A Max-Correlation White Noise Test for Weakly Dependent Time Series*

Jonathan B. Hill † and Kaiji Motegi‡
University of North Carolina   Kobe University

November 7, 2018

Abstract

This paper 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 non-mixing sequences. We show Shao's (2011) 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 as asymptotically valid without exploiting extreme value theory (standard in the literature) or recent Gaussian approximation theory. Finally, we extend Escanciano and Lobato's (2009) 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.

MSC2010 classifications : 62J07, 62F03, 62F40. **JEL classifications** : C12, C52.

Keywords : maximum correlation, white noise test, near epoch dependence, dependent wild bootstrap, automatic lag selection.

*We thank Eric Ghysels, Shigeyuki Hamori, Peter R. Hansen, Yoshihiko Nishiyama, Kenichiro Tamaki, and Kozo Ueda, seminar participants at Kyoto Institute of Economic Research and the University of North Carolina at Chapel Hill, and conference participants at the 10th Spring Meeting of the Japan Statistical Society, 2016 Summer Workshop on Economic Theory, 2016 Asian Meeting of the Econometric Society, 2016 Japanese Joint Statistical Meeting, and 2016 NBER-NSF Time Series Conference for helpful comments. The second author is grateful for financial supports from JSPS KAKENHI (Grant Number 16K17104), Kikawada Foundation, Mitsubishi UFJ Trust Scholarship Foundation, Nomura Foundation, and Suntory Foundation.

†Corresponding author. Department of Economics, University of North Carolina at Chapel Hill. E-mail: jbhill@email.unc.edu; web: https://jbhill.web.unc.edu.

‡Graduate School of Economics, Kobe University. E-mail: motegi@econ.kobe-u.ac.jp

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 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. We combine convergence in finite dimensional distributions of the sample correlation with new theory for handling convergence of arbitrary arrays. The latter is applicable for dealing with the maximum of an increasing sequence of correlations.

The class of time series models considered here is:

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

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 $[\phi_0', \delta_0'] \in \mathbb{R}^{k_\theta}$ and $\delta_0 \in \mathbb{R}^{k_\delta}$ are volatility-specific parameters, $(k_\theta, k_\delta) \geq 0$. 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:

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

Notice ϵ_t need not have a zero conditional mean: we do not require, e.g., $E[\epsilon_t | x_{t-1}] = 0 \ a.s.$ This implies that we do not require $\sigma_t^2(\theta_0)$ to be a conditional variance. Together, (1) allows for model misspeicification. Nevertheless, (1) is assumed correct in some sense, whether H_0 is true or not, in view of $E[\epsilon_t] = 0$. Thus, θ_0 should be thought of as a pseudo-true value that can be identified, often by unconditoinal moment conditions (Kullback and Leibler, 1951, Sawa, 1978).

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 $h \geq 1$:

$$\epsilon_t(\hat{\theta}_n) \equiv \frac{u_t(\hat{\phi}_n)}{\sigma_t(\hat{\theta}_n)} \equiv \frac{y_t - f(x_{t-1}, \hat{\phi}_n)}{\sigma_t(\hat{\theta}_n)} \quad \text{and} \quad \hat{\gamma}_n(h) \equiv \frac{1}{n} \sum_{t=1+h}^{n} \epsilon_t(\hat{\theta}_n) \epsilon_{t-h}(\hat{\theta}_n) \quad \text{and} \quad \hat{\rho}_n(h) \equiv \frac{\hat{\gamma}_n(h)}{\hat{\gamma}_n(0)}.$$

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,

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

where $\hat{\omega}_n(h) > 0$ are possibly stochastic weights with $\hat{\omega}_n(h) \xrightarrow{p} \omega(h) > 0$. 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 sharp size and comparably high power in Monte Carlo simulations.

The number of lags 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 $h \in \{1, ..., \mathcal{L}_n\}$. The limit theory in that case requires more than convergence in finite dimensional distributions based on classic arguments (Hoffmann-Jørgensen, 1984, 1991, e.g.), which is one of the major challenges we address in this paper.

Interest in the maximum of an increasing sequence of deviated covariances $\sqrt{n} \max_{1 \leq h \leq \mathcal{L}_n} |\hat{\gamma}_n(h) - \gamma(h)|$ 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 $\sqrt{n} \max_{1 \leq h \leq \mathcal{L}_n} |\hat{\gamma}_n(h) - \gamma(h)|$ 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 $a_n\{\sqrt{n} \max_{1 \leq h \leq \mathcal{L}_n} |\hat{\gamma}_n(h) - \gamma(h)|/(\sum_{h=0}^{\infty} \gamma(h)^2)^{1/2} - b_n\} \xrightarrow{d} \exp\{-\exp\{-x\}\}$, 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 Hill and Motegi, 2018, 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, 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 $\{0, ..., n-1\}$, 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. Further, a Gaussian approximation theory cannot handle 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, Section 7) allow for β-mixing data, but the above problem

[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 $\max_{1 \leq h \leq \mathcal{L}_n} |1/\sqrt{n} \sum_{t=1}^n z_t(h) - \mathcal{Z}(h)| \xrightarrow{p} 0$ for some Gaussian process $\{\mathcal{Z}(h)\}$. The Gaussian approximation theory of Chernozhukov, Chetverikov, and Kato (2013, 2017) can handle $\max_{1 \leq h \leq \mathcal{L}_n} |1/\sqrt{n} \sum_{t=1}^n z_t(h) - \mathcal{Z}(h)| \xrightarrow{p} 0$ 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.

involving filtered data is not resolved, and our NED environment eclipses a mixing environment (see Section 2.1, below, and see, e.g., Davidson, 1994, Chapter 17).

Compared to 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 $\mathcal{L} \in \mathbb{N}$. Because 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 $\mathcal{L}_n \rightarrow \infty$ and $\mathcal{L}_n = o(n)$, without using extreme value theoretic arguments or Gaussian approximation theory. This is our primary contribution. 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 $\sqrt{n}\max_{1\leq h\leq \mathcal{L}_n} |\hat{\omega}_n(h)\hat{\rho}_n(h)|$.[2]

Our asymptotic theory covers a class of continuous transforms of $[\sqrt{n}\hat{\omega}_n(h)\hat{\rho}_n(h)]_{h=1}^{\mathcal{L}_n}$. This includes 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. Our theory 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.

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 $\{\mathcal{L}_n\}$, we extend Escanciano and Lobato's (2009) automatic maximum lag selection method to our setting. Escanciano and Lobato (2009) develop a Q-test with 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 an asymptotic theory for the new 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), amongst 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 non-mixing processes, hence our setting is more general than Zhu's (2015) for his block-wise random weighting bootstrap.

[2] We cannot provide an upper bound on $\mathcal{L}_n \rightarrow \infty$ similar to the one in Xiao and Wu (2014). This is an unavoidable cost for our (i) having a broad class of dependence under the null; (ii) using residuals and therefore requiring convergence of maxima that are not approximated by a Gaussian process; and (iii) sidestepping extreme value theory arguments.

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 4^{th} moment, as we do, but do not allow for residuals. We show in the supplemental material Hill and Motegi (2018, 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).

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, Lobato, Nankervis, and Savin, 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).

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 that our proposed max-correlation test with Shao's (2011) dependent wild bootstrap and automatic lag (denoed $\hat{\mathcal{T}}^{dw}(\mathcal{L}_n^*)$) dominates a variety of other tests. In this paper, 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 Hill and Motegi (2018, Appendix G), 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 almost no 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. 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 paper is as follows. Section 2 contains the assumptions and main results, automatic lag selection is developed in Section 3, and a Monte Carlo study follows in Section 4. Concluding remarks are left for Section 5. Proofs are gathered in Appendix A and the supplemental material Hill and Motegi (2018, Appendix F), and all figures and tables are placed at the end.

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 $||\epsilon_t - E[\epsilon_t|\mathfrak{V}_{t-m}^{t+m}]||_q = O\left(m^{-\lambda-\iota}\right)$ for tiny $\iota > 0$.[3] 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 non-mixing 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, Chapter 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* $\hat{\omega}_n(h) \xrightarrow{p} \omega(h)$ *for non-random* $\omega(h) \in (0, \infty)$*, for each* h.

Remark 1. 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 an NED process. α-mixing, for example, implies ergodicity, it extends to measurable transforms, and is a sub-class of NED. Lobato, Nankervis, and Savin (2002)

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

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.[4]

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 iid 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

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

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

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

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

$$z_t(h) \equiv r_t(h) - \rho(h)r_t(0) = \frac{\epsilon_t\epsilon_{t-h} - \rho(h)\epsilon_t^2 - (1-\rho(h))\,\mathcal{D}(h)'\mathcal{A}m_t}{E\left[\epsilon_t^2\right]}.$$

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 $j = 1,2,3$; $E[\sup_{\phi\in\mathcal{N}_{\phi_0}} |(\partial/\partial\phi)^j f(x_t,\phi)|^4] < \infty$ for $j = 0,1,2,3$ and some compact set with positive measure $\mathcal{N}_{\phi_0} \subseteq \Phi$ containing ϕ_0.*

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 $\sqrt{n}(\hat{\theta}_n - \theta_0) = \mathcal{A}n^{-1/2}\sum_{t=1}^n m_t(\theta_0) + o_p(1)$, 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*

[4] 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, Theorem 17.9).

$\geq k_\theta$; *and non-stochastic* $\mathcal{A} \in \mathbb{R}^{k_\theta \times k_m}$. *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 2. Smoothness (a) and (b) ensure a stochastic equicontinuity property for uniform laws of large numbers. Non-differentiability can be allowed provided certain other smoothness conditions involving, e.g., bracketing numbers apply (see, e.g. Pakes and Pollard, 1989, Arcones and Yu, 1994).

Remark 3. $E[\sup_{\phi\in\mathcal{N}_{\phi_0}} |(\partial/\partial\phi)^j f(x_t,\phi)|^4] < \infty$ 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 required 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, e.g. GARCH, Quadratic GARCH, GJR-GARCH (Francq and Zakoïan, 2004, 2010).

Remark 4. $\hat{\theta}_n$ under (c) is asymptotically a linear function of some zero mean $\mathcal{F}_t$-measurable process $m_t(\theta_0)$. This includes M-estimators, GMM and (Generalized) Empirical Likelihood with smooth or nonsmooth estimating equations, and estimators with non-smooth criteria and asymptotic expansions like LAD 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, Theorem 17.8), which promotes Gaussian finite dimensional asymptotics for the residuals cross-product.

Remark 5. (d) is a standard nondegeneracy assumption for finite dimensional asymptotics.

The theory developed in this paper extends to a class of 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: 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 sum $\vartheta(a) = \sum_{h=1}^{\mathcal{L}} |a_h|$, 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, for the sum $\tilde{\vartheta}(a) = \sum_{h=1}^{\mathcal{L}} a_h$, $\tilde{\vartheta}([\sqrt{n}\hat{\omega}_n(h)\hat{\rho}_n(h)]_{h=1}^{\mathcal{L}_n})$ 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. We can use the normalization $\mathcal{N}_n \equiv (2\mathcal{L}_n)^{-1/2}\sum_{h=1}^{\mathcal{L}_n}\hat{\omega}_n(h)\{n\hat{\rho}_n^2(h)-1\}$ used in Hong (1996, 2001), but bootstrapping the latter is arithmetically equivalent to bootstrapping $n\sum_{h=1}^{\mathcal{L}_n}\hat{\omega}_n^2(h)\hat{\rho}_n^2(h)$, and contrary to Hong (1996, 2001) we do not require $\mathcal{N}_n$ to converge to a standard normal law under the null.

The following result establishes a key approximation theory for an increasing sequence of serial correlations. See Appendix A for all proofs.

Lemma 2.1. *Let Assumptions 1 and 2 hold. For some non-unique sequence $\{\mathcal{L}_n\}$ of positive integers, where $\mathcal{L}_n \to \infty$ 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:*

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

Remark 6. The sequence $\{\mathcal{L}_n\}$ is not unique because for any other $\{\mathring{\mathcal{L}}_n\}$, $\mathring{\mathcal{L}}_n \to \infty$ and $\limsup_{n\to\infty}\{\mathring{\mathcal{L}}_n/\mathcal{L}_n\} < 1$, 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 $n \to \infty$:

$$\begin{aligned}\vartheta\left(\left[\sqrt{n}\hat{\omega}_n(h)\{\hat{\rho}_n(h)-\rho(h)\} - \omega(h)\mathcal{Z}_n(h)\right]_{h=1}^{\mathring{\mathcal{L}}_n}\right) \\ \leq \vartheta\left(\left[\sqrt{n}\hat{\omega}_n(h)\{\hat{\rho}_n(h)-\rho(h)\} - \omega(h)\mathcal{Z}_n(h)\right]_{h=1}^{\mathcal{L}_n}\right) \xrightarrow{p} 0,\end{aligned} \tag{6}$$

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 (6) 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. We do not highlight it because it does not promote a consistent test.

Remark 7. In our general environment we cannot obtain an upper bound on the maximum lag increase $\mathcal{L}_n \to \infty$. We can only say that the approximation holds over all $1 \leq h \leq \mathcal{L}_n$ for some $\{\mathcal{L}_n\}$, $\mathcal{L}_n = o(n)$ and $\mathcal{L}_n \to \infty$. This arises entirely from our allowing for a filter: Gaussian approximations and extreme value theoretic approaches are not suitable in this general case. In Section 3 we propose a data-dependent automatic lag selection that helps resolve the arbitrariness of lag choice in practice. The theory there, however, requires an upper bound on how fast any feasible $\mathcal{L}_n$ diverges. In Section 4 we show that the automatic lag works very well in practice.

The proof of Lemma 2.1 relies on a two-fold argument. 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 equivelant to pointwise convergence and 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 also Dudley (1978, 1984) and Pollard (1990, Chapters 9-10). We take an approach different from Hoffman-Jorgensen's (1984) notion of weak dependence, based on new theory developed below. 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 and A.2 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.a,b and 2.c,d hold. Let $\{\mathcal{Z}(h) : h \in \mathbb{N}\}$ be 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 $E[\mathcal{Z}(h)\mathcal{Z}(\tilde{h})] = \lim_{n\to\infty} n^{-1}\sum_{s,t=1}^{n} E[z_s(h)z_t(\tilde{h})]$. Then for some $\{\mathcal{Z}(h) : h \in \mathbb{N}\}$ and some non-unique sequence of positive integers $\{\mathcal{L}_n\}$, $\mathcal{L}_n \to \infty$ and $\mathcal{L}_n = o(n)$:*

$$\left|\vartheta\left([\omega(h)\mathcal{Z}_n(h)]_{h=1}^{\mathcal{L}_n}\right) - \vartheta\left([\omega(h)\mathcal{Z}(h)]_{h=1}^{\mathcal{L}_n}\right)\right| \leq \vartheta\left([\omega(h)\mathcal{Z}_n(h) - \omega(h)\mathcal{Z}(h)]_{h=1}^{\mathcal{L}_n}\right) \xrightarrow{p} 0.$$

Remark 8. 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 iid 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 in principle we can bypass our array convergence argument and use the Gaussian approximation argument in, for example, Chernozhukov, Chetverikov, and Kato (2013). However, we do not know if their argument extends to non-independent data, while $\epsilon_t\epsilon_{t-h}$ in the paper is only required to be NED and ergodic. Indeed, the array convergence argument for Lemma A.1 does not rely on probabilistic properties at all. The NED assumption merely ensures convergence in finite dimensional distributions.

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, $|\vartheta([\sqrt{n}\hat{\omega}_n(h)\{\hat{\rho}_n(h) - \rho(h)\}]_{h=1}^{\mathcal{L}_n}) - \vartheta([\omega(h)\mathcal{Z}(h)]_{h=1}^{\mathcal{L}_n})| \xrightarrow{p} 0$ for some 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 $|\vartheta([\sqrt{n}\hat{\omega}_n(h)\hat{\rho}_n(h)]_{h=1}^{\mathcal{L}_n}) - \vartheta([\omega(h)\mathcal{Z}(h)]_{h=1}^{\mathcal{L}_n})| \xrightarrow{p} 0$, 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 $r_t(h) \equiv \{\epsilon_t\epsilon_{t-h} - \mathcal{D}(h)'\mathcal{A}m_t\}/E[\epsilon_t^2]$.*

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

Corollary 2.4. *Under Assumptions 1 and 2,* $|\max_{1\leq h\leq \mathcal{L}_n}|\sqrt{n}\hat{\omega}_n(h)\{\hat{\rho}_n(h) - \rho(h)\}| - \max_{1\leq h\leq \mathcal{L}_n}|\omega(h)\mathcal{Z}(h)|| \xrightarrow{p} 0$ *for some 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 hypothesis* $|\max_{1\leq h\leq \mathcal{L}_n}|\sqrt{n}\hat{\omega}_n(h)\hat{\rho}_n(h)| - \max_{1\leq h\leq \mathcal{L}_n}|\omega(h)\mathcal{Z}(h)|| \xrightarrow{p} 0.$

Remark 9. The conclusions of Theorem 2.3 and Corollary 2.4 do not require $\vartheta([\sqrt{n}\hat{\omega}_n(h)\hat{\rho}_n(h)]_{h=1}^{\mathcal{L}_n})$ to have a well defined limit law under the null. This is decidedly different from the max-correlation literature in which $\lim_{n\to\infty}\max_{1\leq h\leq \mathcal{L}_n}|\omega(h)\mathcal{Z}(h)|$ is characterized under suitable conditions that ensure asymptotic independence $E[\mathcal{Z}(i)\mathcal{Z}(j)] \to 0$ as $|i-j| \to 0$. See, e.g., Leadbetter, Lindgren, and Rootzén (1983, Chapter 6) and Hüsler (1986). We do not require asymptotic independence, nor therefore convergence in law.

2.2 Bootstrapped P-Value Test

We work with the 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

$$\hat{\mathcal{D}}_n(h) \equiv \frac{1}{n}\sum_{t=h+1}^{n}\left\{\left(\epsilon_t(\hat{\theta}_n)s_t(\hat{\theta}_n) + \frac{G_t(\hat{\theta}_n)}{\sigma_t(\hat{\theta}_n)}\right)\epsilon_{t-h}(\hat{\theta}_n) + \epsilon_t(\hat{\theta}_n)\left(\epsilon_{t-h}(\hat{\theta}_n)s_{t-h}(\hat{\theta}_n) + \frac{G_{t-h}(\hat{\theta}_n)}{\sigma_{t-h}(\hat{\theta}_n)}\right)\right\}. \quad (7)$$

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

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), amongst others, lay out sufficient conditions for the QML estimator for a large class of 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 iid and mds sequences (Wu, 1986, Liu, 1988, Hansen, 1996). Shao (2010, 2011) generalizes the idea to allow for dependent sequences. Shao (2010) allows for general dependence by using block-wise iid 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 amongst 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 iid 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:

$$\hat{\rho}_n^{(dw)}(h) \equiv \frac{1}{1/n\sum_{t=1}^n \epsilon_t^2(\hat{\theta}_n)} \frac{1}{n} \sum_{t=1+h}^n \varphi_t \left\{ \widehat{\mathcal{E}}_{n,t,h}(\hat{\theta}_n) - \frac{1}{n} \sum_{s=1+h}^n \widehat{\mathcal{E}}_{n,s,h}(\hat{\theta}_n) \right\}. \tag{8}$$

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:

$$\hat{\rho}_n^{(dw)}(h) \equiv \frac{1}{1/n\sum_{t=1}^n \epsilon_t^2(\hat{\theta}_n)} \frac{1}{n} \sum_{t=1+h}^n \varphi_t \left\{ \epsilon_t(\hat{\theta}_n)\epsilon_{t-h}(\hat{\theta}_n) - \frac{1}{n} \sum_{s=1+h}^n \epsilon_t(\hat{\theta}_n)\epsilon_{t-h}(\hat{\theta}_n) \right\}. \tag{9}$$

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 (9) 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 (8), however, contains the required information.

Shao (2011) imposes Wu's (2005) moment contraction property with an eighth moment, which we denote MC_8 (see Appendix B in Hill and Motegi, 2018, for details). 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 $\int_0^\pi \hat{S}_n^2(\lambda)d\lambda$.[5] 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 Theorem 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, 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: Theorem 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

[5] See, e.g., Politis and Romano (1994) for applications of weak convergence in a Hilbert space to the bootstrap.

are achievable at the expense of more intense notation.[6]

Assumption 2.c′. *$\hat{\theta}_n \in \Theta$ for each n, and for a unique interior point $\theta_0 \in \Theta$ we have $\sqrt{n}(\hat{\theta}_n - \theta_0)$ $= \mathcal{A}n^{-1/2}\sum_{t=1}^{n} m_t(\theta_0) + o_p(1)$, 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 non-stochastic $\mathcal{A} \in \mathbb{R}^{k_\theta \times k_m}$. $m_t(\theta)$ 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.

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 non-unique 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 $P(\hat{p}_{n,M}^{(dw)} < \alpha) \to 1$.*

Remark 10. A similar theory applies to an approximate p-value computed by wild bootstrap where φ_t is iid $N(0,1)$, provided ϵ_t forms a mds under the null.

Remark 11. 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 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 Hill and Motegi (2018, Appendix D, especially Theorem D.1).

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 theoretical terms, as explained above, when using filtered data we cannot pinpoint an upper bound on the rate of increase of $\mathcal{L}_n$ because the Lemma 2.1 expansion cannot rely on a Gaussian approximation theory as in Chernozhukov, Chetverikov, and Kato (2013, 2014, 2015, 2017), nor extreme value theory arguments as in Xiao and Wu (2014). We therefore assume an upper bound $\{\bar{\mathcal{L}}_n\}$ on the growth of $\mathcal{L}_n$. Let $\{\bar{\mathcal{L}}_n\}$ be such that $\bar{\mathcal{L}}_n \to \infty$. 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 $K = 1$. The proof of Theorem 3.1 below requires $\bar{\mathcal{L}}_n = o(n/\ln(n))$ in order to expedite the proof.

[6]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, e.g., Andrews (1987), Arcones and Yu (1994) and Gaenssler and Ziegler (1994).

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 $\hat{\omega}_n(h) \xrightarrow{p} \omega(h) > 0$.

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 6 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

$$\hat{\mathcal{T}}_n^{\mathcal{P}}(\mathcal{L}) \equiv \hat{\mathcal{T}}_n(\mathcal{L}) - \mathcal{P}_n(\mathcal{L}) \text{ where } \hat{\mathcal{T}}_n(\mathcal{L}) \equiv \sqrt{n} \max_{1 \leq h \leq \mathcal{L}} |\hat{\rho}_n(h)| \tag{10}$$

with penalty $\mathcal{P}_n(\cdot)$. The penalty function is:

$$\mathcal{P}_n(\mathcal{L}) = \begin{cases} \sqrt{\mathcal{L}\ln n} & \text{if } \hat{\mathcal{T}}_n(\mathcal{L}) \leq \sqrt{q \ln n} \\ \sqrt{2\mathcal{L}} & \text{if } \hat{\mathcal{T}}_n(\mathcal{L}) > \sqrt{q \ln n} \end{cases} \tag{11}$$

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 $q = 3.25$ 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 $\mathcal{L}_n^*$ is:

$$\mathcal{L}_n^* = \min\left\{\mathcal{L}_n : 1 \leq \mathcal{L}_n \leq \bar{\mathcal{L}}_n : \hat{\mathcal{T}}_n^{\mathcal{P}}(\mathcal{L}_n) \geq \hat{\mathcal{T}}_n^{\mathcal{P}}(l) \text{ for each } l = 1, ..., \bar{\mathcal{L}}_n\right\}. \tag{12}$$

We chose $\{\mathcal{L}_n\}$ from those integer sequences satisfying $\mathcal{L}_n \geq 1$ and $\mathcal{L}_n \leq \bar{\mathcal{L}}_n$ 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 $\mathcal{L}_n \geq 1$ and $\mathcal{L}_n/\bar{\mathcal{L}}_n \to [0,1]$. Notice l may be a function of n, e.g. $l = \bar{\mathcal{L}}_n - 1$. 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.[7]

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

[7] Escanciano and Lobato (2009, second remark following Theorem 2) claim that a diverging maximum lag is possible for their Q-test and maximum lag, but an asymptotic theory is not presented.

Theorem 3.1. *Let Assumptions 1 and 2 hold, and let $\bar{\mathcal{L}}_n = o(n/\ln(n))$. a. Under H_0, $P(\mathcal{L}_n^* = 1) \to 1$; and b. under H_1, $\mathcal{L}_n^* \xrightarrow{p} h^*$.*

Remark 12. 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 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 Theorem 2 in Escanciano and Lobato, 2009).

4 Monte Carlo Experiments

We now perform a Monte Carlo experiment to gauge the merits of the max-correlation test and automatic lag (labeled as $\hat{\mathcal{T}}^{dw}(\mathcal{L}_n^*)$). A main competitor studied here is a Shao's (2011) dependent wild bootstrap spectral Cramér-von Mises test (labeled as CvM^{dw}). See Section 4.1 for the simulation design and Section 4.2 for results. In the supplemental material Hill and Motegi (2018, Appendix G) we study other tests, including the max-correlation with a pre-chosen non-random lag $\mathcal{L}_n$, the Ljung-Box test, Hong's (1996) test is 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 iid $N(0,1)$. We consider iid $e_t = \nu_t$; 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.[8] 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.

[8]Ergodicity follows since each error process is stationary α-mixing. See, e.g., 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.

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.[9]

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 iid 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 iid 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 iid, and semi-strong when e_t is GARCH(1,1) since it is an adapted mds (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 standardizd error ϵ_t $= e_t = y_t/\sigma_t$ which has a finite fourth moment in all cases.

In Scenario #5, however, we test GARCH $\epsilon_t = y_t$ itself. $E[\epsilon_t^4] < \infty$ holds when e_t is iid or MA(2), but it is unknown whether a fourth moment exists when e_t is GARCH(1,1) or AR(1). 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 iid error e_t is used for the following processes in order to focus in 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 the serial dependence.

We draw $J = 1000$ Monte Carlo samples of size $n \in \{100, 250, 500, 1000\}$. 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.

[9] 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.

Table 1: Dependence of Test Variable ϵ_t under Each Scenario, Error e_t, and Filter

	Scenario: Model and Filter						
e_t \ filter	#1 Simple -	#2 Bilinear -	#3 AR(2) AR(2) filter	#4 AR(2) AR(1) filter	#5 GARCH -	#6 GARCH GARCH filter	#7, #8, #9 Remote MA -
iid	**iid**	**wn**	**iid**	corr	**mds**	**iid**	remote corr
GARCH	**mds**	**wn**	**mds**	corr	**mds**	**mds**	not considered
MA(2)	corr	corr	corr	corr	corr	corr	not considered
AR(1)	corr	corr	corr	corr	corr	corr	not considered

wn = non-mds white noise. corr = autocorrelated. remote corr = autocorrelation is present at a remote lag. **bold** text is used to highlight when the null is true.

Our proposed test is the max-correlation test with the dependent wild bootstrap and automatic lag, $\hat{\mathcal{T}}^{dw}(\mathcal{L}_n^*)$. The test statistic is $\hat{\mathcal{T}}_n(\mathcal{L}_n^*) \equiv \sqrt{n}\max_{1\leq h\leq \mathcal{L}_n^*}|\hat{\omega}_n(h)\hat{\rho}_n(h)|$ with weight $\hat{\omega}_n(h) = 1$.[10] We compute the bootstrapped statistic $\hat{\mathcal{T}}_{n,i}^{(dw)}(\mathcal{L}_{n,i}^*) \equiv \sqrt{n}\max_{1\leq h\leq \mathcal{L}_{n,i}^*}|\hat{\rho}_{n,i}^{(dw)}(h)|$ for each bootstrap sample $i \in \{1,\ldots,M\}$ with $M = 500$. $\hat{\rho}_{n,i}^{(dw)}(h)$ is computed via (8) based on the Lemma 2.1 correlation expansion, which correctly accounts for the first order (asymptotic) impact of the i^{th} sample's plug-in $\hat{\theta}_{n,i}$. Note that $\mathcal{L}_{n,i}^*$ is the automatic lag for the i^{th} 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. We therefore use the middle value $b = 1$.[11] The approximate p-value is computed as $\hat{p}_{n,M}^{(dw)} = 1/M\sum_{i=1}^{M} I(\hat{\mathcal{T}}_{n,i}^{(dw)}(\mathcal{L}_{n,i}^*) \geq \hat{\mathcal{T}}_n(\mathcal{L}_n^*))$.

The automatic lag selection requires a choice of the maximum possible lag length $\bar{\mathcal{L}}_n = o(n/\ln(n))$ and the tuning parameter q (cf. (11) and (12)). We set $\bar{\mathcal{L}}_n = [\delta \times n/(\ln n)^{4/3}]$ with $\delta = 1.5$ so that $\bar{\mathcal{L}}_n \in \{19, 38, 65, 114\}$ for $n \in \{100, 250, 500, 1000\}$, respectively. Our choice satisfies the requirement $\bar{\mathcal{L}}_n = o(n/\ln(n))$, and gives a reasonable increase with n. Similar values lead to qualitatively similar results.

In order to choose a plausible value of q, 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 iid error; and size-adjusted power is computed under #4 with an iid error. In Case 2, size is computed under #5 with an iid 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

[10] Other plausible weights include an inverted standard deviation based on a HAC estimator, and/or the Ljung and Box (1978) weights. It is left as a future task to investigate how small sample performance changes under those weights. In the present paper, we demonstrate that the uniform weight leads to sharp size and high power.

[11] In simulations not reported here, we compared $b_n = b\sqrt{n}$ across $b \in \{.5, 1, 2\}$ and found there is little difference in test performance.

the test performance. For each case and sample size, we obtain relatively accurate size and high power around $q = 3.25$. We therefore use $q = 3.25$ throughout.

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 non-standard limit distribution under the null.[12] We then use Shao's (2011, Section 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.

4.2 Simulation Results

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 $\mathcal{L}_n^* \to h^*$, the smallest lag at which the largest correlation occurs. Under Scenarios #1-#6 when the error e_t is iid or GARCH the null is false only for #4. In the latter case, the test variable ϵ_t is AR(1) hence it's $h^* = 1$.

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 analyitic 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 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 abberant result. As suggested in Section 4.1, either of these processes may fail the required moment conditions for the underlying theory surrounding $\mathcal{L}_n^*$.

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 $n \geq 500$.

We now report rejection frequencies associated with nominal size $\alpha \in \{.01, .05, .10\}$. See Table 3 for $\hat{\mathcal{T}}^{dw}(\mathcal{L}_n^*)$ under #1–#6; see Table 4 for CvM^{dw} under #1–#6; and see Table 5 for both tests under

[12] In practice we use a numerical integral based on the midpoint approximation with the increment of .01.

#7–#9.

4.2.1 Empirical Size

We begin with Scenario #1 (simple), $n = 100$, and iid error. The empirical size with respect to nominal sizes $\alpha \in \{.010, .050, .100\}$ is $\{.011, .058, .108\}$ for $\hat{\mathcal{T}}^{dw}(\mathcal{L}_n^*)$ and $\{.023, .081, .138\}$ for CvM^{dw}, hence $\hat{\mathcal{T}}^{dw}(\mathcal{L}_n^*)$ has sharp size, and sharper than CvM^{dw}. A similar implication holds for #2 (bilinear), $n = 100$, and iid error, where the empirical size is $\{.009, .060, .107\}$ for $\hat{\mathcal{T}}^{dw}(\mathcal{L}_n^*)$ and $\{.018, .076, .149\}$ for CvM^{dw}. In general, size f or 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 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 Hill and Motegi (2018, Appendix G) that $\hat{\mathcal{T}}^{dw}(\mathcal{L}_n^*)$ achieves the sharpest size among a varety of tests.[13] 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.

4.2.2 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, $n = 1000$, and AR(1) error, for example. The empirical power with respect to $\alpha \in \{.010, .050, .100\}$ is $\{.732, .822, .856\}$ for $\hat{\mathcal{T}}^{dw}(\mathcal{L}_n^*)$ and $\{.474, .697, .810\}$ for CvM^{dw}. But in #3, with $n = 1000$, and an AR(1) error, power is $\{.616, .841, .913\}$ for $\hat{\mathcal{T}}^{dw}(\mathcal{L}_n^*)$ and $\{.688, .876, .923\}$ 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. The power of $\hat{\mathcal{T}}^{dw}(\mathcal{L}_n^*)$ under #8 (Remote MA(12)), for instance, is $\{.019, .077, .132\}$ for $n = 100$, $\{.029, .151, .249\}$ for $n = 250$, $\{.377, .652, .717\}$ for $n = 500$, and $\{.985, .993, .993\}$ for $n = 1000$. Logically power is better detected 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 $n \geq 500$. The power of CvM^{dw}, by contrast, is $\{.034, .110, .179\}$ for $n = 100$, $\{.025, .087, .155\}$ for $n = 250$, $\{.026, .092, .161\}$ for $n = 500$, and $\{.017, .083, .166\}$ for $n = 1000$. CvM^{dw} has (almost) no power against the remote autocorrelation even when $n = 1000$. In fact, we find in Hill and Motegi (2018, Appendix G) 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.

[13] In Scenario #2 (bilinear) with a GARCH error, the max-correlation test is undersized, even in large samples $n = 1000$. 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. We leave that possibility for a future project.

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 (non-weighted) 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 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 remove correlations.

In some cases against adjacent correlations power is not dominant when a large pre-chosen non-random $\mathcal{L}_n$ is used (see Hill and Motegi, 2018, Appendix G), 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 out-performs 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 high power while the Cramér-von Mises test has almost no power. 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 constrction incorporate many lags into a test statistic (the CvM test gives the greatest weight to low lags, making it useless against remote lags). In future work a deep examination of the max-correlation test with weights other than unity should be performed, since sample autocorrelations with high dispersion weaken the efficacy of the bootstrap and automatic lag selection.

A Appendix: 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 Theorem 17.5 in Davidson (1994) and Theorem 1.6 in McLeish (1975):

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

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, Lemma 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.

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 $\lim_{n\to\infty} \mathcal{A}_{k,n} = 0$ for each fixed k, and $\mathcal{A}_{k,n} \leq \mathcal{A}_{k+1,n}$ for each n and all k. Then $\lim_{n\to\infty} \mathcal{A}_{\mathcal{L}_n,n} = 0$ for some sequence $\{\mathcal{L}_n\}$ of positive integers, $\mathcal{L}_n \to \infty$ that is not unique.*

Proof.

Step 1. We will prove $\lim_{l\to\infty} \mathcal{A}_{\mathcal{L}(n_l),n_l} = 0$ for some sequence of positive integers $\{n_l\}_{l=1}^{\infty}$, $n_l < n_{l+1}$ $\forall l$, and some mapping $\mathcal{L}(n_l) \leq \mathcal{L}(n_{l+1})$, $\mathcal{L}(n_l) \to \infty$ and $n_l \to \infty$ as $l \to \infty$. We use that result in Step 2 to prove the claim.

$\{\mathcal{A}_{k,n} : 1 \leq k \leq \mathcal{I}_n\}_{n\geq 1}$ lies in a first countable topological space and $\lim_{k\to\infty} \lim_{n\to\infty} \mathcal{A}_{k,n} = 0$. Therefore, by Lemma 1 in Boehme and Rosenfeld (1974) there exists a sequence of positive integers $\{\mathcal{L}_i\}_{i=1}^{\infty}$, $\mathcal{L}_i \to \infty$ as $i \to \infty$, and an integer mapping $n(\mathcal{L}) \to \infty$ as $\mathcal{L} \to \infty$ such that $\lim_{i\to\infty} \mathcal{A}_{\mathcal{L}_i,n(\mathcal{L}_i)} = 0$. The relation $n(\mathcal{L}) \to \infty$ as $\mathcal{L} \to \infty$ holds by construction of the array $\{\mathcal{A}_{k,n} : 1 \leq k \leq \mathcal{I}_n\}_{n\geq 1}$ with $\mathcal{I}_n \to \infty$ as $n \to \infty$.

We can always assume monotonicity: $\mathcal{L}_i \leq \mathcal{L}_{i+1}$ $\forall i$. Simply note that $\lim_{i\to\infty} \mathcal{A}_{\mathcal{L}_i,n(\mathcal{L}_i)} = 0$ implies $\lim_{l\to\infty} \mathcal{A}_{\mathcal{L}_{i_l},n(\mathcal{L}_{i_l})} = 0$ for every infinite subsequence $\{i_l\}_{l\geq 1}$ of $\{i\}_{i\geq 1}$. Since $\mathcal{L}_i \to \infty$ as $i \to \infty$, we can find a subsequence $\{i_l^*\}_{l\geq 1}$ such that $i_l^* \leq i_{l+1}^*$ and $\mathcal{L}_{i_l^*} \leq \mathcal{L}_{i_{l+1}^*}$ for each l. This follows from the monotone subsequence theorem, which itself follows from Ramsey's (1930) theorem, cf. Erdös and Szekeres (1935) and Burkill and Mirsky (1973). Now define $\mathcal{L}_l^* \equiv \mathcal{L}_{i_l^*}$, hence $\lim_{l\to\infty} \mathcal{A}_{\mathcal{L}_l^*,n(\mathcal{L}_l^*)} = 0$ where $\mathcal{L}_l^* \leq \mathcal{L}_{l+1}^*$ and $\mathcal{L}_l^* \to \infty$ as $l \to \infty$.

Now let $\{n_i\}_{i=1}^{\infty}$ and $\{\mathcal{L}(n_i)\}_{i=1}^{\infty}$ be any sequences satisfying $n_i = n(\mathcal{L}_i)$ and $\mathcal{L}(n_i) = \mathcal{L}_i$. Hence $\mathcal{L}(n_i) \leq \mathcal{L}(n_i + 1)$, $\mathcal{L}(n_i) \to \infty$ and $n_i \to \infty$, such that $\lim_{i\to\infty} \mathcal{A}_{\mathcal{L}(n_i),n_i} = 0$. Note that $\lim_{i\to\infty} \mathcal{A}_{\mathcal{L}(n_i),n_i} = 0$ *if and only if* $\lim_{l\to\infty} \mathcal{A}_{\mathcal{L}(n_{i_l}),n_{i_l}} = 0$ for every subsequence $\{n_{i_l}\}_{l=1}^{\infty}$ of $\{n_i\}_{i=1}^{\infty}$. Since $n_i \to \infty$ as $i \to \infty$, by the monotone subsequence theorem there exists a strictly monotonically increasing subsequence $\{n_{i_l}\}_{l=1}^{\infty}$. Therefore, as required $\lim_{l\to\infty} \mathcal{A}_{\mathcal{L}(n_l),n_l} = 0$ for some sequence of positive integers $\{n_l\}_{l=1}^{\infty}$, $n_l < n_{l+1}$ $\forall l$, and $\mathcal{L}(n_l) \leq \mathcal{L}(n_{l+1})$, $\mathcal{L}(n_l) \to \infty$ and $n_l \to \infty$ as $l \to \infty$.

Step 2. By assumption $\lim_{n\to\infty} \mathcal{A}_{k,n} = 0$ $\forall k$. Therefore:

$$\lim_{s\to\infty} \mathcal{A}_{k,n_s} = 0 \text{ for every } k \text{ and every infinite subsequence } \{n_s\}_{s\geq 1}. \tag{A.2}$$

Now repeat the Step 1 argument for each $\{\mathcal{A}_{k,n_s}\}_{s\geq 1}$: there exists a strictly monotonically increasing subsequence of positive integers $\{n_{s_l}\}_{l\geq 1}$ and some integer mapping $\mathcal{L}_s(n_{s_l})$ that may depend on s, with $n_{s_l} \to \infty$ and $\mathcal{L}_s(n_{s_l}) \to \infty$ as $l \to \infty$ $\forall s$, such that $\lim_{l\to\infty} \mathcal{A}_{\mathcal{L}_s(n_{s_l}),n_{s_l}} = 0$ $\forall s$. As above, we may take $\mathcal{L}_s(\cdot)$ to be monotonic: $\mathcal{L}_s(\tilde{n}) \leq \mathcal{L}_s(\tilde{n} + 1)$ $\forall \tilde{n}$.

Since monotonic $\mathcal{L}_s(\tilde{n}) \to \infty$ as $\tilde{n} \to \infty$ $\forall s$, there exists an integer mapping $\mathcal{L}(\cdot)$ such that $\mathcal{L}(n) \to \infty$ as $n \to \infty$ and for each s, $\limsup_{n\to\infty}\{\mathcal{L}(n)/\mathcal{L}_s(n)\} < 1$. By monotonicity $\mathcal{A}_{k,n} \leq \mathcal{A}_{k+1,n}$ this mapping satisfies

$$\lim_{l\to\infty} \mathcal{A}_{\mathcal{L}(n_{s_l}),n_{s_l}} \leq \lim_{l\to\infty} \mathcal{A}_{\mathcal{L}_s(n_{s_l}),n_{s_l}} = 0 \ \forall s. \tag{A.3}$$

Notice $\mathcal{L}(\cdot)$ is not unique: for any $\mathcal{L}(\cdot)$ that satisfies (A.3) there exists $\tilde{\mathcal{L}}(n) \to \infty$ such that $\limsup_{n\to\infty} \tilde{\mathcal{L}}(n)/\mathcal{L}(n) < 1$, hence by monotonicity $\lim_{l\to\infty} \mathcal{A}_{\tilde{\mathcal{L}}(n_{s_l}),n_{s_l}} \leq \lim_{l\to\infty} \mathcal{A}_{\mathcal{L}(n_{s_l}),n_{s_l}} = 0$.

Now write $\mathcal{B}_n \equiv \mathcal{A}_{\mathcal{L}(n),n}$. By a direct implication of (A.2) and (A.3), for every subsequence $\{\mathcal{B}_{n_s}\}_{s\geq 1}$ there exists a further subsequence $\{\mathcal{B}_{n_{s_l}}\}_{l\geq 1}$ that converges $\lim_{l\to\infty} \mathcal{B}_{n_{s_l}} = 0$. Therefore $\lim_{n\to\infty} \mathcal{B}_n = 0$ (see Royden, 1988, p. 39). This proves $\lim_{n\to\infty} \mathcal{A}_{\mathcal{L}_n,n} = 0$ with $\mathcal{L}_n = \mathcal{L}(n)$ as required. $\mathcal{QED}$.

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. Recall the continuous mapping $\vartheta : \mathbb{R}^{\mathcal{L}_n} \to [0, \infty)$ that satisfies the following: 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 $\vartheta(a + b) \leq \vartheta(a) + \vartheta(b)$ $\forall a, b \in \mathbb{R}^{\mathcal{L}_n}$.

Lemma A.2. *Let $\{\mathcal{X}_n(i), \mathcal{Y}_n(i) : 1 \leq i \leq \mathcal{I}_n\}_{n\geq 1}$ be arrays of random variables, where $\{\mathcal{I}_n\}_n \geq 1$ is a sequence of positive integers, $\mathcal{I}_n \to \infty$ as $n \to \infty$.*

a. If $\mathcal{X}_n(i) \xrightarrow{p} 0$ for each i then $\vartheta([\mathcal{X}_n(i)]_{i=1}^{\mathcal{L}_n}) \xrightarrow{p} 0$ for some sequence $\{\mathcal{L}_n\}$ of positive integers with $\mathcal{L}_n \to \infty$. Moreover $\mathcal{L}_n = o(n)$ can always be assured.

b. If each $\mathcal{X}_n(i) - \mathcal{Y}_n(i) \xrightarrow{p} 0$ then for some sequence $\{\mathcal{L}_n\}$ of positive integers 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.

Remark 13. $\mathcal{L}_n = o(n)$ is always possible due to monotonicity of ϑ. This is required for sample correlation consistency.

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 $\mathcal{A}_{k,n} \equiv 1 - \exp\{-\vartheta([\mathcal{X}_n(i)]_{i=1}^{k})\}$ and $\mathcal{P}_{k,n} \equiv \int_0^\infty P(\mathcal{A}_{k,n} > \epsilon)d\epsilon$. By construction $\mathcal{A}_{k,n} \in [0,1]$ $a.s.$ $\forall k$. Lebesgue's dominated convergence theorem, and $\mathcal{A}_{k,n} \xrightarrow{p} 0$, therefore yield for each k:

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

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, $\mathcal{L}_n \to \infty$ and $\mathcal{L}_n = o(n)$, such that $\lim_{n\to\infty} \mathcal{P}_{\mathcal{L}_n,n} = \lim_{n\to\infty} \int_0^1 P(\mathcal{A}_{\mathcal{L}_n,n} > \epsilon)d\epsilon = 0$. 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.

The sequence $\{\mathcal{L}_n\}$ is not unique for either of the following reasons: (i) the probability limit is asymptotic hence we can always change $\mathcal{L}_n$ for finitely many n; and (ii) by monotonicity of ϑ any other $\{\mathring{\mathcal{L}}_n\}$ that satisfies $\mathring{\mathcal{L}}_n \to \infty$ and $\limsup_{n\to\infty}\{\mathring{\mathcal{L}}_n/\mathcal{L}_n\} < 1$ satisfies $\vartheta([\mathcal{X}_n(i)]_{i=1}^{\mathring{\mathcal{L}}_n}) \leq \vartheta([\mathcal{X}_n(i)]_{i=1}^{\mathcal{L}_n}) \xrightarrow{p} 0$ as $n \to \infty$. Therefore we can always find $\{\mathcal{L}_n\}$, $\mathcal{L}_n \to \infty$ and $\mathcal{L}_n = o(n)$, that satisfies $\vartheta([\mathcal{X}_n(i)]_{i=1}^{\mathcal{L}_n}) \xrightarrow{p} 0$.

Claim (b). The mapping ϑ satisfies the triangular inequality and $\vartheta(\cdot) \geq 0$. Apply the inequality twice to yield $\vartheta([\mathcal{X}_n(i)]_{i=1}^{\mathcal{L}_n}) \leq \vartheta([\mathcal{Y}_n(i)]_{i=1}^{\mathcal{L}_n}) + \vartheta([\mathcal{X}_n(i) - \mathcal{Y}_n(i)]_{i=1}^{\mathcal{L}_n})$ and $\vartheta([\mathcal{Y}_n(i)]_{i=1}^{\mathcal{L}_n}) \leq \vartheta([\mathcal{X}_n(i)]_{i=1}^{\mathcal{L}_n}) + \vartheta([\mathcal{X}_n(i) - \mathcal{Y}_n(i)]_{i=1}^{\mathcal{L}_n})$, hence $|\vartheta([\mathcal{X}_n(i)]_{i=1}^{\mathcal{L}_n}) - \vartheta([\mathcal{Y}_n(i)]_{i=1}^{\mathcal{L}_n})| \leq \vartheta([\mathcal{X}_n(i) - \mathcal{Y}_n(i)]_{i=1}^{\mathcal{L}_n})$. Now apply (a) to $\mathcal{X}_n(i) - \mathcal{Y}_n(i)$ to yield the desired result. $\mathcal{QED}$.

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

$$\begin{aligned}
&G_t(\phi) \equiv \left[\frac{\partial}{\partial\phi'} f(x_{t-1},\phi), \mathbf{0}'_{k_\delta}\right]' \in \mathbb{R}^{k_\theta} \quad \text{and} \quad s_t(\theta) \equiv \frac{1}{2}\frac{\partial}{\partial\theta}\ln\sigma_t^2(\theta) \\
&\mathcal{D}(h) \equiv E\left[\left(\epsilon_t s_t + G_t/\sigma_t\right)\epsilon_{t-h}\right] + E\left[\epsilon_t\left(\epsilon_{t-h}s_{t-h} + G_{t-h}/\sigma_{t-h}\right)\right] \in \mathbb{R}^{k_\theta} \\
&z_t(h) \equiv r_t(h) - \rho(h)r_t(0) \text{ where } r_t(h) \equiv \frac{\epsilon_t\epsilon_{t-h} - E\left[\epsilon_t\epsilon_{t-h}\right] - \mathcal{D}(h)'\mathcal{A}m_t}{E\left[\epsilon_t^2\right]},
\end{aligned}$$

where m_t are the Assumption 2.c estimating equations. The following two lemmas are based on standard arguments and are therefore proved in Hill and Motegi (2018, Appendix F).

Lemma A.3. *Under Assumptions 1 and 2: $\mathcal{X}_n(h) \equiv |\sqrt{n}\{\hat{\rho}_n(h) - \rho(h)\} - n^{-1/2}\sum_{t=1+h}^{n}\{r_t(h) - \rho(h)r_t(0)\}| \xrightarrow{p} 0$ for each h.*

Recall

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

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

Lemma A.4. *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. Assumption 1.c states $\hat{\omega}_n(h) \xrightarrow{p} \omega(h)$. Property (A.1) applies to $r_t(h) - \rho(h) r_t(0)$ under Assumptions 1 and 2, cf. Theorem 17.8 in Davidson (1994), hence $1/\sqrt{n} \sum_{t=1+h}^n \{r_t(h) - \rho(h) r_t(0)\} = O_p(1)$. Now use Lemma A.3 and the triangle inequality to yield:

$$\begin{aligned} \tilde{\mathcal{X}}_n(h) &\equiv \left| \sqrt{n} \hat{\omega}_n(h) \left\{ \hat{\rho}_n(h) - \rho(h) \right\} - \omega(h) n^{-1/2} \sum\nolimits_{t=1+h}^n \left\{ r_t(h) - \rho(h) r_t(0) \right\} \right| \\ &\leq |\omega(h)| \times \mathcal{X}_n(h) + |\hat{\omega}_n(h) - \omega(h)| \times \mathcal{X}_n(h) \\ &\quad + |\hat{\omega}_n(h) - \omega(h)| \times \left| n^{-1/2} \sum\nolimits_{t=1+h}^n \left\{ r_t(h) - \rho(h) r_t(0) \right\} \right| \xrightarrow{p} 0. \end{aligned} \tag{A.4}$$

The claims now follow by applications of Lemma A.2 to $\tilde{\mathcal{X}}_n(h)$. $\mathcal{QED}$.

Proof of Lemma 2.2. Since $\omega(h) > 0$ are finite, we set $\omega(h) = 1$ without loss of generality.

Step 1. By Step 2, for each h:

$$\mathcal{Z}_n(h) - \tilde{\mathcal{Z}}(h) \xrightarrow{p} 0 \tag{A.5}$$

where $\{\tilde{\mathcal{Z}}(h) : 1 \leq h \leq \mathcal{L}\}$ is a copy of the Lemma A.4 zero mean Gaussian process $\{\mathcal{Z}(h) : 1 \leq h \leq \mathcal{L}\}$. Apply Lemma A.2 to $\mathcal{Z}_n(h) - \tilde{\mathcal{Z}}(h)$ to yield the desired result.

Step 2. We now prove (A.5). Lemma A.4 implies $\mathcal{Z}_n(h) \xrightarrow{d} \mathcal{Z}(h)$ for each h, where $\{\mathcal{Z}(h) : 1 \leq h \leq \mathcal{L}\}$ is a zero mean Gaussian process. Therefore $E[\exp\{i\lambda \mathcal{Z}_n(h)\}] - E[\exp\{i\lambda \tilde{\mathcal{Z}}(h)\}] \to 0$ for all $\lambda \in \mathcal{R}$, where $i = \sqrt{-1}$ and $\{\tilde{\mathcal{Z}}(h) : 1 \leq h \leq \mathcal{L}\}$ is a copy of $\{\mathcal{Z}(h) : 1 \leq h \leq \mathcal{L}\}$. This follows because convergence in distribution holds *if and only if* there is convergence in characteristic functions by the portmanteau theorem (e.g. Billingsley, 1995, Theorem 26.3). Now factor out $E[\exp\{i\lambda \tilde{\mathcal{Z}}(h)\}]$ to yield:

$$E\left[\exp\left\{i\lambda \tilde{\mathcal{Z}}(h)\right\}\right] \times \left\{E\left[\exp\left\{i\lambda\left(\mathcal{Z}_n(h) - \tilde{\mathcal{Z}}(h)\right)\right\}\right] - 1\right\} \to 0. \tag{A.6}$$

But $\tilde{\mathcal{Z}}(h)$ is Gaussian with zero mean and variance $v(h)^2 \in (0, \infty)$ in view of Lemma A.4, hence $E[\exp\{i\lambda \tilde{\mathcal{Z}}(h)\}] = E[\exp\{-(1/2)\lambda^2 / v(h)^2\}] \in (0, \infty)$ for each $\lambda \in \mathcal{R}$. From (A.6) it therefore follows that $E[\exp\{i\lambda(\mathcal{Z}_n(h) - \tilde{\mathcal{Z}}(h))\}] \to 1$ for each $\lambda \in \mathcal{R}$. Thus, the characteristic function of $\mathcal{Z}_n(h) - \tilde{\mathcal{Z}}(h)$ converges to one everywhere on $\mathcal{R}$. But that is only possible if $\mathcal{Z}_n(h) - \tilde{\mathcal{Z}}(h) \xrightarrow{d} 0$ by uniqueness of the characteristic function (Billingsley, 1995, Theorem 26.2). Therefore $\mathcal{Z}_n(h) - \tilde{\mathcal{Z}}(h) \xrightarrow{p} 0$ by application of Theorem 25.3 in Billingsley (1995). This proves (A.5) which completes the proof. $\mathcal{QED}$.

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 Hill and Motegi (2018, Appendix F). Recall m_t are the Assumption 2.c estimating equations.

Lemma A.5. *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, Section 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.6. *Let Assumptions 1 and 2.a,b,c′,d hold.*

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

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

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

Proof.

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

$$\rho_n^*(h) \equiv \frac{1}{E\left[\epsilon_t^2\right]}\frac{1}{n}\sum_{t=1+h}^{n}\varphi_t\left\{\mathcal{E}_{t,h} - E\left[\mathcal{E}_{1,h}\right]\right\} \text{ where } \mathcal{E}_{t,h} \equiv \epsilon_t\epsilon_{t-h} - \mathcal{D}(h)'\mathcal{A}m_t. \tag{A.7}$$

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

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

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

$$\left\{\sqrt{n}\rho_n^*(h) : 1 \leq h \leq \mathcal{L}\right\} \Rightarrow^p \left\{\mathring{\mathcal{Z}}(h) : 1 \leq h \leq \mathcal{L}\right\} \tag{A.8}$$

$$\sqrt{n}\left|\hat{\rho}_n^{(dw)}(h) - \rho_n^*(h)\right| \xrightarrow{p} 0 \text{ for each } h, \tag{A.9}$$

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 iid $N(0,1)$ random variables. The general case is similar, where ξ_i are iid, $E[\xi_i] = 0$, $E[\xi_i^2] = 1$ 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.8). Define $\mathbb{L} \equiv \{1, ..., \mathcal{L}\}$. It suffices to prove weak convergence on a Polish space in the sense of Hoffmann-Jørgensen (1984, 1991), cf. Giné and Zinn (1990, p. 853 and Theorem 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; $\{\sqrt{n}\rho_n^*(h) : 1 \leq h \leq \mathcal{L}\} \stackrel{d}{\to} \{\mathring{\mathcal{Z}}(h) : 1 \leq h \leq \mathcal{L}\}$; and the sequence of distributions governing $\{\sqrt{n}\rho_n^*(h)\}_{n\geq 1}$ 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 Theorem 2).

By construction of φ_t via ξ_t:

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

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, $\{\sqrt{n}\rho_n^*(h) : 1 \leq h \leq \mathcal{L}\}$ 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:

$$\begin{aligned}
&\lim_{n\to\infty} E\left[nE_{\mathfrak{X}_n}\left[\rho_n^*(h)\rho_n^*(\tilde{h})\right]\right] \qquad\qquad \text{(A.10)} \\
&= \frac{1}{\left[E\left[\epsilon_t^2\right]\right]^2} \lim_{n\to\infty} \frac{1}{n} \sum_{s=1}^{n/b_n} \sum_{t,u=(s-1)b_n+1+h}^{sb_n} E\left[\left\{\mathcal{E}_{t,h} - E\left[\mathcal{E}_{1,h}\right]\right\}\left\{\mathcal{E}_{u,\tilde{h}} - E\left[\mathcal{E}_{1,\tilde{h}}\right]\right\}\right] \\
&= \frac{1}{\left[E\left[\epsilon_t^2\right]\right]^2} \sum_{i=0}^{\infty} E\left[\left\{\mathcal{E}_{1,h} - E\left[\mathcal{E}_{1,h}\right]\right\}\left\{\mathcal{E}_{1+i,\tilde{h}} - E\left[\mathcal{E}_{1,\tilde{h}}\right]\right\}\right] \\
&= \lim_{n\to\infty} \frac{1}{n} E\left[\sum_{t=1}^{n} \frac{\left(\mathcal{E}_{t,h} - E\left[\mathcal{E}_{t,h}\right]\right)}{E\left[\epsilon_t^2\right]} \sum_{t=1}^{n} \frac{\left(\mathcal{E}_{t,\tilde{h}} - E\left[\mathcal{E}_{t,\tilde{h}}\right]\right)}{E\left[\epsilon_t^2\right]}\right] = E\left[\mathcal{Z}(h)\mathcal{Z}(\tilde{h})\right].
\end{aligned}$$

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

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

$$P\left(\mathfrak{X}_n \in \mathfrak{X}\right) = 1. \qquad \text{(A.11)}$$

In conjunction with (A.10), it then follows that the finite dimensional distributions of $\{\sqrt{n}\rho_n^*(h) : 1 \leq h \leq \mathcal{L}\}$ converge to those of $\{\mathring{\mathcal{Z}}(h) : 1 \leq h \leq \mathcal{L}\}$, where $\{\mathring{\mathcal{Z}}(h) : 1 \leq h \leq \mathcal{L}\}$ 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 $\{\mathring{\mathcal{Z}}(h) : 1 \leq h \leq \mathcal{L}\}$ is an independent copy of $\{\mathcal{Z}(h) : 1 \leq h \leq \mathcal{L}\}$.

Consider (A.11). 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:

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

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

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

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:

$$\frac{1}{n}\sum_{s=1}^{n/b_n} \mathfrak{E}_{n,h}\mathfrak{E}_{n,\tilde{h}} = \frac{1}{n}\sum_{s=1}^{n/b_n} \mathcal{Y}_{n,s}(h)\mathcal{Y}_{n,s}(\tilde{h}) + o_p(1). \tag{A.12}$$

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

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 $O(h^{-p/(p-2)})$. 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

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

This proves (A.12).

Next, de Jong's (1997: Assumption 2) conditions are satisfied under the given NED property. Hence, by the proof of de Jong's (1997) Theorem 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:

$$\frac{1}{n}\sum_{s=1}^{n/b_n}\mathcal{Y}_{n,s}(h)\mathcal{Y}_{n,s}(\tilde{h}) \xrightarrow{p} \lim_{n\to\infty}\frac{1}{n}E\left[\left(\sum_{t=1}^{n}\{\mathcal{E}_{t,h} - E[\mathcal{E}_{1,h}]\}\right)\left(\sum_{t=1}^{n}\left\{\mathcal{E}_{t,\tilde{h}} - E\left[\mathcal{E}_{1,\tilde{h}}\right]\right\}\right)\right]. \tag{A.13}$$

Property (A.11) is proved since combining (A.10), (A.12) and (A.13) 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.9).

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

$$\frac{1}{\sqrt{n}}\sum_{t=1+h}^{n}\varphi_t\left\{\widehat{\mathcal{E}}_{n,t,h}(\hat{\theta}_n) - \frac{1}{n}\sum_{s=1+h}^{n}\widehat{\mathcal{E}}_{n,s,h}(\hat{\theta}_n)\right\} = \frac{1}{\sqrt{n}}\sum_{t=1+h}^{n}\varphi_t\{\mathcal{E}_{t,h} - E[\mathcal{E}_{t,h}]\} + o_p(1) \tag{A.14}$$

by showing (it is straightforward to show (A.15)-(A.18) imply (A.14)):

$$n^{-1/2}\sum\nolimits_{t=1+h}^{n}\varphi_t\epsilon_t(\hat{\theta}_n)\epsilon_{t-h}(\hat{\theta}_n) = n^{-1/2}\sum\nolimits_{t=1+h}^{n}\varphi_t\epsilon_t\epsilon_{t-h} + o_p(1) \tag{A.15}$$

$$\hat{\mathcal{D}}_n(h)'\widehat{\mathcal{A}}_n n^{-1/2}\sum\nolimits_{t=1+h}^{n}\varphi_t m_t(\hat{\theta}_n) = \mathcal{D}(h)'\mathcal{A}n^{-1/2}\sum\nolimits_{t=1+h}^{n}\varphi_t m_t + o_p(1) \tag{A.16}$$

$$n^{-1/2}\sum\nolimits_{t=1+h}^{n}\varphi_t n^{-1}\sum\nolimits_{s=1+h}^{n}\epsilon_s(\hat{\theta}_n)\epsilon_{s-h}(\hat{\theta}_n) = n^{-1/2}\sum\nolimits_{t=1+h}^{n}\varphi_t E[\epsilon_t\epsilon_{t-h}] + o_p(1) \tag{A.17}$$

$$n^{-1/2}\sum\nolimits_{t=1+h}^{n}\varphi_t\hat{\mathcal{D}}_n(h)'\widehat{\mathcal{A}}_n n^{-1}\sum\nolimits_{t=1+h}^{n}m_t(\hat{\theta}_n) = o_p(1). \tag{A.18}$$

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

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

Under Assumptions 1.b and 2.c$'$, (A.1) applies to $\mathcal{E}_{t,h} - E[\mathcal{E}_{t,h}]$ (Davidson, 1994, Theorems 17.8 and 17.9). Hence $E[(1/\sqrt{b_n}\sum_{t=1}^{b_n}\{\mathcal{E}_{t,h} - E[\mathcal{E}_{t,h}]\})^2] = O(1)$, and therefore:

$$n^{-1/2}\sum\nolimits_{t=1+h}^{n}\varphi_t\{\mathcal{E}_{t,h} - E[\mathcal{E}_{t,h}]\} = O_p(1). \tag{A.19}$$

Further, by application of Lemma A.3, $\sqrt{n}\{\hat{\gamma}_n(0) - \gamma(0)\} = n^{-1/2}\sum_{t=1}^{n}\{\epsilon_t^2 - E[\epsilon_t^2] - \mathcal{D}(0)'\mathcal{A}m_t\} + O_p(1/\sqrt{n})$. Coupled with stationarity, ergodicity and square integrability yields:

$$n^{-1}\sum\nolimits_{t=1}^{n}\epsilon_t^2(\hat{\theta}_n) = E\left[\epsilon_t^2\right] + o_p(1). \tag{A.20}$$

Combine (A.14), (A.19) and (A.20) to yield (A.9) as required:

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

Step 2.2 We now prove (A.15)-(A.18). Consider (A.15). Since φ_t is zero mean Gaussian and independent of the sample, the proof of Lemma 2.1 carries over verbatim to show:

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{t=1+h}^{n}\varphi_t\epsilon_t(\hat{\theta}_n)\epsilon_{t-h}(\hat{\theta}_n) &= \frac{1}{\sqrt{n}}\sum_{t=1+h}^{n}\varphi_t\epsilon_t\epsilon_{t-h} - \sqrt{n}\left(\hat{\theta}_n - \theta_0\right)'\frac{1}{n}\sum_{t=1+h}^{n}\varphi_t\left(\epsilon_t s_t + G_t/\sigma_t\right)\epsilon_{t-h} \\ &\quad -\sqrt{n}\left(\hat{\theta}_n - \theta_0\right)'\frac{1}{n}\sum_{t=1+h}^{n}\varphi_t\epsilon_t\left(\epsilon_{t-h}s_{t-h} + \frac{G_{t-h}}{\sigma_{t-h}}\right) + o_p(1).\end{aligned} \tag{A.21}$$

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

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

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:

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

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.21) yields (A.15).

Next, (A.16). By Lemma A.5:

$$\sup_{\theta\in\Theta}\left\|\frac{1}{n}\sum_{t=1}^{n}\varphi_t\frac{\partial}{\partial\theta}m_t(\theta)\right\| \xrightarrow{p} 0 \text{ and } \frac{1}{\sqrt{n}}\sum_{t=1+h}^{n}\varphi_t m_t = O_p(1). \tag{A.22}$$

Now write:

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

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.22):

$$\left\|n^{-1/2}\sum\nolimits_{t=1+h}^{n}\varphi_t\left\{m_t(\hat{\theta}_n) - m_t\right\}\right\| \leq \sqrt{n}\left\|\hat{\theta}_n - \theta_0\right\| \times \sup_{\theta\in\Theta}\left\|n^{-1}\sum\nolimits_{t=1+h}^{n}\varphi_t\frac{\partial}{\partial\theta}m_t(\theta)\right\| \xrightarrow{p} 0.$$

The latter convergence in probability, combined with (A.22), suffice to prove (A.16).

Proceeding to (A.17), first note that

$$n^{-1/2}\sum\nolimits_{t=1}^{n}\varphi_t = b_n n^{-/12}\sum\nolimits_{s=1}^{n/b_n}\xi_s = \sqrt{b_n}\,(n/b_n)^{-1/2}\sum\nolimits_{s=1}^{n/b_n}\xi_s = O_p\left(\sqrt{b_n}\right). \tag{A.23}$$

Second, by equation (F.5) in the proof of Lemma A.3 in Hill and Motegi (2018):

$$\left|\sqrt{n}\hat{\gamma}_n(h) - n^{-1/2}\sum\nolimits_{t=1+h}^{n}\epsilon_t\epsilon_{t-h} + \sqrt{n}\left(\hat{\theta}_n - \theta_0\right)'\mathcal{D}(h)\right| \xrightarrow{p} 0. \tag{A.24}$$

Use (A.24), 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

$$n^{-1/2}\sum\nolimits_{t=1+h}^{n}\varphi_t n^{-1}\sum\nolimits_{s=1+h}^{n}\epsilon_s(\hat{\theta}_n)\epsilon_{s-h}(\hat{\theta}_n) = n^{-1/2}\sum\nolimits_{t=1+h}^{n}\varphi_t n^{-1}\sum\nolimits_{t=1+h}^{n}\epsilon_t\epsilon_{t-h} + O_p\left(1/\sqrt{n/b_n}\right).$$

It remains to show

$$n^{-1/2}\sum\nolimits_{t=1+h}^{n}\varphi_t n^{-1}\sum\nolimits_{t=1+h}^{n}\epsilon_t\epsilon_{t-h} = n^{-1/2}\sum\nolimits_{t=1+h}^{n}\varphi_t E\left[\epsilon_t\epsilon_{t-h}\right] + o_p(1). \tag{A.25}$$

Under Assumptions 1.b, $\epsilon_t\epsilon_{t-h} - E[\epsilon_t\epsilon_{t-h}]$ satisfies (A.1), hence $E[(1/\sqrt{n}\sum_{t=1}^{n}\{\epsilon_t\epsilon_{t-h} - E\left[\epsilon_t\epsilon_{t-h}\right]\})^2] = O(1)$. Further $1/\sqrt{n}\sum_{t=1+h}^{n}\varphi_t = O_p(\sqrt{b_n})$ from (A.23). Hence

$$n^{-1/2}\sum\nolimits_{t=1+h}^{n}\varphi_t n^{-1}\sum\nolimits_{t=1+h}^{n}\left\{\epsilon_t\epsilon_{t-h} - E\left[\epsilon_t\epsilon_{t-h}\right]\right\} = n^{-1/2}\sum\nolimits_{t=1+h}^{n}\varphi_t \times O_p(1/\sqrt{n}) = O_p\left(1/\sqrt{n/b_n}\right).$$

Since $b_n/n \to 0$, (A.25) follows directly.

Finally, for (A.18), 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:

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

By Lemma A.5: $\sup_{\theta\in\Theta}||1/n\sum_{t=1}^{n}(\partial/\partial\theta)m_t(\theta) - E[(\partial/\partial\theta)m_t(\theta)]|| \xrightarrow{p} 0$, 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 $\mathcal{L} \in \mathbb{N}$:

$$\vartheta\left(\left[\sqrt{n}\hat{\omega}_n(h)\hat{\rho}_n^{(dw)}(h)\right]_{h=1}^{\mathcal{L}}\right) \Rightarrow^p \vartheta\left(\left[\omega(h)\mathring{\mathcal{Z}}(h)\right]_{h=1}^{\mathcal{L}}\right). \tag{A.26}$$

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

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

Now apply arguments used to prove Lemma A.2.a in order to yield $\mathcal{A}_{\mathcal{L}_n,n} \xrightarrow{p} 0$ for some sequence of positive integers $\{\mathcal{L}_n\}_{n\geq 1}$, $\mathcal{L}_n \to \infty$ and $\mathcal{L}_n = o(n)$. $\mathcal{QED}$.

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:

$$\hat{p}_{n,M}^{(dw)} \xrightarrow{p} P\left(\vartheta\left(\left[\sqrt{n}\hat{\rho}_n^{(dw)}(h)\right]_{h=1}^{\mathcal{L}_n}\right) \geq \vartheta\left(\left[\sqrt{n}\hat{\rho}_n(h)\right]_{h=1}^{\mathcal{L}_n}\right)|\mathfrak{X}_n\right) \text{ as } M \to \infty. \tag{A.27}$$

Second, by Theorem 2.3 and Lemma A.6:

$$\left|\vartheta\left(\left[\sqrt{n}\left\{\hat{\rho}_n(h) - \rho(h)\right\}\right]_{h=1}^{\mathcal{L}_n}\right) - \vartheta\left(\left[\mathcal{Z}(h)\right]_{h=1}^{\mathcal{L}_n}\right)\right| \xrightarrow{p} 0 \tag{A.28}$$

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

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

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:

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

Together, (A.27), (A.30) and (A.31) yield for any sequence of positive integers $\{M_n\}$, $M_n \rightarrow \infty$:

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

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 $P(\hat{p}_{n,M_n}^{(dw)} < \alpha) = P(\bar{F}_n^{(0)}\left((\vartheta\left[\mathcal{Z}(h)\right]_{h=1}^{\mathcal{L}_n})\right) < \alpha) + o(1) = \alpha + o(1)$ $\rightarrow \alpha$.

Impose $H_1 : \rho(h) \neq 0$ for some $h \in \mathbb{N}$. Recall ϑ satisfies the triangle inequality, and divergence $\vartheta(a) \rightarrow \infty$ as $||a|| \rightarrow \infty$. 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

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

Combine (A.27), (A.29) and (A.33) to deduce $P(\hat{p}_{n,M_n}^{(dw)} < \alpha) \rightarrow 1$ for any $\alpha \in (0,1)$ because:

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

Proof of Theorem 3.1. Let q be any fixed positive constant. Recall $\mathcal{P}_n(\mathcal{L}) = \sqrt{\mathcal{L}\ln n}$ 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 $\{\mathcal{L}_n\}$, $\mathcal{L}_n \rightarrow (0,\infty]$ and $\mathcal{L}_n/\bar{\mathcal{L}}_n \rightarrow [0,1]$, the the penalty term satisfies:

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

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:

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

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

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

Consider (A.34). By construction of $\mathcal{P}_n(\mathcal{L}_n)$ it suffices to prove $P(\hat{\mathcal{T}}_n(\mathcal{L}_n) > \sqrt{q\ln n}) \rightarrow 0$. Under H_0, $\sqrt{n}\hat{\rho}_n(h) = O_p(1)$ by (A.4) and (A.5), hence $\sqrt{n}\hat{\rho}_n(h)/\sqrt{q\ln n} \xrightarrow{p} 0$ for any fixed $q \in (0, \infty)$. Therefore, for any integer sequence $\{\bar{\mathcal{L}}_n\}$, $\bar{\mathcal{L}}_n \rightarrow (0, \infty)$, or by Lemma A.2 for some $\{\bar{\mathcal{L}}_n\}$, $\{\bar{\mathcal{L}}_n\} \rightarrow \infty$:

$$\frac{\hat{\mathcal{T}}_n(\bar{\mathcal{L}}_n)}{\sqrt{q\ln n}} = \frac{\sqrt{n}\max_{1\leq h\leq\bar{\mathcal{L}}_n}|\hat{\rho}_n(h)|}{\sqrt{q\ln n}} \xrightarrow{p} 0. \tag{A.36}$$

By monotonicity of $\hat{\mathcal{T}}_n(\cdot)$, (A.36) holds for any $\{\mathcal{L}_n\}$, $\mathcal{L}_n \rightarrow (0, \infty]$ and $\mathcal{L}_n/\bar{\mathcal{L}}_n \rightarrow [0, 1]$. 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.36), $\hat{\mathcal{T}}_n(\bar{\mathcal{L}}_n)/\sqrt{\ln n} = o_p(1)$ and therefore $\hat{\mathcal{T}}_n(\mathcal{L}_n) - \hat{\mathcal{T}}_n(l) = o_p(\sqrt{\ln(n)})$ for any $\{\mathcal{L}_n\}$, $\mathcal{L}_n \rightarrow (0, \infty]$ and $\mathcal{L}_n/\bar{\mathcal{L}}_n \rightarrow [0, 1]$, 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 as $n \rightarrow \infty$:

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

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

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

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 $\underline{\mathbb{N}}_1 \equiv \min\{\mathbb{N}_1\}$, 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 $\mathcal{L}_n/\bar{\mathcal{L}}_n \to [0,1]$:

$$P\left(\mathcal{P}_n(\mathcal{L}_n) = \sqrt{2\mathcal{L}_n}\right) \to 1. \tag{A.37}$$

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

$$P\left(\hat{\mathcal{T}}_n(\mathcal{L}_n) \geq \hat{\mathcal{T}}_n(l) + 2(\sqrt{\mathcal{L}_n} - \sqrt{l})\right) \to 1 \textit{ for each } 1 \leq l \leq \bar{\mathcal{L}}_n. \tag{A.38}$$

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 assumptoins, 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.37) holds because $h^* \in [\underline{\mathrm{N}}_1, \infty)$, hence $\hat{\mathcal{T}}_n^{\mathcal{P}}(\mathcal{L}_n) \equiv \hat{\mathcal{T}}_n(\mathcal{L}_.) - \sqrt{2\mathcal{L}_n}$ $ap1$. Since such a sequence implies (A.38), we have $\hat{\mathcal{T}}_n^{\mathcal{P}}(\mathcal{L}_n) \geq \hat{\mathcal{T}}_n^{\mathcal{P}}(l)$ $ap1$ for each $l = 1, ..., \bar{\mathcal{L}}_n$. Conversely, if (A.38) holds then $\mathcal{L}_n/h_n^* \xrightarrow{p} [1,\infty]$. This yields (A.37) because $h^* \in [\underline{\mathrm{N}}_1, \infty)$. Therefore $\hat{\mathcal{T}}_n^{\mathcal{P}}(\mathcal{L}_n) \geq \hat{\mathcal{T}}_n^{\mathcal{P}}(l)$ $ap1$ 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.37). Use (A.4) and (A.5) to deduce $\hat{\rho}_n(h) - \rho(h) \xrightarrow{p} 0$ for each h. Lemma A.2 therefore yields for some integer sequence $\{\bar{\mathcal{L}}_n\}$, $\bar{\mathcal{L}}_n \to \infty$:

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

where $\lim_{n\to\infty} \max_{1 \leq h \leq \bar{\mathcal{L}}_n} |\rho(h)| \in (0,\infty)$. 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:

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

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

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

This proves (A.37) by construction (11) of the penalty term $\mathcal{P}_n(\mathcal{L}_n)$.

Step 2: Next we prove (A.38). First note that by (A.4) and (A.5) $\hat{\mathcal{T}}_n(\mathcal{L}_n)/\sqrt{n} \xrightarrow{p} (0,1)$ 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 $\hat{\mathcal{T}}_n(\mathcal{L}_n) \geq \hat{\mathcal{T}}_n(l)$ for each $l \leq \mathcal{L}_n$, 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 $(l, \mathcal{L}_n) \geq h_n^* \equiv \min\{h_n : h_n = \arg\max_{1 \leq h \leq \bar{\mathcal{L}}_n} |\hat{\rho}_n(h)|\}$ then by construction $\hat{\mathcal{T}}_n(l)/\hat{\mathcal{T}}_n(\mathcal{L}_n) = 1$.

Now suppose $1 \leq l$ and $l/\mathcal{L}_n \to [0,1)$, and $\mathcal{L}_n/h_n^* \xrightarrow{p} [0,1)$, hence $1 \leq l < \mathcal{L}_n < h_n^*$ 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 $\mathcal{L}_n \leq \bar{\mathcal{L}}_n = o(n/\ln(n))$ and $\hat{\mathcal{T}}_n(\mathcal{L}_n)/\sqrt{n/\ln(n)} \xrightarrow{p} \infty$ to yield:

$$
\begin{aligned}
P\left(\hat{\mathcal{T}}_n(\mathcal{L}_n) \geq \hat{\mathcal{T}}_n(l) + 2\left(\sqrt{\mathcal{L}_n} - \sqrt{l}\right)\right) &= P\left(\hat{\mathcal{T}}_n(\mathcal{L}_n)\left(1 - \frac{\hat{\mathcal{T}}_n(l)}{\hat{\mathcal{T}}_n(\mathcal{L}_n)}\right) \geq 2\sqrt{\mathcal{L}_n}\left(1 - \sqrt{\frac{l}{\mathcal{L}_n}}\right)\right) \quad \text{(A.39)} \\
&\geq P\left(\frac{\hat{\mathcal{T}}_n(\mathcal{L}_n)}{\sqrt{n/\ln(n)}}\left(1 - \frac{\hat{\mathcal{T}}_n(l)}{\hat{\mathcal{T}}_n(\mathcal{L}_n)}\right) \geq 2\sqrt{\frac{\mathcal{L}_n}{n/\ln(n)}}\right) \rightarrow 1.
\end{aligned}
$$

Next, consider $1 \leq l$ and $l/h_n^* \xrightarrow{p} [0,1)$, and $\mathcal{L}_n/h_n^* \xrightarrow{p} [1,\infty]$, hence $1 \leq l \leq h_n^* - 1$ *ap*1 and $\mathcal{L}_n \geq h_n^*$ *ap*1. Then $P(\hat{\mathcal{T}}_n(l) = \hat{\mathcal{T}}_n(\mathcal{L}_n) \rightarrow 0$ since by contruction 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.39).

Now let $(l, \mathcal{L}_n) \geq h_n^*$ *ap*1. Then by construction $\hat{\mathcal{T}}_n(\mathcal{L}_n) = \hat{\mathcal{T}}_n(l)$ *ap*1. Trivially if $l < \mathcal{L}_n$ ($l \geq \mathcal{L}_n$) then $\sqrt{\mathcal{L}_n} - \sqrt{l} > 0$ ($\sqrt{\mathcal{L}_n} - \sqrt{l} \leq 0$). Hence $P(\hat{\mathcal{T}}_n(\mathcal{L}_n) \geq \hat{\mathcal{T}}_n(l) + 2[\sqrt{\mathcal{L}_n} - \sqrt{l}]) \rightarrow 1$ *if and only if* $l \geq \mathcal{L}_n$.

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

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

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 $P(\hat{\mathcal{T}}_n(\mathcal{L}_n) \geq \hat{\mathcal{T}}_n(l) + 2(\sqrt{\mathcal{L}_n} - \sqrt{l})) = P(0 \geq 2(\sqrt{\mathcal{L}_n} - \sqrt{l})) \rightarrow 1$ *if and only if* $l \geq \mathcal{L}_n$.

Combining the above results, we deduce $P(\hat{\mathcal{T}}_n(\mathcal{L}_n) \geq \hat{\mathcal{T}}_n(l) + 2[\sqrt{\mathcal{L}_n} - \sqrt{l}]) \rightarrow 1$ *for every* $1 \leq l \leq \bar{\mathcal{L}}_n$ *if and only if* $\mathcal{L}_n \geq h_n^*$, proving (A.38). $\mathcal{QED}$.

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., and 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., and 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*. Wiley-Interscience, 3 edn.

Boehme, T. K., and 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., and 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.

Burkill, H., and L. Mirsky (1973): "Monotonicity," *Journal of Mathematical Analysis and Applications*, 41(391-410).

Carrasco, M., and X. Chen (2002): "Mixing and Moment Properties of Various GARCH and Stochastic Volatility Models," *Econometric Theory*, 18, 17–39.

Chernozhukov, V., D. Chetverikov, and K. Kato (2013): "Gaussian Approximations and Multiplier Bootstrap for Maxima of Sums of High-Dimensional Random Vectors," *Annals of Statistics*, 41, 2786–2819.

——— (2014): "Testing Many Moment Inequalities," arXiv:1312.7614.

——— (2015): "Comparison and Anti-Concentration Bounds for Maxima of Gaussian Random Vectors," *Probability Theory and Related Fields*, 162, 47–70.

——— (2017): "Central Limit Theorems and Bootstrap in High Dimensions," *Annals of Probability*, 45, 2309–2352.

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

de Jong, R. M. (1997): "Central Limit Theorems for Dependent Heterogeneous Random Variables," *Econometric Theory*, 13, 353–367.

Delgado, M. A., J. Hidalgo, and C. Velasco (2005): "Distribution Free Goodness-of-Fit Tests for Linear Processes," *Annals of Statistics*, 33, 2568–2609.

Delgado, M. A., and 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., and T. E. Nijman (1993): "Temporal Aggregation of GARCH Processes," *Econometrica*, 61, 909–927.

Dudley, R. M. (1978): "Central Limit Theorems for Empirical Measures," *Annals of Probability*, 6, 899–929.

——— (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., and 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.

Erdös, P., and G. Szekeres (1935): "A combinatorial problem in geometry," *Compositio Mathematica*, 2, 463–470.

Escanciano, J. C., and I. N. Lobato (2009): "An Automatic Portmanteau Test for Serial Correlation," *Journal of Econometrics*, 151, 140–149.

Francq, C., and J.-M. Zakoïan (2004): "Maximum Likelihood Estimation of Pure GARCH and ARMA-GARCH Processes," *Bernoulli*, 10, 605–637.

——— (2010): *GARCH Models: Structure, Statistical Inference and Financial Applications*. John Wiley & Sons, UK.

Gaenssler, P., and K. Ziegler (1994): "A Uniform Law of Large Numbers for Set-Indexed Processes with Applications to Empirical and Partial-Sum Processes," in *Probability in Banach Spaces*, ed. by J. Hoffmann-Jørgensen, J. Kuelbs, and M. B. Marcus, vol. 9, pp. 385–400. Birkhauser, Boston.

Giné, E., and J. Zinn (1990): "Bootstrapping General Empirical Measures," *Annals of Probability*, 18, 851–869.

Grenander, U., and M. Rosenblatt (1952): "On Spectral Analysis of Stationary Time Series," in *Proceedings of the National Academy of Sciences of the United States of America*, vol. 38, pp. 519–521.

Guay, A., E. Guerre, and 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.

Hill, J. B., and K. Motegi (2018): "Supplemental Material for "A Max-Correlation White Noise Test for Weakly Dependent Time Series"," Dept. of Economics, University of North Carolina at Chapel Hill.

Hoffmann-Jørgensen, J. (1984): "Convergence of Stochastic Processes on Polish Spaces," mimeo.

——— (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.

——— (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., and 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., and Y. A. Rozanov (1960): "On Strong Mixing Conditions for Stationarity Gaussian Processes," *Theory of Probability and its Applications*, 5.

Kuan, C.-M., and 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., and 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.

Leadbetter, M. R., G. Lindgren, and H. Rootzén (1983): *Extremes and Related Properties of Random Sequences and Processes*, Springer Series in Statistics. Springer-Verlag, New York.

Lee, S.-W., and B. E. Hansen (1994): "Asymptotic Theory for the GARCH(1,1) Quasi-Maximum Likelihood Estimator," *Econometric Theory*, 10, 29–52.

Liu, R. Y. (1988): "Bootstrap Procedures under some Non-I.I.D. Models," *Annals of Statistics*, 16, 1696–1708.

Ljung, G. M., and 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, and 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., and 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(2), R4.

Nankervis, J. C., and N. E. Savin (2010): "Testing for Serial Correlation: Generalized Andrews-Ploberger Tests," *Journal of Business & Economic Statistics*, 28, 246–255.

——— (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): "Smooth Test for Goodness of Fit," *Scandinavian Aktuarietidskr*, 20, 149–199.

Pakes, A., and D. Pollard (1989): "Simulation and the Asymptotics of Optimization Estimators," *Econometrica*, 57, 1027–1057.

Politis, D. N., and 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.

——— (1990): "Empirical Processes: Theory and Applications," *NSF-CBMS Regional Conference Series in Probability and Statistics*, 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., and L. A. Thombs (1996): "Inference for Autocorrelations under Weak Assumptions," *Journal of the American Statistical Association*, 91, 590–600.

Royden, H. L. (1988): *Real Analysis*. Prentise Hall, 3rd edn.

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.

——— (2011): "A Bootstrap-Assisted Spectral Test of White Noise under Unknown Dependence," *Journal of Econometrics*, 162, 213–224.

Shao, X., and W. B. Wu (2007): "Asymptotic Spectral Theory for Nonlinear Time Series," *Annals of Statistics*, 35, 1773–1801.

Straumann, D., and T. Mikosch (2006): "Quasi-Maximum-Likelihood Estimation in Conditionally Heteroscedastic Time Series: A Stochastic Recurrence Equations Approach," *Annals of Statistics*, 34.

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," in *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, pp. 14150–14154.

Wu, W. B., and W. Min (2005): "On Linear Processes with Dependent Innovations," *Stochastic Processes and their Applications*, 115, 939–958.

Xiao, H., and 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.

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., and W. K. Li (2015): "A Bootstrapped Spectral Test for Adequacy in Weak ARMA Models," *Journal of Econometrics*, 187, 113–130.

Table 2: Median of Automatic Lags $\mathcal{L}_n^*$

e_t	IID	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	$\{1, 1, 1, 1\}$	$\{1, 1, 1, 1\}$	$\{1, 1, 1, 1\}$	$\{1, 1, 2, 2\}$
	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, $h^* = 24$	-	-	-

#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 IID, GARCH(1,1), MA(2), or AR(1) in Scenarios #1–#6, while it is IID in Scenarios #7–#9. We report the median of automatic lags for actual test statistics, $\mathcal{L}_n^*$, across $J = 1000$ Monte Carlo samples. The largest possible lag length is $\bar{\mathcal{L}}_n = [1.5 \times n/(\ln n)^{4/3}]$. The tuning parameter that affects the penalty term $\mathcal{P}_n(\mathcal{L})$ is $q = 3.25$. H_0 implies the test variable $\{\epsilon_t\}$ is white noise, while H_1 implies ϵ_t is serially correlated.

The smallest lag at which the largest correlation occurs, h^*, is recorded if it can be computed analytically. Otherwise, we report a simulation based $\hat{h}^*$. We use $J = 50,000$ Monte Carlo samples of size $n = 50,000$, and compute sample autocorrelations of $\{\epsilon_t\}$ at $h = 1, \ldots, 20$. The smallest lag at which the largest correlation occurs for the j^{th} sample is $\hat{h}_j^*$, and the reported $\hat{h}^*$ is the median of $\{\hat{h}_1^*, \ldots, \hat{h}_J^*\}$.

Table 3: Rejection Frequencies of Max-Correlation Test with Automatic Lag (Scenarios #1–#6)

IID Error: $e_t = \nu_t$						
	#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	.011, .058, .108	.009, .060, .107	.010, .067, .153	.060, .236, .339	.006, .039, .075	.024, .079, .130
250	.010, .050, .101	.013, .054, .100	.001, .045, .096	.192, .514, .655	.011, .032, .066	.016, .063, .115
500	.010, .050, .088	.008, .030, .068	.004, .038, .087	.566, .851, .909	.009, .040, .081	.006, .052, .082
1000	.008, .058, .103	.011, .041, .076	.009, .049, .099	.935, .990, .993	.010, .052, .093	.013, .054, .103

GARCH(1,1) Error: $e_t = \nu_t w_t$ 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	.005, .035, .078	.017, .026, .053	.012, .064, .138	.043, .155, .267	.000, .003, .008	.023, .084, .138
250	.005, .028, .066	.017, .035, .050	.004, .035, .096	.107, .342, .482	.001, .006, .010	.013, .055, .106
500	.006, .028, .073	.018, .026, .037	.003, .036, .088	.329, .620, .748	.001, .001, .005	.014, .058, .105
1000	.006, .042, .091	.013, .022, .026	.009, .042, .076	.747, .929, .967	.002, .002, .002	.015, .058, .106

MA(2) Error: $e_t = \nu_t + 0.5\nu_{t-1} + 0.25\nu_{t-2}$						
	#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	.696, .911, .954	.566, .758, .839	.014, .076, .148	.241, .621, .780	.469, .692, .783	.901, .970, .984
250	.993, 1.00, 1.00	.851, .932, .966	.005, .052, .101	.698, .968, .991	.701, .830, .872	.990, .991, .992
500	1.00, 1.00, 1.00	.911, .960, .973	.020, .101, .165	.980, 1.00, 1.00	.838, .893, .912	1.00, 1.00, 1.00
1000	1.00, 1.00, 1.00	.983, .988, .992	.070, .166, .243	1.00, 1.00, 1.00	.879, .927, .949	1.00, 1.00, 1.00

AR(1) Error: $e_t = 0.7e_{t-1} + \nu_t$						
	#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	.496, .756, .848	.507, .646, .697	.029, .132, .227	.299, .641, .786	.205, .338, .423	.988, .989, .990
250	.882, .976, .993	.686, .784, .829	.075, .272, .391	.780, .962, .994	.167, .267, .326	.999, .999, .999
500	.997, 1.00, 1.00	.731, .837, .884	.221, .517, .648	.995, 1.00, 1.00	.106, .175, .218	1.00, 1.00, 1.00
1000	1.00, 1.00, 1.00	.732, .822, .856	.616, .841, .913	1.00, 1.00, 1.00	.074, .123, .161	1.00, 1.00, 1.00

Scenario #1: Simple $y_t = e_t$ with a mean filter. Scenario #2: Bilinear $y_t = 0.5e_{t-1}y_{t-2} + e_t$ with a mean filter. Scenario #3: AR(2) $y_t = 0.3y_{t-1} - 0.15y_{t-2} + e_t$ with an AR(2) filter. Scenario #4: AR(2) $y_t = 0.3y_{t-1} - 0.15y_{t-2} + e_t$ with an AR(1) filter. Scenario #5: GARCH(1,1) $y_t = \sigma_t e_t$, $\sigma_t^2 = 1 + 0.2y_{t-1}^2 + 0.5\sigma_{t-1}^2$ without a filter. Scenario #6: GARCH(1,1) $y_t = \sigma_t e_t$, $\sigma_t^2 = 1 + 0.2y_{t-1}^2 + 0.5\sigma_{t-1}^2$ with a GARCH filter. For each scenario, $\nu_t \overset{i.i.d.}{\sim} N(0,1)$. Actual and bootstrapped test statistics are based on their own automatic lag lengths. The largest possible lag length is $\bar{\mathcal{L}}_n = [1.5 \times n/(\ln n)^{4/3}]$, and the tuning parameter that affects the penalty term $\mathcal{P}_n(\mathcal{L})$ is $q = 3.25$. We report rejection frequencies with respect to nominal size $\alpha \in \{0.01, 0.05, 0.10\}$ across $J = 1000$ Monte Carlo samples. *wo* denotes *without a filter*, and *w* denotes *with a filter*.

Table 4: Rejection Frequencies of Cramér-von Mises Test CvM^{dw} in Scenarios #1–#6

IID Error: $e_t = \nu_t$						
	#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	.023, .081, .138	.018, .076, .149	.020, .086, .167	.133, .338, .483	.021, .077, .141	.034, .087, .144
250	.016, .072, .144	.030, .085, .154	.011, .065, .127	.370, .615, .735	.011, .058, .118	.019, .065, .112
500	.010, .051, .102	.014, .072, .124	.012, .059, .132	.710, .882, .939	.009, .053, .103	.016, .072, .141
1000	.008, .060, .108	.016, .063, .106	.010, .049, .102	.974, .991, .993	.015, .058, .107	.013, .057, .103

GARCH(1,1) Error: $e_t = \nu_t w_t$ 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	.017, .081, .149	.002, .030, .070	.026, .086, .168	.118, .287, .430	.006, .049, .103	.036, .100, .168
250	.013, .059, .108	.029, .048, .083	.012, .058, .127	.242, .501, .648	.009, .037, .080	.020, .075, .132
500	.015, .066, .115	.026, .038, .075	.011, .051, .104	.550, .802, .881	.013, .052, .111	.026, .072, .143
1000	.010, .060, .116	.004, .014, .028	.008, .056, .105	.880, .973, .993	.006, .032, .065	.049, .065, .073

MA(2) Error: $e_t = \nu_t + 0.5\nu_{t-1} + 0.25\nu_{t-2}$						
	#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	.898, .984, .995	.450, .743, .866	.029, .113, .182	.570, .805, .898	.681, .908, .969	.878, .927, .940
250	.999, 1.00, 1.00	.769, .924, .968	.019, .086, .189	.951, .996, .999	.903, .979, .994	.983, .989, .991
500	1.00, 1.00, 1.00	.884, .966, .990	.032, .144, .250	1.00, 1.00, 1.00	.959, .994, .995	.995, .998, .998
1000	1.00, 1.00, 1.00	.974, .994, .997	.068, .295, .471	1.00, 1.00, 1.00	.986, .997, 1.00	.998, .998, .998

AR(1) Error: $e_t = 0.7e_{t-1} + \nu_t$						
	#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	.925, .996, 1.00	.282, .567, .741	.064, .193, .299	.472, .741, .849	.564, .818, .923	.958, .970, .973
250	.999, 1.00, 1.00	.341, .572, .718	.136, .341, .465	.935, .991, .999	.680, .849, .912	.984, .987, .988
500	1.00, 1.00, 1.00	.393, .630, .781	.325, .592, .700	.999, 1.00, 1.00	.700, .852, .918	.999, 1.00, 1.00
1000	1.00, 1.00, 1.00	.474, .697, .810	.688, .876, .923	1.00, 1.00, 1.00	.750, .877, .929	.998, .999, .999

Scenario #1: Simple $y_t = e_t$ with a mean filter. Scenario #2: Bilinear $y_t = 0.5e_{t-1}y_{t-2} + e_t$ with a mean filter. Scenario #3: AR(2) $y_t = 0.3y_{t-1} - 0.15y_{t-2} + e_t$ with an AR(2) filter. Scenario #4: AR(2) $y_t = 0.3y_{t-1} - 0.15y_{t-2} + e_t$ with an AR(1) filter. Scenario #5: GARCH(1,1) $y_t = \sigma_t e_t$, $\sigma_t^2 = 1 + 0.2y_{t-1}^2 + 0.5\sigma_{t-1}^2$ without a filter. Scenario #6: GARCH(1,1) $y_t = \sigma_t e_t$, $\sigma_t^2 = 1 + 0.2y_{t-1}^2 + 0.5\sigma_{t-1}^2$ with a GARCH filter. For each scenario, $\nu_t \overset{i.i.d.}{\sim} N(0,1)$. The dependent wild bootstrap is used to compute an approximate p-value. All $\mathcal{L}_n = n - 1$ lags are used by construction. We report rejection frequencies with respect to nominal size $\alpha \in \{0.01, 0.05, 0.10\}$ across $J = 1000$ Monte Carlo samples. *wo* denotes *without a filter*, and *w* denotes *with a filter*.

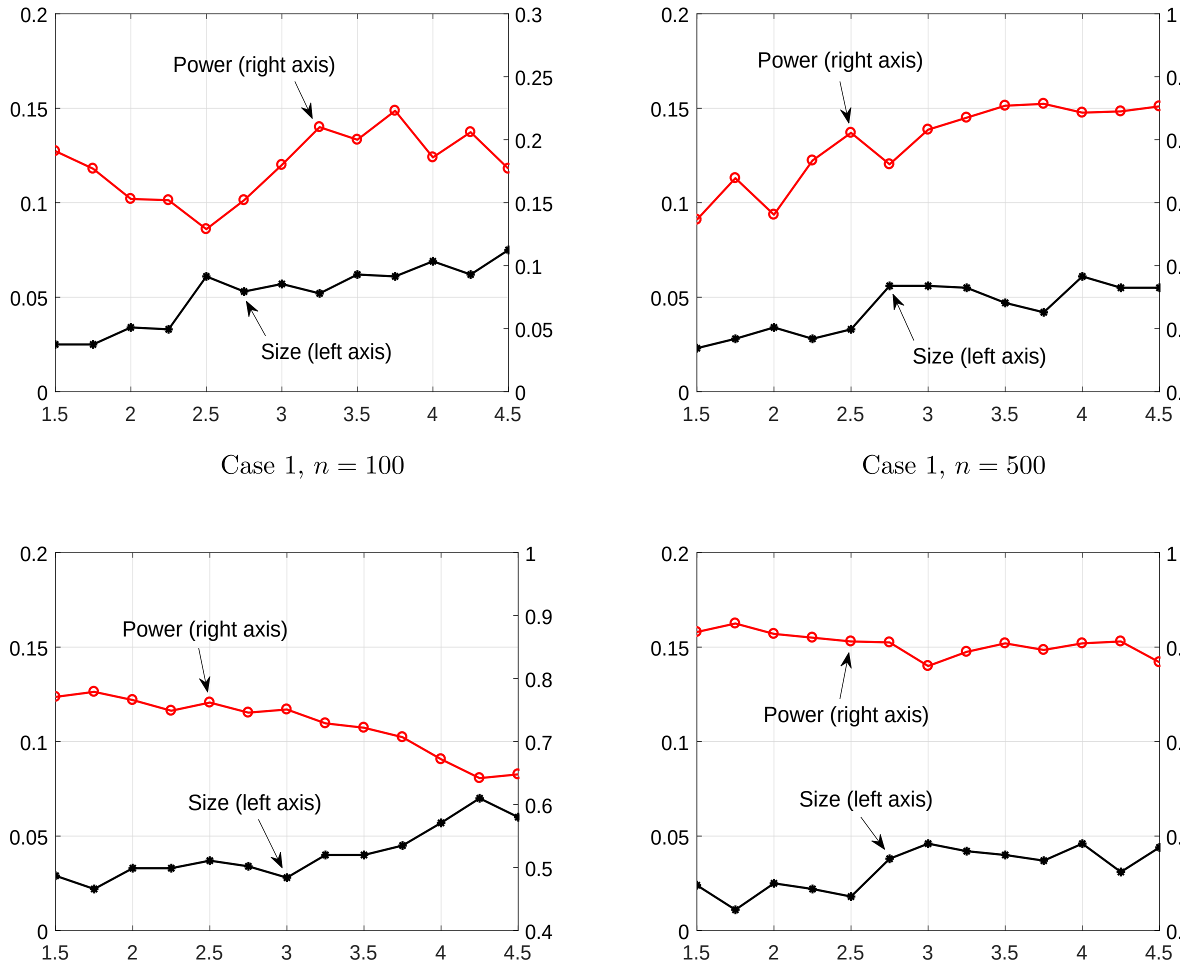
Table 5: Rejection Frequencies in Scenarios #7–#9

Max-Correlation Test with Automatic 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	19	.014	.066	.127	.019	.077	.132	.020	.083	.144
250	38	.159	.261	.315	.029	.151	.249	.016	.068	.117
500	65	.685	.746	.759	.377	.652	.717	.020	.112	.193
1000	114	.998	.998	.999	.985	.993	.993	.615	.864	.920

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	.040	.098	.171	.034	.110	.179	.029	.098	.186
250	249	.026	.080	.142	.025	.087	.155	.022	.088	.143
500	499	.014	.087	.175	.026	.092	.161	.024	.071	.133
1000	999	.038	.160	.320	.017	.083	.166	.028	.079	.144

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, actual and bootstrapped test statistics are based on their own automatic lags with $\bar{\mathcal{L}}_n = [1.5 \times n/(\ln n)^{4/3}]$ and $q = 3.25$. 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 $J = 1000$ Monte Carlo samples.

Figure 1: Empirical Size and Size-Adjusted Power of $\hat{\mathcal{T}}^{dw}(\mathcal{L}_n^*)$ with $\alpha = 0.05$



Case 1, $n = 100$ Case 1, $n = 500$

Case 2, $n = 100$ Case 2, $n = 500$

We plot empirical size and size-adjusted power of the bootstrapped max-correlation test with automatic lag selection with respect to nominal size 5%. In Case 1, the empirical size and empirical quantiles for size adjustment are computed under Scenario #1 (iid 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 $\bar{\mathcal{L}}_n = [1.5 \times n/(\ln n)^{4/3}]$, which implies that $\bar{\mathcal{L}}_{100} = 19$ and $\bar{\mathcal{L}}_{500} = 65$. We generate $J = 1000$ Monte Carlo samples and $M = 500$ bootstrap samples.