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

Jonathan B. Hill † and Kaji Motegi‡
University of North Carolina — Waseda University

May 3, 2016

Abstract

This paper presents bootstrapped p-value white noise tests based on the max-correlation, for a time series that may be weakly dependent under the null hypothesis. The time series may be prefiltered residuals based on a $\sqrt{n}$-convergent estimator. Our test statistic is a scaled maximum sample correlation coefficient $\sqrt{n} \max_{1 \leq h \leq \mathcal{L}_n} |\hat{\rho}_n(h)|$ where 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, and exploit two wild bootstrap methods for p-value computation. We operate either on a first order expansion of the sample correlation, or Delgado and Velasco's (2011) orthogonalized correlation for fixed $\mathcal{L}_n = \mathcal{L}$, both to control for the impact of residual estimation. A numerical study shows the first order expansion is superior, especially when $\mathcal{L}$ is large. When the filter involves a GARCH model then the orthogonalization breaks down, while the first order expansion works quite well. We show Shao's (2011) dependent wild bootstrap is valid for a much larger class of processes than originally considered. Since only the most relevant sample serial correlation is exploited amongst a set of sample correlations that are consistent asymptotically, empirical size tends to be sharp and power is comparatively large for many time series processes. The test has non-trivial local power against $\sqrt{n}$-local alternatives, and can detect very weak and distant serial dependence better than a variety of other tests. Finally, we prove that our bootstrapped p-value leads to a valid test without exploiting extreme value theoretic arguments, the standard in the literature.

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

1 Introduction

We present a bootstrap white noise test based on the maximum serial correlation. The data may be observed, or filtered residuals, and a large class of dependent processes and estimators (for computing

*We thank Hideki Konishi, Masahiro Tanaka, Kozo Ueda, and Ming Yang for helpful comments. We also thank participants at the 10th Spring Meeting of Japan Statistical Society held at Tohoku University in March 2016.

†Department of Economics, University of North Carolina at Chapel Hill. E-mail: jbhill@email.unc.edu

‡Faculty of Political Science and Economics, Waseda University. E-mail: motegi@aoni.waseda.jp

regression residuals) are allowed. A new asymptotic theory approach is used relative to the literature, one that sidesteps deriving the asymptotic distribution of a max-correlation statistic (e.g. Xiao and Wu, 2014). We instead operate solely on the bootstrapped p-value and exploit weak convergence of the correlation process to prove validity and consistency (asymptotic power of one) of the test. Our theory for a dependent wild bootstrap is more general than Shao's (2011), covering a large array of mixing and non-mixing processes.

The class of time series models considered here are

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

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

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

Model (1) is assumed correct in some sense, whether H_0 is true or not. Thus, θ_0 should be thought of as a pseudo-true value that can be identified, usually by a moment condition (Kullback and Leibler, 1951, Sawa, 1978, White, 1982). We ignore the possibility of nuisance parameters that arise under either hypothesis, including ARMA models with common roots, and GARCH volatility with a particular start condition (Andrews and Ploberger, 1996, Andrews, 2001).

Unless $y_t = \epsilon_t$, 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 coefficients:

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

In the pure volatility model set $f(x_{t-1}, \hat{\phi}_n) = 0$, and in the level model set $\sigma_t(\hat{\theta}_n) = 1$. Our primary test statistic is the sample *max-correlation*,

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

Note that $\{\mathcal{L}_n\}$ is a sequence of integers, $\mathcal{L}_n \to \infty$ as $n \to \infty$ ensures a white noise test, and $\mathcal{L}_n = o(n)$ implies $\hat{\gamma}_n(h) = E[\epsilon_t \epsilon_{t-h}] + O_p(1/\sqrt{n})$ for each $h \in \{1, ..., \mathcal{L}_n\}$. The max-correlation delivers only the most informative sample correlation (amongst a set of consistent sample correlations), and therefore ignores all others as though they were simply zeros, much like a shrinkage estimator.

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 Type I extreme value distribution (i.e. Gumbel), 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 tied to the bound on $\mathcal{L}_n$, and in terms of asymptotics it is more restrictive than the Near Epoch Dependence [NED] property used here. See the supplemental material Hill and Motegi (2016, Appendix B) for a comparison between our NED property and their dependence assumption. We only need $\mathcal{L}_n = o(n)$, we allow for a very large class of regression residuals, and we provide a complete theory for the dependent wild bootstrap.

We also use a different asymptotic theory approach. We sidestep Xiao and Wu's (2014) popular extreme value theoretic approach by exploiting only the weak convergence of $\{\sqrt{n}(\hat{\gamma}_n(h) - \gamma(h)) : 1 \leq h \leq \mathcal{L}\}$ to a Gaussian process, for each $\mathcal{L} \in \mathbb{N}$. This suffices since we do not need to have the null extreme value limit distribution of a suitably normalized $\hat{\mathcal{T}}_n$ in order to show that the bootstrapped test statistic and $\hat{\mathcal{T}}_n$ have the same limit distribution function under H_0.

We perform a bootstrap p-value test using Shao's (2011) dependent wild bootstrap. In order to control for the impact of residuals estimation, we either apply the bootstrap to a first order expansion of the sample correlations, or we use Delgado and Velasco's (2011) orthogonal correlation transform which requires a fixed maximum lag $\mathcal{L}$. In general, the first order expansion is superior to the orthogonal transform, both in terms of empirical size and power, in particular when the maximum lag is large. In a numerical study we find the max-correlation test dominates the max-transformed correlation test, and Delgado and Velasco's (2011) Q-test. This is logical since the orthogonal transform requires standardizing the vector $[\hat{\rho}_n(1), ..., \hat{\rho}_n(\mathcal{L})]'$ with a robust variance matrix estimator, while nonparametric estimators are sensitive to bandwidth and are a major source of sampling error when $\mathcal{L}$ is large. Indeed, the orthogonal transform breaks down entirely when we test GARCH filtered residuals for a GARCH process: the orthogonalized correlations are minute, irrespective of serial dependence, the maximum lag used, and whether an identity matrix or kernel estimator are used for variance computation (in the latter case, irrespective of the bandwidth). Thus, both max-transformed correlation and Delgado and Velasco Q-tests yield zero empirical size and power in this case, revealing a non-trivial shortcoming of the test statistic.

Of separate interest, we use a different asymptotic theory than Shao (2011) to show that his dependent wild bootstrap is valid for either the correlation first order expansion or orthogonal transform, in a potentially much less restrictive environment than treated in Shao (2011) or recently Zhu (2015).

The asymptotic arguments are not trivial since we want to impose little more than white noise $E[\epsilon_t \epsilon_{t-h}] = 0 \ \forall h \geq 1$ under the null. Similar challenges are tackled in Hong (1996), Romano and Thombs (1996), Shao (2011), and Guay, Guerre, and Lazarova (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 plug-in estimators than the latter. NED encompasses mixing and non-mixing processes, hence our setting is more general than Zhu's (2015) setting for his block-wise random weighting bootstrap.

Shao (2011), Guay, Guerre, and Lazarova (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. Guay, Guerre, and Lazarova (2013) require a finite 12^{th} moment, hyperbolic memory decay that is fourteen times faster than we require, and impose a joint weak limit theory for the residuals process and plug-estimator that holds for linear processes and least squares. Shao (2011) requires a finite 8^{th} moment, a complicated eighth order cumulant summability 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, and restrict weak dependence based on how fast the maximum lag increases. In the supplemental material Hill and Motegi (2016, Appendix B) we show that our NED setting is potentially far more general then the moment contraction properties employed in Guay, Guerre, and Lazarova (2013) and Shao (2011), and allows for slower moment decay than Xiao and Wu (2014).

Test statistics that combine serial correlations as tests of serial uncorrelatedness, or more generally white noise, have a vast history. The profoundly popular sum of squared correlations up to a fixed maximum lag is due to Box and Pierce (1970). There are many generalizations of the resulting Q-statistic. This includes letting the maximum lag increase with the sample size (Hong, 1996, Hong and Lee, 2003); bootstrapping for size correction under weak dependence (Romano and Thombs, 1996, Horowitz, Lobato, Nankervis, and Savin, 2006, Zhu, 2015); re-scaling for size correction under weak dependence (Lobato, 2001, Francq, Roy, and Zakoïan, 2005, Kuan and Lee, 2006); using a Lagrange Multiplier type statistic to account for weak dependence (e.g. Lobato, Nankervis, and Savin, 2002); exploiting an asymptotic expansion and then projection of sample correlations to produce pivotal statistics (Delgado and Velasco, 2011); and using endogenous maximum lag selection (Escanciano and Lobato, 2009, Guay, Guerre, and Lazarova, 2013).

A related class estimators exploits the periodogram, an increasing sum of sample correlations, dating to Grenander and Rosenblatt (1952) (e.g. Durlauf, 1991, Hong, 1996, Deo, 2000, Delgado, Hidalgo, and Velasco, 2005, Hong and Lee, 2005, Escanciano and Velasco, 2006, 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 omnibus type tests, leading to $\sqrt{n}$-local power. Nevertheless, Guay, Guerre, and Lazarova (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 $\sqrt{n}$, e.g. $\rho(h) = \rho_n(h) = o(n^{-1/2})$, provided an adaptive increasing maximum lag is used. Finally, a weighted sum of correlations also arises in Andrews and Ploberger (1996) sup-LM test couched in an ARMA model (cf. Nankervis and Savin, 2010, 2012).

In all of these cases, multiple or all possible sample correlations are combined. Sample correlations at large lags may be a poor approximation of the true correlation. Bootstrapping, therefore begs the question since at best it yields a better approximation of the finite sample distribution of a combination of possibly noisy estimates.[1]

Our test statistic can operate on residuals from a general regression model, and large class of estimators. It does not require a covariance matrix scale, nor any type of standardization since we exploit a dependent wild bootstrap. Thus, our test is generally not pivotal, as compared to Lobato (2001) and Delgado and Velasco (2011). Lobato (2001) replaces a covariance estimator with a stochastic process, leading to an asymptotically pivotal test, where critical values are computed by simulation. The maximum lag, however, is fixed and only observed data are used. See Kuan and Lee (2006) for extension to account for residuals. Delgado and Velasco (2011) use a least squares projection method for the sample covariances to yield a Q-statistic that is pivotal, but must therefore have a bounded maximum lag.

A simulation study shows that our test works well for unfiltered or filtered data from various time series models, with sample sizes $\{100, 250, 500, 1000\}$ and lags $\mathcal{L}_n$ ranging from 5 to 144 depending on n. The errors are variously iid, martingale difference sequences [mds] and uncorrelated non-mds processes. Our test achieves sharp empirical size and high power for all sample sizes when the wild bootstrap is used, while the dependent wild bootstrap generally results in slightly smaller rejection frequencies. We compare our test with Hong's (1996) standardized spectral test; Delgado and Velasco's (2011) Q-test based on orthogonalized correlations; Shao's (2011) dependent wild bootstrap spectral Cramér-von Mises test, which is proposed for observed data; and Zhu and Li's (2015) block-wise random weighting bootstrap Cramér-von Mises test for linear regression residuals. Only Delgado and Velasco's (2011) Q-test requires a finite maximum lag. In the supplemental material (Hill and Motegi, 2016, Appendix E) we also compare our tests with an asymptotic and bootstrapped Ljung-Box test, and a bootstrapped Andrews and Ploberger's (1996) sup-LM white noise test.

The remainder of the paper is as follows. Section 2 contains the assumptions and main results for the max-correlation white noise test. In Section 3 we work with a fixed maximum lag and Delgado and Velasco's (2011) orthogonal correlation transform, followed by a variety of examples in Section 4. A Monte Carlo study follows in Section 5, and concluding remarks are left for Section 6.

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

[1] We show in an additional simulation study in the supplemental material that at large lags the bootstrapped Q-test is dominated by the max-correlation test.

lie in a complete probability measure space $(\Omega, \mathcal{P}, \mathcal{F})$, hence $\sigma(\cup_{t\in\mathbb{Z}}\mathcal{F}_t) \subseteq \mathcal{F}$.

2 Max-Correlation Test

We first lay out the main assumptions, and then derive the main results for a dependent wild bootstrap test. We complete this section by studying local power.

2.1 Assumptions and Asymptotic Expansion

In order to bootstrap the asymptotic distribution of $\max_{1\leq h\leq \mathcal{L}_n} |\sqrt{n}\hat{\rho}_n(h)|$, we utilize a first order asymptotic expansion of $\epsilon_t(\hat{\theta}_n)$ around θ_0. Recall $\mathcal{L}_n \to \infty$ and $\mathcal{L}_n = o(n)$. We find in general this provides a better small sample approximation to the limit distribution of $\hat{\mathcal{T}}_n$ since it accounts for the first order structure of $\hat{\theta}_n$.

The first order expansion is accomplished under various regularity assumptions which we discuss here. 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 $\alpha_m^{(\upsilon)} = \sup_{t\in\mathbb{Z}} \sup_{\mathcal{A}\subset\mathfrak{V}_t^\infty, \mathcal{B}\subset\mathfrak{V}_{-\infty}^{t-m}} |P(\mathcal{A}\cap\mathcal{B}) - P(\mathcal{A})P(\mathcal{B})|$. We say L_q-bounded $\{\epsilon_t\}$ is stationary L_q-NED with size λ 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 $\iota > 0$. 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. 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_p*-bounded,* $p > 4$, *and* L_4*-NED with size* $1/2$ *on stationary* α*-mixing* $\{\upsilon_t\}$ *with coefficients* $\alpha_h^{(\upsilon)} = O(h^{-p/(p-2)-\iota})$ *for tiny* $\iota > 0$.

Remark 1 Lobato, Nankervis, and Savin (2002) impose a similar NED property for a Lagrange Multiplier type extension of a Q-statistic for observed data. 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.[2]

Remark 2 Ergodicity is not required in principle, but imposed to allow easily for laws of large numbers

[2] 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).

on functions of the general response functions $f(x_t, \phi)$ and $\sigma_t^2(\theta)$ and their derivatives. Indeed, NED does not necessarily carry over to arbitrary measurable transform of an NED process. α-mixing, however, implies ergodicity, it extends to measurable transforms, and is a sub-class of the processes under (b).

If we simply have $y_t = \epsilon_t$ then a plug-in estimator 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 set of assumptions are required only if a filter is used on y_t.

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 $\mathbb{R}^{k_\phi}$, $k_\phi \geq 0$; $f(\cdot, \phi)$ is three times continuously differentiable; $E[\sup_{\phi \in \mathcal{N}_{\phi_0}} |(\partial/\partial\phi)^j f(x_t, \phi)|^4] < \infty$ for $j = 1, 2, 3$ and some compact set $\mathcal{N}_{\phi_0} \subseteq \Phi$ 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 $\mathcal{F}_t$-measurable, continuous, and three times continuously differentiable with $\inf_{\theta \in \Theta} |\sigma_t^2(\theta)| \geq \iota > 0$ $a.s.$ and $E[\sup_{\theta \in \mathcal{N}_0} |(\partial/\partial\theta)^j \ln \sigma_t^2(\theta)|^4] < \infty$ for $j = 1, 2, 3$ and some compact subset $\mathcal{N}_0 \subseteq \Theta$ containing θ_0.*

c. Estimator. *$\hat{\theta}_n \in \Theta$ $a.s.$ for each n, and for a unique interior point $\theta_0 \in \Theta$ we have $\sqrt{n}(\hat{\theta}_n - \theta_0) = \mathcal{A} n^{-1/2} \sum_{t=1}^n m_t + o_p(1)$, where $m_t = [m_{i,t}]_{i=1}^{k_m}$ are stationary and ergodic, $k_m \geq k_\theta$, $E[m_t] = 0$, and $\mathcal{A} \in \mathbb{R}^{k_\theta \times k_m}$. Moreover, m_t is L_r-bounded, $r > 2$, and L_2-NED with size $-1/2$ on a stationary α-mixing process with coefficient decay $O(h^{-r/(r-2)-\iota})$ for tiny $\iota > 0$.*

Remark 3 Smoothness (a) and (b) ensure a stochastic equicontinuity property for well known uniform laws of large numbers (e.g. Newey, 1991). These can be relaxed substantially for specific response functions: see Section 4 for examples.

Remark 4 The fourth moment bounds are standard due to required expansions of residuals cross-products. Note that $E[\sup_{\theta \in \mathcal{N}_0} |(\partial/\partial\theta)^j \ln \sigma_t^2(\theta)|^4] < \infty$ holds for many linear and nonlinear volatility models, including GARCH, Quadratic GARCH, GJR-GARCH, etc. See Francq and Zakoïan (2004, 2010). The bound $E[\sup_{\phi \in \mathcal{N}_{\phi_0}} |(\partial/\partial\phi)^j f(x_t, \phi)|^4] < \infty$ can imply higher moment bounds than in Assumption 1 depending on the nonlinear response. A logistic Smooth Transition Autoregression, for example, has $f(x_t, \phi) = \phi_1' x_t + \phi_2' x_t / (1 + \exp\{-\phi_3' x_t\})$ where x_t contains a constant term and lags of y_t (Terasvirta, 1994). The required bound holds under Assumption 1.

Remark 5 The estimator $\hat{\theta}_n$ asymptotically is a linear function of some zero mean process $\{m_t\}$. This is fairly mild since it includes M-estimators, Generalized Method of Moments, and (Generalized) Empirical Likelihood, and non-smooth estimators with asymptotic expansions like LAD and quantile regression (see Knight, 1998). We ignore the possibility that θ_0 lies on the boundary of Θ to conserve space (boundary cases include GARCH models with zero parameter values, and random coefficient models where some coefficients have a zero variance).

Remark 6 Typically m_t 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. In the pure level response case $y_t = f(x_t, \phi_0) + \epsilon_t$, for example, we may assume ϕ_0 is identified by least squares first order equations $E[\epsilon_t (\partial/\partial\phi) f(x_t, \phi_0)] = 0$, hence assumed estimating equations $m_t = \epsilon_t (\partial/\partial\phi) f(x_t, \phi_0)$ from nonlinear least squares, QML, GMM or GEL have a zero mean.

In the following we drop θ_0. Let $\mathbf{0}_l$ be an l-dimensional zero vector. Define

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

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

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

The required expansion follows. Proofs are presented in Appendix D of Hill and Motegi (2016).

Lemma 2.1 *Under Assumptions 1 and 2* $\max_{1 \leq h \leq \mathcal{L}_n} |\sqrt{n}\{\hat{\rho}_n(h) - \rho(h)\} - \mathcal{Z}_n(h)| \xrightarrow{p} 0$ *where*

$$\mathcal{Z}_n(h) \equiv \frac{1}{\sqrt{n}} \sum_{t=1+h}^{n} \frac{(\epsilon_t \epsilon_{t-h} - E[\epsilon_t \epsilon_{t-h}] - \mathcal{D}'(h)\mathcal{A}m_t)}{E[\epsilon_t^2]}. \tag{3}$$

Remark 7 In Hill and Motegi (2016, Appendix D: Lemma D.1) we provide a Gaussian weak limit theory for the expansion process $\{\mathcal{Z}_n(h) : 1 \leq h \leq \mathcal{L}_n\}$. This is the key limit theory behind the following bootstrap results.

2.2 Bootstrapped P-Value Test

We need to compute a p-value approximation $\hat{p}_n \in [0,1]$ for the max-correlation $\hat{\mathcal{T}}_n$ such that an asymptotically correctly sized test is achieved. That is, under H_0 for any level $\alpha \in (0,1)$ we achieve $P(\hat{p}_n < \alpha) \to \alpha$ under H_0, and $P(\hat{p}_n < \alpha) \to 1$ under H_1. The max-correlation test is therefore:

$$\text{reject } H_0 \text{ at level } \alpha \in (0,1) \text{ when } \hat{p}_n < \alpha.$$

We work with the dependent wild bootstrap for p-value computation since these methods have been widely explored in the literature. $m_t(\theta)$ are the estimating equations for $\hat{\theta}_n$, $\widehat{\mathcal{A}}_n$ is consistent estimator of $\mathcal{A}$ in Assumption 2.c, and define

$$\hat{\mathcal{D}}_n(h) \equiv \frac{1}{n}\sum_{t=h+1}^{n}\left\{\left(\epsilon_t(\hat{\theta}_n)s_t(\hat{\theta}_n) + \frac{G_t(\hat{\theta}_n)}{\sigma_t(\hat{\theta}_n)}\right)\epsilon_{t-h}(\hat{\theta}_n) + \epsilon_t(\hat{\theta}_n)\left(\epsilon_{t-h}(\hat{\theta}_n)s_{t-h}(\hat{\theta}_n) + \frac{G_{t-h}(\hat{\theta}_n)}{\sigma_{t-h}(\hat{\theta}_n)}\right)\right\}. \quad (4)$$

Under Assumptions 1 and 2, arguments used to prove Lemma 2.1 yield $\hat{\mathcal{D}}_n(h) \xrightarrow{p} \mathcal{D}(h)$ where $\mathcal{D}(h)$ is defined in (2). We now operate on $\epsilon_t(\hat{\theta}_n)\epsilon_{t-h}(\hat{\theta}_n) - \hat{\mathcal{D}}_n(h)'\times\widehat{\mathcal{A}}_n\times m_t(\hat{\theta}_n)$, an approximation of an expansion of $\epsilon_t(\hat{\theta}_n)$ around θ_0 under H_0, cf. Lemma 2.1.

The same method and theory promote valid p-value approximations for a Q-test and Andrews and Ploberger's (1996) sup-LM test, and indeed any test statistic that is a measurable function of $\{\hat{\rho}_n(h)\}_{h=1}^{\mathcal{L}_n}$. See Hill and Motegi (2016, Appendix C).

2.3 Dependent Wild Bootstrap

The classic wild bootstrap detailed in Wu (1986) and Liu (1988) was proposed for iid sequences, and Hansen (1996) shows it applies to an adapted martingale difference sequence [mds]. Shao (2010a, 2011) generalizes the wild bootstrap to allow for dependent sequences. Shao (2010a) uses iid random draws as weights, similar to the wild bootstrap, with a covariance function that equals a kernel function. His requirements rule out a truncated kernel, but allow a Bartlett kernel amongst others (Shao, 2010a, Assumption 2.1). We follow Shao (2011) whose random draws effectively have a truncated kernel covariance function. Write compactly.

$$\widehat{\mathcal{E}}_{n,t,h}(\hat{\theta}_n) \equiv \epsilon_t(\hat{\theta}_n)\epsilon_{t-h}(\hat{\theta}_n) - \hat{\mathcal{D}}_n(h)'\times\widehat{\mathcal{A}}_n\times m_t(\hat{\theta}_n) \text{ and } \hat{g}_n(h,\hat{\theta}_n) \equiv \frac{1}{n}\sum_{t=1+h}^{n}\widehat{\mathcal{E}}_{n,t,h}(\hat{\theta}_n). \quad (5)$$

The algorithm is as follows. Set a block size b_n such that $1 \leq b_n < n$. 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 $\omega_t = \xi_s$ if $t \in \mathcal{B}_s$. Compute for $h = 1, \ldots, \mathcal{L}_n$:

$$\hat{\rho}_n^{(dw)}(h) \equiv \frac{1}{1/n \sum_{t=1}^n \epsilon_t^2(\hat{\theta}_n)} \frac{1}{n} \sum_{t=1+h}^n \omega_t \left\{ \widehat{\mathcal{E}}_{n,t,h}(\hat{\theta}_n) - \hat{g}_n(h, \hat{\theta}_n) \right\}, \tag{6}$$

and the max-statistic $\hat{\mathcal{T}}_n^{(dw)} \equiv \sqrt{n} \max_{1 \leq h \leq \mathcal{L}_n} |\hat{\rho}_n^{(dw)}(h)|$. Repeat M times, resulting in bootstrapped max-correlation statistics $\{\hat{\mathcal{T}}_{n,i}^{(dw)}\}_{i=1}^M$. The approximate p-value is $\hat{p}_{n,M}^{(dw)} \equiv 1/M \sum_{i=1}^M I(\hat{\mathcal{T}}_{n,i}^{(dw)} \geq \hat{\mathcal{T}}_n)$.

Remark 8 The key difference between wild and dependent wild bootstraps is the block-wise dependence of the auxiliary variable, and re-centering with the sample mean of $\widehat{\mathcal{E}}_{n,t,h}(\hat{\theta}_n)$.

Remark 9 The auxiliary variable satisfies $\omega_t = \xi_1$ for $t = 1, ..., b_n$, $\omega_t = \xi_2$ for $t = b_n + 1, ..., 2b_n$, etc. This implies b_n-dependence, with perfect dependence within blocks, and $E[\omega_s \omega_t] = I(|s - t| \leq b_n)$, which is just the truncated kernel (cf. Andrews, 1991).

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, 2016, 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 $\int_0^\pi \hat{S}_n^2(\lambda) d\lambda$.[3] Further, 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, e.g., Wu and Shao, 2004, Shao and Wu, 2007). Second, we need weak convergence of $\{\hat{\gamma}_n(h) : 1 \leq h \leq \mathcal{L}_n\}$ which is far easier to handle than weak convergence of $\{\hat{S}_n(\lambda) : \lambda \in [0, \pi]\}$ because tightness on a discrete set is trivial: weak convergence of $\{\hat{\gamma}_n(h) : 1 \leq h \leq \mathcal{L}_n\}$ only requires convergence in finite dimensional distribution. Third, the supremum is not a continuous mapping from the space of square integrable (with respect to

[3] See also Politis and Romano (1994) and Escanciano and Velasco (2006) for applications of weak convergence in a Hilbert space to the bootstrap.

Lebesgue measure) functions on $[0, \pi]$. It is therefore not clear how, or even if, Shao's (2011: Theorem 3.1) proof applies to our max-statistic.

Our proof exploits what are essentially Bernstein (1927) blocks to separate Shao's (2011) blocks. Coupled with an NED assumption, and using Gine and Zinn's (1990: Section 3) notion of *weak convergence in probability*, we exploit a clever argument in de Jong (1997) to prove the bootstrapped correlation process $\{\sqrt{n}\hat{\rho}_n^{(dw)}(h) : 1 \leq h \leq \mathcal{L}_n\}$ converges weakly in probability to the weak limit of $\{\sqrt{n}\hat{\rho}_n(h) : 1 \leq h \leq \mathcal{L}_n\}$. This implies $\hat{\mathcal{T}}_n^{(dw)}$ has the same distribution function as $\hat{\mathcal{T}}_n$ asymptotically with probability approaching one, which yields consistency of the bootstrapped p-value. We exploit an NED assumption, hence Shao's (2011) complicated cumulant condition is not required, and ϵ_t only requires slightly more than a fourth moment, which is sharp under our NED assumption.[4]

Theorem 2.2 *Let Assumptions 1 and 2 hold. Let $\hat{\mathcal{D}}_n(h) \xrightarrow{p} \mathcal{D}(h)$, $\widehat{\mathcal{A}}_n \xrightarrow{p} \mathcal{A}$, the maximum lag $\mathcal{L}_n \to \infty$ and $\mathcal{L}_n = o(n)$, and the number of bootstrap samples $M = M_n \to \infty$. 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 The max-correlation test by dependent wild bootstrap yields an asymptotically correctly sized test, and is consistent against any deviation from the white noise null hypothesis.

Remark 11 A similar theory applies to an approximate p-value computed by wild bootstrap, provided ϵ_t is a mds under the null. The algorithm is similar to the dependent wild bootstrap, except ω_t is iid $N(0, 1)$ for each $t = 1, ..., n$.

2.4 Local Asymptotic Power

Hong (1996) shows that his standardized periodogram yields non-trivial asymptotic power against a sequence of local alternatives applied to the spectrum, with a slower than $\sqrt{n}$ drift. Ultimately the reduced rate arises from an increasing bandwidth parametric. Using a similar spectrum local alternative, but with $\sqrt{n}$ drift, Shao (2011) proves the Cramér-von Mises test achieves non-trivial local power.

[4] If we restrict dependence to be just α-mixing (recall NED encompasses mixing), then $\mathcal{E}_{t,h} \equiv \epsilon_t \epsilon_{t-h} - E[\epsilon_t \epsilon_{t-h}] - \mathcal{D}'\mathcal{A}m_t$ has a spectral density $f(\cdot)$. Further, if $f(0) > 0$ then a Gaussian limit theory applies when $\epsilon_t \epsilon_{t-h}$ and m_t only have a second moment. See Ibragimov (1975, Theorem 2.2) and Bradley (1993, Theorem 1.a). Under dependence, then, ϵ_t only needs a fourth moment. Demonstrating $f(0) > 0$, however, may be generally impossible due to the construction of $\mathcal{E}_{t,h}$ (e.g. if the filter is a nonlinear ARMA-GARCH, or the filter is mis-specified).

Delgado and Velasco (2011) impose local $\sqrt{n}$ drift to the correlations, and show that a weighted average portmanteau statistic yields non-trivial local power.

We operate on the largest (in absolute value) of an increasing sequence of sample correlations. The max-correlation statistic therefore achieves the parametric rate of local asymptotic power against the sequence of alternatives:

$$H_1^L : \rho(h) = \rho_n(h) = \frac{r(h)}{\sqrt{n}} \text{ for each } h\text{, where } |r(h)| \leq \sqrt{n}.$$

Note that $r(h)$ is a fixed constant for each h, where $|r(h)| \leq \sqrt{n}$ ensures $|\rho_n(h)| \leq 1$.

Theorem 2.3 *Let Assumptions 1 and 2 hold. Let $\hat{\mathcal{D}}_n(h) \xrightarrow{p} \mathcal{D}(h)$, $\widehat{\mathcal{A}}_n \xrightarrow{p} \mathcal{A}$, $\mathcal{L}_n \to \infty$ and $\mathcal{L}_n = o(n)$, and $M = M_n \to \infty$. Under H_1^L, $\lim_{n\to\infty} P(\hat{p}_{n,M}^{(dw)} < \alpha) > \alpha$ if $r(h) \neq 0$ for some $h \in \mathbb{N}$. Specifically $\lim_{n\to\infty} P(\hat{p}_{n,M}^{(dw)} < \alpha) \nearrow 1$ monotonically in $|r(h)| \nearrow \infty$.*

3 Max-Transformed Correlation Test

We use expansion (3) for a better approximation of the residuals cross-product due to the impact of $\hat{\theta}_n$ on small and large sample dynamics. We then apply a bootstrap due to a non-standard limit theory under the maximum transform. Delgado and Velasco (2011), however, take a different approach by exploiting an orthogonal projection of the sample correlations. This requires an inverted covariance matrix for robustness to weak dependence, and therefore a finite maximum lag. Dependence robust estimators are also known to be sensitive to user chosen nuisance parameters, including bandwidth (Newey and West, 1987, Andrews, 1991) and related parameters (e.g. Shao, 2010b), possibly leading both to size distortions and diminished power, even for comparatively small maximum lags (see Delgado and Velasco, 2011, Section 4).

We are nevertheless interesting in seeing if the transformation can lead to a better first order approximation to the sample correlation dependence structure, in particular when a filter is used, as compared to the use of the first order expansion Lemma 2.1. Operating on the maximum transformed correlation should also provide an advantage over Delgado and Velasco's (2011) Q-statistic, since we operate on only the most informative sample statistic for correlation detection. The statistic and related theory are developed below. A Monte Carlo experiment in Section 5, however, shows bootstrapping based on the Lemma 2.1 expansion is generally superior.

We first characterize the transformed correlations. Define estimator functions of θ:

$$\hat{\gamma}_{n,\theta}(h) \equiv \frac{1}{n}\sum_{t=1+h}^{n} \epsilon_t(\theta)\epsilon_{t-h}(\theta) \text{ and } \hat{\rho}_{n,\theta}(h) \equiv \frac{\hat{\gamma}_{n,\theta}(h)}{\hat{\gamma}_{n,\theta}(0)} \text{ and } \hat{\boldsymbol{\rho}}_{n,\theta}^{(\mathcal{L})} \equiv [\hat{\rho}_{n,\theta}(h)]_{h=1}^{\mathcal{L}}.$$

Notice $\hat{\rho}_{n,\hat{\theta}_n}(h) = \hat{\rho}_n(h)$ by definition. Under Assumptions 1 and 2, and H_0, $\sqrt{n}\hat{\boldsymbol{\rho}}_{n,\theta_0}^{(\mathcal{L})} \xrightarrow{d} N(0, \mathcal{V}^{(\mathcal{L})})$ for any fixed $\mathcal{L}$, and $\mathcal{V}^{(\mathcal{L})} = [\mathcal{V}_{i,j}^{(\mathcal{L})}]_{i,j=1}^{\mathcal{L}}$, $||\mathcal{V}^{(\mathcal{L})}|| < \infty$. In general $\mathcal{V}_{i,j}^{(\mathcal{L})} = (E[\epsilon_t^2])^{-2}\sum_{l=-\infty}^{\infty} E[\epsilon_t\epsilon_{t-i}\epsilon_{t-l}\epsilon_{t-l-j}]$. If ϵ_t is iid then $\mathcal{V}^{(\mathcal{L})} = I_{\mathcal{L}}$, and if ϵ_t and $\epsilon_t^2 - E[\epsilon_t^2]$ are martingale differences then $\mathcal{V}^{(\mathcal{L})}$ is diagonal with $\mathcal{V}_{i,i}^{(\mathcal{L})} = (E[\epsilon_t^2])^{-2}E[\epsilon_t^2\epsilon_{t-i}^2]$. The claim follows from the arguments used to prove Lemma D.1 in Hill and Motegi (2016, Appendix D).

Now assume a positive definite $\hat{\mathcal{V}}_{n,\theta}^{(\mathcal{L})}$ exists such that $\hat{\mathcal{V}}_{n,\hat{\theta}_n}^{(\mathcal{L})} \xrightarrow{p} \tilde{\mathcal{V}}^{(\mathcal{L})}$ where $\tilde{\mathcal{V}}^{(\mathcal{L})} = \mathcal{V}^{(\mathcal{L})}$ under H_0, and define the standardized sample correlation vector:

$$\tilde{\boldsymbol{\rho}}_{n,\theta}^{(\mathcal{L})} = \left[\tilde{\rho}_{n,\theta}^{(\mathcal{L})}(1), ..., \tilde{\rho}_{n,\theta}^{(\mathcal{L})}(\mathcal{L})\right]' \equiv \hat{\mathcal{V}}_{n,\theta}^{(\mathcal{L})-1/2}\hat{\boldsymbol{\rho}}_{n,\theta}^{(\mathcal{L})}.$$

An example is a kernel estimator, e.g. Lobato, Nankervis, and Savin (2002, Section 4) or Delgado and Velasco (2011, Section 4). Define $\boldsymbol{D}_{\theta}^{(\mathcal{L})} = [\mathcal{D}_{\theta}(1), ..., \mathcal{D}_{\theta}(\mathcal{L})]' \in \mathbb{R}^{\mathcal{L}\times k_\theta}$ and:

$$\boldsymbol{\xi}_{\theta}^{(\mathcal{L})} = [\xi_{\theta}^{(\mathcal{L})}(1)', ..., \xi_{\theta}^{(\mathcal{L})}(\mathcal{L})']' \equiv \mathcal{V}_{\theta}^{(\mathcal{L})-1/2}\boldsymbol{D}_{\theta}^{(\mathcal{L})} \in \mathbb{R}^{\mathcal{L}\times k_\theta} \text{ where } \frac{\partial}{\partial\theta}\hat{\rho}_{n,\theta}(h) \xrightarrow{p} \mathcal{D}_{\theta}(h) \in \mathbb{R}^{k_\theta}.$$

By the line of proof of Lemma 2.1, $\mathcal{D}(h)$ in (2) is identically $\mathcal{D}_{\theta_0}(h)$. Under Assumptions 1 and 2, arguments similar to those used to prove Lemma 2.1 yield (cf. Delgado and Velasco, 2011, Proposition 1): $\tilde{\boldsymbol{\rho}}_{n,\hat{\theta}_n}^{(\mathcal{L})} = \tilde{\boldsymbol{\rho}}_{n,\theta_0}^{(\mathcal{L})} + \boldsymbol{\xi}_{\theta_0}^{(\mathcal{L})}(\hat{\theta}_n - \theta_0) + o_p(1/\sqrt{n})$.

Define the operator $\mathcal{P}^{(\mathcal{L})}$ for any sequence $\{\eta(h)\}_{h=1}^{\mathcal{L}}$ by the forward recursive residuals of its least squares projection on $\{\xi_{\theta}^{(\mathcal{L})}(h)\}_{h=1}^{\mathcal{L}}$:

$$\mathcal{P}^{(\mathcal{L})}\eta(h) = \eta(h) - \xi_{\theta_0}^{(\mathcal{L})}(h)\left(\sum_{j=h+1}^{\mathcal{L}}\xi_{\theta_0}^{(\mathcal{L})}(j)'\xi_{\theta_0}^{(\mathcal{L})}(j)\right)^{-1}\sum_{j=h+1}^{\mathcal{L}}\xi_{\theta_0}^{(\mathcal{L})}(j)'\eta(j).$$

By construction, the correlation expansion projection $\mathcal{P}^{(\mathcal{L})}(\tilde{\rho}_{n,\theta_0}(h) + \xi_{\theta_0}^{(\mathcal{L})}(h)(\hat{\theta}_n - \theta_0)) = \mathcal{P}^{(\mathcal{L})}\tilde{\rho}_{n,\theta_0}(h)$, hence it is free of $\hat{\theta}_n$. Further, $\sqrt{n}\tilde{\rho}_{n,\theta_0}^{(\mathcal{L})} \xrightarrow{d} N(0, I_{\mathcal{L}})$ under H_0, hence $[\mathcal{P}^{(\mathcal{L})}\tilde{\rho}_{n,\theta_0}(h)]_{h=1}^{\mathcal{L}} \xrightarrow{d} N(0, \Sigma^{(\mathcal{L})})$ where $\Sigma^{(\mathcal{L})}$ is diagonal, with entries $\sigma^2(h) = 1 + \xi_{\theta_0}^{(\mathcal{L})}(h)(\sum_{j=h+1}^{\mathcal{L}}\xi_{\theta_0}^{(\mathcal{L})}(j)'\xi_{\theta_0}^{(\mathcal{L})}(j))^{-1}\xi_{\theta_0}^{(\mathcal{L})}(h)'$.

Now define sample versions of the above key components:

$$\hat{\boldsymbol{D}}_{n,\hat{\theta}_n}^{(\mathcal{L})} \equiv [\hat{\mathcal{D}}_n(1), ..., \hat{\mathcal{D}}_n(\mathcal{L})]' \text{ and } \hat{\boldsymbol{\xi}}_{n,\hat{\theta}_n}^{(\mathcal{L})} \equiv \hat{\mathcal{V}}_{n,\hat{\theta}_n}^{(\mathcal{L})-1/2} \hat{\boldsymbol{D}}_{n,\hat{\theta}_n}^{(\mathcal{L})}$$
$$\hat{\sigma}_n^2(h) = 1 + \hat{\xi}_{n,\hat{\theta}_n}^{(\mathcal{L})}(h) \left(\sum_{j=h+1}^{\mathcal{L}} \hat{\xi}_{n,\hat{\theta}_n}^{(\mathcal{L})}(j)' \hat{\xi}_{n,\hat{\theta}_n}^{(\mathcal{L})}(j) \right)^{-1} \hat{\xi}_{n,\hat{\theta}_n}^{(\mathcal{L})}(h)'$$
$$\hat{\mathcal{P}}_n^{(\mathcal{L})} \eta(h) = \eta(h) - \hat{\xi}_{n,\hat{\theta}_n}^{(\mathcal{L})}(h) \left(\sum_{j=h+1}^{\mathcal{L}} \hat{\xi}_{n,\hat{\theta}_n}^{(\mathcal{L})}(j)' \hat{\xi}_{n,\hat{\theta}_n}^{(\mathcal{L})}(j) \right)^{-1} \sum_{j=h+1}^{\mathcal{L}} \hat{\xi}_{n,\hat{\theta}_n}^{(\mathcal{L})}(j)' \eta(j),$$

where $\hat{\mathcal{D}}_n(h)$ are defined in (4). Delgado and Velasco's (2011) proposed transformation is:

$$\bar{\rho}_{n,\theta}^{(\mathcal{L})}(h) \equiv \frac{\hat{\mathcal{P}}_n^{(\mathcal{L})} \tilde{\rho}_{n,\theta}^{(\mathcal{L})}(h)}{\hat{\sigma}_n(h)} \text{ for } h \leq \mathcal{L} - k_\theta. \tag{7}$$

The number of lags is restricted to $\mathcal{L} - k_\theta$ due to the restricted degrees of freedom arising for the plug-in estimator. The scale $\hat{\sigma}_n(h)$ ensures an asymptotically standard normal limit distribution under the null. The max-transformed correlation statistic is:

$$\bar{\mathcal{T}}_n \equiv \max_{1 \leq h \leq \mathcal{L} - k_\theta} \left| \sqrt{n} \bar{\rho}_{n,\theta}^{(\mathcal{L})}(h) \right|.$$

Let $\bar{p}_{n,M}^{(dw)}$ be the dependent wild bootstrap approximate p-value for $\bar{\mathcal{T}}_n$. The limit theory for $\bar{p}_{n,M}^{(dw)}$ follows from the theory for $\hat{\mathcal{T}}_n$ because $\hat{\mathcal{P}}_n^{(\mathcal{L})} \tilde{\rho}_{n,\theta}^{(\mathcal{L})}(h)$ is just a stochastic weighted average of $\{\hat{\rho}_{n,\theta}(i)\}_{i=1}^{\mathcal{L}}$. A similar theory extends to the wild bootstrap for mds data under the null, and the bootstrapped test has non-trivial local power against $\sqrt{n}$-local alternatives by arguments used to prove Theorem 2.3.

Theorem 3.1 *Let Assumptions 1 and 2 hold, and let $M = M_n \to \infty$. Let $\hat{\mathcal{V}}_{n,\hat{\theta}_n}^{(\mathcal{L})} \xrightarrow{p} \tilde{\mathcal{V}}^{(\mathcal{L})}$, $||\tilde{\mathcal{V}}^{(\mathcal{L})}|| < \infty$, and $\tilde{\mathcal{V}}^{(\mathcal{L})} = \mathcal{V}^{(\mathcal{L})}$ under H_0. Under H_0, $P(\bar{p}_{n,M}^{(dw)} < \alpha) \to \alpha$, and if H_0 is false then $P(\bar{p}_{n,M}^{(dw)} < \alpha) \to 1$.*

4 Examples

In order to verify the assumptions, we give several examples of models under (1). We also need the form of expansion (3) in order to compute the bootstrapped p-value. Expansion (3) can be simplified depending on whether ϵ_t is assumed independent under the null, the regressors $\{x_t\}$ are independent of the sequence $\{\epsilon_t\}$, the stochastic volatility component σ_t is estimated, and the level response f is linear.

Example 1 (level response) The level response model is $y_t = f(x_{t-1}, \phi_0) + u_t$. Assume $f(\cdot, \phi)$ is three times continuously differentiable in $\phi \in \mathbb{R}^{k_\phi}$, $E[u_t^2] < \infty$, and $E[G_t G_t']$ is finite and positive definite where $G_t = G_t(\phi_0) = [G_{t,i}(\phi_0)]_{i=1}^k \equiv (\partial/\partial\phi) f(x_{t-1}, \phi_0)$. Define nonlinear least squares estimating equations $m_t(\phi) = (y_t - f(x_{t-1}, \phi)) \times (\partial/\partial\phi) f(x_t, \phi)$. Assume $E[m_t(\phi)] = 0$ *if and only if* $\phi = \phi_0$, a unique interior point of compact Φ.

Assume $\{u_t, x_t\}$ are stationary L_p-bounded, $p > 4$, α-mixing with coefficients $\alpha_h = O(h^{-p/(p-2)}/\ln(h))$. Assume $E[\sup_{\phi \in \mathcal{N}_{\phi_0}} |G_{t,i}(\phi)|^r] < \infty$ for each i, some $r > 4$, and some compact $\mathcal{N}_{\phi_0} \subseteq \Phi$ containing ϕ_0. Many nonlinear response functions satisfy this condition under L_p-boundedness of $\{u_t, x_t\}$, including linear, logistic, and trigonometric functions. Then m_t is stationary, L_p-bounded, and α-mixing. Sufficient conditions for stationary geometric ergodicity of nonlinear AR-GARCH with iid innovations are in Meitz and Saikkonen (2008), amongst others, cf. Doukhan (1994, Chapt. 2.4.2).

Define the non-linear least squares estimator $\hat{\phi}_n = \arg\min_{\phi \in \Phi}\{1/n \sum_{t=1}^n (y_t - f(x_{t-1}, \phi))^2\}$. By construction $s_t = .5(\partial/\partial\theta) \ln \sigma_t^2 = 0$ since $\sigma_t^2 = 1$, and $\mathcal{D}(h) = E[u_{t-h} G_t] + E[u_t G_{t-h}] = E[u_{t-h} G_t]$. Then Assumptions 1 and 2 hold, with $m_t = u_t G_t$ and $\mathcal{A} = (E[G_t G_t'])^{-1}$, hence

$$\sqrt{n}\hat{\rho}_n(h) = \frac{1}{\sqrt{n}} \sum_{t=1+h}^{n} \frac{u_t \left\{u_{t-h} - E[u_{t-h} G_t]' \left(E\left[G_t G_t'\right]\right)^{-1} G_t\right\}}{E\left[u_t^2\right]} + o_p(1). \tag{8}$$

If additionally u_t is independent of the sequence $\{x_t\}$, then $E[u_{t-h} G_t] = 0$, hence $\sqrt{n}\hat{\rho}_n(h) = 1/\sqrt{n} \sum_{t=1+h}^n u_t u_{t-h}/E[u_t^2] + o_p(1)$, the well known result that $\hat{\phi}_n$ does not impact the limit distribution of $\sqrt{n}\hat{\rho}_n(h)$ (cf. Wooldridge, 1990).

Example 2 (linear response with least squares) The model is $y_t = \phi_0' x_{t-1} + u_t$, $E[u_t] = 0$. Let $E[(y_t - \phi_0' x_{t-1}) u_t] = 0$ for a unique interior $\phi_0 \in \Phi$, and assume $E[x_t x_t']$ is positive definite. Assume $\{x_t, u_t\}$ are stationary and ergodic, L_p-bounded, $p > 4$, and L_4-NED on an α-mixing base with coefficient decay $O(h^{-p/(p-2)-\zeta})$. An AR process with an iid error that has a continuous bounded distribution is geometrically α-mixing and therefore geometrically NED. This extends to linear or nonlinear GARCH errors (see, e.g., Doukhan, 1994, Meitz and Saikkonen, 2008).

By construction $G_t = x_{t-1}$ and $s_t = 0$ hence $\mathcal{D}(h) \equiv E[u_{t-h} x_{t-1}]$. If $\hat{\phi}_n$ is least squares then

Assumptions 1 and 2 are satisfied, with $\mathcal{A} = (E[x_t x_t'])^{-1}$ and $m_t = u_t x_{t-1}$. Therefore:

$$\sqrt{n}\hat{\rho}_n(h) = \frac{1}{E\left[u_t^2\right]} \frac{1}{\sqrt{n}} \sum_{t=1+h}^{n} u_t \left\{u_{t-h} - E[u_{t-h} x_{t-1}]'(E[x_t x_t'])^{-1} x_{t-1}\right\} + o_p(1).$$

If $u_t^2 - E[u_t^2]$ is an adapted mds then $E[m_{i,t}^2] = E[u_t^2]E[x_{i,t}^2] < \infty$ and $E[(u_t u_{t-h})^2] = (E[u_t^2])^2 < \infty$, hence we only need $\{x_t, u_t\}$ to be L_p-bounded, $p > 2$, and α-mixing.

Example 3 (mean model with sample mean) The mean model is $y_t = E[y_t] + u_t$, hence $f(x_{t-1}, \phi_0)$ $= E[y_t]$, $\sigma_t = 1$ and $E[u_t] = 0$. Assume $\{y_t\}$ is stationary and ergodic, and L_p-bounded, $p > 4$, and L_4-NED on and α-mixing base with decay $O(h^{-p/(p-2)-\iota})$. Then $G_t = 1$ and $s_t = 0$, hence $\mathcal{D}(h) \equiv E[u_t u_{t-h}] + E[u_t] = E[(y_t - E[y_t])(y_{t-h} - E[y_{t-h}])]$. The plug-in estimator is the sample mean $\hat{\theta}_n = 1/n \sum_{t=1}^{n} y_t$, so that $m_t = u_t$ and $\mathcal{A} = 1$. Assumptions 1 and 2 are satisfied, hence from (8) we obtain

$$\sqrt{n}\hat{\rho}_n(h) = \frac{1}{\sqrt{n}} \sum_{t=1+h}^{n} \frac{(y_t - E[y_t])\,(y_{t-h} - E[y_{t-h}])}{E\left[(y_t - E[y_t])^2\right]} + o_p(1).$$

Example 4 (GARCH(1,1) with QML) The model is GARCH(1,1) $y_t = \sigma_t \epsilon_t$ with $\sigma_t^2 = \omega_0 + \alpha_0 y_{t-1}^2$ $+ \beta_0 \sigma_{t-1}^2$, $\omega_0, \alpha_0, \beta_0 > 0$, $E[\epsilon_t] = 0$ and $E[\epsilon_t^2] = 1$ (Bollerslev, 1986). We ignore boundary cases by assuming $\alpha_0, \beta_0 > 0$. The model includes weak, semi-strong or strong GARCH (see Drost and Nijman, 1993), in which case the model is correct in some sense since the errors are assumed to be serially uncorrelated. Conditions for strict stationarity in the case of iid or mds ϵ_t are given in Nelson (1990) and Lee and Hansen (1994), and Boussama (2006) proves geometric ergodicity.

Let $\theta \equiv [\omega, \alpha, \beta]'$, and $\Theta = [\iota_\omega, u_\omega] \times [0, 1-\iota] \times [0, 1-\iota]$, where $u_\omega > \iota_\omega > 0$ and $\iota \in (0,1)$. Define the unobserved volatility process $\sigma_t^2(\theta) = \omega + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2(\theta)$ on Θ, and define the iterated process used for estimation: $\tilde{\sigma}_0^2(\theta) = \omega$ and $\tilde{\sigma}_t^2(\theta) = \omega + \alpha y_{t-1}^2 + \beta \tilde{\sigma}_{t-1}^2(\theta)$ for $t \geq 1$. Let θ_0 be the unique interior point of Θ such that $\sigma_t^2(\theta_0) = \sigma_t^2$ and $E[(y_t^2/\sigma_t^2(\theta_0) - 1)(\partial/\partial\theta)\ln(\sigma_t^2(\theta_0))] = 0$, the QML first order moment condition. The feasible QML estimator is $\hat{\theta}_n \equiv \arg\inf_{\theta\in\Theta}\{\sum_{t=1}^{n}\{\ln \tilde{\sigma}_t^2(\theta) + \sum_{t=1}^{n} y_t^2/\tilde{\sigma}_t^2(\theta)\}\}$.[5] See Francq and Zakoïan (2004) for refined QML asymptotics when ϵ_t is iid, and see Lee and Hansen (1994) for the semi-strong case.

[5]Since we assume the start condition $\tilde{\sigma}_0^2(\theta) = \omega$ we avoid the case where $\alpha_0 = 0$, β_0 is not identified and is therefore a nuisance parameter, and there are no GARCH effects (see Andrews, 2001). We do not allow nuisance parameters for brevity, but their inclusion is straightforward, although beyond this paper's scope.

Since our assumptions must hold whether ϵ_t is white noise or not, under potentially much weaker conditions than weak-GARCH (Drost and Nijman, 1993), we assume $\{y_t, \epsilon_t\}$ are stationary and ergodic, $(E|y_t|^\iota, E|\sigma_t^2|^\iota) < \infty$ for some $\iota > 0$, $\inf_{\theta \in \Theta} |\sigma_t^2(\theta)| \geq \iota > 0$ $a.s.$, and $\{\epsilon_t, (\partial/\partial\theta)^i \ln(\sigma_t^2(\theta_0)) : i = 0, 1, 2, 3\}$ are stationary geometrically α-mixing. Further, for each $\theta \in \Theta$ unique stationary and ergodic solutions exist for the iterated process and its derivatives $\{(\partial/\partial\theta)^j \tilde{\sigma}_i^2(\theta) : j = 0, 1, 2, 3\}_{i=0}^t$ as $t \to \infty$ at a geometric rate.[6] We also require for some compact subset $\mathcal{N}_0 \subseteq \Theta$ containing θ_0:

$$E\left[s_t(\theta_0)s_t'(\theta_0)\right] - E\left[\left(\frac{y_t^2}{\sigma_t^2(\theta_0)} - 1\right)\frac{\partial^2}{\partial\theta\partial\theta'}\ln(\sigma_t^2(\theta_0))\right] \text{ is non-singular} \tag{9}$$

$$E\left[\sup_{\theta \in \mathcal{N}_0}\left|\left(\frac{\partial}{\partial\theta}\right)^j \ln\sigma_t^2(\theta)\right|^4\right] < \infty \text{ for each } j = 1, 2, 3. \tag{10}$$

If ϵ_t is iid, or $\{\epsilon_t, \epsilon_t^2 - 1\}$ are martingale differences adapted to some sequence of sigma fields $\{\mathcal{G}_t\}$, then stationary solutions exist respectively when $E[\ln(\omega_0 + \alpha_0\epsilon_t^2)] < 0$ and $E[\ln(\omega_0 + \alpha_0\epsilon_t^2)|\mathcal{G}_t] < 0$ $a.s.$ (Nelson, 1990, Lee and Hansen, 1994). Write $s_t(\theta) \equiv 0.5 \times (\partial/\partial\theta)\ln(\sigma_t^2(\theta))$. If ϵ_t is iid or $\{\epsilon_t, \epsilon_t^2 - 1\}$ are martingale differences then (9) holds, and (10) holds by arguments in Francq and Zakoïan (2004, Section 4.2). The latter assume an iid error, but their proofs of (10) do not make use of independence. See, e.g., their equation (4.28).

Under the above conditions, Assumptions 1 and 2 hold. Assuming θ_0 does not lie on the boundary of Θ, plug-in estimator Assumption 2.c holds with $m_t = (\epsilon_t^2 - 1)s_t$ and $\mathcal{A} = \{2E[s_t s_t'] - E[(\epsilon_t^2 - 1)(\partial/\partial\theta)s_t(\theta_0)]\}^{-1}$. Finally, $G_t = 0$ hence $\mathcal{D}(h) \equiv E[\epsilon_t\epsilon_{t-h}(s_t + s_{t-h})]$, and expansion (3) holds.

5 Monte Carlo Experiments

We now perform a Monte Carlo experiment to gauge the merits of the max-correlation test. We simulate 1000 samples of size $n \in \{100, 250, 500, 1000\}$ from the following processes:

Simple : $y_t = e_t$	Bilinear : $y_t = .5e_{t-1}y_{t-2} + e_t$
AR(2) : $y_t = .3y_{t-1} - .15y_{t-2} + e_t$	GARCH(1,1) : $y_t = \sigma_t e_t$ with $\sigma_t^2 = 1 + .2y_{t-1}^2 + .5\sigma_{t-1}^2$

[6] It is plausible that stationarity can be dropped at the expense of deeper technical details and high order assumptions, but to date it appears an asymptotic theory only exists for iid or mds innovations (e.g. Jensen and Rahbek, 2004).

Let ν_t be iid $N(0,1)$. We consider four processes for e_t: iid $e_t = \nu_t$; GARCH(1,1) $e_t = \nu_t w_t$ with $w_t^2 = 1 + .2e_{t-1}^2 + .5w_{t-1}^2$; MA(2) $e_t = \nu_t + .5\nu_{t-1} + .25\nu_{t-2}$; and AR(1) $e_t = .7e_{t-1} + \nu_t$.[7] The error e_t is standardized when y_t is GARCH(1,1) so that $E[e_t^2] = 1$. For each DGP and error, we draw $2n$ observations and retain the last n for analysis.

The first three processes with each error are stationary. The GARCH process is strong when e_t is iid, and semi-strong when e_t itself is GARCH since it is an adapted mds (Drost and Nijman, 1993), hence in those cases GARCH y_t is stationary (Nelson, 1990, Lee and Hansen, 1994). If e_t is MA or AR, then both $\{e_t, y_t\}$ are serially correlated. In the MA case since the feedback structure is finite it can be verified that GARCH y_t is stationary. It is unknown whether GARCH y_t with an AR error has a stationary solution.

All of our chosen tests require a finite fourth moment on the tested variable, and in all cases $E[e_t^4] < \infty$. In all models except GARCH, $E[y_t^4] < \infty$ holds for any error e_t. In the GARCH case $E[y_t^4] < \infty$ holds when e_t is iid or MA(2). Hence, in general test results for the unfiltered GARCH y_t with an AR or GARCH error should be interpreted with some caution.

We estimate a *mean* filter $\epsilon_t = y_t - E[y_t]$ for the mean and bilinear y_t with $\hat{\phi}_n = 1/n\sum_{t=1}^n y_t$. We estimate AR(2) and AR(1) filters $\epsilon_t = y_t - \phi_0' x_t$ for the AR(2) y_t. Finally, for the GARCH(1,1) y_t we either do not use a filter $y_t = \epsilon_t$, or we estimate a GARCH(1,1) filter $\epsilon_t = y_t/\sigma_t$. AR and GARCH filters are computed by least squares and QML, respectively.[8] Finally, we also investigate power under lagged weak dependence with MA errors. The process is $y_t = e_t$ with MA(q) error $e_t = .25\nu_{t-q} + \nu_t$, with order $q \in \{12, 24, 48\}$. We estimate a *mean* filter $\epsilon_t = y_t - E[y_t]$. Notice that any test with a maximum lag under q cannot detect dependence in this process. Summary details are presented in Table 1.

5.1 Test Statistics and P-Values

Max-Correlation Tests We perform the max-correlation test with the standard correlation, and Delgado and Velasco's (2011) orthogonally transformed correlations. We also perform Hong's (1996) standardized spectral test, Delgado and Velasco's (2011) Q-test, Shao's (2011) spectral Cramér-von Mises [CvM] test, and Zhu and Li's (2015) spectral CvM test. In the supplemental material, we also compare our test with the Ljung-Box Q-test and Andrews and Ploberger's (1996) sup-LM test.

[7] In separate simulations we tried weaker degrees of persistence: MA(2) with $e_t = \nu_t + .3\nu_{t-1} + .1\nu_{t-2}$ and AR(1) with $e_t = .2e_{t-1} + \nu_t$. Power is logically lower across tests in these cases, and therefore not reported for the sake of brevity.

[8] The GARCH model is estimated 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 $t = 2, ..., n$. We impose $(\omega, \alpha, \beta) > 0$ and $\alpha + \beta \leq 1$ during estimation.

Table 1: Data Generating Processes and Filter: Memory in Test Variable (corr = correlated)

	Data Generating Process & Filter				Memory in Test Variable			
	y_t	Filter	Test Variable		e_t iid	e_t GARCH(1,1)	e_t MA(2)	e_t AR(1)
1	e_t	Mean	$\hat{\epsilon}_t = y_t - \hat{\phi}_n$		iid	mds	corr	corr
2	Bilinear	Mean	$\hat{\epsilon}_t = y_t - \hat{\phi}_n$		non-mds wn	non-mds wn	corr	corr
3	AR(2)	AR(2)	$\hat{\epsilon}_t = y_t - \hat{\phi}_n' x_t$		iid	mds	corr	corr
4	AR(2)	AR(1)	$\hat{\epsilon}_t = y_t - \hat{\phi}_n' x_t$		corr	corr	corr	corr
5	GARCH	No Filter	y_t		mds	mds	corr	corr
6	GARCH	GARCH	$\hat{\epsilon}_t = y_t / \sigma_t(\hat{\theta}_n)$		mds	mds	corr	corr
7	e_t	Mean	$\hat{\epsilon}_t = y_t - \hat{\phi}_n$		e_t iid	e_t MA(12)	e_t MA(24)	e_t MA(48)

wn = white noise; corr = correlated.

The max-correlation require a lag length $\mathcal{L}_n$. We use a fixed length at 5 or sample-size dependent length $\mathcal{L}_n = [\delta n / \ln(n)]$ with $\delta \in \{.5, 1\}$, where $[\cdot]$ truncates to an integer value. We have $\mathcal{L}_n \in \{5, 10, 21\}$ for $n = 100$; $\mathcal{L}_n \in \{5, 22, 45\}$ for $n = 250$; $\mathcal{L}_n \in \{5, 40, 80\}$ for $n = 500$; and $\mathcal{L}_n \in \{5, 72, 144\}$ for $n = 1000$. We compute the max-correlation statistics $\sqrt{n}\max_{1 \leq h \leq \mathcal{L}_n} |\hat{\rho}_n(h)|$ and $\sqrt{n}\max_{1 \leq h \leq \mathcal{L}_n} |\bar{\rho}_{n,\hat{\theta}_n}^{(\mathcal{L}_n)}(h)|$ where $\bar{\rho}_{n,\hat{\theta}_n}^{(\mathcal{L}_n)}(h)$ is Delgado and Velasco's (2011) transformed correlation. The latter requires a robust covariance matrix, hence choice of kernel function and bandwidth: we use either an identity or kernel matrix estimator. See the details below for Delgado and Velasco's (2011) Q-test.

P-values for the max-correlation test are computed by wild bootstrap [WB] and Shao's (2011) dependent wild bootstrap [DWB]. We bootstrap $\sqrt{n}\max_{1 \leq h \leq \mathcal{L}_n} |\hat{\rho}_n(h)|$ using the Lemma 2.1 correlation expansion, or simply $\hat{\rho}_n(h)$ itself. Conversely, we bootstrap $\sqrt{n}\max_{1 \leq h \leq \mathcal{L}_n} |\bar{\rho}_{n,\hat{\theta}_n}^{(\mathcal{L}_n)}(h)|$ using $\bar{\rho}_{n,\hat{\theta}_n}^{(\mathcal{L}_n)}(h)$ itself, since the orthogonalized correlation accounts for the impact of residuals estimation. In each case we generate $M = 500$ bootstrap samples. The DWB requires a block size b_n, while Shao (2011) uses $b_n = b\sqrt{n}$ with $b \in \{.5, 1, 2\}$, leading to qualitatively similar results. We therefore use $b_n = \sqrt{n}$.[9]

Hong's Standardized Periodogram Test Hong's (1996) test is based on a standardized periodogram. If the periodogram is computed with a truncated kernel, then the statistic is just a standardized Box-Pierce statistic. In order to make a comparison with the Ljung-Box test, we use a standardized Ljung-

[9] 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.

Box $\mathcal{N}_n \equiv (2\mathcal{L}_n)^{-1/2} \sum_{h=1}^{\mathcal{L}_n} w_n(h)\{n\hat{\rho}_n^2(h) - 1\}$ with $w_n(h) = (n+2)/(n-h)$. This is asymptotically equivalent to a standardized Ljung-Box statistic under the null and Hong's (1996) Assumptions 1.a, 2-3. If the null is true and asymptotically $\{\sqrt{n}\hat{\rho}_n^2(h)\}_{h=1}^{\mathcal{L}_n}$ are independent then under Hong's Assumptions 1.a, 2 and 3 $\mathcal{N}_n \xrightarrow{d} N(0,1)$ else $\mathcal{N}_n \xrightarrow{p} \infty$. Asymptotic independence generally requires independence of ϵ_t under the null hypothesis, which fails for our bilinear and GARCH models, and models with mds errors. Thus, the test is expected to result in empirical size distortions when $\mathcal{L}_n$ is small (e.g. $\mathcal{L}_n = 5$), and when ϵ_t is dependent under H_0. We perform an asymptotic test based on the $N(0,1)$ distribution, and bootstrapped tests with and without the correlation first order expansion. Observe that for a given $\mathcal{L}_n$, Hong's (1996) $\mathcal{N}_n$ is just an affine transformation of the Ljung-Box statistic. Thus, a bootstrapped test based on $\mathcal{N}_n$ is identical to a bootstrap Q-test when each is computed with $\mathcal{L}_n$ lags.

Shao's Cramér von-Mises Test Shao's (2011) test is based on the sample spectral distribution function $F_n(\lambda) = \int_0^\lambda I_n(\omega)d\omega$ with periodogram $I_n(\omega) = (2\pi)^{-1}\sum_{h=1-n}^{n-1}\hat{\gamma}_n(h)e^{-h\omega}$. The core statistic is $S_n(\lambda) \equiv \sqrt{n}(F_n(\lambda) - \hat{\gamma}_n(0)\psi_0(\lambda)) = \sum_{h=1}^{n-1}\sqrt{n}\hat{\gamma}_n(h)\psi_h(\lambda)$ 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 $\mathcal{C}_n = \int_0^\pi S_n^2(\lambda)d\lambda$. The statistic has a non-standard limit distribution, hence we use the WB and DWB, applied both to the correlation first order expansion and to $\hat{\rho}_n(h)$ itself.

Shao (2011) does not verify if his bootstrapped CvM test is valid for regression residuals, although it likely is (we have not verified it). We therefore also perform Shao's (2011) CvM test using Zhu and Li's (2015) block-wise random weighting bootstrap [BRWB]. This affords a first time comparison between DWB and BRWB for the CvM test, which is of separate interest.[10]

The BRWB algorithm (without the first-order expansion here for simplicity) is as follows. Suppose that the objective function to be minimized is written as $1/n\sum_{t=1}^n l_t(\theta)$. Set a block size b_n, $1 \leq b_n < n$, and denote the blocks by $\mathcal{B}_s = \{(s-1)b_n + 1, \ldots, sb_n\}$ with $s = 1, \ldots, n/b_n$. Assume n/b_n is an integer for simplicity. Generate positive i.i.d. random numbers $\{\delta_1, \ldots, \delta_{n/b_n}\}$ from a common distribution with mean 1 and variance 1.[11] Define an auxiliary variable $\omega_t^* = \delta_s$ if $t \in \mathcal{B}_s$. Calculate $\hat{\theta}_n^* = \text{argmin}_{\theta\in\Theta} 1/n\sum_{t=1}^n \omega_t^* l_t(\theta)$. Compute $\hat{\gamma}_n^*(h) = 1/n\sum_{t=1+h}^n \omega_t^*\epsilon_t(\hat{\theta}_n^*)\epsilon_{t-h}(\hat{\theta}_n^*)$ and $S_n^*(\lambda) = \sum_{h=1}^{n-1}\sqrt{n}\hat{\gamma}_n^*(h)\psi_h(\lambda)$, where $\psi_h(\lambda) = (h\pi)^{-1}\sin(h\lambda)$. Define the bootstrapped process

[10] We apply the BRWB for both AR residuals by least squares and GARCH residuals by QML, although Zhu and Li (2015) only verify the validity of their test for least squares residuals from an ARMA model.

[11] Following Zhu and Li (2015), we use the Bernoulli distribution with $P[\delta_t = 0.5 \times (3 - \sqrt{5})] = (2\sqrt{5})^{-1} \times (1 + \sqrt{5})$ and $P[\delta_t = 0.5 \times (3 + \sqrt{5})] = 1 - (2\sqrt{5})^{-1} \times (1 + \sqrt{5})$.

$\Delta_n(\lambda) = S_n^*(\lambda) - S_n(\lambda) - Z_n(\lambda)$, where $Z_n(\lambda) = n^{-1/2}\sum_{h=1}^{n-1}\{\sum_{t=1+h}^{n}\omega_t^* - n + h\}\hat{\gamma}_n(h)\psi_h(\lambda)$. Then compute the bootstrapped test statistic $\mathcal{C}_n^* = \int_0^\pi \{\Delta_n(\lambda)\}^2 d\lambda$. Repeat M times, resulting in the sequence $\{\mathcal{C}_{n,j}^*\}_{j=1}^M$ and approximate p-value $1/M\sum_{j=1}^M I(\mathcal{C}_{n,j}^* \geq \mathcal{C}_n)$.

When we implement the first-order expansion, the bootstrapped autocovariance is computed as $\hat{\gamma}_n^*(h)$ $= 1/n\sum_{t=1+h}^n \omega_t^*[\epsilon_t(\hat{\theta}_n^*)\epsilon_{t-h}(\hat{\theta}_n^*) - \hat{\mathcal{D}}_n^{*\prime}(h) \times \hat{\mathcal{A}}_n \times m_t(\hat{\theta}_n^*)]$.

Delgado and Velasco's Q-Test Delgado and Velasco's (2011) Q-statistic is $n\sum_{h=1}^{\mathcal{L}-k_\theta}(\bar{\rho}_{n,\hat{\theta}_n}^{(\mathcal{L}_n)}(h))^2$ with the transformed correlations. Computation of $\bar{\rho}_{n,\hat{\theta}_n}^{(\mathcal{L}_n)}(h)$ requires a consistent estimator of $\mathcal{V}^{(\mathcal{L})} = (E[\epsilon_t^2])^{-2}\mathcal{S}^{(\mathcal{L})}$ under H_0, where $\mathcal{S}^{(\mathcal{L})} \equiv [\sum_{l=-\infty}^{\infty} E[\epsilon_t\epsilon_{t-i}\epsilon_{t-l}\epsilon_{t-l-j}]]_{i,j=1}^{\mathcal{L}}$. Define $\hat{\gamma}_n^2(0) \equiv 1/n\sum_{t=1}^n \epsilon_t^2(\hat{\theta}_n)$, and let $\hat{\mathcal{V}}_n^{(\mathcal{L})}$ be an estimator of $\mathcal{V}^{(\mathcal{L})}$. We use $\hat{\mathcal{V}}_n^{(\mathcal{L})} = I_{\mathcal{L}}$, which is appropriate when ϵ_t is iid; or $\hat{\mathcal{V}}_n^{(\mathcal{L})} = \hat{\gamma}_n^{-2}(0)\widehat{\mathcal{S}}_n^{(\mathcal{L})}$ with the kernel estimator $\widehat{\mathcal{S}}_n^{(\mathcal{L})}$ in Lobato, Nankervis, and Savin (2002, eq. 5) with Bartlett kernel and bandwidth $2(n/100)^{1/3} \approx .431n^{1/3}$, and no prewhitening, exactly as in Delgado and Velasco (2011, p. 951). The Q-statistic has an asymptotic $\chi^2(\mathcal{L} - k_\theta)$ distribution under Assumptions 1 and 2. The orthogonal transform is used to correct for the impact of weak dependence and a plug-in estimator, but their simulation experiments show size distortions for even small $\mathcal{L}$ (and small k_θ), and lower power especially at higher lags when a fully robust covariance matrix is used. We therefore also bootstrap the test with WB and DWB. The test is a fixed lag length Q-test by construction. In order to make a fair comparison across tests, we use $\mathcal{L} = \mathcal{L}_n$.

Under the null hypothesis, the wild bootstrap is appropriate when the test series is iid or mds. In the case of a bilinear y_t with a mean filter, and iid or GARCH errors, ϵ_t is white noise but not mds. This suggests size distortions may arise in that case. Under the alternative hypothesis, the wild bootstrap cannot in theory accurately approximate the dependence structure in any model when ϵ_t is MA or AR, hence power should be smaller in small samples in such cases. As stated above, however, we find the WB works roughly on par with the DWB, at least for our chosen design. Finally, we have from Theorem 3.1 asymptotic validity of the (dependent) wild bootstrap under the appropriate assumptions.

5.2 Simulation Results

See Tables 2-9 for rejection frequencies. In order to save space, we only report dependent wild bootstrap results; for bootstrap tests that can be applied to expanded or non-expanded correlations, we only report results based on the expansion; and we only look at sample sizes 100 and 500. The supplemental material Hill and Motegi (2016, Appendix G) contains all simulation results. The WB typically results in size and power on par with the DWB, with higher power than the DWB in some cases.

5.2.1 Empirical Size

We use short hand notation for brevity. The *simple process* is $y_t = e_t$; *max-trans* denotes the max-correlation test with the transformed correlations; *max-corr* denotes the test with the non-transformed correlations; *DV* is the Q-test in Delgado and Velasco (2011); *CvM* is the Cramér-von Mises test. Rejection frequencies are stated for $\{1\%, 5\%, 10\%\}$ nominal levels. There are several major points to note.

1. The max-trans test generally works as well as the max-corr test at the lowest maximum lag $\mathcal{L}_n = 5$. As $\mathcal{L}_n$ increases the max-trans test performs precipitously worse in terms of extreme under-sizedness. The distortions tend to be worse when the errors are GARCH. Both properties are not implausible. A large $\mathcal{L}_n$ implies we need to invert a large dimensional covariance matrix for standardization, and kernel variance estimators need not adequately match finite sample dependence properties. The latter is compounded under GARCH errors.[12] As we discuss below, the DV test exhibits similar size distortions.

Overall, the standardization is precarious for large dimensional problems, making the transformation far less attractive than the first order expansion as a means for handling the impact of plug-in estimators.

2. The orthogonalization completely breaks down when the process is GARCH and a GARCH filter is estimated: the transformed correlations are minute, irrespective of serial dependence, for any sample size n and any lag $\mathcal{L}$. The problem exists irrespective of whether the identity matrix or kernel variance are used, and irrespective of the chosen bandwidth. The problem therefore is the orthogonalization itself, rather than the standardization. See Appendix F in Hill and Motegi (2016) for a numerical example that shows the under-workings of the transformed correlation failure. Note that Delgado and Velasco (2011, Section 4) only consider ARCH(1) errors in their simulation study and use a different estimator (Whittle maximum likelihood). Our study therefore reveals that a more realistic error dependence structure is challenging for the orthogonalized correlations to deal with. In general, therefore, the first order asymptotic expansion is superior for size control when filtered residuals are used.

[12] In addition to the bandwidth $2(n/100)^{1/3} \approx .431n^{1/3}$, we tried a variety of other bandwidths $\lambda n^{1/3}$ for $\lambda \in (0, 1)$. Size distortions for the max-trans test are not alleviated. Similarly, the DV test performs equally as well for other chosen bandwidths.

3. Hong's (1996) test generally exhibits size distortions, for both asymptotic and bootstrapped tests. Often the asymptotic test is over-sized and the bootstrapped test is under-sized. For example, in the simple process case with iid error and $(n, \mathcal{L}_n) = (100, 5)$, asymptotic and bootstrap test sizes are $\{.024, .040, .058\}$ and $\{.012, .048, .111\}$, and at $\mathcal{L}_n = 21$ they are $\{.024, .097, .202\}$ and $\{.002, .021, .062\}$. The asymptotic test requires the lag length to increase to infinity. However, for the simple process with iid error and $(n, \mathcal{L}_n) = (500, 80)$ size is not improved: asymptotic and bootstrap test sizes yield $\{.037, .136, .231\}$ and $\{.000, .005, .022\}$. In the bilinear case with GARCH errors and $(n, \mathcal{L}_n) = (100, 21)$ we obtain $\{.135, .220, .312\}$ and $\{.000, .006, .029\}$, and for GARCH process with iid error and $(n, \mathcal{L}_n)$ $= (500, 80)$ the respective frequencies are $\{.043, .139, .219\}$ and $\{.000, .001, .017\}$. Similar rates appear for the AR(2) process with AR(2) filter, and GARCH processes with GARCH filter.

4. Asymptotic and bootstrapped DV tests in many cases exhibit asymmetric size distortions. The simple process with iid error and $(n, \mathcal{L}_n) = (100, 5)$ yields respective rates $\{.002, .027, .082\}$ and $\{.009, .055, .145\}$. Compare this to the max-corr and max-trans test: $\{.013, .051, .123\}$ and $\{.004, .045, .103\}$. If the errors are GARCH the DV tests result in $\{.001, .024, .058\}$ and $\{.005, .054, .111\}$, compared to max-corr and max-trans $\{.008, .059, .133\}$ and $\{.010, .053, .134\}$.

At higher lags there are greater distortions in many cases. If the process is simple with an iid error, when $(n, \mathcal{L}_n) = (100, 21)$ the asymptotic and bootstrap DV tests yield $\{.008, .025, .046\}$ and $\{.000, .001, .020\}$, compared to the max-corr and max-trans $\{.003, .030, .085\}$ and $\{.002, .018, .071\}$.

In the bilinear case with GARCH error and $(n, \mathcal{L}_n) = (100, 21)$ the DV rates are $\{.206, .238, .255\}$ and $\{.000, .000, .008\}$, suggesting that neither the asymptotic distribution, nor the bootstrap, can well approximate this test statistic's null distribution. By comparison, the max-corr test and max-trans test yield $\{.000, .005, .044\}$ and $\{.000, .005, .024\}$. Both are undersized, with a larger distortion by max-trans, but both beat the DV test. At a higher sample size $n = 500$ the pattern repeats.

At very high lags, e.g. $(n, \mathcal{L}_n) = (500, 80)$, the DV test performs very poorly, with size near zero in several cases (simple-iid, AR(2)-iid or GARCH with AR(2) filter). Finally, as detailed above the DV test breaks down when a GARCH filter is used, evidently due to the orthogonalization itself. The transformed correlations are tiny irrespective of sample size, lag length, dependence, or variance estimator. Hence, rejection frequencies are simply zero in all cases. See Hill and Motegi (2016, Appendix F) for numerical examples.

5. The CvM test generally exhibits larger size distortions than the max-corr test, using either the dependent wild or block-wise random weighting bootstrap. This includes the simple process with iid error $(n = 100)$ or GARCH error; bilinear process with iid or GARCH error; AR(2) process with AR(2) filter and iid or GARCH error; and GARCH process with or without a GARCH filter, and iid or GARCH errors. The size distortions typically result in a modestly over-sized test. See Table 9.

6. The max-corr test generally results in the sharpest empirical size when all cases are considered. It is evidently the most easily bootstrapped, even across the most difficult cases in this study. The arguable reason is the simplicity of the test statistic. The more challenging processes are bilinear with GARCH error and the AR(2) with GARCH error and AR(2) filter. Further, a white noise test requires an increasing lag, and at all lags the test results in competitive empirical size. In cases where the test exhibits a size distortion, it is usually under-sized, and in most of those cases it yields a smaller distortion than other tests.

As examples, consider the bilinear process with GARCH error and $(n, \mathcal{L}_n) = (100, 5)$ (Table 3). The rejection rates are $\{.004, .027, .089\}$, while Hong's bootstrapped test yields $\{.002, .018, .060\}$, the DV test without and with bootstrap yield $\{.019, .035, .068\}$ and $\{.000, .017, .069\}$, and the CvM test results in $\{.002, .030, .070\}$ (recall the CvM test does not have the maximum lag $\mathcal{L}_n$). At a higher lag $\mathcal{L}_n = 21$, the max-corr test is under-sized with rates $\{.000, .005, .044\}$. By comparison, the bootstrapped Hong test is similarly under-sized $\{.000, .006, .029\}$; the asymptotic DV test is heavily over-sized with $\{.206, .238, .255\}$; the bootstrapped DV test is heavily under-sided with $\{.000, .000, .008\}$.

In the AR(2) case with AR(2) filter, the max-corr test is slightly over-sized at low lags and small n (Table 4). When the error is iid or GARCH and $(n, \mathcal{L}_n) = (100, 5)$ the rates are $\{.031, .095, .159\}$ and $\{.017, .079, .169\}$. At lag $\mathcal{L}_n = 21$ the rates are $\{.002, .037, .105\}$ and $\{.005, .038, .107\}$. The rejection rates are sharper at larger sample sizes.

5.2.2 Empirical Power

Since max-trans, Hong's (1996) asymptotic, and Delgado and Velasco's (2011) bootstrapped tests yield the largest size distortions, we do not discuss (or give cursory discussion of) their power.

1. Hong's (1996) bootstrapped test yields high power at low lags, but has comparatively lower power at higher lags in most cases. The largest discrepancies arise in the simple, bilinear, GARCH (with and without a filter) processes with MA or AR errors. The differentials are smaller or do not exist in the AR process with either filter. However, the test is incapable of detecting remote dependence at small sample sizes (Table 8). When $n = 500$ and the lag is large enough then power is non-trivial, but substantially smaller than the max-corr test power.

2. The DV test has competitive power in many cases at low maximum lags, but struggles at higher lags in every case. The orthogonal transform spread over a larger window of lags seems to add sufficient sampling noise such that the true non-zero correlations cannot be detected well. Indeed, power is typically a small fraction of the max-corr test at larger lags.

Further, the test is incapable of detecting remote dependence in the MA(q) model, for any $q \in \{12, 24, 48\}$ and at any sample size. Indeed, the test fails altogether when $(n, \mathcal{L}_n)$ are large, e.g. MA(48), $n = 1000$ and $\mathcal{L}_n \in \{72, 144\}$. See Hill and Motegi (2016, Table 22).

3. The CvM test is very competitive in terms of raw or size corrected power (the latter is not shown in the tables). Its main shortcoming is its complete inability to detect remote dependence in the MA(q) model, for any $q \in \{12, 24, 48\}$ and at any sample size. This holds for each bootstrap method under consideration. See the bottom panel of Table 9, and see Hill and Motegi (2016, Table 29).

4. The max-corr test has the highest power in many cases, and competitive power in the remaining cases, and does not exhibit a sharp decline in power at higher lags. In general, the max-trans test yields lower (much lower in some cases) power. As examples, in the bilinear model with MA error and $(n, \mathcal{L}_n) = (100, 21)$, max-corr rejection rates are $\{.240, .569, .765\}$, but Hong's (1996) bootstrapped test yields $\{.016, .144, .333\}$, the asymptotic and bootstrapped DV tests bring $\{.185, .242, .293\}$ and $\{.000, .032, .147\}$ and the CvM test yields $\{.450, .743, .866\}$.

Across cases the CvM test is the best competitor with the max-corr test. However, as stated above the CvM test cannot detect remote serial correlation by any of the three bootstrap methods considered. The max-corr test, however, is easily capable of detecting remote dependence once the maximum lag is large enough, in particular for $n \geq 250$ (Hill and Motegi, 2016, Tables 22.B-D). As an example, for the MA(24) model with $(n, \mathcal{L}_n) = (250, 45)$ the max-corr test with dependent wild bootstrap yields $\{.220, .526, .664\}$, Hong's bootstrapped test yields $\{.019, .133, .311\}$, and asymptotic and bootstrapped DV tests result in $\{.012, .036, .061\}$ and $\{.000, .000, .005\}$ (Hill and Motegi, 2016, Table 22.B). The CvM test with wild, dependent wild, and random weighting bootstrap brings $\{.010, .065, .121\}$,

$\{.022, .088, .143\}$ and $\{.017, .083, .137\}$ (Hill and Motegi, 2016, Table 29).

5. The max-correlation test dominates Ljung-Box and Andrews and Ploberger's (1996) sup-LM tests when all cases are considered (note that in any finite sample bootstrapped Ljung-Box and Hong 1996 tests are equivalent). The strongest competitor there is the sup-LM test, but it is incapable of detecting remote serial dependence of the types considered. See Hill and Motegi (2016).

6 Conclusion

We present a bootstrap max-correlation test of the white noise hypothesis. 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. We apply a bootstrap to a first order correlation expansion in order to account for the impact of a plug-in estimator, or we use Delgado and Velasco's (2011) orthogonalized correlation with a fixed maximum lag. We prove Shao's (2011) dependent wild bootstrap yields a valid test using either sample correlation, in a more general environment than Shao (2011) or Xiao and Wu (2014) used. The limit distribution of a suitably normalized max-correlation is not required to show that the original and bootstrapped test statistics have the same limit distribution under the null, allowing us to bypass the extreme value theory approach altogether. We allow for a general class of models and broad range of estimators in order to derive filtered residuals.

A simulation study shows that the max-correlation test, where the bootstrap operates on the first order correlation expansion, generally dominates each test considered: the spectrogram-based test in Hong (1996), the Q-test in Delgado and Velasco (2011) with orthogonalized correlations, and the Cramér-von Mises test in Shao (2011) with either Shao's (2011) dependent wild bootstrap or Zhu and Li's (2015) block-wise random weighting bootstrap. The strongest competitor with the max-correlation test in this paper is the Cramér-von Mises test, although that test leads to larger size distortions than the max-correlation test. Further, in terms of power against possibly very remote serial dependence, the max-correlation test dominates all tests, and the Cramér-von Mises exhibits roughly trivial power based on all three bootstrap methods considered (wild, dependent wild, random weighting). In the supplemental material we compare our test with the Ljung-Box Q-test and Andrews and Ploberger's (1996) sup-LM test. The max-correlation test dominates when all cases are considered.

We only study the maximum correlation for a white noise test. Another obvious application is a test of many zero restrictions in a regression model, when possibly infinitely many regressors exist. Many possible applications exist, including a white noise test couched in an AR(p_n) where $p_n \to \infty$ as $n \to \infty$, but also tests where many redundant parameters can lead to a poor sized asymptotic test, or lower power bootstrapped test. Penalized estimators like lasso can impart shrinkage based on a sparsity assumption, but inference on a parameter subset may also be desired, including a target subset set to zero by the penalty function. We leave this idea for future consideration.

References

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

——— (2001): "Testing when a Parameter Is on the Boundary of the Maintained Hypothesis," *Econometrica*, 69, 683–734.

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.

BERMAN, S. M. (1964): "Limit Theorems for the Maximum Term in Stationary Sequences," *Annals of Mathematical Statistics*, 35, 502–516.

BERNSTEIN, S. (1927): "Sur l'Extension du Théoréme Limite du Calcul des Probabilités aux Sommes de Quantités Dependantes," *Mathematische Annalen*, 97, 1–59.

BOLLERSLEV, T. (1986): "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, 307–327.

BOUSSAMA, F. (2006): "Ergodicité des Chaînes de Markov à Valeurs dans une Variété Algébrique: Application aux Modèles GARCH Multivariés," *Comptes Rendus Mathematique*, 343, 275–278.

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.

BRADLEY, R. C. (1993): "Equivalent Mixing Condition for Random Fields," *Annals of Probability*, 21, 1921–1926.

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.

DOUKHAN, P. (1994): *Mixing: Properties and Examples*. Springer-Verlag New York.

DROST, F. C., AND T. E. NIJMAN (1993): "Temporal Aggregation of GARCH Processes," *Econometrica*, 61, 909–927.

DURLAUF, S. N. (1991): "Spectral Based Testing of the Martingale Hypothesis," *Journal of Econometrics*, 50, 355–376.

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

ESCANCIANO, J. C., AND C. VELASCO (2006): "Generalized Spectral Tests for the Martingale Difference Hypothesis," *Journal of Econometrics*, 134, 151–185.

FRANCQ, C., R. ROY, AND J.-M. ZAKOÏAN (2005): "Diagnostic Checking in ARMA Models with Uncorrelated Errors," *Journal of the American Statistical Association*, 100, 532–544.

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.

GINE, 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. LAZAROVA (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 (2016): "Supplemental Material for "A Max-Correlation White Noise Test for Weakly Dependent Time Series"," Dept. of Economics, University of North Carolina at Chapel Hill.

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

HONG, Y., AND T. H. LEE (2003): "Inference on Predictability of Foreign Exchange Rate Changes via Generalized Spectrum and Nonlinear Time Series Models.," *Review of Economics and Statistics*, 85, 1048–1062.

HONG, Y., AND Y. J. LEE (2005): "Generalized Spectral Tests for Conditional Mean Models in Time Series with Conditional Heteroscedasticity of Unknown Form," *Review of Economic Studies*, 72, 499–541.

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.

IBRAGIMOV, I. A. (1975): "A Note on the Central Limit Theorems for Dependent Random Variables," *Theory of Probability and its Applications*, 20, 135–141.

JENSEN, S. T., AND A. RAHBEK (2004): "Asymptotic Inference for Nonstationary GARCH," *Econometric Theory*, 20, 1203–1226.

JIRAK, M. (2011): "On the Maximum of Covariance Estimators," *Journal of Multivariate Analysis*, 102, 1032–1046.

KNIGHT, K. (1998): "Limiting Distributions for L_1 Regression Estimators under General Conditions," *Annals of Statistics*, 26, 755–770.

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.

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.

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.

MEITZ, M., AND P. SAIKKONEN (2008): "Stability of Nonlinear AR-GARCH Models," *Journal of Time Series Analysis*, 29, 453–475.

NANKERVIS, J. C., AND N. E. SAVIN (2010): "Testing for Serial Correlation: Generalized Andrews-Ploberger Tests," *Journal of Business and 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.

NEWEY, W., AND K. WEST (1987): "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

NEWEY, W. K. (1991): "Uniform Convergence in Probability and Stochastic Equicontinuity," *Econo-*

metrica, 59, 1161–1167.

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.

ROMANO, J. P., AND L. A. THOMBS (1996): "Inference for Autocorrelations under Weak Assumptions," *Journal of the American Statistical Association*, 91, 590–600.

SAWA, T. (1978): "Information Criteria for Discriminating among Alternative Regression Models," *Econometrica*, 46, 1273–1291.

SHAO, X. (2010a): "The Dependent Wild Bootstrap," *Journal of the American Statistical Association*, 105, 218–235.

——— (2010b): "A Self-Normalized Approach to Confidence Interval Construction in Time Series," *Journal of the Royal Statistical Society Series B*, 72, 343–366.

——— (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.

TERASVIRTA, T. (1994): "Specification, Estimation, and Evaluation of Smooth Transition Autoregressive Models," *Journal of the American Statistical Association*, 89, 208 218.

WHITE, H. (1982): "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50, 1–25.

WOOLDRIDGE, J. M. (1990): "A Unified Approach to Robust, Regression-Based Specification Tests," *Econometric Theory*, 6, 17–43.

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.

WU, W. B., AND X. SHAO (2004): "Limit Theorems for Iterated Random Functions," *Journal of Applied Probability*, 41, 425–436.

XIAO, H., AND W. B. WU (2014): "Portmanteau Test and Simultaneous Inference for Serial Covariances," *Statistica Sinica*, 24, 577–600.

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: Rejection Frequencies – Simple $y_t = e_t$, Mean Plug-in

$n = 100$

e_t is iid				e_t is GARCH: $w_t^2 = 1.0 + 0.2e_{t-1}^2 + 0.5w_{t-1}^2$			
	Lag = 5	Lag = 10	Lag = 21		Lag = 5	Lag = 10	Lag = 21
	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%		1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%
$\hat{\mathcal{T}}_{ex}^{dw}$	.013, .051, .123	.008, .042, .096	.003, .030, .085	$\hat{\mathcal{T}}_{ex}^{dw}$	.008, .059, .133	.007, .033, .099	.002, .024, .073
$\bar{\mathcal{T}}_{o:K}^{dw}$	.004, .045, .103	.007, .042, .111	.002, .018, .071	$\bar{\mathcal{T}}_{o:K}^{dw}$	.010, .053, .134	.003, .028, .092	.001, .011, .050
N	.024, .040, .058	.023, .047, .089	.024, .097, .202	N	.060, .093, .120	.039, .076, .121	.044, .109, .186
N_{ex}^{dw}	.012, .048, .111	.005, .029, .100	.002, .021, .062	N_{ex}^{dw}	.001, .033, .095	.001, .029, .071	.001, .008, .061
$DV_{o:K}$	.002, .027, .082	.001, .012, .036	.008, .025, .046	$DV_{o:K}$	.001, .024, .058	.005, .021, .051	.043, .077, .100
$DV_{o:K}^{dw}$	.009, .055, .145	.001, .022, .071	.000, .001, .020	$DV_{o:K}^{dw}$	.005, .054, .111	.002, .011, .059	.000, .002, .013
e_t is MA(2): $e_t = \nu_t + 0.50\nu_{t-1} + 0.25\nu_{t-2}$				e_t is AR(1): $e_t = 0.7e_{t-1} + \nu_t$			
	Lag = 5	Lag = 10	Lag = 21		Lag = 5	Lag = 10	Lag = 21
	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%		1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%
$\hat{\mathcal{T}}_{ex}^{dw}$	.824, .971, .987	.714, .946, .979	.628, .924, .970	$\hat{\mathcal{T}}_{ex}^{dw}$	.946, .997, 1.00	.929, .998, 1.00	.922, 1.00, 1.00
$\bar{\mathcal{T}}_{o:K}^{dw}$	.728, .935, .977	.617, .914, .977	.346, .743, .873	$\bar{\mathcal{T}}_{o:K}^{dw}$	.906, .997, .999	.909, .993, .999	.742, .945, .988
N	.955, .978, .986	.883, .932, .955	.730, .815, .860	N	1.00, 1.00, 1.00	1.00, 1.00, 1.00	.994, .999, .999
N_{ex}^{dw}	.458, .823, .949	.141, .584, .831	.052, .278, .522	N_{ex}^{dw}	.731, .961, .997	.460, .881, .974	.164, .580, .858
$DV_{o:K}$	.462, .849, .935	.108, .412, .640	.102, .206, .301	$DV_{o:K}$	.806, .986, .995	.272, .718, .884	.172, .278, .383
$DV_{o:K}^{dw}$	.538, .886, .964	.109, .576, .829	.004, .098, .340	$DV_{o:K}^{dw}$	.779, .983, .997	.364, .874, .964	.017, .275, .630

$n = 500$

e_t is iid				e_t is GARCH: $w_t^2 = 1.0 + 0.2e_{t-1}^2 + 0.5w_{t-1}^2$			
	Lag = 5	Lag = 40	Lag = 80		Lag = 5	Lag = 40	Lag = 80
	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%		1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%
$\hat{\mathcal{T}}_{ex}^{dw}$	.008, .047, .114	.004, .029, .078	.000, .027, .068	$\hat{\mathcal{T}}_{ex}^{dw}$	.008, .056, .123	.000, .026, .074	.002, .025, .059
$\bar{\mathcal{T}}_{o:K}^{dw}$	.013, .052, .111	.007, .032, .062	.001, .013, .055	$\bar{\mathcal{T}}_{o:K}^{dw}$	.005, .053, .105	.000, .022, .052	.000, .006, .031
N	.023, .050, .074	.011, .054, .110	.037, .136, .231	N	.074, .126, .152	.025, .079, .141	.034, .114, .199
N_{ex}^{dw}	.007, .061, .117	.000, .013, .050	.000, .005, .022	N_{ex}^{dw}	.002, .032, .094	.000, .005, .035	.000, .001, .015
$DV_{o:K}$	.007, .042, .092	.000, .000, .005	.001, .004, .009	$DV_{o:K}$	.008, .036, .084	.003, .007, .013	.014, .025, .043
$DV_{o:K}^{dw}$	.013, .055, .136	.000, .000, .004	.000, .000, .001	$DV_{o:K}^{dw}$	.006, .043, .102	.000, .000, .006	.000, .000, .001
e_t is MA(2): $e_t = \nu_t + 0.50\nu_{t-1} + 0.25\nu_{t-2}$				e_t is AR(1): $e_t = 0.7e_{t-1} + \nu_t$			
	Lag = 5	Lag = 40	Lag = 80		Lag = 5	Lag = 40	Lag = 80
	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%		1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%
$\hat{\mathcal{T}}_{ex}^{dw}$	1.00, 1.00, 1.00	1.00, 1.00, 1.00	1.00, 1.00, 1.00	$\hat{\mathcal{T}}_{ex}^{dw}$	1.00, 1.00, 1.00	1.00, 1.00, 1.00	1.00, 1.00, 1.00
$\bar{\mathcal{T}}_{o:K}^{dw}$	.998, .998, .999	1.00, 1.00, 1.00	1.00, 1.00, 1.00	$\bar{\mathcal{T}}_{o:K}^{dw}$	1.00, 1.00, 1.00	1.00, 1.00, 1.00	1.00, 1.00, 1.00
N	1.00, 1.00, 1.00	1.00, 1.00, 1.00	1.00, 1.00, 1.00	N	1.00, 1.00, 1.00	1.00, 1.00, 1.00	1.00, 1.00, 1.00
N_{ex}^{dw}	1.00, 1.00, 1.00	.912, .999, .999	.283, .889, .984	N_{ex}^{dw}	1.00, 1.00, 1.00	.963, 1.00, 1.00	.784, .996, 1.00
$DV_{o:K}$	.999, .999, 1.00	.500, .905, .974	.051, .137, .223	$DV_{o:K}$	1.00, 1.00, 1.00	.636, .955, .994	.077, .163, .247
$DV_{o:K}^{dw}$	.999, 1.00, 1.00	.417, .976, .998	.000, .070, .369	$DV_{o:K}^{dw}$	1.00, 1.00, 1.00	.824, .998, 1.00	.009, .324, .749

Rejection frequencies (1%, 5%, 10%). $\hat{\mathcal{T}}$ is the max-correlation test; $\bar{\mathcal{T}}$ is the max-transformed test; N is Hong's (1996) test; DV is the Q-test by Delgado and Velasco (2011); "dw" implies the dependent wild bootstrap is used; "ex" signifies the correlation first order expansion is used for the bootstrap; "o:K" signifies an orthogonalized correlation with a kernel-based covariance matrix; lag length is 5, $[.5n/\ln(n)]$, or $[n/\ln(n)]$.

Table 3: Rejection Frequencies – Bilinear $y_t = 0.50e_{t-1}y_{t-2} + e_t$, Mean Plug-in

$n = 100$

e_t is iid				e_t is GARCH: $w_t^2 = 1.0 + 0.2e_{t-1}^2 + 0.5w_{t-1}^2$			
	Lag = 5	Lag = 10	Lag = 21		Lag = 5	Lag = 10	Lag = 21
	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%		1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%
$\hat{\mathcal{T}}_{ex}^{dw}$	.009, .067, .127	.003, .033, .104	.002, .027, .087	$\hat{\mathcal{T}}_{ex}^{dw}$	.004, .027, .089	.000, .026, .063	.000, .005, .044
$\bar{\mathcal{T}}_{o:K}^{dw}$	.007, .063, .131	.005, .044, .110	.001, .016, .057	$\bar{\mathcal{T}}_{o:K}^{dw}$	.005, .030, .081	.001, .011, .047	.000, .005, .024
N	.074, .120, .148	.054, .094, .145	.039, .121, .212	N	.325, .397, .449	.267, .346, .406	.135, .220, .312
N_{ex}^{dw}	.005, .039, .109	.002, .022, .075	.001, .020, .061	N_{ex}^{dw}	.002, .018, .060	.000, .007, .034	.000, .006, .029
$DV_{o:K}$	.002, .029, .071	.006, .019, .056	.047, .071, .098	$DV_{o:K}$	.019, .035, .068	.068, .092, .108	.206, .238, .255
$DV_{o:K}^{dw}$	.009, .058, .139	.001, .021, .086	.000, .002, .018	$DV_{o:K}^{dw}$	.000, .017, .069	.000, .006, .042	.000, .000, .008
e_t is MA(2): $e_t = \nu_t + 0.50\nu_{t-1} + 0.25\nu_{t-2}$				e_t is AR(1): $e_t = 0.7e_{t-1} + \nu_t$			
	Lag = 5	Lag = 10	Lag = 21		Lag = 5	Lag = 10	Lag = 21
	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%		1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%
$\hat{\mathcal{T}}_{ex}^{dw}$	.341, .655, .827	.296, .603, .772	.240, .569, .765	$\hat{\mathcal{T}}_{ex}^{dw}$	.233, .511, .679	.251, .491, .673	.241, .495, .672
$\bar{\mathcal{T}}_{o:K}^{dw}$	.298, .597, .772	.233, .539, .704	.109, .345, .512	$\bar{\mathcal{T}}_{o:K}^{dw}$	.211, .464, .642	.169, .405, .585	.111, .316, .472
N	.852, .904, .924	.737, .839, .875	.594, .675, .725	N	.984, .993, .993	.950, .966, .973	.902, .924, .939
N_{ex}^{dw}	.175, .479, .699	.053, .308, .563	.016, .144, .333	N_{ex}^{dw}	.169, .452, .640	.083, .348, .566	.035, .214, .448
$DV_{o:K}$	.264, .591, .760	.104, .263, .440	.185, .242, .293	$DV_{o:K}$	.300, .596, .746	.131, .272, .390	.232, .277, .318
$DV_{o:K}^{dw}$	.263, .631, .796	.046, .309, .553	.000, .032, .147	$DV_{o:K}^{dw}$	.155, .472, .660	.040, .231, .434	.002, .044, .148

$n = 500$

e_t is iid				e_t is GARCH: $w_t^2 = 1.0 + 0.2e_{t-1}^2 + 0.5w_{t-1}^2$			
	Lag = 5	Lag = 40	Lag = 80		Lag = 5	Lag = 40	Lag = 80
	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%		1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%
$\hat{\mathcal{T}}_{ex}^{dw}$	.007, .063, .111	.003, .019, .066	.003, .023, .063	$\hat{\mathcal{T}}_{ex}^{dw}$	.008, .022, .053	.007, .026, .048	.002, .013, .032
$\bar{\mathcal{T}}_{o:K}^{dw}$	.011, .061, .114	.000, .029, .069	.003, .006, .030	$\bar{\mathcal{T}}_{o:K}^{dw}$	.002, .016, .055	.000, .003, .020	.001, .003, .008
N	.101, .158, .201	.030, .078, .142	.033, .110, .183	N	.771, .805, .827	.537, .610, .653	.383, .487, .553
N_{ex}^{dw}	.006, .056, .112	.000, .006, .044	.000, .004, .021	N_{ex}^{dw}	.008, .011, .027	.002, .005, .017	.001, .003, .005
$DV_{o:K}$	.006, .054, .103	.000, .002, .006	.009, .021, .035	$DV_{o:K}$	.008, .012, .019	.099, .112, .121	.225, .245, .253
$DV_{o:K}^{dw}$	.008, .056, .135	.000, .000, .008	.000, .000, .002	$DV_{o:K}^{dw}$	.000, .006, .032	.000, .001, .005	.000, .000, .000
e_t is MA(2): $e_t = \nu_t + 0.50\nu_{t-1} + 0.25\nu_{t-2}$				e_t is AR(1): $e_t = 0.7e_{t-1} + \nu_t$			
	Lag = 5	Lag = 40	Lag = 80		Lag = 5	Lag = 40	Lag = 80
	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%		1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%
$\hat{\mathcal{T}}_{ex}^{dw}$	.901, .979, .991	.896, .972, .992	.891, .968, .985	$\hat{\mathcal{T}}_{ex}^{dw}$	.376, .624, .760	.370, .614, .729	.396, .635, .769
$\bar{\mathcal{T}}_{o:K}^{dw}$	.906, .982, .992	.808, .937, .970	.666, .854, .911	$\bar{\mathcal{T}}_{o:K}^{dw}$	.405, .651, .765	.249, .458, .591	.170, .325, .438
N	1.00, 1.00, 1.00	.999, .999, .999	.991, .998, .998	N	1.00, 1.00, 1.00	1.00, 1.00, 1.00	1.00, 1.00, 1.00
N_{ex}^{dw}	.750, .917, .963	.327, .786, .920	.067, .482, .779	N_{ex}^{dw}	.328, .593, .752	.177, .503, .699	.091, .384, .613
$DV_{o:K}$	.969, .992, .997	.148, .339, .481	.174, .232, .277	$DV_{o:K}$	.556, .754, .826	.105, .137, .167	.223, .246, .255
$DV_{o:K}^{dw}$	.954, .990, .996	.034, .434, .750	.000, .010, .081	$DV_{o:K}^{dw}$	.454, .700, .820	.004, .106, .251	.000, .002, .020

Rejection frequencies (1%, 5%, 10%). $\hat{\mathcal{T}}$ is the max-correlation test; $\bar{\mathcal{T}}$ is the max-transformed test; N is Hong's (1996) test; DV is the Q-test by Delgado and Velasco (2011); "dw" implies the dependent wild bootstrap is used; "ex" signifies the correlation first order expansion is used for the bootstrap; "o:K" signifies an orthogonalized correlation with a kernel-based covariance matrix; lag length is 5, $[.5n/\ln(n)]$, or $[n/\ln(n)]$.

Table 4: Rejection Frequencies – AR(2) $y_t = 0.30y_{t-1} - 0.15y_{t-2} + e_t$, AR(2) Plug-in

$n = 100$

e_t is iid				e_t is GARCH: $w_t^2 = 1.0 + 0.2e_{t-1}^2 + 0.5w_{t-1}^2$			
	Lag = 5	Lag = 10	Lag = 21		Lag = 5	Lag = 10	Lag = 21
	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%		1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%
$\hat{\mathcal{T}}_{ex}^{dw}$	.031, .095, .159	.014, .072, .156	.002, .037, .105	$\hat{\mathcal{T}}_{ex}^{dw}$	.017, .079, .169	.009, .058, .147	.005, .038, .107
$\bar{\mathcal{T}}_{o:K}^{dw}$	.021, .096, .171	.008, .054, .121	.001, .023, .066	$\bar{\mathcal{T}}_{o:K}^{dw}$	.020, .076, .146	.005, .044, .125	.000, .022, .076
N	.004, .009, .014	.006, .030, .104	.015, .109, .229	N	.006, .015, .023	.005, .023, .095	.019, .110, .230
N_{ex}^{dw}	.028, .087, .174	.006, .055, .136	.002, .031, .091	N_{ex}^{dw}	.015, .086, .174	.004, .053, .127	.001, .015, .064
$DV_{o:K}$	.001, .033, .080	.001, .014, .036	.012, .035, .053	$DV_{o:K}$	.003, .028, .072	.002, .015, .039	.031, .057, .084
$DV_{o:K}^{dw}$	.019, .087, .170	.000, .025, .089	.000, .001, .016	$DV_{o:K}^{dw}$	.017, .074, .169	.000, .017, .081	.000, .000, .022

e_t is MA(2): $e_t = \nu_t + 0.50\nu_{t-1} + 0.25\nu_{t-2}$				e_t is AR(1): $e_t = 0.7e_{t-1} + \nu_t$			
	Lag = 5	Lag = 10	Lag = 21		Lag = 5	Lag = 10	Lag = 21
	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%		1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%
$\hat{\mathcal{T}}_{ex}^{dw}$	.037, .139, .239	.014, .086, .159	.007, .044, .126	$\hat{\mathcal{T}}_{ex}^{dw}$	.043, .124, .235	.022, .081, .184	.008, .049, .116
$\bar{\mathcal{T}}_{o:K}^{dw}$	.027, .102, .179	.005, .050, .146	.001, .024, .074	$\bar{\mathcal{T}}_{o:K}^{dw}$	.026, .116, .209	.010, .065, .148	.003, .025, .075
N	.012, .022, .032	.019, .041, .114	.021, .114, .219	N	.014, .029, .041	.018, .051, .110	.029, .097, .202
N_{ex}^{dw}	.033, .119, .221	.011, .075, .162	.001, .033, .092	N_{ex}^{dw}	.039, .148, .245	.011, .071, .180	.003, .041, .109
$DV_{o:K}$	.003, .040, .107	.002, .016, .050	.011, .032, .059	$DV_{o:K}$	.012, .066, .153	.002, .022, .070	.004, .020, .048
$DV_{o:K}^{dw}$	.025, .100, .183	.003, .053, .135	.000, .003, .017	$DV_{o:K}^{dw}$	.027, .114, .228	.003, .049, .134	.000, .003, .024

$n = 500$

e_t is iid				e_t is GARCH: $w_t^2 = 1.0 + 0.2e_{t-1}^2 + 0.5w_{t-1}^2$			
	Lag = 5	Lag = 40	Lag = 80		Lag = 5	Lag = 40	Lag = 80
	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%		1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%
$\hat{\mathcal{T}}_{ex}^{dw}$	.018, .064, .126	.004, .028, .075	.003, .030, .074	$\hat{\mathcal{T}}_{ex}^{dw}$	.010, .060, .116	.004, .032, .069	.002, .021, .067
$\bar{\mathcal{T}}_{o:K}^{dw}$	.015, .066, .124	.005, .035, .077	.001, .014, .044	$\bar{\mathcal{T}}_{o:K}^{dw}$	.010, .055, .113	.002, .022, .055	.003, .017, .040
N	.007, .012, .019	.007, .074, .146	.045, .154, .247	N	.012, .024, .030	.014, .063, .133	.045, .154, .256
N_{ex}^{dw}	.016, .068, .135	.001, .012, .039	.000, .005, .022	N_{ex}^{dw}	.013, .067, .128	.000, .008, .031	.000, .003, .020
$DV_{o:K}$	.007, .048, .091	.000, .001, .008	.000, .001, .004	$DV_{o:K}$	.007, .034, .067	.000, .000, .004	.011, .021, .035
$DV_{o:K}^{dw}$	.014, .067, .134	.000, .002, .019	.000, .000, .000	$DV_{o:K}^{dw}$	.006, .056, .113	.000, .000, .005	.000, .000, .000

e_t is MA(2): $e_t = \nu_t + 0.50\nu_{t-1} + 0.25\nu_{t-2}$				e_t is AR(1): $e_t = 0.7e_{t-1} + \nu_t$			
	Lag = 5	Lag = 40	Lag = 80		Lag = 5	Lag = 40	Lag = 80
	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%		1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%
$\hat{\mathcal{T}}_{ex}^{dw}$	.090, .270, .406	.010, .053, .119	.006, .040, .090	$\hat{\mathcal{T}}_{ex}^{dw}$	.134, .365, .504	.010, .070, .158	.007, .046, .099
$\bar{\mathcal{T}}_{o:K}^{dw}$	.085, .248, .395	.004, .042, .095	.001, .030, .059	$\bar{\mathcal{T}}_{o:K}^{dw}$	.182, .378, .521	.022, .092, .199	.009, .053, .116
N	.079, .134, .171	.016, .072, .143	.036, .120, .221	N	.146, .225, .289	.042, .100, .155	.036, .112, .194
N_{ex}^{dw}	.100, .264, .396	.003, .030, .093	.000, .008, .036	N_{ex}^{dw}	.189, .433, .574	.003, .050, .140	.000, .010, .057
$DV_{o:K}$	.082, .238, .353	.000, .004, .016	.000, .006, .013	$DV_{o:K}$	.158, .394, .537	.000, .004, .019	.001, .001, .006
$DV_{o:K}^{dw}$	.084, .247, .400	.000, .006, .025	.000, .000, .002	$DV_{o:K}^{dw}$	.178, .425, .573	.000, .004, .046	.000, .000, .002

Rejection frequencies (1%, 5%, 10%). $\hat{\mathcal{T}}$ is the max-correlation test; $\bar{\mathcal{T}}$ is the max-transformed test; N is Hong's (1996) test; DV is the Q-test by Delgado and Velasco (2011); "dw" implies the dependent wild bootstrap is used; "ex" signifies the correlation first order expansion is used for the bootstrap; "o:K" signifies an orthogonalized correlation with a kernel-based covariance matrix; lag length is 5, $[.5n/\ln(n)]$, or $[n/\ln(n)]$.

Table 5: Rejection Frequencies – AR(2) $y_t = 0.30y_{t-1} - 0.15y_{t-2}$ + e_t, AR(1) Plug-in

$n = 100$

e_t is iid				e_t is GARCH: $w_t^2 = 1.0 + 0.2e_{t-1}^2 + 0.5w_{t-1}^2$			
	Lag = 5	Lag = 10	Lag = 21		Lag = 5	Lag = 10	Lag = 21
	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%		1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%
$\hat{\mathcal{T}}_{ex}^{dw}$	.069, .222, .370	.032, .158, .299	.017, .077, .184	$\hat{\mathcal{T}}_{ex}^{dw}$	.058, .198, .341	.034, .129, .242	.007, .073, .172
$\bar{\mathcal{T}}_{o:K}^{dw}$	.055, .222, .344	.018, .111, .239	.004, .051, .125	$\bar{\mathcal{T}}_{o:K}^{dw}$	.033, .156, .277	.015, .087, .184	.005, .041, .109
N	.093, .134, .179	.059, .102, .148	.054, .132, .208	N	.102, .155, .193	.076, .136, .176	.044, .106, .183
N_{ex}^{dw}	.054, .205, .347	.019, .113, .235	.006, .047, .151	N_{ex}^{dw}	.041, .164, .312	.017, .102, .205	.003, .047, .136
$DV_{o:K}$	.027, .123, .248	.005, .043, .099	.030, .057, .090	$DV_{o:K}$	.018, .106, .215	.007, .036, .078	.054, .092, .128
$DV_{o:K}^{dw}$	.043, .205, .351	.003, .059, .197	.000, .006, .041	$DV_{o:K}^{dw}$	.030, .148, .290	.002, .034, .140	.000, .001, .018
e_t is MA(2): $e_t = \nu_t + 0.50\nu_{t-1} + 0.25\nu_{t-2}$				e_t is AR(1): $e_t = 0.7e_{t-1} + \nu_t$			
	Lag = 5	Lag = 10	Lag = 21		Lag = 5	Lag = 10	Lag = 21
	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%		1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%
$\hat{\mathcal{T}}_{ex}^{dw}$	.217, .542, .722	.089, .321, .518	.035, .184, .399	$\hat{\mathcal{T}}_{ex}^{dw}$	.148, .407, .587	.039, .231, .425	.013, .110, .272
$\bar{\mathcal{T}}_{o:K}^{dw}$	.331, .635, .781	.177, .480, .639	.029, .174, .315	$\bar{\mathcal{T}}_{o:K}^{dw}$	.236, .564, .713	.120, .372, .564	.024, .153, .295
N	.543, .661, .714	.391, .490, .553	.288, .386, .468	N	.330, .457, .526	.200, .320, .394	.144, .215, .277
N_{ex}^{dw}	.282, .627, .773	.083, .351, .548	.019, .160, .331	N_{ex}^{dw}	.201, .490, .668	.045, .236, .445	.006, .083, .194
$DV_{o:K}$	.184, .511, .702	.036, .229, .405	.062, .132, .192	$DV_{o:K}$	.139, .463, .645	.023, .162, .334	.043, .109, .154
$DV_{o:K}^{dw}$	.231, .597, .770	.021, .246, .495	.000, .014, .101	$DV_{o:K}^{dw}$	.198, .548, .727	.017, .204, .443	.000, .020, .101

$n = 500$

e_t is iid				e_t is GARCH: $w_t^2 = 1.0 + 0.2e_{t-1}^2 + 0.5w_{t-1}^2$			
	Lag = 5	Lag = 40	Lag = 80		Lag = 5	Lag = 40	Lag = 80
	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%		1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%
$\hat{\mathcal{T}}_{ex}^{dw}$	.524, .768, .865	.199, .438, .587	.170, .383, .520	$\hat{\mathcal{T}}_{ex}^{dw}$	.356, .650, .776	.175, .403, .565	.127, .329, .469
$\bar{\mathcal{T}}_{o:K}^{dw}$	.428, .721, .829	.129, .327, .474	.088, .271, .404	$\bar{\mathcal{T}}_{o:K}^{dw}$	.277, .589, .722	.076, .229, .348	.033, .139, .255
N	.567, .690, .748	.160, .272, .353	.108, .195, .256	N	.589, .694, .745	.183, .287, .358	.119, .200, .259
N_{ex}^{dw}	.501, .756, .847	.013, .106, .242	.001, .025, .103	N_{ex}^{dw}	.321, .632, .758	.002, .096, .235	.000, .013, .073
$DV_{o:K}$	.468, .737, .840	.001, .022, .055	.002, .006, .016	$DV_{o:K}$	.293, .585, .715	.005, .012, .046	.024, .039, .055
$DV_{o:K}^{dw}$	.432, .738, .845	.000, .018, .103	.000, .000, .000	$DV_{o:K}^{dw}$	.272, .592, .741	.000, .005, .048	.000, .000, .001
e_t is MA(2): $e_t = \nu_t + 0.50\nu_{t-1} + 0.25\nu_{t-2}$				e_t is AR(1): $e_t = 0.7e_{t-1} + \nu_t$			
	Lag = 5	Lag = 40	Lag = 80		Lag = 5	Lag = 40	Lag = 80
	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%		1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%
$\hat{\mathcal{T}}_{ex}^{dw}$	.979, 1.00, 1.00	.825, .984, .997	.775, .960, .984	$\hat{\mathcal{T}}_{ex}^{dw}$	.973, .998, .999	.699, .953, .987	.615, .927, .973
$\bar{\mathcal{T}}_{o:K}^{dw}$	.999, 1.00, 1.00	.969, .999, 1.00	.818, .970, .988	$\bar{\mathcal{T}}_{o:K}^{dw}$	.995, 1.00, 1.00	.948, .991, .999	.844, .972, .989
N	1.00, 1.00, 1.00	.963, .985, .988	.846, .922, .948	N	.999, 1.00, 1.00	.819, .906, .930	.571, .716, .783
N_{ex}^{dw}	.999, 1.00, 1.00	.294, .827, .958	.045, .374, .682	N_{ex}^{dw}	.984, 1.00, 1.00	.108, .559, .802	.008, .164, .397
$DV_{o:K}$	1.00, 1.00, 1.00	.062, .339, .579	.017, .055, .103	$DV_{o:K}$	.995, .999, .999	.080, .318, .524	.012, .048, .087
$DV_{o:K}^{dw}$	.998, 1.00, 1.00	.019, .408, .748	.000, .002, .038	$DV_{o:K}^{dw}$	.996, .999, 1.00	.016, .339, .671	.000, .000, .034

Rejection frequencies (1%, 5%, 10%). $\hat{\mathcal{T}}$ is the max-correlation test; $\bar{\mathcal{T}}$ is the max-transformed test; N is Hong's (1996) test; DV is the Q-test by Delgado and Velasco (2011); "dw" implies the dependent wild bootstrap is used; "ex" signifies the correlation first order expansion is used for the bootstrap; "o:K" signifies an orthogonalized correlation with a kernel-based covariance matrix; lag length is 5, $[.5n/\ln(n)]$, or $[n/\ln(n)]$.

Table 6: Rejection Frequencies – GARCH(1,1) $y_t = \sigma_t e_t$, $\sigma_t^2 = 1.0 + 0.2y_{t-1}^2 + 0.5\sigma_{t-1}^2$, No Plug-in

$n = 100$								
e_t is iid					e_t is GARCH: $w_t^2 = 1.0 + 0.2e_{t-1}^2 + 0.5w_{t-1}^2$			
	Lag = 5	Lag = 10	Lag = 21			Lag = 5	Lag = 10	Lag = 21
	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%			1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%
$\hat{\mathcal{T}}^{dw}_{ex}$	.006, .052, .112	.000, .033, .090	.002, .021, .091		$\hat{\mathcal{T}}^{dw}_{ex}$	.003, .034, .097	.004, .026, .073	.001, .016, .057
$\bar{\mathcal{T}}^{dw}_{o:K}$	.004, .047, .118	.003, .031, .092	.001, .011, .047		$\bar{\mathcal{T}}^{dw}_{o:K}$	.005, .039, .091	.001, .018, .068	.000, .009, .040
N	.059, .094, .124	.036, .060, .101	.033, .105, .188		N	.183, .251, .288	.149, .196, .245	.074, .165, .248
N^{dw}_{ex}	.006, .032, .095	.004, .021, .075	.000, .007, .038		N^{dw}_{ex}	.001, .028, .076	.000, .009, .044	.000, .006, .024
$DV_{o:K}$	.003, .024, .065	.007, .021, .048	.043, .068, .104		$DV_{o:K}$	.013, .025, .038	.046, .066, .081	.123, .155, .174
$DV^{dw}_{o:K}$	.005, .039, .106	.000, .017, .073	.000, .000, .014		$DV^{dw}_{o:K}$	.002, .031, .073	.000, .006, .026	.000, .000, .010
e_t is MA(2): $e_t = \nu_t + 0.50\nu_{t-1} + 0.25\nu_{t-2}$					e_t is AR(1): $e_t = 0.7e_{t-1} + \nu_t$			
	Lag = 5	Lag = 10	Lag = 21			Lag = 5	Lag = 10	Lag = 21
	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%			1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%
$\hat{\mathcal{T}}^{dw}_{ex}$	.619, .890, .969	.565, .855, .952	.523, .855, .942		$\hat{\mathcal{T}}^{dw}_{ex}$	.573, .838, .919	.593, .841, .929	.556, .827, .933
$\bar{\mathcal{T}}^{dw}_{o:K}$	.460, .800, .926	.339, .709, .855	.154, .457, .662		$\bar{\mathcal{T}}^{dw}_{o:K}$	.517, .781, .901	.440, .729, .860	.264, .571, .738
N	.989, .995, .996	.938, .964, .971	.826, .887, .914		N	1.00, 1.00, 1.00	.997, .999, .999	.998, .998, .998
N^{dw}_{ex}	.306, .715, .871	.104, .500, .760	.026, .193, .461		N^{dw}_{ex}	.436, .786, .911	.248, .657, .869	.096, .442, .737
$DV_{o:K}$	.216, .623, .807	.089, .237, .429	.178, .234, .290		$DV_{o:K}$	.333, .667, .797	.124, .304, .476	.228, .299, .345
$DV^{dw}_{o:K}$	.229, .669, .866	.031, .297, .557	.000, .035, .201		$DV^{dw}_{o:K}$	.269, .629, .816	.072, .378, .624	.001, .068, .255

$n = 500$								
e_t is iid					e_t is GARCH: $w_t^2 = 1.0 + 0.2e_{t-1}^2 + 0.5w_{t-1}^2$			
	Lag = 5	Lag = 40	Lag = 80			Lag = 5	Lag = 40	Lag = 80
	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%			1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%
$\hat{\mathcal{T}}^{dw}_{ex}$	.008, .041, .099	.002, .025, .059	.000, .021, .067		$\hat{\mathcal{T}}^{dw}_{ex}$	.008, .031, .084	.003, .020, .052	.006, .014, .031
$\bar{\mathcal{T}}^{dw}_{o:K}$	.008, .053, .107	.001, .028, .062	.002, .013, .038		$\bar{\mathcal{T}}^{dw}_{o:K}$	.005, .024, .059	.002, .009, .032	.002, .007, .015
N	.075, .118, .166	.035, .082, .135	.043, .139, .219		N	.530, .603, .641	.303, .363, .403	.166, .281, .354
N^{dw}_{ex}	.005, .030, .090	.000, .009, .035	.000, .001, .017		N^{dw}_{ex}	.005, .021, .058	.000, .006, .012	.002, .004, .009
$DV_{o:K}$	.005, .032, .075	.001, .003, .007	.016, .032, .042		$DV_{o:K}$	.004, .012, .030	.044, .056, .064	.139, .160, .170
$DV^{dw}_{o:K}$	.005, .040, .091	.000, .000, .007	.000, .000, .001		$DV^{dw}_{o:K}$	.000, .016, .053	.000, .000, .005	.000, .000, .000
e_t is MA(2): $e_t = \nu_t + 0.50\nu_{t-1} + 0.25\nu_{t-2}$					e_t is AR(1): $e_t = 0.7e_{t-1} + \nu_t$			
	Lag = 5	Lag = 40	Lag = 80			Lag = 5	Lag = 40	Lag = 80
	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%			1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%
$\hat{\mathcal{T}}^{dw}_{ex}$	.973, .992, .997	.964, .990, .997	.972, .991, .998		$\hat{\mathcal{T}}^{dw}_{ex}$	.730, .865, .915	.723, .871, .934	.733, .867, .922
$\bar{\mathcal{T}}^{dw}_{o:K}$	.959, .985, .997	.915, .973, .988	.844, .941, .971		$\bar{\mathcal{T}}^{dw}_{o:K}$	.673, .827, .899	.609, .776, .828	.515, .683, .754
N	1.00, 1.00, 1.00	1.00, 1.00, 1.00	1.00, 1.00, 1.00		N	1.00, 1.00, 1.00	1.00, 1.00, 1.00	1.00, 1.00, 1.00
N^{dw}_{ex}	.910, .980, .992	.618, .908, .972	.216, .769, .949		N^{dw}_{ex}	.681, .839, .893	.508, .772, .865	.306, .704, .858
$DV_{o:K}$	.981, .996, .997	.137, .365, .581	.145, .204, .244		$DV_{o:K}$	.704, .820, .869	.115, .189, .258	.196, .235, .257
$DV^{dw}_{o:K}$	.959, .991, .997	.045, .527, .829	.000, .009, .134		$DV^{dw}_{o:K}$	.620, .793, .866	.045, .297, .520	.000, .019, .109

Rejection frequencies (1%, 5%, 10%). $\hat{\mathcal{T}}$ is the max-correlation test; $\bar{\mathcal{T}}$ is the max-transformed test; N is Hong's (1996) test; DV is the Q-test by Delgado and Velasco (2011); "dw" implies the dependent wild bootstrap is used; "ex" signifies the correlation first order expansion is used for the bootstrap; "o:K" signifies an orthogonalized correlation with a kernel-based covariance matrix; lag length is 5, $[.5n/\ln(n)]$, or $[n/\ln(n)]$.

Table 7: Rejection Frequencies – GARCH(1,1) $y_t = \sigma_t e_t$, $\sigma_t^2 = 1.0 + 0.2y_{t-1}^2 + 0.5\sigma_{t-1}^2$, QML Plug-in

$n = 100$

e_t is iid					e_t is GARCH: $w_t^2 = 1.0 + 0.2e_{t-1}^2 + 0.5w_{t-1}^2$			
	Lag = 5	Lag = 10	Lag = 21			Lag = 5	Lag = 10	Lag = 21
	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%			1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%
$\hat{\mathcal{T}}_{ex}^{dw}$	.009, .050, .109	.002, .031, .095	.001, .030, .075		$\hat{\mathcal{T}}_{ex}^{dw}$	.006, .045, .089	.007, .039, .085	.002, .029, .077
$\bar{\mathcal{T}}_{o:K}^{dw}$	.000, .000, .000	.000, .000, .000	.000, .000, .000		$\bar{\mathcal{T}}_{o:K}^{dw}$	.000, .000, .000	.000, .000, .000	.000, .000, .000
N	.024, .055, .078	.019, .043, .083	.023, .096, .175		N	.028, .043, .064	.019, .047, .084	.032, .113, .184
N_{ex}^{dw}	.005, .041, .114	.001, .029, .088	.003, .017, .069		N_{ex}^{dw}	.006, .039, .099	.000, .021, .060	.001, .014, .044
$DV_{o:K}$	.000, .000, .000	.000, .000, .000	.000, .000, .000		$DV_{o:K}$	.000, .000, .000	.000, .000, .000	.000, .000, .000
$DV_{o:K}^{dw}$	.000, .000, .000	.000, .000, .000	.000, .000, .000		$DV_{o:K}^{dw}$	.000, .000, .000	.000, .000, .000	.000, .000, .000
e_t is MA(2): $e_t = \nu_t + 0.50\nu_{t-1} + 0.25\nu_{t-2}$					e_t is AR(1): $e_t = 0.7e_{t-1} + \nu_t$			
	Lag = 5	Lag = 10	Lag = 21			Lag = 5	Lag = 10	Lag = 21
	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%			1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%
$\hat{\mathcal{T}}_{ex}^{dw}$	.670, .880, .924	.547, .854, .909	.415, .770, .866		$\hat{\mathcal{T}}_{ex}^{dw}$	.806, .875, .895	.750, .834, .873	.727, .847, .870
$\bar{\mathcal{T}}_{o:K}^{dw}$	.000, .000, .000	.000, .000, .000	.000, .000, .000		$\bar{\mathcal{T}}_{o:K}^{dw}$	.000, .000, .000	.000, .000, .000	.000, .000, .000
N	.920, .958, .967	.809, .895, .928	.574, .691, .749		N	.999, .999, 1.00	.994, .996, 1.00	.972, .987, .992
N_{ex}^{dw}	.400, .753, .878	.132, .474, .702	.030, .201, .397		N_{ex}^{dw}	.673, .849, .889	.341, .681, .816	.107, .390, .598
$DV_{o:K}$	.000, .000, .000	.000, .000, .000	.000, .000, .000		$DV_{o:K}$	.000, .000, .000	.000, .000, .000	.000, .000, .000
$DV_{o:K}^{dw}$	.000, .000, .000	.000, .000, .000	.000, .000, .000		$DV_{o:K}^{dw}$	.000, .000, .000	.000, .000, .000	.000, .000, .000

$n = 500$

e_t is iid					e_t is GARCH: $w_t^2 = 1.0 + 0.2e_{t-1}^2 + 0.5w_{t-1}^2$			
	Lag = 5	Lag = 40	Lag = 80			Lag = 5	Lag = 40	Lag = 80
	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%			1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%
$\hat{\mathcal{T}}_{ex}^{dw}$	.016, .056, .124	.002, .032, .087	.004, .025, .065		$\hat{\mathcal{T}}_{ex}^{dw}$	.005, .041, .098	.004, .030, .085	.002, .014, .056
$\bar{\mathcal{T}}_{o:K}^{dw}$	.000, .000, .000	.000, .000, .000	.000, .000, .000		$\bar{\mathcal{T}}_{o:K}^{dw}$	.000, .000, .000	.000, .000, .000	.000, .000, .000
N	.023, .056, .072	.023, .064, .120	.031, .132, .203		N	.016, .040, .057	.013, .058, .105	.037, .125, .222
N_{ex}^{dw}	.008, .039, .093	.000, .005, .028	.000, .007, .026		N_{ex}^{dw}	.009, .056, .111	.000, .008, .036	.000, .001, .022
$DV_{o:K}$	.000, .000, .000	.000, .000, .000	.000, .000, .000		$DV_{o:K}$	.000, .000, .000	.000, .000, .000	.000, .000, .000
$DV_{o:K}^{dw}$	.000, .000, .000	.000, .000, .000	.000, .000, .000		$DV_{o:K}^{dw}$	.000, .000, .000	.000, .000, .000	.000, .000, .000
e_t is MA(2): $e_t = \nu_t + 0.50\nu_{t-1} + 0.25\nu_{t-2}$					e_t is AR(1): $e_t = 0.7e_{t-1} + \nu_t$			
	Lag = 5	Lag = 40	Lag = 80			Lag = 5	Lag = 40	Lag = 80
	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%			1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%
$\hat{\mathcal{T}}_{ex}^{dw}$	.975, .981, .984	.970, .980, .982	.972, .979, .982		$\hat{\mathcal{T}}_{ex}^{dw}$	.879, .918, .930	.890, .915, .922	.861, .888, .904
$\bar{\mathcal{T}}_{o:K}^{dw}$	.000, .000, .000	.000, .000, .000	.000, .000, .000		$\bar{\mathcal{T}}_{o:K}^{dw}$	.000, .000, .000	.000, .000, .000	.000, .000, .000
N	1.00, 1.00, 1.00	1.00, 1.00, 1.00	.999, 1.00, 1.00		N	1.00, 1.00, 1.00	1.00, 1.00, 1.00	1.00, 1.00, 1.00
N_{ex}^{dw}	.972, .982, .986	.713, .964, .981	.151, .739, .924		N_{ex}^{dw}	.854, .886, .899	.775, .871, .888	.493, .767, .810
$DV_{o:K}$	.000, .000, .000	.000, .000, .000	.000, .000, .000		$DV_{o:K}$	.000, .000, .000	.000, .000, .000	.000, .000, .000
$DV_{o:K}^{dw}$	.000, .000, .000	.000, .000, .000	.000, .000, .000		$DV_{o:K}^{dw}$	.000, .000, .000	.000, .000, .000	.000, .000, .000

Rejection frequencies (1%, 5%, 10%). $\hat{\mathcal{T}}$ is the max-correlation test; $\bar{\mathcal{T}}$ is the max-transformed test; N is Hong's (1996) test; DV is the Q-test by Delgado and Velasco (2011); "dw" implies the dependent wild bootstrap is used; "ex" signifies the correlation first order expansion is used for the bootstrap; "o:K" signifies an orthogonalized correlation with a kernel-based covariance matrix; lag length is 5, $[.5n/\ln(n)]$, or $[n/\ln(n)]$.

Table 8: Rejection Frequencies – Simple $y_t = e_t$, Mean Plug-in (Remote MA Errors)

$n = 100$

e_t is iid				e_t is MA(12): $e_t = \nu_t + 0.25\nu_{t-12}$			
	Lag = 5	Lag = 10	Lag = 21		Lag = 5	Lag = 10	Lag = 21
	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%		1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%
$\hat{\mathcal{T}}^{dw}_{ex}$	.014, .069, .144	.003, .041, .108	.001, .038, .100	$\hat{\mathcal{T}}^{dw}_{ex}$	.018, .081, .163	.018, .066, .138	.039, .155, .285
$\bar{\mathcal{T}}^{dw}_{o:K}$	.013, .061, .131	.003, .029, .113	.001, .019, .065	$\bar{\mathcal{T}}^{dw}_{o:K}$	.012, .071, .171	.010, .059, .133	.000, .028, .090
N	.020, .035, .048	.025, .045, .097	.023, .101, .174	N	.037, .064, .086	.042, .081, .126	.111, .167, .242
N^{dw}_{ex}	.007, .053, .129	.002, .024, .089	.002, .023, .072	N^{dw}_{ex}	.011, .082, .165	.009, .058, .149	.014, .109, .233
$DV_{o:K}$	.000, .021, .056	.001, .012, .043	.007, .026, .048	$DV_{o:K}$	.005, .039, .097	.003, .020, .064	.025, .067, .104
$DV^{dw}_{o:K}$	.017, .071, .137	.001, .024, .081	.000, .002, .014	$DV^{dw}_{o:K}$	.026, .078, .155	.001, .043, .106	.000, .006, .033
e_t is MA(24): $e_t = \nu_t + 0.25\nu_{t-24}$				e_t is MA(48): $e_t = \nu_t + 0.25\nu_{t-48}$			
	Lag = 5	Lag = 10	Lag = 21		Lag = 5	Lag = 10	Lag = 21
	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%		1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%
$\hat{\mathcal{T}}^{dw}_{ex}$	.006, .067, .131	.011, .055, .117	.007, .060, .138	$\hat{\mathcal{T}}^{dw}_{ex}$	.013, .073, .157	.008, .053, .125	.003, .036, .110
$\bar{\mathcal{T}}^{dw}_{o:K}$	.015, .088, .180	.006, .053, .124	.001, .026, .069	$\bar{\mathcal{T}}^{dw}_{o:K}$	.011, .068, .151	.003, .053, .115	.000, .016, .060
N	.031, .064, .086	.037, .070, .116	.041, .119, .201	N	.043, .076, .098	.028, .059, .105	.031, .092, .179
N^{dw}_{ex}	.014, .089, .167	.005, .059, .147	.007, .047, .117	N^{dw}_{ex}	.011, .078, .164	.006, .039, .101	.003, .028, .079
$DV_{o:K}$	.004, .042, .093	.000, .017, .049	.020, .042, .071	$DV_{o:K}$	.002, .030, .084	.002, .010, .036	.011, .029, .052
$DV^{dw}_{o:K}$	.011, .072, .169	.000, .030, .096	.000, .000, .025	$DV^{dw}_{o:K}$	.008, .063, .154	.000, .031, .085	.000, .006, .021

$n = 500$

e_t is iid				e_t is MA(12): $e_t = \nu_t + 0.25\nu_{t-12}$			
	Lag = 5	Lag = 40	Lag = 80		Lag = 5	Lag = 40	Lag = 80
	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%		1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%
$\hat{\mathcal{T}}^{dw}_{ex}$	.010, .054, .119	.007, .035, .070	.006, .022, .044	$\hat{\mathcal{T}}^{dw}_{ex}$	.011, .079, .144	.903, .986, .994	.839, .947, .975
$\bar{\mathcal{T}}^{dw}_{o:K}$	.010, .066, .141	.000, .017, .058	.001, .015, .054	$\bar{\mathcal{T}}^{dw}_{o:K}$	.012, .067, .134	.651, .874, .923	.539, .799, .886
N	.016, .037, .056	.022, .076, .141	.029, .113, .224	N	.031, .072, .096	.616, .772, .834	.354, .507, .580
N^{dw}_{ex}	.007, .048, .097	.000, .010, .033	.000, .006, .020	N^{dw}_{ex}	.009, .049, .106	.075, .425, .691	.014, .161, .364
$DV_{o:K}$	.006, .050, .099	.000, .001, .007	.001, .005, .012	$DV_{o:K}$	.014, .049, .105	.016, .121, .275	.007, .024, .048
$DV^{dw}_{o:K}$	.011, .057, .111	.000, .001, .010	.000, .000, .000	$DV^{dw}_{o:K}$	.009, .067, .134	.001, .079, .333	.000, .000, .005
e_t is MA(24): $e_t = \nu_t + 0.25\nu_{t-24}$				e_t is MA(48): $e_t = \nu_t + 0.25\nu_{t-48}$			
	Lag = 5	Lag = 40	Lag = 80		Lag = 5	Lag = 40	Lag = 80
	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%		1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%
$\hat{\mathcal{T}}^{dw}_{ex}$	.014, .071, .139	.873, .970, .988	.811, .942, .971	$\hat{\mathcal{T}}^{dw}_{ex}$	.013, .082, .154	.007, .055, .141	.764, .925, .957
$\bar{\mathcal{T}}^{dw}_{o:K}$	.014, .083, .149	.395, .682, .796	.297, .574, .713	$\bar{\mathcal{T}}^{dw}_{o:K}$	.013, .091, .164	.003, .039, .095	.057, .200, .323
N	.032, .082, .108	.607, .734, .795	.356, .493, .564	N	.041, .077, .097	.061, .137, .201	.302, .457, .542
N^{dw}_{ex}	.010, .074, .146	.131, .532, .757	.023, .224, .470	N^{dw}_{ex}	.017, .064, .136	.002, .030, .117	.013, .210, .448
$DV_{o:K}$	.013, .068, .126	.001, .064, .189	.003, .014, .041	$DV_{o:K}$	.014, .057, .115	.000, .002, .009	.006, .022, .031
$DV^{dw}_{o:K}$	.017, .081, .160	.000, .078, .288	.000, .000, .002	$DV^{dw}_{o:K}$	.017, .078, .139	.000, .002, .027	.000, .000, .003

Rejection frequencies (1%, 5%, 10%). $\hat{\mathcal{T}}$ is the max-correlation test; $\bar{\mathcal{T}}$ is the max-transformed test; N is Hong's (1996) test; DV is the Q-test by Delgado and Velasco (2011); "dw" implies the dependent wild bootstrap is used; "ex" signifies the correlation first order expansion is used for the bootstrap; "o:K" signifies an orthogonalized correlation with a kernel-based covariance matrix; lag length is 5, $[.5n/\ln(n)]$, or $[n/\ln(n)]$.

Table 9: Rejection Frequencies – Cramér-von Mises Tests with First Order Expansion

Model 1: Simple $y_t = e_t$, Mean Plug-in									
$n = 100$					$n = 500$				
e_t	IID	GARCH	MA(2)	AR(1)	e_t	IID	GARCH	MA(2)	AR(1)
	DW, BRW	DW, BRW	DW, BRW	DW, BRW		DW, BRW	DW, BRW	DW, BRW	DW, BRW
1%	.023, .028	.017, .028	.898, .929	.925, .981	1%	.010, .023	.015, .013	1.00, 1.00	1.00, 1.00
5%	.081, .082	.081, .079	.984, .991	.996, 1.00	5%	.051, .064	.066, .052	1.00, 1.00	1.00, 1.00
10%	.138, .142	.149, .140	.995, .995	1.00, 1.00	10%	.102, .115	.115, .105	1.00, 1.00	1.00, 1.00
Model 2: Bilinear $y_t = 0.50e_{t-1}y_{t-2} + e_t$, Mean Plug-in									
$n = 100$					$n = 500$				
e_t	IID	GARCH	MA(2)	AR(1)	e_t	IID	GARCH	MA(2)	AR(1)
	DW, BRW	DW, BRW	DW, BRW	DW, BRW		DW, BRW	DW, BRW	DW, BRW	DW, BRW
1%	.018, .045	.002, .013	.450, .712	.282, .564	1%	.014, .027	.026, .003	.884, .955	.393, .584
5%	.076, .115	.030, .056	.743, .841	.567, .744	5%	.072, .070	.038, .021	.966, .980	.630, .730
10%	.149, .186	.070, .113	.866, .899	.741, .830	10%	.124, .134	.075, .056	.990, .989	.781, .803
Model 3: AR(2) $y_t = 0.30y_{t-1} - 0.15y_{t-2} + e_t$, AR(2) Plug-in									
$n = 100$					$n = 500$				
e_t	IID	GARCH	MA(2)	AR(1)	e_t	IID	GARCH	MA(2)	AR(1)
	DW, BRW	DW, BRW	DW, BRW	DW, BRW		DW, BRW	DW, BRW	DW, BRW	DW, BRW
1%	.020, .009	.026, .009	.029, .018	.064, .044	1%	.012, .014	.011, .011	.032, .031	.325, .365
5%	.086, .041	.086, .032	.113, .058	.193, .118	5%	.059, .057	.051, .063	.144, .131	.592, .603
10%	.167, .076	.168, .061	.182, .110	.299, .192	10%	.132, .099	.104, .116	.250, .232	.700, .713
Model 4: AR(2) $y_t = 0.30y_{t-1} - 0.15y_{t-2} + e_t$, AR(1) Plug-in									
$n = 100$					$n = 500$				
e_t	IID	GARCH	MA(2)	AR(1)	e_t	IID	GARCH	MA(2)	AR(1)
	DW, BRW	DW, BRW	DW, BRW	DW, BRW		DW, BRW	DW, BRW	DW, BRW	DW, BRW
1%	.133, .123	.118, .098	.570, .587	.472, .520	1%	.710, .712	.550, .583	1.00, 1.00	.999, 1.00
5%	.338, .276	.287, .209	.805, .802	.741, .748	5%	.882, .879	.802, .802	1.00, 1.00	1.00, 1.00
10%	.483, .390	.430, .300	.898, .884	.849, .845	10%	.939, .934	.881, .882	1.00, 1.00	1.00, 1.00
Model 5: GARCH(1,1) $y_t = \sigma_t e_t$, $\sigma_t^2 = 1.0 + 0.2y_{t-1}^2 + 0.5\sigma_{t-1}^2$, No Plug-in									
$n = 100$					$n = 500$				
e_t	IID	GARCH	MA(2)	AR(1)	e_t	IID	GARCH	MA(2)	AR(1)
	DW, BRW	DW, BRW	DW, BRW	DW, BRW		DW, BRW	DW, BRW	DW, BRW	DW, BRW
1%	.021, .034	.006, .015	.681, .791	.564, .687	1%	.009, .009	.013, .008	.959, .971	.700, .794
5%	.077, .076	.049, .060	.908, .932	.818, .862	5%	.053, .052	.052, .029	.994, .988	.852, .876
10%	.141, .139	.103, .113	.969, .967	.923, .921	10%	.103, .103	.111, .065	.995, .992	.918, .911
Model 6: GARCH(1,1) $y_t = \sigma_t e_t$, $\sigma_t^2 = 1.0 + 0.2y_{t-1}^2 + 0.5\sigma_{t-1}^2$, QML Plug-in									
$n = 100$					$n = 500$				
e_t	IID	GARCH	MA(2)	AR(1)	e_t	IID	GARCH	MA(2)	AR(1)
	DW, BRW	DW, BRW	DW, BRW	DW, BRW		DW, BRW	DW, BRW	DW, BRW	DW, BRW
1%	.026, .038	.024, .021	.801, .537	.785, .298	1%	.017, .017	.012, .013	.973, .830	.852, .580
5%	.078, .092	.066, .068	.897, .857	.853, .750	5%	.081, .061	.059, .066	.979, .987	.893, .989
10%	.143, .167	.119, .124	.927, .950	.874, .891	10%	.156, .113	.118, .106	.983, .997	.907, .998
Model 7: Simple $y_t = e_t$, Mean Plug-in (Remote MA Errors)									
$n = 100$					$n = 500$				
e_t	IID	MA(12)	MA(24)	MA(48)	e_t	IID	MA(12)	MA(24)	MA(48)
	DW, BRW	DW, BRW	DW, BRW	DW, BRW		DW, BRW	DW, BRW	DW, BRW	DW, BRW
1%	.022, .036	.034, .056	.029, .038	.022, .034	1%	.013, .018	.026, .020	.024, .019	.022, .028
5%	.087, .088	.110, .125	.098, .091	.082, .093	5%	.066, .064	.092, .068	.071, .078	.093, .089
10%	.148, .146	.179, .189	.186, .143	.158, .159	10%	.136, .119	.161, .128	.133, .142	.155, .143

Rejection frequencies of Cramér-von Mises tests (1%, 5%, 10%). "DW" implies Shao's (2011) dependent wild bootstrap is used; "BRW" implies Zhu and Li's (2015) block-wise random weighting bootstrap is used. The correlation first order expansion is used for both bootstraps.