n\to\infty} \int_0^{1} P\left(\mathcal{A}_{k,n} > \epsilon\right) d\epsilon$$
$$= \int_0^{1} \lim_{n\to\infty} P\left(\mathcal{A}_{k,n} > \epsilon\right) d\epsilon = 0.$$

Hence, $\lim_{n\to\infty} \mathcal{P}_{k,n} = 0$ for each k, and therefore $\lim_{k\to\infty} \lim_{n\to\infty} \mathcal{P}_{k,n} = 0$.

Now apply Lemma A.1 to $\mathcal{P}_{k,n}$ to deduce that there exists a positive integer sequence $\{\mathcal{L}_n\}$ that is not unique and monotonically increasing, where $\mathcal{L}_n \to \infty$, and $\mathcal{L}_n = o(n)$ is always feasible, such that $\lim_{n\to\infty} \mathcal{P}_{\mathcal{L}_n,n} = \lim_{n\to\infty} \int_0^1 P(\mathcal{A}_{\mathcal{L}_n,n} > \epsilon)d\epsilon = 0$. Furthermore, the limit holds for any other monotonic sequence of positive integers $\{\tilde{\mathcal{L}}_n\}$ with $\tilde{\mathcal{L}}_n \to \infty$ and $\lim_{n\to\infty} \tilde{\mathcal{L}}_n/\mathcal{L}_n \leq 1$. Therefore, by construction $E[\mathcal{A}_{\mathcal{L}_n,n}] = \int_0^1 P(\mathcal{A}_{\mathcal{L}_n,n} > \epsilon)d\epsilon \to 0$. Hence $\mathcal{A}_{\mathcal{L}_n,n} \xrightarrow{p} 0$ by Markov's inequality, which yields $\vartheta([\mathcal{X}_n(i)]_{i=1}^{\mathcal{L}_n}) \xrightarrow{p} 0$ as claimed.

Now consider an upper bound on $\mathcal{L}_n \to \infty$ when ϑ is the maximum transform $\vartheta([\mathcal{X}_n(i)]_{i=1}^{\mathcal{L}_n}) = \max_{1\leq i\leq \mathcal{L}_n} |\mathcal{X}_n(i)|$. By assumption, $g(n)|\mathcal{X}_n(i)| \xrightarrow{p} 0$ $\forall i \in \mathbb{N}$ for some nonrandom positive $g(n) \to \infty$. Bonferroni and Markov inequalities yield

$$P\left(\max_{1\leq i\leq \mathcal{L}_n} |\mathcal{X}_n(i)| > \eta\right) = P\left(\bigcup_{i=1}^{\mathcal{L}_n} |\mathcal{X}_n(i)| > \eta\right) \leq \sum_{i=1}^{\mathcal{L}_n} P\left(|\mathcal{X}_n(i)| > \eta\right) \tag{A.3}$$
$$\leq \frac{1}{\eta} \sum_{i=1}^{\mathcal{L}_n} E|\mathcal{X}_n(i)| = \frac{\mathcal{L}_n}{\eta g(n)} \frac{1}{\mathcal{L}_n} \sum_{i=1}^{\mathcal{L}_n} g(n) E|\mathcal{X}_n(i)|.$$

https://doi.org/10.1017/S0266466619000367 Published online by Cambridge University Press

Since $g(n)|\mathcal{X}_n(i)| \xrightarrow{p} 0\ \forall i \in \mathbb{N}$, if $g(n)\mathcal{X}_n(i)$ is also uniformly integrable for each i then $g(n)E|\mathcal{X}_n(i)| \to 0\ \forall i \in \mathbb{N}$ (e.g., Billingsley, 1999, Thm. 3.5). Thus $1/\mathcal{L}_n \sum_{i=1}^{\mathcal{L}_n} g(n) E|\mathcal{X}_n(i)| \to 0$, hence $P(\max_{1\le i\le \mathcal{L}_n}|\mathcal{X}_n(i)| > \eta) = o(\mathcal{L}_n/g(n))$. Therefore, any $\mathcal{L}_n = O(g(n))$ yields $\max_{1\le i\le \mathcal{L}_n}|\mathcal{X}_n(i)| \xrightarrow{p} 0$ as required.

Claim (b). The mapping ϑ satisfies the triangle inequality and $\vartheta(\cdot) \ge 0$. Apply the inequality twice to yield

$$\vartheta\left([\mathcal{X}_n(i)]_{i=1}^{\mathcal{L}_n}\right) \le \vartheta\left([\mathcal{Y}_n(i)]_{i=1}^{\mathcal{L}_n}\right) + \vartheta\left([\mathcal{X}_n(i) - \mathcal{Y}_n(i)]_{i=1}^{\mathcal{L}_n}\right)$$

and

$$\vartheta\left([\mathcal{Y}_n(i)]_{i=1}^{\mathcal{L}_n}\right) \le \vartheta\left([\mathcal{X}_n(i)]_{i=1}^{\mathcal{L}_n}\right) + \vartheta\left([\mathcal{X}_n(i) - \mathcal{Y}_n(i)]_{i=1}^{\mathcal{L}_n}\right),$$

hence

$$\left|\vartheta\left([\mathcal{X}_n(i)]_{i=1}^{\mathcal{L}_n}\right) - \vartheta\left([\mathcal{Y}_n(i)]_{i=1}^{\mathcal{L}_n}\right)\right| \le \vartheta\left([\mathcal{X}_n(i) - \mathcal{Y}_n(i)]_{i=1}^{\mathcal{L}_n}\right).$$

Now apply (a) to $\mathcal{X}_n(i) - \mathcal{Y}_n(i)$ to yield $\vartheta([\mathcal{X}_n(i) - \mathcal{Y}_n(i)]_{i=1}^{\mathcal{L}_n}) \xrightarrow{p} 0$. The upper bound on $\mathcal{L}_n$ follows from the Claim (a) argument. $\mathcal{QED}$. □

Lemma A.3 similarly uses Lemma A.1 as the basis for *in distribution* convergence of a function of an increasing set of random variables. This result is used to prove the Gaussian approximation Lemma 2.2. The following result, however, does not use any distributional assumptions other than convergence.

LEMMA A.3. *Let* $\{\mathcal{X}_n(i) : i \in \mathbb{N}\}_{n\ge 1}$ *be an array of random variables, and assume* $\{\mathcal{X}_n(i) : 1 \le i \le \mathcal{L}\} \xrightarrow{d} \{\mathcal{X}(i) : 1 \le i \le \mathcal{L}\}$ *for each* $\mathcal{L} \in \mathbb{N}$, *where* $\{\mathcal{X}(i) : 1 \le i \le \infty\}$ *is a stochastic process. Then* $\vartheta([\mathcal{X}_n(i)]_{i=1}^{\mathcal{L}_n}) \xrightarrow{d} \vartheta([\mathcal{X}(i)]_{i=1}^{\infty})$ *for some nonunique monotonic sequence of positive integers* $\{\mathcal{L}_n\}$, *where* $\mathcal{L}_n \to \infty$ *and* $\mathcal{L}_n = o(n)$. *Moreover,* $\vartheta([\mathcal{X}_n(i)]_{i=1}^{\tilde{\mathcal{L}}_n}) \xrightarrow{d} \vartheta([\mathcal{X}(i)]_{i=1}^{\infty})$ *for any other monotonic sequence of positive integers* $\{\tilde{\mathcal{L}}_n\}$ *with* $\tilde{\mathcal{L}}_n \to \infty$ *and* $\lim_{n\to\infty} \tilde{\mathcal{L}}_n/\mathcal{L}_n \le 1$.

Proof.

Step 1. Convergence in finite dimensional distributions, continuity of $\vartheta(\cdot) \ge 0$, and the mapping theorem yield $\vartheta([\mathcal{X}_n(i)]_{i=1}^{k}) \xrightarrow{d} \vartheta([\mathcal{X}(i)]_{i=1}^{k})$ for each $k \in \mathbb{N}$. By the definition of distribution convergence:

$$P\left(\vartheta\left([\mathcal{X}_n(i)]_{i=1}^{k}\right) \ge x\right) \to P\left(\vartheta\left([\mathcal{X}(i)]_{i=1}^{k}\right) \ge x\right) \text{ for each } k \in \mathbb{N} \text{ and } x \in [0,\infty).$$

Similarly, by the mapping theorem $P(\exp\{-\vartheta([\mathcal{X}_n(i)]_{i=1}^{k})\} \ge x) \to P(\exp\{-\vartheta([\mathcal{X}(i)]_{i=1}^{k})\} \ge x)$ for each $x \in [0,1]$. Now define

$$\mathcal{I}_{k,n} \equiv \int_0^1 P\left(\exp\left\{-\vartheta\left([\mathcal{X}_n(i)]_{i=1}^{k}\right)\right\} \ge x\right) dx,$$

$$\mathcal{I}_k \equiv \int_0^1 P\left(\exp\left\{-\vartheta\left([\mathcal{X}(i)]_{i=1}^{k}\right)\right\} \ge x\right) dx \quad \text{and}$$

$$\mathcal{I} \equiv \int_0^1 P\left(\exp\left\{-\vartheta\left([\mathcal{X}(i)]_{i=1}^{\infty}\right)\right\} \ge x\right) dx.$$

https://doi.org/10.1017/S0266466619000367 Published online by Cambridge University Press

Notice $\mathcal{I}_{k,n}$, $\mathcal{I}_k$, and $\mathcal{I}$ are well defined for any k and n because $\vartheta(\cdot) \geq 0$ and $\exp\{-\vartheta(\cdot)\} \in [0, 1]$. Then by Lebesgue's dominated convergence theorem

$$\lim_{n\to\infty} \int_0^1 \left| P\left(\exp\left\{-\vartheta\left([\mathcal{X}_n(i)]_{i=1}^k\right)\right\} \geq x\right) \right.$$
$$\left. - P\left(\exp\left\{-\vartheta\left([\mathcal{X}(i)]_{i=1}^k\right)\right\} \geq x\right) \right| dx = 0$$

and

$$\lim_{k\to\infty} \int_0^1 \left| P\left(\exp\left\{-\vartheta\left([\mathcal{X}(i)]_{i=1}^k\right)\right\} \geq x\right) \right.$$
$$\left. - P\left(\exp\left\{-\vartheta\left([\mathcal{X}(i)]_{i=1}^\infty\right)\right\} \geq x\right) \right| dx = 0.$$

Scheffé's lemma now yields $\lim_{n\to\infty} \mathcal{I}_{k,n} = \mathcal{I}_k$ for each k, and $\lim_{k\to\infty} \mathcal{I}_k = \mathcal{I}$.

Now apply Lemma A.1 to $\mathcal{I}_{k,n}$ to yield $\lim_{n\to\infty} \mathcal{I}_{\mathcal{L}_n,n} = \mathcal{I}$ for some nonunique, monotonically increasing positive integer sequence $\{\mathcal{L}_n\}$, where $\mathcal{L}_n \to \infty$. Therefore,

$$\lim_{n\to\infty} \int_0^1 P\left(\exp\left\{-\vartheta\left([\mathcal{X}_n(i)]_{i=1}^{\mathcal{L}_n}\right)\right\} \geq x\right) dx$$
$$= \int_0^1 P\left(\exp\left\{-\vartheta\left([\mathcal{X}(i)]_{i=1}^\infty\right)\right\} \geq x\right) dx,$$

hence identically

$$\lim_{n\to\infty} E[\exp\{-\vartheta([\mathcal{X}_n(i)]_{i=1}^{\mathcal{L}_n})\}] = E\left[\exp\left\{-\vartheta\left([\mathcal{X}(i)]_{i=1}^\infty\right)\right\}\right]. \tag{A.4}$$

Moreover, in view of Lemma A.1, (A.4) holds for any other monotonic sequence of positive integers $\{\tilde{\mathcal{L}}_n\}_{n=1}^\infty$, $\tilde{\mathcal{L}}_n \to \infty$ that satisfies $\lim_{n\to\infty} \tilde{\mathcal{L}}_n/\mathcal{L}_n \leq 1$, hence $\mathcal{L}_n = o(n)$ is always feasible.

Step 2. Repeat Step 1 to deduce that for each $s \in \mathbb{N}$ and some nonunique, positive, monotonically increasing integer sequences $\{\mathcal{L}_s(n)\}$, $\mathcal{L}_s(n) = o(n)$ and $\mathcal{L}_s(n) \to \infty$:

$$\lim_{n\to\infty} E\left[\exp\left\{-\vartheta\left([\mathcal{X}_n(i)]_{i=1}^{\mathcal{L}_s(n)}\right)\right\}^s\right] = E\left[\exp\left\{-\vartheta\left([\mathcal{X}(i)]_{i=1}^\infty\right)\right\}^s\right]. \tag{A.5}$$

Each $\{\mathcal{L}_s(n)\}$ is monotonic, and any other sequence $\{\tilde{\mathcal{L}}_n\}_{n=1}^\infty$, $\tilde{\mathcal{L}}_n \to \infty$ that satisfies $\lim_{n\to\infty} \tilde{\mathcal{L}}_n/\mathcal{L}_s(n) \leq 1$ and $\tilde{\mathcal{L}}_n = o(n)$ is also valid. Hence, there exists a nonunique monotonic sequence of positive integers $\{\mathcal{L}_n\}$ such that $\mathcal{L}_n \to \infty$, $\mathcal{L}_n = o(n)$, and $\limsup_{n\to\infty}\{\mathcal{L}_n/\mathcal{L}_s(n)\} < 1$ for each s, that satisfies (A.5) for each s. Therefore,

$$\lim_{n\to\infty} E\left[\exp\left\{-\vartheta\left([\mathcal{X}_n(i)]_{i=1}^{\mathcal{L}_n}\right)\right\}^s\right] = E\left[\exp\left\{-\vartheta\left([\mathcal{X}(i)]_{i=1}^\infty\right)\right\}^s\right] \quad \forall s \in \mathbb{N}. \tag{A.6}$$

Property (A.6) implies $\exp\{-\vartheta([\mathcal{X}_n(i)]_{i=1}^{\mathcal{L}_n})\} \xrightarrow{d} \exp\{-\vartheta([\mathcal{X}(i)]_{i=1}^\infty)\}$ (Billingsley, 1995, Thm. 30.2). The claim now follows by the mapping theorem. $\mathcal{QED}$.

□

https://doi.org/10.1017/S0266466619000367 Published online by Cambridge University Press

Let $h \geq 0$. Recall $\rho(h) \equiv E[\epsilon_t \epsilon_{t-h}]/E[\epsilon_t^2]$ and

$$G_t(\phi) \equiv \left[\frac{\partial}{\partial \phi'} f(x_{t-1}, \phi), \mathbf{0}'_{k_\delta}\right]' \in \mathbb{R}^{k_\theta} \text{ and } s_t(\theta) \equiv \frac{1}{2}\frac{\partial}{\partial \theta} \ln \sigma_t^2(\theta)$$

$$\mathcal{D}(h) \equiv E\left[(\epsilon_t s_t + G_t/\sigma_t)\epsilon_{t-h}\right] + E\left[\epsilon_t\left(\epsilon_{t-h}s_{t-h} + G_{t-h}/\sigma_{t-h}\right)\right] \in \mathbb{R}^{k_\theta}$$

$$z_t(h) \equiv r_t(h) - \rho(h) r_t(0) \text{ where } r_t(h) \equiv \frac{\epsilon_t \epsilon_{t-h} - E[\epsilon_t \epsilon_{t-h}] - \mathcal{D}(h)' \mathcal{A} m_t}{E\left[\epsilon_t^2\right]},$$

where m_t and $\mathcal{A}$ appear in plug-in expansion Assumption 2.c: $\sqrt{n}(\hat{\theta}_n - \theta_0) = \mathcal{A} n^{-1/2} \sum_{t=1}^n m_t(\theta_0) + O_p(n^{-\zeta})$ for some $\zeta > 0$. The following two lemmas are based on standard arguments and are therefore proved in Appendix F.

LEMMA A.4. *Under Assumptions 1 and 2: for some $\zeta > 0$ that appears in Assumption 2.c, $\mathcal{X}_n(h) \equiv |\sqrt{n}\{\hat{\rho}_n(h) - \rho(h)\} - 1/\sqrt{n}\sum_{t=1+h}^n \{r_t(h) - \rho(h) r_t(0)\}| = O_p(1/n^{\min\{\zeta, 1/2\}})$ for each h.*

LEMMA A.5. *Let Assumptions 1 and 2 hold, and write $\mathcal{Z}_n(h) \equiv 1/\sqrt{n}\sum_{t=1+h}^n z_t(h)$. For each $\mathcal{L} \in \mathbb{N}$: $\{\mathcal{Z}_n(h) : 1 \leq h \leq \mathcal{L}\} \xrightarrow{d} \{\mathcal{Z}(h) : 1 \leq h \leq \mathcal{L}\}$, where $\{\mathcal{Z}(h) : 1 \leq h \leq \mathcal{L}\}$ is a zero-mean Gaussian process with variance $\lim_{n\to\infty} n^{-1} \sum_{s,t=1}^n E[z_s(h) z_t(h)] \in (0, \infty)$, and covariance function $\lim_{n\to\infty} n^{-1} \sum_{s,t=1}^n E[z_s(h) z_t(\tilde{h})]$.*

Proof of Lemma 2.1. Recall Assumption 1.c states $\hat{\omega}_n(h) = \omega(h) + O_p(n^{-\kappa})$ for some $\kappa > 0$, and under Assumption 2.c, $\sqrt{n}(\hat{\theta}_n - \theta_0) = \mathcal{A} n^{-1/2} \sum_{t=1}^n m_t(\theta_0) + O_p(n^{-\zeta})$ for some $\zeta > 0$.

Property (A.1) applies to $r_t(h) - \rho(h) r_t(0)$ under Assumptions 1 and 2, cf. Thm. 17.8 in Davidson (1994), hence $1/\sqrt{n}\sum_{t=1+h}^n \{r_t(h) - \rho(h) r_t(0)\} = O_p(1)$. By Lemma A.4,

$$\mathcal{X}_n(h) \equiv \left|\sqrt{n}\left\{\hat{\rho}_n(h) - \rho(h)\right\} - \frac{1}{\sqrt{n}} \sum_{t=1+h}^n \{r_t(h) - \rho(h) r_t(0)\}\right| = O_p\left(1/n^{\min\{\zeta, 1/2\}}\right).$$

Therefore,

$$\begin{aligned}
\left|\tilde{\mathcal{X}}_n(h)\right| &\equiv \left|\sqrt{n}\hat{\omega}_n(h)\left\{\hat{\rho}_n(h) - \rho(h)\right\} - \omega(h)\frac{1}{\sqrt{n}} \sum_{t=1+h}^n \{r_t(h) - \rho(h) r_t(0)\}\right| \qquad \textbf{(A.7)} \\
&\leq |\omega(h)| \times \mathcal{X}_n(h) + \left|\hat{\omega}_n(h) - \omega(h)\right| \times \mathcal{X}_n(h) \\
&\quad + \left|\hat{\omega}_n(h) - \omega(h)\right| \times \left|\frac{1}{\sqrt{n}} \sum_{t=1+h}^n \{r_t(h) - \rho(h) r_t(0)\}\right| \\
&= O_p\left(1/n^{\min\{\zeta, 1/2\}}\right) + O_p\left(1/n^{\kappa}\right) \\
&= O_p\left(1/n^{\min\{\zeta, \kappa, 1/2\}}\right),
\end{aligned}$$

https://doi.org/10.1017/S0266466619000367 Published online by Cambridge University Press

which implies $(n^{\min\{\zeta,\kappa,1/2\}}/\ln(n))|\tilde{\mathcal{X}}_n(h)| \xrightarrow{p} 0$. Setting $\hat{\rho}_n(h) = 0$ for any $h \notin \{0,...,n\}$, the claims now follow by applications of Lemma A.2 to $\{|\tilde{\mathcal{X}}_n(h)|\}_{h\in\mathbb{N}}$. In particular, by (A.7) and Lemma A.2, if $(n^{\min\{\zeta,\kappa,1/2\}}/\ln(n))\tilde{\mathcal{X}}_n(h)$ for each h is uniformly integrable then $\mathcal{L}_n = O(n^{\min\{\zeta,\kappa,1/2\}}/\ln(n))$. $\mathcal{QED}$.

Proof of Lemma 2.2. Lemma A.5 implies $\{\mathcal{Z}_n(h) : 1 \le h \le \mathcal{L}\} \xrightarrow{d} \{\mathcal{Z}(h) : 1 \le h \le \mathcal{L}\}$ for each $\mathcal{L} \in \mathbb{N}$, where $\{\mathcal{Z}(h) : 1 \le h \le \mathcal{L}\}$ is a zero-mean Gaussian process. Now apply Lemma A.3 to prove the claim. $\mathcal{QED}$. □

The proof of Theorem 2.5 requires the following uniform laws and probability bound, and weak convergence for the bootstrapped p-value. The first result is rudimentary and therefore proved in Appendix F. Recall m_t are the Assumption 2.c′ estimating equations.

LEMMA A.6. *Under Assumptions 1 and 2.a,b,c′,d* $\sup_{\theta\in\Theta}||1/n\sum_{t=1}^n \omega_t(\partial/\partial\theta)m_t(\theta)|| \xrightarrow{p} 0, \sup_{\theta\in\Theta}||1/n\sum_{t=1}^n(\partial/\partial\theta)m_t(\theta) - E[(\partial/\partial\theta)m_t(\theta)]|| \xrightarrow{p} 0$, *and* $1/\sqrt{n}\sum_{t=1+h}^n \omega_t m_t = O_p(1)$.

Let $\Rightarrow^p$ denote weak convergence in probability on l_∞ (the space of bounded functions) as defined in Giné and Zinn (1990, Sect. 3). Recall by Lemma 2.2 that $|\vartheta([\mathcal{Z}_n(h)]_{h=1}^{\mathcal{L}_n}) - \vartheta([\mathcal{Z}(h)]_{h=1}^{\mathcal{L}_n})| \xrightarrow{p} 0$ for some zero-mean Gaussian process $\{\mathcal{Z}(h) : h \in \mathbb{N}\}$ with variance $\lim_{n\to\infty} n^{-1}\sum_{s,t=1}^n E[z_s(h)z_t(h)] < \infty$. Define the sample

$$\mathfrak{X}_n \equiv \{m_t, x_t, y_t\}_{t=1}^n.$$

LEMMA A.7. *Let Assumptions 1 and 2.a,b,c′,d hold.*

a. *For each* $\mathcal{L} \in \mathbb{N}$, $\{\sqrt{n}\hat{\rho}_n^{(dw)}(h) : 1 \le h \le \mathcal{L}\} \Rightarrow^p \{\mathring{\mathcal{Z}}(h) : 1 \le h \le \mathcal{L}\}$, *where* $\{\mathring{\mathcal{Z}}(h) : h \in \mathbb{N}\}$ *is an independent copy of* $\{\mathcal{Z}(h) : h \in \mathbb{N}\}$.

b. *For some monotonic sequence of positive integers* $\{\mathcal{L}_n\}$, $\mathcal{L}_n \to \infty$ *and* $\mathcal{L}_n = o(n)$:

$$\sup_{c>0}\left|P\left(\vartheta\left(\left[\hat{\omega}_n(h)\sqrt{n}\hat{\rho}_n^{(dw)}(h)\right]_{h=1}^{\mathcal{L}_n}\right) \le c|\mathfrak{X}_n\right) - P\left(\vartheta\left(\left[\omega(h)\mathring{\mathcal{Z}}(h)\right]_{h=1}^{\mathcal{L}_n}\right) \le c\right)\right| \xrightarrow{p} 0.$$

Finally, $\mathcal{L}_n = O(n^{\min\{\zeta,\kappa,1/2\}}/\ln(n))$ *must be satisfied.*

Proof.

Claim (a). Let $\{\varphi_t\}_{t=1}^n$ be a draw of the auxiliary variables, and write

$$\rho_n^*(h) \equiv \frac{1}{E\left[\epsilon_t^2\right]}\frac{1}{n}\sum_{t=1+h}^n \varphi_t\left\{\mathcal{E}_{t,h} - E\left[\mathcal{E}_{1,h}\right]\right\} \text{ where } \mathcal{E}_{t,h} \equiv \epsilon_t\epsilon_{t-h} - \mathcal{D}(h)'\mathcal{A}m_t. \tag{A.8}$$

Recall $\widehat{\mathcal{E}}_{n,t,h}(\hat{\theta}_n) \equiv \epsilon_t(\hat{\theta}_n)\epsilon_{t-h}(\hat{\theta}_n) - \hat{\mathcal{D}}_n(h)'\hat{\mathcal{A}}_n m_t(\hat{\theta}_n)$, and

$$\hat{\rho}_n^{(dw)}(h) \equiv \frac{1}{1/n\sum_{t=1}^n \epsilon_t^2(\hat{\theta}_n)}\frac{1}{n}\sum_{t=1+h}^n \varphi_t\left\{\widehat{\mathcal{E}}_{n,t,h}(\hat{\theta}_n) - \frac{1}{n}\sum_{s=1+h}^n \widehat{\mathcal{E}}_{n,s,h}(\hat{\theta}_n)\right\}.$$

https://doi.org/10.1017/S0266466619000367 Published online by Cambridge University Press

Let $\{\mathcal{Z}(h) : h \in \mathbb{N}\}$ be the Lemma 2.2 Gaussian process. It suffices to show

$$\left\{\sqrt{n}\rho_n^*(h) : 1 \le h \le \mathcal{L}\right\} \Rightarrow^p \left\{\mathring{\mathcal{Z}}(h) : 1 \le h \le \mathcal{L}\right\}, \tag{A.9}$$

$$\sqrt{n}\left|\hat{\rho}_n^{(dw)}(h) - \rho_n^*(h)\right| \xrightarrow{p} 0 \text{ for each } h, \tag{A.10}$$

where $\{\mathring{\mathcal{Z}}(h) : h \in \mathbb{N}\}$ is an independent copy of $\{\mathcal{Z}(h) : h \in \mathbb{N}\}$. We shorten the proof by letting $\{\xi_1, \ldots, \xi_{n/b_n}\}$ be i.i.d. $N(0, 1)$ random variables. The general case is similar, where ξ_i are i.i.d., $E[\xi_i] = 0$, $E[\xi_i^2] = 1$ and $E[\xi_i^4] < \infty$, except statements about conditional distribution normality must be replaced with added steps to show asymptotic convergence in conditional distribution.

Step 1. Consider (A.9). Define $\mathbb{L} \equiv \{1, ..., \mathcal{L}\}$. It suffices to prove weak convergence on a Polish space in the sense of Hoffmann-Jrgensen (1984, 1991), cf. Giné and Zinn (1990, p. 853, Thm. 3.1.a). The latter holds *if and only if* there exists a pseudo metric d on $\mathbb{L}$ such that $(\mathbb{L}, d)$ is a totally bounded pseudo metric space; $\{\sqrt{n}\rho_n^*(h) : 1 \le h \le \mathcal{L}\} \xrightarrow{d} \{\mathring{\mathcal{Z}}(h) : 1 \le h \le \mathcal{L}\}$; and the sequence of distributions governing $\{\sqrt{n}\rho_n^*(h)\}_{n\ge1}$ are stochastically equicontinuous on $\mathbb{L}$. $\mathbb{L}$ is compact, so pick the sup-norm d. Stochastic equicontinuity is trivial because $\mathbb{L}$ is discrete and bounded. It now suffices to prove convergence in finite dimensional distributions. We follow an argument given in Hansen (1996, proof of Thm. 2).

By construction of φ_t via ξ_t,

$$\begin{aligned}\rho_n^*(h) &= \frac{1}{E\left[\epsilon_t^2\right]}\frac{1}{n/b_n}\sum_{s=1}^{n/b_n}\xi_s\frac{1}{b_n}\sum_{t=(s-1)b_n+1+h}^{sb_n}\left\{\mathcal{E}_{t,h} - E\left[\mathcal{E}_{1,h}\right]\right\} \\ &= \frac{1}{E\left[\epsilon_t^2\right]}\frac{1}{n/b_n}\sum_{s=1}^{n/b_n}\xi_s\frac{1}{b_n}\mathfrak{E}_{n,h},\end{aligned}$$

say, where $\mathfrak{E}_{n,h} \equiv \sum_{t=(s-1)b_n+1+h}^{sb_n}\{\mathcal{E}_{t,h} - E\left[\mathcal{E}_{1,h}\right]\}$. Operate conditionally on $\mathfrak{X}_n \equiv \{m_t, x_t, y_t\}_{t=1}^n$, and write $E_{\mathfrak{X}_n}[\cdot] \equiv E[\cdot|\mathfrak{X}_n]$. By joint Gaussianicity and independence of ξ_s, $\{\sqrt{n}\rho_n^*(h) : 1 \le h \le \mathcal{L}\}$ is for each $\mathcal{L} \in \mathbb{N}$ a zero-mean Gaussian process with covariance function $nE_{\mathfrak{X}_n}[\rho_n^*(h)\rho_n^*(\tilde{h})] = 1/n\sum_{s=1}^{n/b_n}\mathfrak{E}_{n,h}\mathfrak{E}_{n,\tilde{h}}/(E[\epsilon_t^2])^2$. Observe

$$\begin{aligned}&\lim_{n\to\infty} E\left[nE_{\mathfrak{X}_n}\left[\rho_n^*(h)\rho_n^*(\tilde{h})\right]\right] \\ &= \frac{1}{\left[E\left[\epsilon_t^2\right]\right]^2}\lim_{n\to\infty}\frac{1}{n}\sum_{s=1}^{n/b_n}\sum_{t,u=(s-1)b_n+1+h}^{sb_n} E\left[\left\{\mathcal{E}_{t,h} - E\left[\mathcal{E}_{1,h}\right]\right\}\left\{\mathcal{E}_{u,\tilde{h}} - E\left[\mathcal{E}_{1,\tilde{h}}\right]\right\}\right] \\ &= \frac{1}{\left[E\left[\epsilon_t^2\right]\right]^2}\sum_{i=0}^{\infty} E\left[\left\{\mathcal{E}_{1,h} - E\left[\mathcal{E}_{1,h}\right]\right\}\left\{\mathcal{E}_{1+i,\tilde{h}} - E\left[\mathcal{E}_{1,\tilde{h}}\right]\right\}\right] \\ &= \lim_{n\to\infty}\frac{1}{n}E\left[\sum_{t=1}^{n}\frac{\left(\mathcal{E}_{t,h} - E\left[\mathcal{E}_{t,h}\right]\right)}{E\left[\epsilon_t^2\right]}\sum_{t=1}^{n}\frac{\left(\mathcal{E}_{t,\tilde{h}} - E\left[\mathcal{E}_{t,\tilde{h}}\right]\right)}{E\left[\epsilon_t^2\right]}\right] = E\left[\mathcal{Z}(h)\mathcal{Z}(\tilde{h})\right].\end{aligned} \tag{A.11}$$

The final equality follows directly from the definition of $\mathcal{Z}(h)$ in Lemma 2.2.

https://doi.org/10.1017/S0266466619000367 Published online by Cambridge University Press

Let $\mathfrak{X}$ be the set of samples $\tilde{\mathfrak{X}}_n$ such that $nE_{\tilde{\mathfrak{X}}_n}[\rho_n^*(h)\rho_n^*(\tilde{h})] \xrightarrow{p} \lim_{n\to\infty} E[nE_{\tilde{\mathfrak{X}}_n}[\rho_n^*(h)\rho_n^*(\tilde{h})]] = E[\mathcal{Z}(h)\mathcal{Z}(\tilde{h})]$. We will prove

$$P(\mathfrak{X}_n \in \mathfrak{X}) = 1. \tag{A.12}$$

In conjunction with (A.11), it then follows that the finite dimensional distributions of $\{\sqrt{n}\rho_n^*(h) : 1 \le h \le \mathcal{L}\}$ converge to those of $\{\mathring{\mathcal{Z}}(h) : 1 \le h \le \mathcal{L}\}$, where $\{\mathring{\mathcal{Z}}(h) : 1 \le h \le \mathcal{L}\}$ is a zero-mean Gaussian process with covariance function $E[\mathcal{Z}(h)\mathcal{Z}(\tilde{h})]$. Independence of ξ_s with respect to the sample $\mathfrak{X}_n$, Gaussianicity, and the fact that Gaussian processes are completely determined by their mean and covariance structure, together imply $\{\mathring{\mathcal{Z}}(h) : 1 \le h \le \mathcal{L}\}$ is an independent copy of $\{\mathcal{Z}(h) : 1 \le h \le \mathcal{L}\}$.

Consider (A.12). The following exploits arguments presented in de Jong (1997, Appendix). Let $\{l_n\}$ be any sequence of integers $l_n \in \{1,..., b_n\}$ such that $l_n \to \infty$ and $l_n = o(b_n)$. Define

$$\begin{aligned}\mathcal{Y}_{n,s}(h) &\equiv \sum_{t=(s-1)b_n+l_n+1}^{sb_n} \left\{\mathcal{E}_{t,h} - E\left[\mathcal{E}_{1,h}\right]\right\},\ \mathcal{U}_{n,s}(h) \\ &\equiv \sum_{t=(s-1)b_n+1}^{(s-1)b_n+l_n} \left\{\mathcal{E}_{t,h} - E\left[\mathcal{E}_{1,h}\right]\right\}, \mathcal{R}(h) \equiv -\sum_{t=1}^{h}\{\mathcal{E}_{t,h} - E[\mathcal{E}_{1,h}]\}.\end{aligned}$$

By construction $\sum_{t=(s-1)b_n+1+h}^{sb_n}\{\mathcal{E}_{t,h} - E[\mathcal{E}_{1,h}]\} = \mathcal{Y}_{n,s}(h) + \mathcal{U}_{n,s}(h) + \mathcal{R}(h)$, hence

$$\begin{aligned}&\frac{1}{n}\sum_{s=1}^{n/b_n} \mathfrak{E}_{n,h}\mathfrak{E}_{n,\tilde{h}} \\ &= \frac{1}{n}\sum_{s=1}^{n/b_n} \mathcal{Y}_{n,s}(h)\mathcal{Y}_{n,s}(\tilde{h}) + \frac{1}{n}\sum_{s=1}^{n/b_n} \mathcal{U}_{n,s}(h)\mathcal{U}_{n,s}(\tilde{h}) + \frac{1}{b_n}\mathcal{R}(h)\mathcal{R}(\tilde{h}) \\ &\quad + \frac{1}{n}\sum_{s=1}^{n/b_n} \mathcal{Y}_{n,s}(h)\mathcal{U}_{n,s}(\tilde{h}) + \frac{1}{n}\sum_{s=1}^{n/b_n} \mathcal{Y}_{n,s}(\tilde{h})\mathcal{U}_{n,s}(h) + \frac{1}{n}\sum_{s=1}^{n/b_n} \mathcal{Y}_{n,s}(h)\mathcal{R}(\tilde{h}) \\ &\quad + \frac{1}{n}\sum_{s=1}^{n/b_n} \mathcal{Y}_{n,s}(\tilde{h})\mathcal{R}(h) + \frac{1}{n}\sum_{s=1}^{n/b_n} \mathcal{U}_{n,s}(h)\mathcal{R}(\tilde{h}) + \frac{1}{n}\sum_{s=1}^{n/b_n} \mathcal{U}_{n,s}(\tilde{h})\mathcal{R}(h).\end{aligned}$$

We will prove all terms are $o_p(1)$ save $1/n\sum_{s=1}^{n/b_n} \mathcal{Y}_{n,s}(h)\mathcal{Y}_{n,s}(\tilde{h})$ hence

$$\frac{1}{n}\sum_{s=1}^{n/b_n} \mathfrak{E}_{n,h}\mathfrak{E}_{n,\tilde{h}} = \frac{1}{n}\sum_{s=1}^{n/b_n} \mathcal{Y}_{n,s}(h)\mathcal{Y}_{n,s}(\tilde{h}) + o_p(1). \tag{A.13}$$

First, under Assumptions 1 and 2, $\mathcal{E}_{t,h}$ is stationary, ergodic, and L_2-bounded. Therefore, $||\mathcal{R}(\tilde{h})||_2 \le \sum_{t=1}^{\tilde{h}} ||\mathcal{E}_{t,h} - E[\mathcal{E}_{1,h}]||_2 \le K$ for each finite $\tilde{h}$, hence by the Cauchy–Schwartz inequality $E|b_n^{-1}\mathcal{R}(h)\mathcal{R}(\tilde{h})| \le K/b_n \to 0$.

Second, the NED and moment properties of ϵ_t and m_t in Assumptions 1 and 2 imply $\mathcal{E}_{t,h} \equiv \epsilon_t\epsilon_{t-h} - \mathcal{D}(h)'\mathcal{A}m_t$ is L_p-bounded, $p > 2$, L_2-NED on an α-mixing base with decay

https://doi.org/10.1017/S0266466619000367 Published online by Cambridge University Press

$O(h^{-p/(p-2)})$. Therefore $||1/\sqrt{b_n}\mathcal{Y}_{n,1}(h)||_2$ and $||1/\sqrt{l_n}\mathcal{U}_{n,1}(\tilde{h})||_2$ are $O(1)$ by (A.1). Multiply and divide $\mathcal{Y}_{n,s}(h)$ and $\mathcal{U}_{n,s}(\tilde{h})$ by b_n and l_n, respectively, and use stationarity, Minkowski, and Cauchy–Schwartz inequalities, and $l_n/b_n = o(1)$ to yield

$$\left\| \frac{1}{n}\sum_{s=1}^{n/b_n} \mathcal{Y}_{n,s}(h)\mathcal{U}_{n,s}(\tilde{h}) \right\|_1 = O\left(\left(\frac{l_n}{b_n}\right)^{1/2} \left\| \frac{1}{\sqrt{b_n}}\mathcal{Y}_{n,1}(h) \right\|_2 \left\| \frac{1}{\sqrt{l_n}}\mathcal{U}_{n,1}(\tilde{h}) \right\|_2 \right)$$

$$= O\left((l_n/b_n)^{1/2} \right) = o(1),$$

$$\left\| \frac{1}{n}\sum_{s=1}^{n/b_n} \mathcal{Y}_{n,s}(h)\mathcal{R}_n(\tilde{h}) \right\|_1 = O\left(\left\| \frac{1}{\sqrt{b_n}}\mathcal{Y}_{n,1}(h) \right\|_2 \left\| \frac{1}{\sqrt{b_n}}\mathcal{R}_n(\tilde{h}) \right\|_2 \right) = o(1),$$

$$\left\| \frac{1}{n}\sum_{s=1}^{n/b_n} \mathcal{U}_{n,s}(h)\mathcal{U}_{n,s}(\tilde{h}) \right\|_1 = O\left(\frac{l_n}{b_n} \left\| \frac{1}{\sqrt{l_n}}\mathcal{U}_{n,1}(h) \right\|_2 \left\| \frac{1}{\sqrt{l_n}}\mathcal{U}_{n,1}(\tilde{h}) \right\|_2 \right) = o(1),$$

$$\left\| \frac{1}{n}\sum_{s=1}^{n/b_n} \mathcal{U}_{n,s}(h)\mathcal{R}_n(\tilde{h}) \right\|_1 = O\left(\left(\frac{l_n}{b_n}\right)^{1/2} \left\| \frac{1}{\sqrt{l_n}}\mathcal{U}_{n,1}(h) \right\|_2 \right) = o(1).$$

This proves (A.13).

Next, de Jong's (1997, Assumption 2) conditions are satisfied under the given NED property. Hence, by the proof of de Jong's (1997) Thm. 2: $1/n\sum_{s=1}^{n/b_n} \mathcal{Y}_{n,s}^2(h) \xrightarrow{p} \lim_{n\to\infty} n^{-1}E[(\sum_{t=1}^{n}\{\mathcal{E}_{t,h} - E[\mathcal{E}_{1,h}]\})^2]$. An identical argument can be used to prove that the product $\mathcal{Y}_{n,s}(h)\mathcal{Y}_{n,s}(\tilde{h})$ satisfies:

$$\frac{1}{n}\sum_{s=1}^{n/b_n} \mathcal{Y}_{n,s}(h)\mathcal{Y}_{n,s}(\tilde{h}) \xrightarrow{p} \lim_{n\to\infty} \frac{1}{n} E\left[\left(\sum_{t=1}^{n} \{\mathcal{E}_{t,h} - E[\mathcal{E}_{1,h}]\} \right) \left(\sum_{t=1}^{n} \left\{ \mathcal{E}_{t,\tilde{h}} - E\left[\mathcal{E}_{1,\tilde{h}}\right] \right\} \right) \right]. \tag{A.14}$$

Property (A.12) is proved since combining (A.11), (A.13), and (A.14) yields

$$nE_{\mathfrak{X}_n}\left[\rho_n^*(h)\rho_n^*(\tilde{h})\right] \xrightarrow{p} \lim_{n\to\infty} E\left[nE_{\mathfrak{X}_n}\left[\rho_n^*(h)\rho_n^*(\tilde{h})\right] \right] = E\left[\mathcal{Z}(h)\mathcal{Z}(\tilde{h})\right].$$

Step 2. Now turn to (A.10).

Step 2.1 Recall $\mathcal{E}_{t,h} \equiv \epsilon_t\epsilon_{t-h} - \mathcal{D}(h)'\mathcal{A}m_t$ and $\widehat{\mathcal{E}}_{n,t,h}(\hat{\theta}_n) \equiv \epsilon_t(\hat{\theta}_n)\epsilon_{t-h}(\hat{\theta}_n) - \hat{\mathcal{D}}_n(h)'\hat{\mathcal{A}}_n m_t(\hat{\theta}_n)$. We will prove in Step 2.2 that

$$\frac{1}{\sqrt{n}}\sum_{t=1+h}^{n} \varphi_t \left\{ \widehat{\mathcal{E}}_{n,t,h}(\hat{\theta}_n) - \frac{1}{n}\sum_{s=1+h}^{n} \widehat{\mathcal{E}}_{n,s,h}(\hat{\theta}_n) \right\}$$

$$= \frac{1}{\sqrt{n}}\sum_{t=1+h}^{n} \varphi_t \left\{ \mathcal{E}_{t,h} - E\left[\mathcal{E}_{t,h}\right] \right\} + o_p(1) \tag{A.15}$$

https://doi.org/10.1017/S0266466619000367 Published online by Cambridge University Press

by showing (it is straightforward to show (A.16)–(A.19) imply (A.15)):

$$\frac{1}{\sqrt{n}}\sum_{t=1+h}^{n}\varphi_t\epsilon_t(\hat{\theta}_n)\epsilon_{t-h}(\hat{\theta}_n)=\frac{1}{\sqrt{n}}\sum_{t=1+h}^{n}\varphi_t\epsilon_t\epsilon_{t-h}+o_p(1), \tag{A.16}$$

$$\hat{\mathcal{D}}_n(h)'\hat{\mathcal{A}}_n\frac{1}{\sqrt{n}}\sum_{t=1+h}^{n}\varphi_t m_t(\hat{\theta}_n)=\mathcal{D}(h)'\mathcal{A}\frac{1}{\sqrt{n}}\sum_{t=1+h}^{n}\varphi_t m_t+o_p(1), \tag{A.17}$$

$$\frac{1}{\sqrt{n}}\sum\nolimits_{t=1+h}^{n}\varphi_t\frac{1}{n}\sum_{s=1+h}^{n}\epsilon_s(\hat{\theta}_n)\epsilon_{s-h}(\hat{\theta}_n)=\frac{1}{\sqrt{n}}\sum_{t=1+h}^{n}\varphi_t E[\epsilon_t\epsilon_{t-h}]+o_p(1), \tag{A.18}$$

$$\frac{1}{\sqrt{n}}\sum_{t=1+h}^{n}\varphi_t\hat{\mathcal{D}}_n(h)'\hat{\mathcal{A}}_n\frac{1}{n}\sum_{t=1+h}^{n}m_t(\hat{\theta}_n)=o_p(1). \tag{A.19}$$

By the construction of φ_t, for i.i.d. ξ_s distributed $N(0,1)$:

$$\begin{aligned}
&E\left[\left(\frac{1}{\sqrt{n}}\sum_{t=1+h}^{n}\varphi_t\left\{\mathcal{E}_{t,h}-E\left[\mathcal{E}_{t,h}\right]\right\}\right)^2\right]\\
&\quad=E\left[\left(\frac{1}{\sqrt{n}}\sum_{s=1}^{n/b_n}\xi_s\sum_{t=(s-1)b_n+1}^{sb_n}\varphi_t\left\{\mathcal{E}_{t,h}-E\left[\mathcal{E}_{t,h}\right]\right\}\right)^2\right]\\
&\quad=E\left[\left(\frac{1}{\sqrt{b_n}}\sum_{t=1}^{b_n}\left\{\mathcal{E}_{t,h}-E\left[\mathcal{E}_{t,h}\right]\right\}\right)^2\right].
\end{aligned}$$

Under Assumptions 1.b and 2.c′, (A.1) applies to $\mathcal{E}_{t,h}-E[\mathcal{E}_{t,h}]$ (Davidson, 1994, Thms. 17.8 and 17.9). Hence, $E[(1/\sqrt{b_n}\sum_{t=1}^{b_n}\{\mathcal{E}_{t,h}-E\left[\mathcal{E}_{t,h}\right]\})^2]=O(1)$, and therefore,

$$\frac{1}{\sqrt{n}}\sum_{t=1+h}^{n}\varphi_t\left\{\mathcal{E}_{t,h}-E\left[\mathcal{E}_{t,h}\right]\right\}=O_p(1). \tag{A.20}$$

Furthermore, by application of Lemma A.4, $\sqrt{n}\{\hat{\gamma}_n(0)-\gamma(0)\}=n^{-1/2}\sum_{t=1}^{n}\{\epsilon_t^2-E[\epsilon_t^2]-\mathcal{D}(0)'\mathcal{A}m_t\}+O_p(1/\sqrt{n})$. Coupled with stationarity, ergodicity, and square integrability yields

$$\frac{1}{n}\sum_{t=1}^{n}\epsilon_t^2(\hat{\theta}_n)=E\left[\epsilon_t^2\right]+o_p(1). \tag{A.21}$$

Combine (A.15), (A.20), and (A.21) to yield (A.10) as required:

$$\begin{aligned}
\sqrt{n}\hat{\rho}_n^{(dw)}(h)&=\frac{1}{1/n\sum_{t=1}^{n}\epsilon_t^2(\hat{\theta}_n)}\frac{1}{\sqrt{n}}\sum_{t=1+h}^{n}\varphi_t\left\{\mathcal{E}_{t,h}-E\left[\mathcal{E}_{t,h}\right]\right\}+o_p(1)\\
&=\frac{1}{E\left[\epsilon_t^2\right]}\frac{1}{\sqrt{n}}\sum_{t=1+h}^{n}\varphi_t\left\{\mathcal{E}_{t,h}-E\left[\mathcal{E}_{t,h}\right]\right\}+o_p(1).
\end{aligned}$$

https://doi.org/10.1017/S0266466619000367 Published online by Cambridge University Press

Step 2.2 We now prove (A.16)–(A.19). Consider (A.16). Since φ_t is zero-mean Gaussian and independent of the sample, the proof of Lemma 2.1 carries over verbatim to show:

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{t=1+h}^{n}\varphi_t\epsilon_t(\hat{\theta}_n)\epsilon_{t-h}(\hat{\theta}_n)\\
&\quad=\frac{1}{\sqrt{n}}\sum_{t=1+h}^{n}\varphi_t\epsilon_t\epsilon_{t-h}-\sqrt{n}\left(\hat{\theta}_n-\theta_0\right)'\frac{1}{n}\sum_{t=1+h}^{n}\varphi_t\left(\epsilon_t s_t+G_t/\sigma_t\right)\epsilon_{t-h}\\
&\quad-\sqrt{n}\left(\hat{\theta}_n-\theta_0\right)'\frac{1}{n}\sum_{t=1+h}^{n}\varphi_t\epsilon_t\left(\epsilon_{t-h}s_{t-h}+\frac{G_{t-h}}{\sigma_{t-h}}\right)+o_p(1). \qquad \textbf{(A.22)}
\end{aligned}$$

By the stated moment bounds and the construction of φ_t, we have

$$\begin{aligned}
\frac{1}{n}\sum_{t=1+h}^{n}\varphi_t\left(\epsilon_t s_t+G_t/\sigma_t\right)\epsilon_{t-h}&=\frac{1}{n}\sum_{t=1}^{n}\varphi_t\left(\epsilon_t s_t+G_t/\sigma_t\right)\epsilon_{t-h}+o_p(1)\\
&=\frac{1}{n}\sum_{s=1}^{n/b_n}\xi_s\sum_{t=(s-1)b_n+1}^{sb_n}\left(\epsilon_t s_t+G_t/\sigma_t\right)\epsilon_{t-h}+o_p(1).
\end{aligned}$$

Stationarity, independence of ξ_s, and $E[(\epsilon_t s_t+G_t/\sigma_t)^2\epsilon_{t-h}^2]<\infty$ under Assumptions 1.b and 2.a,b yield

$$\begin{aligned}
&E\left[\left(\frac{1}{n}\sum_{s=1}^{n/b_n}\xi_s\left\{\sum_{t=(s-1)b_n+1+h}^{sb_n}\left(\epsilon_t s_t+\frac{G_t}{\sigma_t}\right)\epsilon_{t-h}\right\}\right)^2\right]\\
&\quad=\frac{b_n}{n}E\left[\left\{\frac{1}{b_n}\sum_{t=1}^{b_n}\left(\epsilon_t s_t+\frac{G_t}{\sigma_t}\right)\epsilon_{t-h}\right\}^2\right]\\
&\quad\le\frac{b_n}{n}\left(\left\|\left(\epsilon_t s_t+\frac{G_t}{\sigma_t}\right)\epsilon_{t-h}\right\|_2\right)^2=o(1).
\end{aligned}$$

Hence, $1/n\sum_{t=1+h}^{n}\varphi_t(\epsilon_t s_t+G_t/\sigma_t)\epsilon_{t-h}\xrightarrow{p}0$. Combining that with $\sqrt{n}(\hat{\theta}_n-\theta_0)=O_p(1)$ and (A.22) yields (A.16).

Next, (A.17). By Lemma A.6,

$$\sup_{\theta\in\Theta}\left\|\frac{1}{n}\sum_{t=1}^{n}\varphi_t\frac{\partial}{\partial\theta}m_t(\theta)\right\|\xrightarrow{p}0 \text{ and } \frac{1}{\sqrt{n}}\sum_{t=1+h}^{n}\varphi_t m_t=O_p(1). \qquad \textbf{(A.23)}$$

Now write

$$\begin{aligned}
&\hat{\mathcal{D}}_n(h)'\widehat{\mathcal{A}}_n\frac{1}{\sqrt{n}}\sum_{t=1+h}^{n}\varphi_t m_t(\hat{\theta}_n)\\
&\quad=\mathcal{D}(h)'\mathcal{A}\frac{1}{\sqrt{n}}\sum_{t=1+h}^{n}\varphi_t m_t+\mathcal{D}(h)'\mathcal{A}\frac{1}{\sqrt{n}}\sum_{t=1+h}^{n}\varphi_t\left\{m_t(\hat{\theta}_n)-m_t\right\}
\end{aligned}$$

https://doi.org/10.1017/S0266466619000367 Published online by Cambridge University Press

$$+\left\{\hat{\mathcal{D}}_n(h)'\widehat{\mathcal{A}}_n-\mathcal{D}(h)'\mathcal{A}\right\}\frac{1}{\sqrt{n}}\sum_{t=1+h}^{n}\varphi_t m_t$$

$$+\left\{\hat{\mathcal{D}}_n(h)'\widehat{\mathcal{A}}_n-\mathcal{D}(h)'\mathcal{A}\right\}\frac{1}{\sqrt{n}}\sum_{t=1+h}^{n}\varphi_t\left\{m_t(\hat{\theta}_n)-m_t\right\}.$$

Note that $\hat{\mathcal{D}}_n(h)\xrightarrow{p}\mathcal{D}(h)$ by arguments in the proof of Lemma 2.1, and by supposition $\widehat{\mathcal{A}}_n\xrightarrow{p}\mathcal{A}$. Moreover, by a mean value theorem argument, Assumption 2.c′, and (A.23):

$$\left\|\frac{1}{\sqrt{n}}\sum_{t=1+h}^{n}\varphi_t\left\{m_t(\hat{\theta}_n)-m_t\right\}\right\|\leq\sqrt{n}\left\|\hat{\theta}_n-\theta_0\right\|\times\sup_{\theta\in\Theta}\left\|\frac{1}{n}\sum_{t=1+h}^{n}\varphi_t\frac{\partial}{\partial\theta}m_t(\theta)\right\|\xrightarrow{p}0.$$

The latter convergence in probability, combined with (A.23), suffice to prove (A.17).

Proceeding to (A.18), first note that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\varphi_t=b_n\frac{1}{\sqrt{n}}\sum_{s=1}^{n/b_n}\xi_s=\sqrt{b_n}\,(n/b_n)^{-1/2}\sum_{s=1}^{n/b_n}\xi_s=O_p\left(\sqrt{b_n}\right). \tag{A.24}$$

Second, by eqn. (F.5) in the proof of Lemma A.4 in supplemental material,

$$\left|\sqrt{n}\hat{\gamma}_n(h)-n^{-1/2}\sum_{t=1+h}^{n}\epsilon_t\epsilon_{t-h}+\sqrt{n}\left(\hat{\theta}_n-\theta_0\right)'\mathcal{D}(h)\right|\xrightarrow{p}0. \tag{A.25}$$

Use (A.25) and $\hat{\theta}_n=\theta_0+O_p(1/\sqrt{n})$ to deduce $1/n\sum_{s=1+h}^{n}\epsilon_s(\hat{\theta}_n)\epsilon_{s-h}(\hat{\theta}_n)=1/n\sum_{t=1+h}^{n}\epsilon_t\epsilon_{t-h}+O_p(1/\sqrt{n})$. Therefore,

$$\frac{1}{\sqrt{n}}\sum_{t=1+h}^{n}\varphi_t\frac{1}{n}\sum_{s=1+h}^{n}\epsilon_s(\hat{\theta}_n)\epsilon_{s-h}(\hat{\theta}_n)=\frac{1}{\sqrt{n}}\sum_{t=1+h}^{n}\varphi_t\frac{1}{n}\sum_{t=1+h}^{n}\epsilon_t\epsilon_{t-h}+O_p\left(1/\sqrt{n/b_n}\right).$$

It remains to show

$$\frac{1}{\sqrt{n}}\sum_{t=1+h}^{n}\varphi_t\frac{1}{n}\sum_{t=1+h}^{n}\epsilon_t\epsilon_{t-h}=\frac{1}{\sqrt{n}}\sum_{t=1+h}^{n}\varphi_t E\left[\epsilon_t\epsilon_{t-h}\right]+o_p(1). \tag{A.26}$$

Under Assumptions 1.b, $\epsilon_t\epsilon_{t-h}-E[\epsilon_t\epsilon_{t-h}]$ satisfies (A.1), hence $E[(1/\sqrt{n}\sum_{t=1}^{n}\{\epsilon_t\epsilon_{t-h}-E[\epsilon_t\epsilon_{t-h}]\})^2]=O(1)$. Furthermore, $1/\sqrt{n}\sum_{t=1+h}^{n}\varphi_t=O_p(\sqrt{b_n})$ from (A.24). Hence,

$$\frac{1}{\sqrt{n}}\sum_{t=1+h}^{n}\varphi_t\frac{1}{n}\sum_{t=1+h}^{n}\left\{\epsilon_t\epsilon_{t-h}-E\left[\epsilon_t\epsilon_{t-h}\right]\right\}=\frac{1}{\sqrt{n}}\sum_{t=1+h}^{n}\varphi_t\times O_p(1/\sqrt{n})=O_p\left(1/\sqrt{n/b_n}\right).$$

Since $b_n/n\to 0$, (A.26) follows directly.

https://doi.org/10.1017/S0266466619000367 Published online by Cambridge University Press

Finally, for (A.19), since $1/\sqrt{n}\sum_{t=1+h}^{n}\varphi_t = O_p(\sqrt{b_n})$ and $\hat{\mathcal{D}}_n(h)'\widehat{\mathcal{A}}_n \xrightarrow{p} \mathcal{D}(h)'\mathcal{A}$, we need only show $1/n\sum_{t=1}^{n} m_t(\hat{\theta}_n) = o_p(1/\sqrt{b_n})$. A first-order expansion and the mean value theorem yield

$$\left\| \frac{1}{n}\sum_{t=1+h}^{n} m_t(\hat{\theta}_n) - \frac{1}{n}\sum_{t=1+h}^{n} m_t \right\| \leq \sup_{\theta\in\Theta} \left\| \frac{1}{n}\sum_{t=1+h}^{n} \frac{\partial}{\partial\theta} m_t(\theta_n^*) \right\| \left\| \hat{\theta}_n - \theta_0 \right\|.$$

By Lemma A.6, $\sup_{\theta\in\Theta}||1/n\sum_{t=1}^{n}(\partial/\partial\theta)m_t(\theta) - E[(\partial/\partial\theta)m_t(\theta)]|| \xrightarrow{p} 0$, and $\sup_{\theta\in\Theta}||E[(\partial/\partial\theta)m_t(\theta)]|| < \infty$ and $\hat{\theta}_n - \theta_0 = O_p(1/\sqrt{n})$ under Assumption 2.c′. Moreover, by Assumption 2.c′, $m_t = [m_{i,t}]_{i=1}^{k_m}$ satisfies (A.1), hence $E[(1/\sqrt{n}\sum_{t=1}^{n} m_{i,t}^2] = O(1)$. This yields $1/n\sum_{t=1+h}^{n} m_t(\hat{\theta}_n) = 1/n\sum_{t=1+h}^{n} m_t + O_p(1/\sqrt{n}) = O_p(1/\sqrt{n})$. Since $b_n = o(n)$ the proof is complete.

Claim (b). Weak convergence in probability Claim (a), the mapping theorem and Slutsky's theorem yield for each $\mathcal{L} \in \mathbb{N}$:

$$\vartheta\left(\left[\sqrt{n}\hat{\omega}_n(h)\hat{\rho}_n^{(dw)}(h)\right]_{h=1}^{\mathcal{L}}\right) \Rightarrow^{p} \vartheta\left(\left[\omega(h)\mathring{\mathcal{Z}}(h)\right]_{h=1}^{\mathcal{L}}\right). \tag{A.27}$$

Therefore (see, e.g., Giné and Zinn, 1990, eqn. (3.4)):

$$\mathcal{X}_n(\mathcal{L}) \equiv \sup_{c>0}\left| P\left(\vartheta\left(\left[\sqrt{n}\hat{\omega}_n(h)\hat{\rho}_n^{(dw)}(h)\right]_{h=1}^{\mathcal{L}}\right) \leq c|\mathfrak{X}_n\right) - P\left(\vartheta\left(\left[\omega(h)\mathring{\mathcal{Z}}(h)\right]_{h=1}^{\mathcal{L}}\right) \leq c\right)\right| \xrightarrow{p} 0.$$

Apply Lemma A.2.a to $\mathcal{X}_n(\mathcal{L})$ to yield for some monotonic sequence of positive integers $\{\mathcal{L}_n\}_{n\geq 1}$, $\mathcal{L}_n \to \infty$ and $\mathcal{L}_n = o(n)$

$$\mathcal{X}_n(\mathcal{L}_n) \leq \max_{1\leq\mathcal{L}\leq\mathcal{L}_n} \mathcal{X}_n(\mathcal{L}) \xrightarrow{p} 0.$$

Therefore, $\mathcal{X}_n(\mathcal{L}_n) \xrightarrow{p} 0$. Finally, since $\mathcal{X}_n(\mathcal{L})$ is bounded, it is uniformly integrable. Hence, $\mathcal{L}_n = O(n^{\min\{\zeta,\kappa,1/2\}}/\ln(n))$ must be satisfied for $\max_{1\leq\mathcal{L}\leq\mathcal{L}_n} \mathcal{X}_n(\mathcal{L}) \xrightarrow{p} 0$ and therefore for $\mathcal{X}_n(\mathcal{L}_n) \xrightarrow{p} 0$, cf. Lemma A.2.a. $\mathcal{QED}$. □

Proof of Theorem 2.5. Assume the weights $\hat{\omega}_n(h) = 1$ to conserve notation, without loss of generality. Operate conditionally on $\mathfrak{X}_n \equiv \{m_t, x_t, y_t\}_{t=1}^{n}$, and recall $\hat{p}_{n,M}^{(dw)} \equiv 1/M\sum_{i=1}^{M} I(\hat{\mathcal{T}}_{n,i}^{(dw)} \geq \hat{\mathcal{T}}_n)$. First, by the Glivenko–Cantelli theorem,

$$\hat{p}_{n,M}^{(dw)} \xrightarrow{p} P\left(\vartheta\left(\left[\sqrt{n}\hat{\rho}_n^{(dw)}(h)\right]_{h=1}^{\mathcal{L}_n}\right) \geq \vartheta\left(\left[\sqrt{n}\hat{\rho}_n(h)\right]_{h=1}^{\mathcal{L}_n}\right)|\mathfrak{X}_n\right) \text{ as } M \to \infty. \tag{A.28}$$

Second, by Lemma A.7,

$$\sup_{c>0}\left| P\left(\vartheta\left(\left[\sqrt{n}\hat{\rho}_n^{(dw)}(h)\right]_{h=1}^{\mathcal{L}_n}\right) \leq c|\mathfrak{X}_n\right) - P\left(\vartheta\left(\left[\mathring{\mathcal{Z}}(h)\right]_{h=1}^{\mathcal{L}_n}\right) \leq c\right)\right| \xrightarrow{p} 0, \tag{A.29}$$

where $\{\mathcal{Z}(h) : h \in \mathbb{N}\}$ is a zero-mean Gaussian process with variance $E[\mathcal{Z}(h)^2] < \infty$, $\{\mathring{\mathcal{Z}}(h) : h \in \mathbb{N}\}$ is an independent copy of $\{\mathcal{Z}(h) : h \in \mathbb{N}\}$, and $\{\mathcal{L}_n\}$ is a monotonic sequence of

https://doi.org/10.1017/S0266466619000367 Published online by Cambridge University Press

positive integers, $\mathcal{L}_n \to \infty$ and $\mathcal{L}_n = o(n)$. In particular, $\mathcal{L}_n = O(n^{\min\{\zeta,\kappa,1/2\}}/\ln(n))$ must be satisfied.

Impose $H_0 : \rho(h) = 0\ \forall h \in \mathbb{N}$. Define $\bar{F}_n^{(0)}(c) \equiv P(\vartheta([\mathring{\mathcal{Z}}(h)]_{h=1}^{\mathcal{L}_n}) > c)$. Note that (A.29) implies:

$$P\left(\vartheta\left(\left[\sqrt{n}\hat{\rho}_n^{(dw)}(h)\right]_{h=1}^{\mathcal{L}_n}\right) \geq \vartheta\left(\left[\sqrt{n}\hat{\rho}_n(h)\right]_{h=1}^{\mathcal{L}_n}\right)|\mathfrak{X}_n\right)$$
$$-P\left(\vartheta\left(\left[\mathring{\mathcal{Z}}(h)\right]_{h=1}^{\mathcal{L}_n}\right) \geq \vartheta\left(\left[\sqrt{n}\hat{\rho}_n(h)\right]_{h=1}^{\mathcal{L}_n}\right)\right) \xrightarrow{p} 0.$$

Since $[\mathring{\mathcal{Z}}(h)]_{h=1}^{\mathcal{L}_n}$ is independent of the sample $\mathfrak{X}_n$, we therefore have

$$P\left(\vartheta\left(\left[\sqrt{n}\hat{\rho}_n^{(dw)}(h)\right]_{h=1}^{\mathcal{L}_n}\right) \geq \vartheta\left(\left[\sqrt{n}\hat{\rho}_n(h)\right]_{h=1}^{\mathcal{L}_n}\right)|\mathfrak{X}_n\right)$$
$$-\bar{F}_n^{(0)}\left(\vartheta\left(\left[\sqrt{n}\hat{\rho}_n(h)\right]_{h=1}^{\mathcal{L}_n}\right)\right) \xrightarrow{p} 0. \tag{A.30}$$

$\bar{F}_n^{(0)}$ is continuous by Gaussianicity. Theorem 2.3 and Slutsky's theorem therefore yield

$$\left|\bar{F}_n^{(0)}\left(\vartheta\left(\left[\sqrt{n}\hat{\rho}_n(h)\right]_{h=1}^{\mathcal{L}_n}\right)\right) - \bar{F}_n^{(0)}\left(\vartheta\left(\left[\mathcal{Z}(h)\right]_{h=1}^{\mathcal{L}_n}\right)\right)\right| \xrightarrow{p} 0. \tag{A.31}$$

Together, (A.28), (A.30), and (A.31) yield for any sequence of positive integers $\{M_n\}$, $M_n \to \infty$:

$$\hat{p}_{n,M_n}^{(dw)} = P\left(\vartheta\left(\left[\sqrt{n}\hat{\rho}_n^{(dw)}(h)\right]_{h=1}^{\mathcal{L}_n}\right) \geq \vartheta\left(\left[\sqrt{n}\hat{\rho}_n(h)\right]_{h=1}^{\mathcal{L}_n}\right)|\mathfrak{X}_n\right) + o_p(1) \tag{A.32}$$
$$= \bar{F}_n^{(0)}\left(\vartheta\left(\left[\mathcal{Z}(h)\right]_{h=1}^{\mathcal{L}_n}\right)\right) + o_p(1).$$

Since$\{\mathring{\mathcal{Z}}(h) : h \in \mathbb{N}\}$ is an independent copy of $\{\mathcal{Z}(h) : h \in \mathbb{N}\}$, $\bar{F}_n^{(0)}(\vartheta([\mathcal{Z}(h)]_{h=1}^{\mathcal{L}_n}))$ is distributed uniform on $[0, 1]$. Now use (A.32) to conclude $P(\hat{p}_{n,M_n}^{(dw)} < \alpha) = P(\bar{F}_n^{(0)}\left((\vartheta\left[\mathcal{Z}(h)\right]_{h=1}^{\mathcal{L}_n})\right) < \alpha) + o(1) = \alpha + o(1) \to \alpha$.

Impose $H_1 : \rho(h) \neq 0$ for some $h \in \mathbb{N}$. Recall ϑ satisfies the triangle inequality, and divergence $\vartheta(a) \to \infty$ as $||a|| \to \infty$. Theorem 2.3 therefore yields: $\vartheta([\sqrt{n}\hat{\rho}_n(h)]_{h=1}^{\mathcal{L}_n}) \leq \vartheta([\sqrt{n}\{\hat{\rho}_n(h) - \rho(h)\}]_{h=1}^{\mathcal{L}_n}) + \vartheta([\sqrt{n}\rho(h)]_{h=1}^{\mathcal{L}_n}) = \vartheta([\mathcal{Z}(h)]_{h=1}^{\mathcal{L}_n}) + \vartheta([\sqrt{n}\rho(h)]_{h=1}^{\mathcal{L}_n}) + o_p(1)$, and $\vartheta([\sqrt{n}\rho(h)]_{h=1}^{\mathcal{L}_n}) \leq \vartheta([\sqrt{n}\{\hat{\rho}_n(h) - \rho(h)\}]_{h=1}^{\mathcal{L}_n}) + \vartheta([\sqrt{n}\hat{\rho}_n(h)]_{h=1}^{\mathcal{L}_n}) = \vartheta([\mathcal{Z}(h)]_{h=1}^{\mathcal{L}_n}) + \vartheta([\sqrt{n}\hat{\rho}_n(h)]_{h=1}^{\mathcal{L}_n}) + o_p(1) \xrightarrow{p} \infty$. Hence,

$$\infty \xleftarrow{p} \vartheta\left(\left[\mathcal{Z}(h)\right]_{h=1}^{\mathcal{L}_n}\right) + \vartheta\left(\left[\sqrt{n}\rho(h)\right]_{h=1}^{\mathcal{L}_n}\right) + o_p(1) \geq \vartheta\left(\left[\sqrt{n}\rho(h)\right]_{h=1}^{\mathcal{L}_n}\right)$$
$$-\vartheta\left(\left[\mathcal{Z}(h)\right]_{h=1}^{\mathcal{L}_n}\right) + o_p(1) \xrightarrow{p} \infty. \tag{A.33}$$

Combine (A.28), (A.29), and (A.33) to deduce $P(\hat{p}_{n,M_n}^{(dw)} < \alpha) \to 1$ for any $\alpha \in (0, 1)$ because

https://doi.org/10.1017/S0266466619000367 Published online by Cambridge University Press

$$\hat{p}_{n,M_n}^{(dw)} = P\left(\vartheta\left(\left[\sqrt{n}\hat{\rho}_n^{(dw)}(h)\right]_{h=1}^{\mathcal{L}_n}\right) \geq \vartheta\left(\left[\sqrt{n}\hat{\rho}_n(h)\right]_{h=1}^{\mathcal{L}_n}\right)|\mathfrak{X}_n\right) + o_p(1)$$
$$= P\left(\vartheta\left(\left[\mathring{\mathcal{Z}}(h)\right]_{h=1}^{\mathcal{L}_n}\right) \geq \vartheta\left(\left[\sqrt{n}\hat{\rho}_n(h)\right]_{h=1}^{\mathcal{L}_n}\right)\right) + o_p(1)$$
$$= \bar{F}_n^{(0)}\left(\vartheta\left(\left[\sqrt{n}\hat{\rho}_n(h)\right]_{h=1}^{\mathcal{L}_n}\right)\right) + o_p(1) \xrightarrow{p} 0. \ \mathcal{QED}.$$

Proof of Theorem 3.1. Let q be any fixed positive constant. Recall that the penalty $\mathcal{P}_n(\mathcal{L}) = \sqrt{\mathcal{L}\ln n}$ if $\hat{\mathcal{T}}_n(\mathcal{L}) \leq \sqrt{q\ln n}$, else $\mathcal{P}_n(\mathcal{L}) = \sqrt{2\mathcal{L}}$.

Claim (a). Let H_0 be true. It suffices to prove the following. First, for any $\{\mathcal{L}_n\}$, $\mathcal{L}_n \to (0,\infty]$ and $\mathcal{L}_n/\bar{\mathcal{L}}_n \to [0,1]$, the penalty term satisfies:

$$P\left(\mathcal{P}_n(\mathcal{L}_n) = \sqrt{\mathcal{L}_n\ln(n)}\right) \to 1. \tag{A.34}$$

Hence, $\hat{\mathcal{T}}_n^{\mathcal{P}}(\mathcal{L}) \equiv \hat{\mathcal{T}}_n(\mathcal{L}) - \sqrt{\mathcal{L}\ln n}$ asymptotically with probability approaching one. Second, for such $\{\mathcal{L}_n\}$ the following holds:

$$P\left(\hat{\mathcal{T}}_n(\mathcal{L}_n) - \hat{\mathcal{T}}_n(l) \geq \left(\sqrt{\mathcal{L}_n} - \sqrt{l}\right)\sqrt{\ln(n)}\right) \to 1 \text{ if } l \geq \mathcal{L}_n \tag{A.35}$$
$$P\left(\hat{\mathcal{T}}_n(\mathcal{L}_n) - \hat{\mathcal{T}}_n(l) \geq \left(\sqrt{\mathcal{L}_n} - \sqrt{l}\right)\sqrt{\ln(n)}\right) \to 0 \text{ for fixed } l = 1,...,\mathcal{L}_n - 1.$$

Together, (A.34) and (A.35) prove the claim $P(\mathcal{L}_n^* = 1) \to 1$ since the following holds *for every* $l = 1,...,\bar{\mathcal{L}}_n$ *if and only if* $\mathcal{L}_n \to 1$

$$\lim_{n\to\infty} P\left(\hat{\mathcal{T}}_n^{\mathcal{P}}(\mathcal{L}_n) \geq \hat{\mathcal{T}}_n^{\mathcal{P}}(l)\right) = \lim_{n\to\infty} P\left(\hat{\mathcal{T}}_n(\mathcal{L}_n) - \hat{\mathcal{T}}_n(l)\right.$$
$$\left. \geq \left(\sqrt{\mathcal{L}_n} - \sqrt{l}\right)\sqrt{\ln(n)}\right) = 1, \tag{A.36}$$

while $\mathcal{L}_n^*$ is the least of sequences that satisfy (A.36) *for every* $l = 1,...,\bar{\mathcal{L}}_n$.

Now consider (A.34). By construction of $\mathcal{P}_n(\mathcal{L}_n)$, it suffices to prove $P(\hat{\mathcal{T}}_n(\mathcal{L}_n) > \sqrt{q\ln n}) \to 0$. Under H_0, $\sqrt{n}\hat{\rho}_n(h) = O_p(1)$ by (A.7), hence $\sqrt{n}\hat{\rho}_n(h)/\sqrt{q\ln n} \xrightarrow{p} 0$ for any fixed $q \in (0,\infty)$. Therefore, by Lemma A.2 for some $\{\bar{\mathcal{L}}_n\}$ that satisfies $\bar{\mathcal{L}}_n \to \infty$ and $\bar{\mathcal{L}}_n = o(n)$:

$$\frac{\hat{\mathcal{T}}_n(\bar{\mathcal{L}}_n)}{\sqrt{q\ln n}} = \frac{\sqrt{n}\max_{1\leq h\leq\bar{\mathcal{L}}_n}\left|\hat{\rho}_n(h)\right|}{\sqrt{q\ln n}} \xrightarrow{p} 0. \tag{A.37}$$

By the same lemma, if $(n^{\min\{\zeta,\kappa,1/2\}}/\ln(n))\tilde{\mathcal{X}}_n(h)$ for all h is uniformly integrable, where $\tilde{\mathcal{X}}_n(h)$ is defined in (6), then $\bar{\mathcal{L}}_n = O(n^{\min\{\zeta,\kappa,1/2\}}/\ln(n))$ must hold. By monotonicity of $\hat{\mathcal{T}}_n(\cdot) \geq 0$, (A.37) holds for any $\{\mathcal{L}_n\}$, $\mathcal{L}_n \to (0,\infty]$ and $\mathcal{L}_n/\bar{\mathcal{L}}_n \to [0,1]$. Thus, $\hat{\mathcal{T}}_n(\mathcal{L}_n)/\sqrt{q\ln n} \xrightarrow{p} 0$ for all such $\{\mathcal{L}_n\}$.

Now consider (A.35). Suppose $l > \mathcal{L}_n$. By (A.37), $\hat{\mathcal{T}}_n(\bar{\mathcal{L}}_n)/\sqrt{\ln n} = o_p(1)$ and therefore $\hat{\mathcal{T}}_n(\mathcal{L}_n) - \hat{\mathcal{T}}_n(l) = o_p(\sqrt{\ln(n)})$ for any $\{\mathcal{L}_n\}$, $\mathcal{L}_n \to (0,\infty]$ and $\mathcal{L}_n/\bar{\mathcal{L}}_n \to [0,1]$, and any $1 \leq l \leq \bar{\mathcal{L}}_n$. Now use (A.34), monotonicity of $\hat{\mathcal{T}}_n(\cdot)$, and $\inf_{n\geq 1}\{\sqrt{l} - \sqrt{\mathcal{L}_n}\} > 0$, to yield that as $n \to \infty$:

https://doi.org/10.1017/S0266466619000367 Published online by Cambridge University Press

$$P\left(\hat{\mathcal{T}}_n(\mathcal{L}_n)-\hat{\mathcal{T}}_n(l)\geq\left(\sqrt{\mathcal{L}_n}-\sqrt{l}\right)\sqrt{\ln(n)}\right)$$
$$=P\left(\frac{\hat{\mathcal{T}}_n(\mathcal{L}_n)-\hat{\mathcal{T}}_n(l)}{\sqrt{\ln(n)}}\geq\sqrt{\mathcal{L}_n}-\sqrt{l}\right)=P\left(\sqrt{l}-\sqrt{\mathcal{L}_n}\geq\frac{\hat{\mathcal{T}}_n(l)-\hat{\mathcal{T}}_n(\mathcal{L}_n)}{\sqrt{\ln(n)}}\right)\to 1.$$

Similarly, if $l=\mathcal{L}_n$ then $\sqrt{l}-\sqrt{\mathcal{L}_n}=0$ and $\hat{\mathcal{T}}_n(l)-\hat{\mathcal{T}}_n(\mathcal{L}_n)=0$ hence the above limit holds.

Conversely, suppose $l\in\{1,...,\mathcal{L}_n-1\}$ and $\mathcal{L}_n>1$. Then from $\hat{\mathcal{T}}_n(\mathcal{L}_n)=o_p(\sqrt{q\ln n})$ and $1-\sqrt{l/\mathcal{L}_n}>0$ it follows

$$P\left(\hat{\mathcal{T}}_n(\mathcal{L}_n)-\hat{\mathcal{T}}_n(l)\geq\left(\sqrt{\mathcal{L}_n}-\sqrt{l}\right)\sqrt{\ln(n)}\right)=P\left(\frac{\hat{\mathcal{T}}_n(\mathcal{L}_n)-\hat{\mathcal{T}}_n(l)}{\sqrt{\mathcal{L}_n}\sqrt{\ln(n)}}\geq\left(1-\sqrt{\frac{l}{\mathcal{L}_n}}\right)\right)\to 0.$$

Claim (A.35) follows directly.

Claim (b). Let H_1 hold. Let *ap*1 denote *asymptotically with probability approaching one*. Define $h_n^*\equiv\min\{h_n : h_n=\arg\max_{1\leq h\leq\bar{\mathcal{L}}_n}|\hat{\rho}_n(h)|\}$, the smallest lag at which the largest sample correlation in magnitude over lags $1\leq h\leq\bar{\mathcal{L}}_n$ occurs.

Define $\mathbb{N}_1\equiv\{h\in\mathbb{N}:E[\epsilon_t\epsilon_{t-h}]\neq 0\}$ and $\underline{\mathrm{N}}_1\equiv\min\{\mathbb{N}_1\}$, the smallest lag at which the autocorrelation is not zero. We prove in Step 1 that for any integer sequence $\{\mathcal{L}_n\}$ such that $\mathcal{L}_n\to[\underline{\mathrm{N}}_1,\infty]$ and $\mathcal{L}_n/\bar{\mathcal{L}}_n\to[0,1]$,

$$P\left(\mathcal{P}_n(\mathcal{L}_n)=\sqrt{2\mathcal{L}_n}\right)\to 1. \tag{A.38}$$

We then prove in Step 2 that *if and only if* $\mathcal{L}_n/h_n^*\xrightarrow{p}[1,\infty]$,

$$P\left(\hat{\mathcal{T}}_n(\mathcal{L}_n)\geq\hat{\mathcal{T}}_n(l)+2(\sqrt{\mathcal{L}_n}-\sqrt{l})\right)\to 1\text{for each}1\leq l\leq\bar{\mathcal{L}}_n. \tag{A.39}$$

Moreover, $h_n^*\xrightarrow{p}h^*\equiv\min\{h:h=\arg\max_{1\leq h\leq\infty}|\rho(h)|\}$ is an easy consequence of $\bar{\mathcal{L}}_n\to\infty$, consistency of the sample correlation under the stated assumptions, and Slutsky's theorem. Notice $h^*\in[\underline{\mathrm{N}}_1,\infty)$ by construction of $\underline{\mathrm{N}}_1$.

The proof of the claim then proceeds as follows. Take any integer sequence $\{\mathcal{L}_n\}$, $\mathcal{L}_n/h_n^*\xrightarrow{p}[1,\infty]$ and $\mathcal{L}_n/\bar{\mathcal{L}}_n\to[0,1]$. Then (A.38) holds because $h^*\in[\underline{\mathrm{N}}_1,\infty)$, hence $\hat{\mathcal{T}}_n^{\mathcal{P}}(\mathcal{L}_n)\equiv\hat{\mathcal{T}}_n(\mathcal{L}_\cdot)-\sqrt{2\mathcal{L}_n}$ *ap*1. Since such a sequence implies (A.39), we have $\hat{\mathcal{T}}_n^{\mathcal{P}}(\mathcal{L}_n)\geq\hat{\mathcal{T}}_n^{\mathcal{P}}(l)$ *ap*1 for each $l=1,...,\bar{\mathcal{L}}_n$. Conversely, if (A.39) holds then $\mathcal{L}_n/h_n^*\xrightarrow{p}[1,\infty]$. This yields (A.38) because $h^*\in[\underline{\mathrm{N}}_1,\infty)$. Therefore $\hat{\mathcal{T}}_n^{\mathcal{P}}(\mathcal{L}_n)\geq\hat{\mathcal{T}}_n^{\mathcal{P}}(l)$ *ap*1 for each $l=1,...,\bar{\mathcal{L}}_n$*if and only if* $\mathcal{L}_n/h_n^*\xrightarrow{p}[1,\infty]$. Since the optimal $\{\mathcal{L}_n^*\}$ is the least of such sequences, the selection $\mathcal{L}_n^*$ satisfies $\mathcal{L}_n/h_n^*\xrightarrow{p}1$. Together, $\mathcal{L}_n/h_n^*\xrightarrow{p}1$ and $h_n^*\xrightarrow{p}h^*$ prove the claim.

Step 1: Consider (A.38). Use (A.7) to deduce $\hat{\rho}_n(h)-\rho(h)\xrightarrow{p}0$ for each h. Lemma A.2 therefore yields for some integer sequence $\{\bar{\mathcal{L}}_n\}$, $\bar{\mathcal{L}}_n\to\infty$:

$$\left|\max_{1\leq h\leq\bar{\mathcal{L}}_n}\left|\hat{\rho}_n(h)\right|-\max_{1\leq h\leq\bar{\mathcal{L}}_n}|\rho(h)|\right|\leq\left|\max_{1\leq h\leq\bar{\mathcal{L}}_n}\left|\hat{\rho}_n(h)-\rho(h)\right|\right|\xrightarrow{p}0,$$

https://doi.org/10.1017/S0266466619000367 Published online by Cambridge University Press

where $\lim_{n\to\infty}\max_{1\le h\le\bar{\mathcal{L}}_n}|\rho(h)|\in(0,\infty)$. By monotonicity, for any $\{\mathcal{L}_n\}$, $\mathcal{L}_n\to(0,\infty]$ and $\mathcal{L}_n/\bar{\mathcal{L}}_n\to[0,1]$, and sufficiently large n,

$$\left|\max_{1\le h\le\mathcal{L}_n}\left|\hat{\rho}_n(h)\right|-\max_{1\le h\le\mathcal{L}_n}|\rho(h)|\right|\le\left|\max_{1\le h\le\mathcal{L}_n}\left|\hat{\rho}_n(h)-\rho(h)\right|\right|$$
$$\le\left|\max_{1\le h\le\bar{\mathcal{L}}_n}\left|\hat{\rho}_n(h)-\rho(h)\right|\right|\xrightarrow{p}0.$$

Therefore, for any $\{\mathcal{L}_n\}$, $\mathcal{L}_n\to[\underline{\mathrm{N}}_1,\infty]$, and $\mathcal{L}_n/\bar{\mathcal{L}}_n\to[0,1]$,

$$\frac{\hat{\mathcal{T}}_n(\mathcal{L}_n)}{\sqrt{q\ln n}}=\frac{\sqrt{n}\max_{1\le h\le\mathcal{L}_n}\left|\hat{\rho}_n(h)\right|}{\sqrt{q\ln n}}\xrightarrow{p}\infty.$$

This proves (A.38) by construction (13) of the penalty term $\mathcal{P}_n(\mathcal{L}_n)$.

Step 2: Next we prove (A.39). First, note that by Theorem 2.3 $\hat{\mathcal{T}}_n(\mathcal{L}_n)/\sqrt{n}\xrightarrow{p}(0,1)$ for any $\{\mathcal{L}_n\}$, $\mathcal{L}_n\to[\underline{\mathrm{N}}_1,\infty]$, and $\mathcal{L}_n/\bar{\mathcal{L}}_n\to[0,1]$. Hence $\hat{\mathcal{T}}_n(\mathcal{L}_n)/\sqrt{n/\ln(n)}\xrightarrow{p}\infty$ for any $\mathcal{L}_n\to[\underline{\mathrm{N}}_1,\infty]$, where $\mathcal{L}_n=o(n/\ln(n))$ by assumption. Monotonicity ensures $\hat{\mathcal{T}}_n(\mathcal{L}_n)\ge\hat{\mathcal{T}}_n(l)$ for each $l\le\mathcal{L}_n$, hence $\hat{\mathcal{T}}_n(l)/\hat{\mathcal{T}}_n(\mathcal{L}_n)=[\hat{\mathcal{T}}_n(l)/\sqrt{n}]/[\hat{\mathcal{T}}_n(\mathcal{L}_n)/\sqrt{n}]\xrightarrow{p}[0,1]$ for such l. Indeed, if both $(l,\mathcal{L}_n)\ge h_n^*\equiv\min\{h_n:h_n=\arg\max_{1\le h\le\bar{\mathcal{L}}_n}|\hat{\rho}_n(h)|\}$ then by construction $\hat{\mathcal{T}}_n(l)/\hat{\mathcal{T}}_n(\mathcal{L}_n)=1$.

Now suppose $1\le l$ and $l/\mathcal{L}_n\to[0,1)$, and $\mathcal{L}_n/h_n^*\xrightarrow{p}[0,1)$, hence $1\le l<\mathcal{L}_n<h_n^*$ as $n\to\infty ap1$. Then $\hat{\mathcal{T}}_n(l)/\hat{\mathcal{T}}_n(\mathcal{L}_n)\xrightarrow{p}[0,1)$ by monotonicity and the construction of h_n^*. Now use $\mathcal{L}_n\le\bar{\mathcal{L}}_n=o(n/\ln(n))$ by assumption, and $\hat{\mathcal{T}}_n(\mathcal{L}_n)/\sqrt{n/\ln(n)}\xrightarrow{p}\infty$ to yield:

$$P\left(\hat{\mathcal{T}}_n(\mathcal{L}_n)\ge\hat{\mathcal{T}}_n(l)+2\left(\sqrt{\mathcal{L}_n}-\sqrt{l}\right)\right)$$
$$=P\left(\hat{\mathcal{T}}_n(\mathcal{L}_n)\left(1-\frac{\hat{\mathcal{T}}_n(l)}{\hat{\mathcal{T}}_n(\mathcal{L}_n)}\right)\ge2\sqrt{\mathcal{L}_n}\left(1-\sqrt{\frac{l}{\mathcal{L}_n}}\right)\right)$$
$$\ge P\left(\frac{\hat{\mathcal{T}}_n(\mathcal{L}_n)}{\sqrt{n/\ln(n)}}\left(1-\frac{\hat{\mathcal{T}}_n(l)}{\hat{\mathcal{T}}_n(\mathcal{L}_n)}\right)\ge2\sqrt{\frac{\mathcal{L}_n}{n/\ln(n)}}\right)\to1. \tag{A.40}$$

Next, consider $1\le l$ and $l/h_n^*\xrightarrow{p}[0,1)$, and $\mathcal{L}_n/h_n^*\xrightarrow{p}[1,\infty]$, hence $1\le l\le h_n^*-1\ ap1$ and $\mathcal{L}_n\ge h_n^*ap1$. Then, $P(\hat{\mathcal{T}}_n(l)=\hat{\mathcal{T}}_n(\mathcal{L}_n)\to0$ since by construction h_n^* is the smallest lag at which the maximum correlation occurs. Monotonicity therefore yields $\hat{\mathcal{T}}_n(l)/\hat{\mathcal{T}}_n(\mathcal{L}_n)\xrightarrow{p}[0,1)$, and again we deduce (A.40).

Now let $(l,\mathcal{L}_n)\ge h_n^*ap1$. Then by construction $\hat{\mathcal{T}}_n(\mathcal{L}_n)=\hat{\mathcal{T}}_n(l)\ ap1$. Trivially, if $l<\mathcal{L}_n$ ($l\ge\mathcal{L}_n$) then $\sqrt{\mathcal{L}_n}-\sqrt{l}>0$ ($\sqrt{\mathcal{L}_n}-\sqrt{l}\le0$). Hence $P(\hat{\mathcal{T}}_n(\mathcal{L}_n)\ge\hat{\mathcal{T}}_n(l)+2[\sqrt{\mathcal{L}_n}-\sqrt{l}])\to1$ *if and only if* $l\ge\mathcal{L}_n$.

Next, let $\mathcal{L}_n<h_n^*\le l\ ap1$ such that $\hat{\mathcal{T}}_n(l)=\hat{\mathcal{T}}_n(h_n^*)ap1$. Use $\mathcal{L}_n/l\to[0,1)$, $l=o(n/\ln(n))$, $\hat{\mathcal{T}}_n(h_n^*)/\sqrt{n/\ln(n)}\xrightarrow{p}\infty$, and $\hat{\mathcal{T}}_n(\mathcal{L}_n)/\hat{\mathcal{T}}_n(h_n^*)\xrightarrow{p}[0,1)$ to yield

https://doi.org/10.1017/S0266466619000367 Published online by Cambridge University Press

$$P\left(\hat{\mathcal{T}}_n(\mathcal{L}_n) \geq \hat{\mathcal{T}}_n(l) + 2\left(\sqrt{\mathcal{L}_n} - \sqrt{l}\right)\right)$$
$$= P\left(2\left(1 - \sqrt{\frac{\mathcal{L}_n}{l}}\right)\sqrt{\frac{l}{n/\ln(n)}} \geq \frac{\hat{\mathcal{T}}_n(h_n^*)}{\sqrt{n/\ln(n)}}\left(1 - \frac{\hat{\mathcal{T}}_n(\mathcal{L}_n)}{\hat{\mathcal{T}}_n(h_n^*)}\right)\right) \to 0.$$

Finally, generally $\hat{\mathcal{T}}_n(l) = \hat{\mathcal{T}}_n(\mathcal{L}_n)$*a.s.* for some $\{l, \mathcal{L}_n\}$ and all but a finite number of n is possible. For example when $l = \mathcal{L}_n$. In this case, $P(\hat{\mathcal{T}}_n(\mathcal{L}_n) \geq \hat{\mathcal{T}}_n(l) + 2(\sqrt{\mathcal{L}_n} - \sqrt{l})) = P(0 \geq 2(\sqrt{\mathcal{L}_n} - \sqrt{l})) \to 1$ *if and only if* $l \geq \mathcal{L}_n$.

Combining the above results, we deduce $P(\hat{\mathcal{T}}_n(\mathcal{L}_n) \geq \hat{\mathcal{T}}_n(l) + 2[\sqrt{\mathcal{L}_n} - \sqrt{l}]) \to 1$ *for every* $1 \leq l \leq \bar{\mathcal{L}}_n$ *if and only if* $\mathcal{L}_n \geq h_n^*$, proving (A.39). $\mathcal{QED}$. □

https://doi.org/10.1017/S0266466619000367 Published online by Cambridge University Press