An Estimating Equation Approach for Robust Confidence Intervals for Autocorrelations of Stationary Time Series (longer version with additional details and simulation results)

Taeyoon Hwang* and Timothy J. Vogelsang†‡

August 4, 2024

Abstract

This paper develops an estimating equation approach to construct confidence intervals for autocorrelations for time series with general stationary serial correlation structures. Inference is heteroskedasticity and autocorrelation robust (HAR). It is well known that the Bartlett formula Bartlett (1946) can provide invalid inference when innovations are not independent and identically distributed (i.i.d.). Romano and Thombs (1996) derive the asymptotic distribution of sample autocorrelations under weak assumptions although they avoid estimation of the variances of sample autocorrelations and suggest resampling schemes to obtain confidence intervals. As an alternative we provide an easy to implement estimating equation approach for estimating autocorrelations and their variances. The asymptotic variances take sandwich forms which can be estimated using well known HAR variance estimators. Resulting t-statistics can be implemented with fixed-smoothing critical values. Confidence intervals using null imposed variance estimators are included in the analysis. Monte Carlo simulations show our approach is robust to innovations that are not i.i.d. and works reasonably well across various serial correlation structures. We confirm that using fixed-smoothing critical values provides size improvements especially in small samples. Consistent with Lazarus, Lewis, Stock and Watson (2018) we find that imposing the null when estimating the asymptotic variance can reduce finite sample size distortions. An empirical illustration using S&P 500 index returns shows that conclusions about market efficiency and volatility clustering during pre and post-Covid periods using our approach contrast with conclusions using traditional (and often incorrectly used) methods.

Keywords: Sample Autocorrelation, White Noise, Long Run Variance, Robust Inference, Fixed-smoothing Asymptotics

1 Introduction

The autocorrelation function is a fundamental quantity in time series analysis with the sample autocovariance routinely computed for observed time series. Approximating the sampling distribution of the

*Department of Economics, Michigan State University East Lansing, MI 48824, email: hwangta5@msu.edu

†Corresponding author, Department of Economics, 486 W. Circle Drive, 110 Marshall-Adams Hall, Michigan State University East Lansing, MI 48824-1038. Phone: 517-353-4582, Fax: 517–432-1068, email: tjv@msu.edu

‡We thank participants at the 2023 Midwest Econometrics Group Conference (Cleveland FED) and seminar participants at Michigan State University for helpful comments and suggestions.

estimated autocorrelation is a key tool in understanding the potential population autocorrelation and the underlying dynamics of a time series. The seminal work by Bartlett (1946) derived a formula, known as the 'Bartlett formula', for the asymptotic covariance matrix of sample autocorrelations under the assumption that the underlying time series is covariance stationary with independent, identically, distributed (i.i.d.) innovations. For a given parametric specification of the autocorrelation function, the Bartlett formula enables one to compute feasible confidence intervals and conduct hypothesis testing for autocorrelations. However, it has been pointed out in the literature that inference using the Bartlett formula is invalid when the i.i.d. innovation assumption is relaxed. See Romano and Thombs (1996) and references.

Upon relaxing the i.i.d. assumption, Romano and Thombs (1996) derived the asymptotic distribution of sample autocorrelations when the underlying innovations are only uncorrelated. Allowing innovations to be uncorrelated but otherwise dependent permits many stationary nonlinear processes frequently used in time series analysis. Another advantage of the approach of Romano and Thombs (1996) is that it does not depend on any particular structure for generating the stationary processes. However, to compute confidence intervals for sample autocorrelations they suggest using the moving block bootstrap and subsampling schemes that may have been viewed as computationally intensive at the time the Romano and Thombs (1996) paper was written. This may be the reason their methods have not been adopted by widely used software packages. In contrast to resampling methods, Lobato (2001) employed nonparametric kernel estimators of the asymptotic variance of sample autocorrelations. A recent paper by Wang and Sun (2020) used a similar approach but with orthonormal series variance estimators. Both of those papers focused on tests of zero autocorrelation and not the construction of generally valid confidence intervals for estimated autocorrelations.

There is a related strand of the literature that focuses on extending Bartlett's asymptotic variance formula that is valid for uncorrelated but potentially dependent innovations. Francq and Zakoïan (2009) derive a generalized Bartlett formula for the case where innovations of the time series process are weak white noise process. The formula obtained by Francq and Zakoïan (2009) can be viewed as a closed-form version of the general asymptotic variance given by Romano and Thombs (1996) that is represented in terms of the autocorrelation function of the time series, the autocorrelation of the square of the innovations, and a kurtosis parameter. Their formula also relies on a symmetry assumption for the fourth moments of the innovations. Implementation of the generalized Bartlett formula is relatively straightforward for simple autocorrelation structures like moving average models but is very complicated in general. It is likely for this reason that the generalized Bartlett variance formula has not been implemented in standard software packages.

While the literature has highlighted the dependence of the original Bartlett formula on the assumption of i.i.d. innovations, many modern statistical packages still rely on Bartlett formula for deriving variance estimators of sample autocorrelations and for inference about autocorrelations. Furthermore, even if the assumption of i.i.d. innovations is valid, many software packages implement a version of the Bartlett formula that is not valid for general stationary serial correlation structures. For example, in Stata's manual, the formula for the estimated variance of, $\widehat{\rho}_k$, the sample autocorrelation at lag k, is given by

$$V\widehat{a}r\left(\widehat{\rho}_k\right) = \begin{cases} 1/T & k=1 \\ \frac{1}{T}\left\{1+2\sum_{i=1}^{k-1}\widehat{\rho}_i^2\right\} & k>1, \end{cases} \tag{1}$$

where T is the sample size. This formula assumes, for the purposes of computing an estimated variance for $\widehat{\rho}_k$ and conducting inference about ρ_k, that the true time series is a moving average process with lag $k-1$, i.e. MA($k-1$). This is equivalent to carrying out a sequence of tests where $\widehat{\rho}_1$ is used to test hypothesis about ρ_1 conditional on the series being i.i.d. (MA(0)), $\widehat{\rho}_2$ is used to test hypothesis about ρ_2 conditional on the series being MA(1), ..., $\widehat{\rho}_k$ is used to test hypothesis about ρ_k conditional on the series being is MA($k-1$). Suppose the series is MA(3). Then the variance formulas for $\widehat{\rho}_1$, $\widehat{\rho}_2$ and $\widehat{\rho}_3$ are invalid along with corresponding confidence intervals. What is missing in the statistical packages is a method for computing confidence intervals for $\widehat{\rho}_k$ (values of ρ_k that cannot be rejected by a test), that are valid for general stationary serial correlation structures **and** do not require the assumption of i.i.d. innovations.

In this paper we develop a simple estimating equation approach for computing confidence intervals for estimated autocorrelations. The estimating equation approach extends the Lobato (2001) and Wang and Sun (2020) approaches to the general stationary serial correlation case. Except in narrow special cases, the asymptotic variances of the estimated autocorrelations take a sandwich form and well known heteroskedasticity autocorrelation robust (HAR) variance estimators can be used in a straightforward manner. We focus on kernel and orthonormal series HAR estimators and use fixed-smoothing theory (Kiefer and Vogelsang (2005), Sun (2013)) to generate critical values for computing confidence intervals. Following Lazarus et al. (2018) we consider HAR variance estimators that impose the null leading to more reliable inference. Confidence intervals using null-imposed HAR variance estimators are obtained using similar methods as used by Vogelsang and Nawaz (2017). Our approach is easy to implement and can be viewed as a method for operationalizing Romano and Thombs (1996) without needing resampling methods for valid first order asymptotic inference.

The paper is organized as follows. Section 2 reviews estimation and inference of/for the autocorrelation function of a stationary time series. In section 3 we develop a simple estimating equation approach using HAR tests for inference. We show that fixed-smoothing asymptotics applies to the test statistics. Our theory allows innovations of the time series to be white noise driven by random variables whose distributions are potentially skewed. We show how to calculate confidence intervals when the null is imposed on the variance estimator. Section 4 provides a simulation study that documents finite sample null rejection probabilities and power for various data generating processes (DGPs). Comparisons are made to existing approaches. Section 5 provides an empirical application using returns of the S&P 500 stock index. Some implications about market efficiency and volatility clustering of the S&P 500 index during pre- and post-Covid periods are obtained. Section 6 concludes the paper.

2 Preliminaries

Consider a real-valued covariance stationary time series, $\{y_t\}$, with mean $E(y_t) = \mu$. The autocovariance and autocorrelation functions for y_t are given as

$$\gamma_k = E\left[\left(y_t - \mu\right)\left(y_{t-k} - \mu\right)\right], \quad k = 0, \pm 1, \pm 2, \ldots,$$

$$\rho_k = \gamma_k / \gamma_0.$$

For a sample of T observations $\{y_1, y_2, ..., y_T\}$ define the sample autocovariance function as

$$\widehat{\gamma}_k = T^{-1} \sum_{t=k+1}^{T} (y_t - \bar{y})(y_{t-k} - \bar{y}), \qquad k = 0, 1, 2, \ldots, T-1,$$

where $\bar{y} = T^{-1} \sum_{t=1}^{T} y_t$, and define the sample autocorrelation function as

$$\widehat{\rho}_k = \widehat{\gamma}_k / \widehat{\gamma}_0. \tag{2}$$

The seminal work of Bartlett (1946) provided a formula, now known as Bartlett's formula, for the asymptotic variances and covariances of $\widehat{\rho}_k$ when y_t is a stationary linear time series driven by i.i.d. innovations. Let y_t be expressed by the Wold decomposition,

$$y_t - \mu = \sum_{m=-\infty}^{\infty} \phi_m \epsilon_{t-m},$$

where ϵ_t is an $i.i.d.(0, \sigma^2)$ innovation. Then the vector of sample autocorrelations up to lag m, $\widehat{\boldsymbol{\rho}} = (\widehat{\rho}_1, \ldots, \widehat{\rho}_m)'$, asymptotically follows a normal distribution with mean $\boldsymbol{\rho}$, the vector of corresponding population autocorrelations up to lag m. The asymptotic variance-covariance matrix of $\widehat{\boldsymbol{\rho}}$ is given by $T^{-1}\mathbf{V}_B$ with $v_{i,j}^B$, the ij^{th} element of the $m \times m$ matrix $\mathbf{V}_B$, given by Bartlett's formula:

$$v_{i,j}^B = \sum_{\ell=-\infty}^{\infty} \{\rho_{\ell+i}\rho_{\ell+j} + \rho_{\ell-i}\rho_{\ell+j} + 2\rho_i\rho_j\rho_\ell^2 - 2\rho_i\rho_\ell\rho_{\ell+j} - 2\rho_j\rho_\ell\rho_{\ell+i}\}.$$

Despite its wide usage in textbooks and statistical packages, Bartlett's formula is only valid when ϵ_t is i.i.d. Use of Bartlett's formula for inference is potentially invalid when ϵ_t is an uncorrelated process (e.g. white noise process), but not i.i.d.. Specifically, using mixing conditions that allow white noise innovations, Romano and Thombs (1996) derived an asymptotic normality result for $\sqrt{T}(\widehat{\boldsymbol{\rho}} - \boldsymbol{\rho})$ with asymptotic variance-covariance matrix $\mathbf{V}_{RT}$ with ij^{th} elements given by

$$v_{i,j}^{RT} = \gamma_0^{-2} \left[c_{i+1,j+1} - \rho_i c_{1,j+1} - \rho_j c_{1,i+1} + \rho_i\rho_j c_{1,1}\right],$$

where $c_{i+1,j+1} = \sum_{d=-\infty}^{\infty} \text{cov}(y_0 y_i, y_d y_{d+j})$. Note that Romano and Thombs (1996) showed that $c_{i+1,j+1}$ is the $(i+1, j+1)^{th}$ element of the asymptotic variance-covariance matrix of $\sqrt{T}(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})$ where $\boldsymbol{\gamma} = (\gamma_0, ..., \gamma_m)'$

and $\widehat{\boldsymbol{\gamma}} = (\widehat{\gamma}_0, ..., \widehat{\gamma}_m)'$. Given the complicated nature of $v_{i,j}^{RT}$, Romano and Thombs (1996) propose resampling methods for constructing confidence intervals for ρ_k. For tests of zero autocorrelation Lobato (2001) proposed nonparametric kernel estimators of $c_{i+1,j+1}$ and Wang and Sun (2020) used series to estimate $c_{i+1,j+1}$. Neither study focused on confidence intervals for ρ_k when the time series has autocorrelation.

Closed form formulas for $v_{i,j}^{RT}$ were obtained by Francq and Zakoïan (2009) for some models of y_t with ϵ_t being white noise with a symmetry condition imposed on the fourth moments of ϵ_t. Francq and Zakoïan (2009) label these formulas 'generalized Bartlett' formulas. For example, suppose that y_t is a weak white noise process (i.e. $y_t = \epsilon_t$ where ϵ_t is a weak white noise process). The generalized Bartlett formula is given by $v_{i,j}^{GB} = v_{i,j}^{B} + v_{i,j}^{B^*}$ where

$$v_{i,i}^{B} = 1, \; v_{i,i}^{B^*} = \frac{\gamma_{\epsilon^2}(i)}{[\gamma_\epsilon(0)]^2}, \tag{3}$$

and $v_{i,j}^{B} = v_{i,j}^{B^*} = 0$ if $i \neq j$ with $\gamma_{\epsilon^2}(i)$ being the autocovariance function of ϵ_t^2 at lag i and $\gamma_\epsilon^2(0)$ being the variance of ϵ_t. When the data generating process of y_t is an MA(q) model, Francq and Zakoïan (2009) show that

$$v_{i,i}^{B} = \sum_{\ell=-q}^{q} \rho_\ell^2, \quad v_{i,i}^{B^*} = \frac{1}{[\gamma_\epsilon(0)]^2} \sum_{\ell=-q}^{q} \gamma_{\epsilon^2}(i-\ell)\rho_\ell^2,$$

for all $i > q$. Francq and Zakoïan (2009) do not provide formulas for $i \leq q$.

While the results of Lobato (2001), Wang and Sun (2020) and Francq and Zakoïan (2009) are useful in specific contexts, they are not comprehensive enough to be used to construct confidence intervals for ρ_k. Therefore, we develop a systematic and simple approach to the construction of confidence intervals that does not require resampling methods. Because our approach is based on the inversion of t-statistics, resampling methods could be used to obtain critical values for the construction of confidence intervals. We leave such an investigation to future research.

3 Theory

3.1 An Estimating Equation Approach For Autocorrelation Inference

In this section we develop an estimating equation approach that uses HAR t-statistics for inference regarding autocorrelations where we relax the assumption that the innovations, ϵ_t, are i.i.d.. There are a few advantages of this approach. First, the HAR tests we use are well known and easy to apply in practice. Second, we show that fixed-smoothing asymptotics can be used for the test statistics providing critical values that depend on tuning parameters used to estimate variances. Third, it is straightforward to construct confidence intervals for both the cases where the null hypothesis about the autocorrelation is **a)** imposed and **b)** not imposed on the variance estimator. As we show, imposing the null on the variance estimator can help reduce distortions in finite sample rejections under the null similar to what was found for stationary time series regressions by Lazarus et al. (2018) and Vogelsang (2018).

Consider the following estimation equation for a stationary time series y_t:

$$y_t = c + \rho_k y_{t-k} + \eta_t^{(k)}, \tag{4}$$

where $c = \mu(1 - \rho_k)$ and $t = k+1, k+2, \ldots, T$. Regression (4) allows consistent estimation of c and ρ_k because

$$E\left(\eta_t^{(k)}\right) = 0, \qquad E\left(y_{t-k}\eta_t^{(k)}\right) = 0.$$

These conditions are easy to establish as follows. Taking the mean of both sides of (4) gives

$$E\left(y_t\right) = c + \rho_k E\left(y_{t-k}\right) + E\left(\eta_t^{(k)}\right).$$

Replacing $E\left(y_t\right)$ and $E\left(y_{t-k}\right)$ with μ, and using $c = \mu(1-\rho_k)$ gives

$$\mu = \mu(1-\rho_k) + \rho_k\mu + E\left(\eta_t^{(k)}\right) = \mu + E\left(\eta_t^{(k)}\right),$$

in which case it follows that $E\left(\eta_t^{(k)}\right) = 0$. To show that $E\left(y_{t-k}\eta_t^{(k)}\right) = 0$, calculate $cov\left(y_{t-k}, y_t\right)$ giving

$$\begin{aligned} cov\left(y_{t-k}, y_t\right) &= cov\left(y_{t-k}, c + \rho_k y_{t-k} + \eta_t^{(k)}\right) \\ &= \rho_k cov\left(y_{t-k}, y_{t-k}\right) + cov\left(y_{t-k}, \eta_t^{(k)}\right), \end{aligned}$$

or equivalently

$$\gamma_k = \rho_k\gamma_0 + cov\left(y_{t-k}, \eta_t^{(k)}\right) = \frac{\gamma_k}{\gamma_0}\gamma_0 + cov\left(y_{t-k}, \eta_t^{(k)}\right) = \gamma_k + cov\left(y_{t-k}, \eta_t^{(k)}\right).$$

It then directly follows that $cov\left(y_{t-k}, \eta_t^{(k)}\right) = 0$. Because $E\left(\eta_t^{(k)}\right) = 0$, it must also be the case that $E\left(y_{t-k}\eta_t^{(k)}\right) = 0$.

Except in certain special cases, $\eta_t^{(k)}$ will have serial correlation. By construction $\eta_t^{(k)}$ is given by

$$\eta_t^{(k)} = y_t - c - \rho_k y_{t-k} = (y_t - \mu) - \rho_k (y_{t-k} - \mu). \tag{5}$$

Suppose y_t is a finite order autoregressive moving average process ($ARMA(p, q)$) given by

$$\phi(L)(y_t - \mu) = \theta(L)\epsilon_t,$$

where $\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p$, $\theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \ldots + \theta_p L^q$ and L is the lag operator. Applying the $\phi(L)$ lag polynomial to both sides of (5) gives

$$\begin{aligned}
\phi(L)\eta_t^{(k)} &= \phi(L)(y_t - \mu) - \phi(L)\rho_k (y_{t-k} - \mu) = \phi(L)(y_t - \mu) - \rho_k \phi(L) L^k (y_t - \mu) \\
&= \phi(L)(y_t - \mu) - \rho_k L^k \phi(L)(y_t - \mu) = \theta(L)\epsilon_t - \rho_k L^k \theta(L)\epsilon_t \\
&= \left(1 - \rho_k L^k\right)\theta(L)\epsilon_t.
\end{aligned} \tag{6}$$

We see from (6) that $\eta_t^{(k)}$ is an $ARMA(p, q + k)$ process.

Suppose that y_t in uncorrelated. Then $p = q = 0$ and $\rho_k = 0$, and it follows that (6) simplifies to $\eta_t^{(k)} = \epsilon_t$ in which case $\eta_t^{(k)}$ is uncorrelated. Whether or not $y_{t-k}\eta_t^{(k)}$ has serial correlation is more complicated and depends on k, the serial correlation in y_t, **and** whether ϵ_t has dependence in higher order moments. Cases where $y_{t-k}\eta_t^{(k)}$ has no serial correlation should be viewed as exceptions, and inference based on estimation of (4) should be made robust to serial correlation (and conditional heteroskedasticity).

It is convenient to rewrite the estimation equation (4) as

$$y_t = \mathbf{x}_{t-k}'\beta + \eta_t^{(k)} \tag{7}$$

where $\mathbf{x}_{t-k} = \begin{bmatrix} 1 & y_{t-k} \end{bmatrix}'$ and $\beta = \begin{bmatrix} c & \rho_k \end{bmatrix}'$. The ordinary least squares (OLS) estimator of β from (7) is given by the usual formula

$$\widetilde{\beta} = \begin{bmatrix} \widetilde{c} \\ \widetilde{\rho}_k \end{bmatrix} = \left(\sum_{t=k+1}^{T} \mathbf{x}_{t-k}\mathbf{x}_{t-k}'\right)^{-1} \sum_{t=k+1}^{T} \mathbf{x}_{t-k} y_t.$$

Using the Frisch-Waugh-Lovell Theorem, $\widetilde{\rho}_k$ can be equivalently expressed as

$$\widetilde{\rho}_k = \frac{\sum_{t=k+1}^{T} \left(y_{t-k} - \bar{y}_{\{1,T-k\}}\right)\left(y_t - \bar{y}_{\{k+1,T\}}\right)}{\sum_{t=k+1}^{T} \left(y_{t-k} - \bar{y}_{\{1,T-k\}}\right)^2},$$

where

$$\bar{y}_{\{1,T-k\}} = \frac{1}{T-k}\sum_{t=1}^{T-k} y_t, \qquad \bar{y}_{\{k+1,T\}} = \frac{1}{T-k}\sum_{t=k+1}^{T} y_t.$$

Define the 2×1 vector, $\mathbf{v}_t^{(k)}$, as

$$\mathbf{v}_t^{(k)} = \mathbf{x}_{t-k}\eta_t^{(k)} = \begin{bmatrix} \eta_t^{(k)} \\ y_{t-k}\eta_t^{(k)} \end{bmatrix},$$

and its partial sum process

$$\mathbf{S}_{[rT]}^{(k)} = \sum_{t=k+1}^{[rT]} \mathbf{v}_t^{(k)},$$

where $[rT]$ is the integer part of rT with $r \in [0,1]$. Using standard calculations,

$$\begin{aligned}\sqrt{T}\left(\widetilde{\beta} - \beta\right) = \begin{bmatrix} \sqrt{T}\,(\widetilde{c} - c) \\ \sqrt{T}\,(\widetilde{\rho}_k - \rho_k) \end{bmatrix} &= \left(T^{-1}\sum_{t=k+1}^{T} \mathbf{x}_{t-k}\mathbf{x}_{t-k}'\right)^{-1} T^{-1/2}\sum_{t=k+1}^{T} \mathbf{x}_{t-k}\eta_t^{(k)} \\ &= \left(T^{-1}\sum_{t=k+1}^{T} \mathbf{x}_{t-k}\mathbf{x}_{t-k}'\right)^{-1} T^{-1/2}\sum_{t=k+1}^{T} \mathbf{v}_t^{(k)} = \left(T^{-1}\sum_{t=k+1}^{T} \mathbf{x}_{t-k}\mathbf{x}_{t-k}'\right)^{-1} T^{-1/2}\mathbf{S}_T^{(k)}.\end{aligned}$$

The asymptotic variance of $\widetilde{\beta}$ depends on the probability limit of $T^{-1}\sum_{t=k+1}^{T} \mathbf{x}_{t-k}\mathbf{x}_{t-k}'$ and the long run variance of $\mathbf{v}_t^{(k)}$ which we denote by

$$\mathbf{\Omega}^{(k)} = \mathbf{\Gamma}_0^{(k)} + \sum_{j=1}^{\infty}\left(\mathbf{\Gamma}_j^{(k)} + \mathbf{\Gamma}_j^{(k)\prime}\right),$$

where $\mathbf{\Gamma}_j^{(k)} = E(\mathbf{v}_t^{(k)}\mathbf{v}_{t-j}^{(k)\prime})$.

The following two assumptions are sufficient to obtain an asymptotic normality result for $\sqrt{T}\left(\widetilde{\beta} - \beta\right)$. We use the symbol $\Rightarrow$ to denote weak convergence in distribution.

Assumption 1 $T^{-1/2}\sum_{t=k+1}^{[rT]} \mathbf{v}_t^{(k)} = T^{-1/2}\mathbf{S}_{[rT]}^{(k)} \Rightarrow \mathbf{\Lambda}^{(k)}\mathbf{W}_2(r)$, *where* $\mathbf{\Lambda}^{(k)}$ *is the matrix square root of* $\mathbf{\Omega}^{(k)}$, *i.e.* $\mathbf{\Omega}^{(k)} = \mathbf{\Lambda}^{(k)}\mathbf{\Lambda}^{(k)\prime}$, $r \in [0,1]$, *and* $\mathbf{W}_2(r)$ *is a* 2×1 *vector of independent Wiener processes (*$\mathbf{W}_2(r) \sim N(0, r\mathbf{I}_2)$ *where* $\mathbf{I}_2$ *is a* 2×2 *identity matrix).*

Assumption 2 $T^{-1}\sum_{t=k+1}^{[rT]}\mathbf{x}_{t-k}\mathbf{x}'_{t-k}\xrightarrow{p} r\mathbf{Q} = r\begin{bmatrix} 1 & \mu \\ \mu & \gamma_0+\mu^2 \end{bmatrix}$, *where* $r\in[0,1]$.

Assumption 1 is a functional central limit theorem (FCLT) for the scaled partial sums of $\mathbf{v}_t^{(k)}$. Assumption 1 is stronger than what is needed for an asymptotic normality result for $\sqrt{T}\left(\widetilde{\beta}-\beta\right)$ but is used to obtain fixed-smoothing results for HAR test statistics. Inference is discussed in the next section. A primitive condition for Assumption 1 to hold is that y_t is near epoch dependence (L_2-NED) with sufficient α-mixing. See Lobato (2001) for details for the case of zero autocovariance tests. Additional details on sufficient conditions for FCLTs using NED and mixing can be found in de Jong and Davidson (2000). Note that because **i)** $\mathbf{v}_t^{(k)}$ involves the product of y_{t-k} and $\eta_t^{(k)}$, and **ii)** $\eta_t^{(k)}$ is a filtered version of y_{t-k}, properties of transformations of NED processes play a role in primitive conditions sufficient for Assumption 1; see Davidson (1994). Assumption 2 holds as long as y_{t-k} is a second order stationary process. As long as $\gamma_0>0$ it follows that $\mathbf{Q}^{-1}$ exists.

We can directly derive the asymptotic distribution of $\sqrt{T}\left(\widetilde{\beta}-\beta\right)$ under Assumptions 1 and 2 as

$$\sqrt{T}\left(\widetilde{\beta}-\beta\right)=\begin{bmatrix}\sqrt{T}\left(\widetilde{c}-c\right)\\ \sqrt{T}\left(\widetilde{\rho}_k-\rho_k\right)\end{bmatrix}\Rightarrow \mathbf{Q}^{-1}\mathbf{\Lambda}\mathbf{W}_2(1)\sim N\left(\mathbf{0},\mathbf{Q}^{-1}\mathbf{\Omega}^{(k)}\mathbf{Q}^{-1}\right)\equiv N(\mathbf{0},\mathbf{V}^{(k)}).$$

The asymptotic variance of $\widetilde{\rho}_k$ is $\mathbf{V}_{22}^{(k)}$, which is the (2,2) element of $\mathbf{V}^{(k)}$. Straightforward calculations can be used to show that $\mathbf{V}_{22}^{(k)}$ is the same as the asymptotic variance for $\widehat{\rho}_k$ obtained by Romano and Thombs (1996) (see their equation (6)). Therefore, $\widetilde{\rho}_k$ is asymptotically equivalent to $\widehat{\rho}_k$. The advantage of using $\widetilde{\rho}_k$ via the regression (4) is that inference about ρ_k can be carried out using well known estimators for $\mathbf{V}^{(k)}$ that are simple to implement in practice.

The asymptotic variance, $\mathbf{V}^{(k)}$, is estimated as follows. The natural estimator of $\mathbf{Q}$ is given by

$$\widetilde{\mathbf{Q}}=(T-k)^{-1}\sum_{t=k+1}^{T}\mathbf{x}_{t-k}\mathbf{x}'_{t-k}.$$

Because the middle matrix of $\mathbf{V}^{(k)}$ is the long-run variance-covariance matrix of $\mathbf{v}_t^{(k)}$, we can use a nonparametric kernel estimator of the form

$$\widetilde{\mathbf{\Omega}}^{(k)} = \widetilde{\mathbf{\Gamma}}_0^{(k)} + \sum_{j=1}^{T-k-1} k\left(\frac{j}{M}\right)\left(\widetilde{\mathbf{\Gamma}}_j^{(k)} + \widetilde{\mathbf{\Gamma}}_j^{(k)\prime}\right),$$

$$\widetilde{\mathbf{\Gamma}}_j^{(k)} = (T-k)^{-1}\sum_{t=k+j+1}^{T} \widetilde{\mathbf{v}}_t^{(k)}\widetilde{\mathbf{v}}_{t-j}^{(k)\prime},$$

where

$$\widetilde{\mathbf{v}}_t^{(k)} = \mathbf{x}_{t-k}\widetilde{\eta}_t^{(k)}, \quad \widetilde{\eta}_t^{(k)} = y_t - \mathbf{x}_{t-k}'\widetilde{\beta} = y_t - \widetilde{c} - \widetilde{\rho}_k y_{t-k}, \tag{8}$$

$k(x)$ is a kernel function, and M is a truncation lag or bandwidth. $\widetilde{\mathbf{\Omega}}^{(k)}$ is the usual kernel HAR long run variance estimator using OLS residuals, $\widetilde{\eta}_t^{(k)}$. This leads to an estimator of $\mathbf{V}^{(k)}$ given by

$$\widetilde{\mathbf{V}}^{(k)} = \widetilde{\mathbf{Q}}^{-1}\widetilde{\mathbf{\Omega}}^{(k)}\widetilde{\mathbf{Q}}^{-1}.$$

We also consider a variant of $\widetilde{\mathbf{\Omega}}^{(k)}$ that imposes the null hypothesis being tested about ρ_k. Suppose we are interested in testing the null hypothesis

$$H_0 : \rho_k = a,$$

where a is a given number in the $(-1, 1)$ range. Define the null-imposed residuals for (4) as

$$\widetilde{\eta}_t^{(k)*} = y_t - \left(\bar{y}_{\{k+1,T\}} - a\bar{y}_{\{1,T-k\}}\right) - ay_{t-k} = \left(y_t - \bar{y}_{\{k+1,T\}}\right) - a\left(y_{t-k} - \bar{y}_{\{1,T-k\}}\right)$$

The null-imposed kernel estimator of $\mathbf{\Omega}^{(k)}$ uses $\widetilde{\mathbf{v}}_t^{(k)*} = \mathbf{x}_{t-k}\widetilde{\eta}_t^{(k)*} - \frac{1}{T-k}\sum_{s=k+1}^{T}\mathbf{x}_{s-k}\widetilde{\eta}_s^{(k)*}$ in place of $\widetilde{\mathbf{v}}_t^{(k)}$ and is given by

$$\widetilde{\mathbf{\Omega}}^{(k)*} = \widetilde{\mathbf{\Gamma}}_0^{(k)*} + \sum_{j=1}^{T-k-1} k\left(\frac{j}{M}\right)\left(\widetilde{\mathbf{\Gamma}}_j^{(k)*} + \widetilde{\mathbf{\Gamma}}_j^{(k)*\prime}\right),$$

$$\widetilde{\mathbf{\Gamma}}_j^{(k)*} = (T-k)^{-1}\sum_{t=k+j+1}^{T} \widetilde{\mathbf{v}}_t^{(k)*}\widetilde{\mathbf{v}}_{t-j}^{(k)*\prime}.$$

Notice that $\widetilde{\mathbf{v}}_t^{(k)*}$ is the demeaned version of $\mathbf{x}_{t-k}\widetilde{\eta}_t^{(k)*}$. This simple demeaning was found to be important for power by Lazarus et al. (2018) and Vogelsang (2018) when imposing the null on the variance estimator. The null-imposed estimator of $\mathbf{V}^{(k)}$ is given by

$$\widetilde{\mathbf{V}}^{(k)*} = \widetilde{\mathbf{Q}}^{-1}\widetilde{\mathbf{\Omega}}^{(k)*}\widetilde{\mathbf{Q}}^{-1}.$$

Lastly, there is one thing to point out about the bandwidth M. In practice data dependent methods are often used to choose M. Those formulas are functions of the proxy used for $\mathbf{v}_t^{(k)}$ when estimating $\mathbf{\Omega}^{(k)}$. For $\widetilde{\mathbf{\Omega}}^{(k)}$ data dependent bandwidths are functions of $\widetilde{\mathbf{v}}_t^{(k)}$. For $\widetilde{\mathbf{\Omega}}^{(k)*}$ data dependent bandwidths would typically be functions of $\widetilde{\mathbf{v}}_t^{(k)*}$ and would depend on a through $\widetilde{\eta}_t^{(k)*}$. Having the bandwidth depend on the null value of ρ_k complicates the computation of confidence intervals. Things are much simpler when $\widetilde{\mathbf{\Omega}}^{(k)*}$ uses the same data dependent bandwidth as $\widetilde{\mathbf{\Omega}}^{(k)}$. Details are provided in Section 3.3.

3.2 Inference about ρ_k

In this section we focus on simple tests of the autocorrelation for a given lag, k. We propose HAR t-tests using the variance estimators $\widetilde{\mathbf{V}}^{(k)}$ and $\widetilde{\mathbf{V}}^{*(k)}$ and an additional variant of those estimators. Our tests are valid for covariance stationary y_t driven by weak white noise innovations. The case of i.i.d. innovations is automatically handled.

For a given lag value, k, suppose we want to test the simple hypothesis

$$H_0 : \rho_k = a,$$

where, because ρ_k is a correlation parameter, a is a given value in the range $(-1, 1)$. The test could be two-sided or one-sided using the appropriate rejection rule. We analyze the following two t-statistics:

$$\widetilde{t}^{(k)} = \frac{(\widetilde{\rho}_k - a)}{\sqrt{\frac{1}{T-k}\widetilde{V}_{22}^{(k)}}}, \qquad \widetilde{t}^{(k)*} = \frac{(\widetilde{\rho}_k - a)}{\sqrt{\frac{1}{T-k}\widetilde{V}_{22}^{(k)*}}} \tag{9}$$

where $\widetilde{V}_{22}^{(k)}$ and $\widetilde{V}_{22}^{(k)*}$ are the (2,2) elements of the respective variance matrix estimators.

Rather than seek sufficient conditions under which $\widetilde{\mathbf{V}}^{(k)}$ and $\widetilde{\mathbf{V}}^{(k)*}$ are consistent estimators, we adopt the fixed-smoothing asymptotic approach (often called fixed-b asymptotics in the context of kernel variance estimators). We do this to generate reference distributions for $\widetilde{t}^{(k)}$ and $\widetilde{t}^{(k)*}$ that depend on the choice of kernel and bandwidth and capture, to some extent, the impact of the sampling distribution of the variance estimators on the t-statistics. As has been documented in the time series econometrics literature (Kiefer and Vogelsang (2005), Sun, Phillips and Jin (2008), Gonçalves and Vogelsang (2011), Zhang and Shao (2013), Lazarus et al. (2018) and Lazarus, Lewis and Stock (2021)), more accurate inference is obtained using critical values from fixed-b reference distributions. Fixed-b asymptotic results are derived using an asymptotic nesting where the bandwidth to sample size ratio, $b = M/T \in (0, 1]$, is held fixed as $T \to \infty$.

The following Theorem gives the fixed-b limits of the kernel variance estimators under Assumptions 1 and 2.

Theorem 1 *Let $M = bT$ where $b \in (0, 1]$ is fixed. Under Assumptions 1 and 2, as $T \to \infty$, the fixed-b limits of $\widetilde{\mathbf{\Omega}}^{(k)}$, and $\widetilde{\mathbf{\Omega}}^{(k)*}$ are given by*

$$\widetilde{\mathbf{\Omega}}^{(k)} \Rightarrow \mathbf{\Lambda}^{(k)}\widetilde{\mathbf{P}}_2(b)\mathbf{\Lambda}^{(k)\prime}, \qquad \widetilde{\mathbf{\Omega}}^{(k)*} \Rightarrow \mathbf{\Lambda}^{(k)}\widetilde{\mathbf{P}}_2(b)\mathbf{\Lambda}^{(k)\prime},$$

where $\widetilde{\mathbf{P}}_2(b)$ is a 2×2 stochastic matrix that is a function of the 2×1 vector of Brownian bridges, $\widetilde{\mathbf{W}}_2(r) = \mathbf{W}_2(r) - r\mathbf{W}_2(1)$ and the form of $\widetilde{\mathbf{P}}_2(b)$ depends on $k(x)$.

Notice that the fixed-b limits of $\widetilde{\mathbf{\Omega}}^{(k)}$ and $\widetilde{\mathbf{\Omega}}^{(k)*}$ are the same[1]. Furthermore, the limits are the same those obtained by Kiefer and Vogelsang (2005) in stationary time series regressions. Kiefer and Vogelsang (2005) provide details on how the form of $\widetilde{\mathbf{P}}_2(b)$ depends on the shape of the kernel. In our simulations we use the Parzen kernel

$$k(x) = \begin{cases} 1 - 6x^2 + 6|x|^3 \text{ for } |x| \leq \frac{1}{2} \\ 2(1 - |x|)^3 \text{ for } \frac{1}{2} \leq |x| \leq 1 \\ 0 \text{ for } |x| > 1, \end{cases}$$

giving

$$\widetilde{\mathbf{P}}_2(b) = -\iint_{|r-s|<b} \frac{1}{b^2} k'' \left(\frac{r-s}{b} \right) \widetilde{\mathbf{W}}_2(r) \widetilde{\mathbf{W}}_2(r)' drds,$$

where $k''(x)$ is the second derivative of $k(x)$.

Using Theorem 1, the fixed-b limits of the t-statistics immediately follow from arguments in Kiefer and Vogelsang (2005) and are given by

$$\widetilde{t}^{(k)} \Rightarrow \frac{W_1(1)}{\sqrt{\widetilde{P}_1(b)}}, \qquad \widetilde{t}^{(k)*} \Rightarrow \frac{W_1(1)}{\sqrt{\widetilde{P}_1(b)}},$$

[1]It was first pointed out by Lazarus et al. (2018) that demeaning $\widetilde{v}_t^{(k)*}$ gives the same fixed-b limit for the null-imposed long run variance estimator as for the null-not-imposed long run variance estimator.

where $\widetilde{P}_1(b)$ is a scalar version of $\widetilde{\mathbf{P}}_2(b)$ defined in terms of the scalar standard Wiener process $W_1(r)$ in place of $\mathbf{W}_2(r)$. The fixed-b limiting distributions are nonstandard but the critical values are easily tabulated using simulation methods. The following formula can be used to compute right tail fixed-b critical values:

$$cv_{\alpha/2}(b) = z_{\alpha/2} + \lambda_1(b \cdot z_{\alpha/2}) + \lambda_2(b \cdot z_{\alpha/2}^2) + \lambda_3(b \cdot z_{\alpha/2}^3) + \lambda_4(b^2 \cdot z_{\alpha/2}) + \lambda_5(b^2 \cdot z_{\alpha/2}^2)$$
$$+ \lambda_6(b^2 \cdot z_{\alpha/2}^3) + \lambda_7(b^3 \cdot z_{\alpha/2}) + \lambda_8(b^3 \cdot z_{\alpha/2}^2) + \lambda_9(b^3 \cdot z_{\alpha/2}^3),$$

where $z_{\alpha/2}$ is the right tail critical value from a standard normal distribution and the λ coefficients depend on the kernel. Left tail critical values follow by symmetry around zero.[2] Notice that the critical values reduce to the $N(0,1)$ distribution as $b \to 0$. This follows from the result, shown by Kiefer and Vogelsang (2005), that $p\lim_{b\to 0} \widetilde{P}_1(b) = 1$. Table 1 gives the λ coefficients for the Parzen kernel.

There are other methods for estimating long run variances. An alternative to the kernel approach is the orthonormal series (OS) approach of Müller (2007) and Sun (2013) which has been applied to tests of zero autocorrelation tests by Wang and Sun (2020). The OS long run variance estimator uses a finite set of orthonormal functions $\Phi_\ell(\cdot)$, $\ell = 1, 2, ..., K$ with the following properties (Assumption 3.1.(b) of Sun (2013)):

Assumption 3 *For $\ell = 1, 2, \ldots, K$, the basis functions $\Phi_\ell(\cdot)$ are continuously differentiable and orthonormal in $L^2[0,1]$ and satisfy $\int_0^1 \Phi_\ell(x)dx = 0$.*

Define $\widetilde{\mathbf{\Lambda}}_\ell = \frac{1}{\sqrt{T-k}} \sum_{t=k+1}^{T} \Phi_\ell\left(\frac{t}{T}\right) \widetilde{\mathbf{v}}_t^{(k)}$ and $\widetilde{\mathbf{\Lambda}}_\ell^* = \frac{1}{\sqrt{T-k}} \sum_{t=k+1}^{T} \Phi_\ell\left(\frac{t}{T}\right) \widetilde{\mathbf{v}}_t^{(k)*}$. The null-not-imposed and the null-imposed OS long run variance estimators of $\mathbf{\Omega}^{(k)}$ are given by

$$\widetilde{\mathbf{\Omega}}_{OS}^{(k)} = \frac{1}{K} \sum_{\ell=1}^{K} \widetilde{\mathbf{\Lambda}}_\ell \widetilde{\mathbf{\Lambda}}_\ell', \quad \widetilde{\mathbf{\Omega}}_{OS}^{(k)*} = \frac{1}{K} \sum_{\ell=1}^{K} \widetilde{\mathbf{\Lambda}}_\ell^* \widetilde{\mathbf{\Lambda}}_\ell^{*\prime}$$

giving the variance estimators

$$\widetilde{\mathbf{V}}_{OS}^{(k)} = \widetilde{\mathbf{Q}}^{-1} \widetilde{\mathbf{\Omega}}_{OS}^{(k)} \widetilde{\mathbf{Q}}^{-1}, \quad \widetilde{\mathbf{V}}_{OS}^{(k)*} = \widetilde{\mathbf{Q}}^{-1} \widetilde{\mathbf{\Omega}}_{OS}^{(k)*} \widetilde{\mathbf{Q}}^{-1}.$$

[2] Kiefer and Vogelsang (2005) show that $W_1(1)$, which is distributed $N(0,1)$, is independent of $\widetilde{P}_1(b)$ in which case $W_1(1)/\sqrt{\widetilde{P}_1(b)}$ has a mixture normal distribution and therefore has a density symmetric around zero.

The corresponding t-statistics are given by

$$\widetilde{t}_{OS}^{(k)} = \frac{(\widetilde{\rho}_k - a)}{\sqrt{\frac{1}{T-k}\widetilde{V}_{OS,22}^{(k)}}}, \qquad \widetilde{t}_{OS}^{(k)*} = \frac{(\widetilde{\rho}_k - a)}{\sqrt{\frac{1}{T-k}\widetilde{V}_{OS,22}^{(k)*}}}.$$

Following Sun (2013), we use asymptotic limits for $\widetilde{t}_{OS}^{(k)}$ and $\widetilde{t}_{OS}^{(k)*}$ where K is held fixed as $T \to \infty$. This is another example of fixed-smoothing asymptotics, called fixed-K asymptotics, that generates reference distributions that, in this case, capture the number of orthonormal series and the impact, to some extent, of the sampling distribution of the variance estimators on the t-statistics. Our assumptions allow direct application of results in Sun (2013) giving

$$\widetilde{t}_{OS}^{(k)} \Rightarrow t_K, \qquad \widetilde{t}_{OS}^{(k)*} \Rightarrow t_K,$$

where t_K is a standard t-distribution with K degrees of freedom. A nice feature of the OS approach is that the fixed-K limit is a well known distribution and critical values are easily calculated using standard statistical software. For a given set of orthonormal series, the value K needs to be chosen in practice. As in the kernel variance estimator case, we use data dependent methods based on $\widetilde{\mathbf{v}}_t^{(k)}$, the null-not-imposed proxy for $\mathbf{v}_t^{(k)}$, for both $\widetilde{t}_{OS}^{(k)}$ and $\widetilde{t}_{OS}^{(k)*}$.

3.3 Computation of Confidence Intervals

When the null is not imposed on the variance estimator, a $(1-\alpha)\%$ two-tail confidence interval can be computed in the usual way as

$$\widetilde{\rho}_k \pm cv_{\alpha/2} \cdot \sqrt{\frac{1}{T-k}\widetilde{V}_{22}^{(k)}},$$

where $cv_{\alpha/2}$ is the critical value taken from the relevant reference distribution (standard normal or fixed-b). In contrast, when the null is imposed on the variance estimator, computation of confidence intervals is more complicated because the variance estimator depends on the null value of ρ_k. Fortunately, the end points of the confidence interval can be computed using the roots of a second order polynomial. The calculation is very similar to the confidence intervals obtained by Vogelsang and Nawaz (2017) for trend ratio parameters.

Recall the formula for the null-imposed t-statistic given by (9). A two tailed $(1-\alpha)\%$ confidence interval is the collection of values of a such that the null hypothesis is not rejected using the inequality

$$\left| \frac{(\widetilde{\rho}_k - a)}{\sqrt{\frac{1}{T-k}\widetilde{V}_{22}^{(k)*}}} \right| \leq cv_{\alpha/2}.$$

What complicates the calculation is that $\widetilde{V}_{22}^{(k)*}$ depends on a as we now show.

It is convenient to write $\widetilde{V}_{22}^{(k)*}$ in terms of quantities from the estimating equation (4) with the intercept projected out using the Frisch-Waugh-Lovell Theorem. Let $\ddot{y}_t$ and $\ddot{y}_{t-k}$ denote demeaned values where $\ddot{y}_t = y_t - \bar{y}_{\{k+1,T\}}$ and $\ddot{y}_{t-k} = y_{t-k} - \bar{y}_{\{1,T-k\}}$. Then $\widetilde{\rho}_k$ can be written as

$$\widetilde{\rho}_k = \frac{\sum_{t=k+1}^{T} \ddot{y}_{t-k}\ddot{y}_t}{\sum_{t=k+1}^{T} \ddot{y}_{t-k}^2},$$

and $\widetilde{\eta}_t^{(k)}$ can be written as

$$\widetilde{\eta}_t^{(k)} = \ddot{y}_t - a\ddot{y}_{t-k}.$$

Define

$$\ddot{v}_t^{(k)*} = \ddot{y}_{t-k}(\ddot{y}_t - a\ddot{y}_{t-k}) = \ddot{y}_{t-k}\ddot{y}_t - a\ddot{y}_{t-k}^2.$$

Then we rewrite $\widetilde{V}_{22}^{(k)*}$ equivalently as

$$\widetilde{V}_{22}^{(k)*} = \ddot{Q}^{-1}\ddot{\Omega}^{(k)*}\ddot{Q}^{-1},$$

where $\ddot{Q} = \frac{1}{T-k}\sum_{t=k+1}^{T} \ddot{y}_{t-k}^2$ and $\ddot{\Omega}^{(k)*}$ is the kernel long run variance estimator computed using the scalar process $\ddot{v}_t^{(k)*}$. It is well known in the literature that kernel long run variance estimators can be equivalently written as a quadratic form. For $\ddot{\Omega}^{(k)*}$ the quadratic form is

$$\begin{aligned}
\ddot{\Omega}^{(k)*} &= (T-k)^{-1} \sum_{t=k+1}^{T} \sum_{s=k+1}^{T} \ddot{v}_t^{(k)*} k_{ts} \ddot{v}_s^{(k)*} \\
&= (T-k)^{-1} \sum_{t=k+1}^{T} \sum_{s=k+1}^{T} (\ddot{y}_{t-k}\ddot{y}_t - a\ddot{y}_{t-k}^2) k_{ts} (\ddot{y}_{s-k}\ddot{y}_s - a\ddot{y}_{s-k}^2),
\end{aligned}$$

where $k_{ts} = k\left(\frac{|t-s|}{M}\right)$. Rearranging $\ddot{\Omega}^{(k)*}$ gives

$$\ddot{\Omega}^{(k)*} = \ddot{\Omega}_{11}^{(k)*} - 2a\ddot{\Omega}_{12}^{(k)*} + a^2\ddot{\Omega}_{22}^{(k)*}, \tag{10}$$

where

$$\ddot{\Omega}_{11}^{(k)*} = (T-k)^{-1} \sum_{t=k+1}^{T} \sum_{s=k+1}^{T} \ddot{y}_{t-k}\ddot{y}_t k_{ts} \ddot{y}_{s-k}\ddot{y}_s,$$
$$\ddot{\Omega}_{12}^{(k)*} = (T-k)^{-1} \sum_{t=k+1}^{T} \sum_{s=k+1}^{T} \ddot{y}_{t-k}\ddot{y}_t k_{ts} \ddot{y}_{s-k}^2,$$
$$\ddot{\Omega}_{22}^{(k)*} = (T-k)^{-1} \sum_{t=k+1}^{T} \sum_{s=k+1}^{T} \ddot{y}_{t-k}^2 k_{ts} \ddot{y}_{s-k}^2.$$

Using these variance formulas, we obtain an equivalent formula for $\widetilde{t}^{(k)*}$ given by

$$\widetilde{t}^{(k)*} = \frac{(\widetilde{\rho}_k - a)}{\sqrt{\frac{1}{T-k}\ddot{Q}^{-2}\left(\ddot{\Omega}_{11}^{(k)*} - 2a\ddot{\Omega}_{12}^{(k)*} + a^2\ddot{\Omega}_{22}^{(k)*}\right)}}.$$

The confidence interval for ρ_k is the values of a such that

$$\left| \frac{(\widetilde{\rho}_k - a)}{\sqrt{\frac{1}{T-k}\ddot{Q}^{-2}\left(\ddot{\Omega}_{11}^{(k)*} - 2a\ddot{\Omega}_{12}^{(k)*} + a^2\ddot{\Omega}_{22}^{(k)*}\right)}} \right| \leq cv_{\alpha/2},$$

or equivalently

$$\left(\frac{(\widetilde{\rho}_k - a)}{\sqrt{\frac{1}{T-k}\ddot{Q}^{-1}\left(\ddot{\Omega}_{11}^{(k)*} - 2a\ddot{\Omega}_{12}^{(k)*} + a^2\ddot{\Omega}_{22}^{(k)*}\right)}} \right)^2 \leq cv_{\alpha/2}^2. \tag{11}$$

The inequality (11) can be rewritten as

$$c_2 a^2 + 2c_1 a + c_0 \leq 0, \tag{12}$$

where

$$c_2 = 1 - \frac{1}{T-k}\ddot{Q}^{-2}\ddot{\Omega}_{22}^{(k)*} \cdot cv_{\alpha/2}^2,$$
$$c_1 = \frac{1}{T-k}\ddot{Q}^{-2}\ddot{\Omega}_{12}^{(k)*} \cdot cv_{\alpha/2}^2 - \widetilde{\rho}_k,$$
$$c_0 = \widetilde{\rho}_k^2 - \frac{1}{T-k}\ddot{Q}^{-2}\ddot{\Omega}_{11}^{(k)*} \cdot cv_{\alpha/2}^2.$$

Notice the importance of using a bandwidth rule for M that does not depend on a. Otherwise $\ddot{\Omega}_{11}^{(k)*}$, $\ddot{\Omega}_{12}^{(k)*}$ and $\ddot{\Omega}_{22}^{(k)*}$ would depend on a greatly complicating the solution to (12).

The values of a satisfying the inequality (12) are determined by the roots of the polynomial

$$p(a) = c_2 a^2 + 2c_1 a + c_0.$$

This polynomial has a similar form to the polynomial analyzed by Vogelsang and Nawaz (2017). Let r_1 and r_2 be the roots of $p(a)$ and order them $r_1 \leq r_2$ when they are real roots. The discriminant of the quadratic equation $p(a)$ is given by $c_1^2 - c_2 c_0$, so the shape of the confidence interval for a depends on the signs of c_2 and $c_1^2 - c_2 c_0$.

There are four cases. Case 1 has $c_2 > 0$ and $c_1^2 - c_2 c_0 \geq 0$ in which case the roots are real and $a \in [r_1, r_2]$. Case 2 has $c_2 > 0$ and $c_1^2 - c_2 c_0 < 0$ in which case $p(a)$ opens upward and its vertex is above zero giving roots that are complex numbers and an empty confidence interval. Case 3 has $c_2 < 0$ and $c_1^2 - c_2 c_0 > 0$ in which case the roots are real and $a \in (-1, r_1] \cup [r_2, 1)$ given that $p(a)$ opens downward and its vertex is above zero. Case 4 has $c_2 < 0$ and $c_1^2 - c_2 c_0 \leq 0$ in which case $a \in (-1, 1)$. It is important to note that Case 2 is *impossible* because the confidence interval cannot be empty given that it *always* contains the value $a = \widetilde{\rho}_k$ because in this case we have $\widetilde{t}^{(k)*} = 0$ and a non-rejection is obtained. The other cases are possible although it is not easy to find intuition as to the likelihood of each case.

First examine the sign of c_2. One can show that $\frac{1}{T-k}\ddot{Q}^{-2}\ddot{\Omega}_{22}^{(k)*} < 1$, however, $cv_{\alpha/2}^2$ will be greater than 1 for commonly used significance levels. Therefore, the sign of c_2 is inconclusive. Whether or not c_2 is positive depends on the kernel, bandwidth, the significance level, and the data. As T increases, $\frac{1}{T-k}\ddot{Q}^{-2}\ddot{\Omega}_{22}^{(k)*}$ converges to zero in which case it is more likely that c_2 is positive. Next examine the sign of $c_1^2 - c_2 c_0$. Algebra gives

$$\begin{aligned} c_1^2 - c_2 c_0 = &\left(\ddot{\Omega}_{12}^{(k)*} - \ddot{\Omega}_{11}^{(k)*}\ddot{\Omega}_{22}^{(k)*}\right)\left(\frac{1}{T-k}\ddot{Q}^{-2} \cdot cv_{\alpha/2}^2\right)^2 \\ &+ \left(\ddot{\Omega}_{11}^{(k)*} - 2\widetilde{\rho}_k \ddot{\Omega}_{12}^{(k)*} + \widetilde{\rho}_k^2 \ddot{\Omega}_{22}^{(k)*}\right)\left(\frac{1}{T-k}\ddot{Q}^{-2} \cdot cv_{\alpha/2}^2\right). \end{aligned}$$

We see that $c_1^2 - c_2 c_0$ is expressed as the sum of the two terms. The second term is the formula for $\ddot{\Omega}^{(k)*}$ with $a = \widetilde{\rho}_k$ in (10) and is scaled by a positive quantity. With appropriate choice of kernel, long run variances like $\ddot{\Omega}^{(k)*}$ are non-negative as argued by Priestley (1981) and Newey and West (1987). Therefore, the second term is non-negative. However, the first term is inconclusive because $\ddot{\Omega}_{12}^{(k)*} - \ddot{\Omega}_{11}^{(k)*}\ddot{\Omega}_{22}^{(k)*}$ can be positive or negative. Therefore, the sign of $c_1^2 - c_2 c_0$ is also inconclusive.

Confidence intervals can be computed using the orthonormal series variance estimator analogously with $\ddot{\Omega}_{11}^{(k)*}$, $\ddot{\Omega}_{12}^{(k)*}$, and $\ddot{\Omega}_{22}^{(k)*}$ replaced, respectively, by $\ddot{\Omega}_{OS,11}^{(k)*} = \frac{1}{K}\sum_{\ell=1}^{K} \ddot{\Lambda}_{\ell,1}^{*}\ddot{\Lambda}_{\ell,1}^{*}$, $\ddot{\Omega}_{OS,12}^{(k)*} = \frac{1}{K}\sum_{\ell=1}^{K} \ddot{\Lambda}_{\ell,1}^{*}\ddot{\Lambda}_{\ell,2}^{*}$, and $\ddot{\Omega}_{OS,22}^{(k)*} = \frac{1}{K}\sum_{\ell=1}^{K} \ddot{\Lambda}_{\ell,2}^{*}\ddot{\Lambda}_{\ell,2}^{*}$ where $\ddot{\Lambda}_{\ell,1}^{*} = \frac{1}{\sqrt{T-k}}\sum_{t=k+1}^{T} \Phi_\ell\left(\frac{t}{T}\right)\ddot{y}_{t-k}\ddot{y}_t$ and $\ddot{\Lambda}_{\ell,2}^{*} = \frac{1}{\sqrt{T-k}}\sum_{t=k+1}^{T} \Phi_\ell\left(\frac{t}{T}\right)\ddot{y}_{t-k}^2$.

4 Monte Carlo Simulations

In this section we study finite sample properties of the proposed t-statistics for testing

$$H_0 : \rho_k = a$$

through extensive Monte Carlo simulations. We use 5,000 replications in all cases. We compare our t-statistics, $\widetilde{t}^{(k)}$, $\widetilde{t}^{(k)*}$, $\widetilde{t}_{OS}^{(k)}$ and $\widetilde{t}_{OS}^{(k)*}$ with each other and with some existing approaches. Fixed-b critical values are used for $\widetilde{t}^{(k)}$, $\widetilde{t}^{(k)*}$ and critical values from the t_K distribution (fixed-K critical values) are used for $\widetilde{t}_{OS}^{(k)}$ and $\widetilde{t}_{OS}^{(k)*}$. We also provide some results for $\widetilde{t}^{(k)*}$ using $N(0,1)$ critical values to show the value of using fixed-b critical values. For $\widetilde{t}^{(k)}$ we used the data dependent bandwidth, denoted by $\widetilde{M}$, proposed by Sun et al. (2008) that balances size distortions and power of the tests, the 'test-optimal-M'. The weighting parameter that balances type 1 and type 2 errors is set to 10. The null-imposed statistic, $\widetilde{t}^{(k)*}$, also uses $\widetilde{M}$ so that its bandwidth does not depend on the value of the null being tested. For $\widetilde{t}_{OS}^{(k)}$ we used the data dependent smoothing parameter, denoted by $\widetilde{K}$, proposed by Phillips (2005) that minimizes the mean square error of the variance estimator, the 'MSE-optimal-K'. The null-imposed statistic, $\widetilde{t}_{OS}^{(k)*}$, also uses $\widetilde{K}$ to avoid dependence on the value of the null being tested. For both $\widetilde{M}$ and $\widetilde{K}$ we use well known AR(1) plug-in methods (see Andrews (1991)) that are functions of $\widetilde{\mathbf{v}}_t^{(k)}$, the null-not-imposed proxy for $\mathbf{v}_t^{(k)}$ given by equation (8).

Results are given for a broad set of data generating processes (DGPs) where y_t follows the $ARMA(1,1)$ process

$$y_t = \phi y_{t-1} + \epsilon_t + \theta\epsilon_{t-1}, \tag{13}$$

where $\mu = 0$ without loss of generality given that we include in a intercept in the estimating equation (4). Special cases include uncorrelated y_t ($\phi = 0$, $\theta = 0$) and $AR(1)$ ($\theta = 0$) and $MA(1)$ ($\phi = 0$) processes. Results are given for nine DGPs of the innovation process, ϵ_t, ranging from i.i.d. to cases with increasing dependence in higher moments.

DGP 1: IID : $\epsilon_t = u_t \sim i.i.d.N(0, 1)$.

DGP 2: MDS : $\epsilon_t = u_t u_{t-1}$, $u_t \sim i.i.d.N(0, 1)$..

DGP 3: GARCH : $\epsilon_t = h_t u_t$ and $h_t^2 = 0.1 + 0.09\epsilon_{t-1}^2 + 0.9h_{t-1}^2$, $u_t \sim i.i.d.N(0, 1)$.

DGP 4: WN-1 : $\epsilon_t = u_t + u_{t-1}u_{t-2}$, $u_t \sim i.i.d.N(0, 1)$.

DGP 5: WN-2: $\epsilon_t = u_t^2 u_{t-1}$, $u_t \sim i.i.d.N(0, 1)$.

DGP 6: WN-NLMA: $\epsilon_t = u_{t-2}u_{t-1}(u_{t-2} + u_t + 1)$, $u_t \sim i.i.d.N(0, 1)$.

DGP 7: WN-BILIN: $\epsilon_t = u_t + 0.5u_{t-1}\epsilon_{t-2}$, $u_t \sim i.i.d.N(0, 1)$.

DGP 8: WN-GAM1 : $\epsilon_t = u_t + u_{t-1}u_{t-2}$, $u_t = \zeta_t - E[\zeta_t]$, $\zeta_t \sim i.i.d.Gamma(0.3, 0.4)$.

DGP 9: WN-GAM2 : $\epsilon_t = u_t - u_{t-1}u_{t-2}$, $u_t = \zeta_t - E[\zeta_t]$, $\zeta_t \sim i.i.d.Gamma(0.3, 0.4)$.

DGP 1 is an i.i.d. Gaussian innovation and serves as a benchmark given that all approaches are valid for this case. DGP 2 relaxes the i.i.d. assumption and ϵ_t is a martingale difference sequence (MDS) innovation that has been studied in the literature. See Romano and Thombs (1996) and Francq and Zakoïan (2009). DGP 3 is a $GARCH(1, 1)$ innovation typical in financial time series. DGPs 4-9 are white noise processes with stronger dependence than the MDS case. DGP 4 is from Hansen (2022) and ϵ_t follows a white noise process that is a function of an underlying i.i.d. Gaussian process. DGP 5 is a white noise process from Wang and Sun (2020). DGPs 6 and 7 are white noise processes from Lobato (2001). DGPs 8 and 9 build white noise process using independent centered Gamma random variables generating some skewness in u_t.

4.1 Null Rejections for Uncorrelated Time Series

We first focus on the case where y_t is uncorrelated, i.e. $\rho_k = 0$ or equivalently $\phi = 0$, $\theta = 0$ in (13). For this case we focus on the first order autocorrelation ($k = 1$) and examine tests of the null hypothesis

$$H_0 : \rho_1 = 0.$$

We consider the (original) Bartlett formula, the generalized Bartlett formula, and White standard errors for constructing t-statistics that we compare to our proposed t-statistics. We carry out two-tailed tests with a nominal significance level of 0.05. The original Bartlett formula always uses $v^B_{1,1} = 1$ whether or not y_t is i.i.d. For the generalized Bartlett formula, we use the formula (3) from Francq and Zakoïan (2009) for a white noise process. White standard errors are a special case of $\widetilde{\Omega}^{(k)}$ where only the $\widetilde{\Gamma}_0^{(k)}$ term is used. Because testing $\rho_1 = 0$ is a zero autocorrelation test for the lag one autocorrelation, we also include the zero autocorrelation test of Taylor (1984) which has recently been extended by Dalla, Giraitis and Phillips (2022). The Taylor (1984) $\widetilde{\tau}_1$ t-statistic is given by

$$\widetilde{\tau}_1 = \frac{\sum_{t=2}^{n} e_{t1}}{\left(\sum_{t=2}^{n} e_{t1}^2\right)^{1/2}}, \quad e_{t1} = \left(y_t - \bar{y}\right)\left(y_{t-1} - \bar{y}\right).$$

Dalla et al. (2022) provide conditions under which $\widetilde{\tau}_1$ is asymptotically standard normally distributed. We also report results using the bootstrap method suggested by Romano and Thombs (1996) where the bootstrapped version of $\widehat{\rho}_k$ is centered around $\widehat{\rho}_k$ but is not standardized (see their equation (11) on page 594). We report results using the moving block bootstrap with block length equal to $\sqrt{T}$. For the case where the DGP for ϵ_t is i.i.d. we also report results using block length equal to 1 (the i.i.d. bootstrap). We obtained results using the stationary bootstrap and the circular bootstrap but exclude them from reporting because they give similar results and patterns as the moving block bootstrap. We also obtained results using subsampling but found those results less accurate than the bootstrap and those results are omitted.

Figures 1.1 through 1.9 plot empirical null rejection probabilities for each of the nine cases for ϵ_t. Results are given for sample sizes $T = 100$, 200, 500 and 2000. The labels Fixed-b (SPJ) and Fixed-b-H_0 (SPJ) correspond to $\widetilde{t}^{(1)}$ and $\widetilde{t}^{(1)*}$ respectively using fixed-b critical values. $N(0,1)$-H_0 (SPJ) corresponds to $\widetilde{t}^{(1)*}$ using $N(0,1)$ critical values. The (SPJ) label indicates that the same data dependent bandwidth, $\widetilde{M}$, was used for all three tests. The labels OS (MSE) and OS-H_0 (MSE) correspond to $\widetilde{t}_{OS}^{(1)}$ and $\widetilde{t}_{OS}^{(1)*}$ using the same $\widetilde{K}$ smoothing parameter.

To understand many of the patterns in Figures 1.1 - 1.9, it is useful to keep in mind that $v_t^{(1)} = \epsilon_{t-1}\epsilon_t$ when y_t is uncorrelated. For the IID, MDS and GARCH DGPs, $v_t^{(1)}$ is obviously uncorrelated. While not as obvious, $v_t^{(1)}$ is uncorrelated for the white noise processes WN-1, WN-2, WN-NLMA and WN-BILIN. In contrast, $v_t^{(1)}$ is positively autocorrelated for the WN-GAM1 DGP because one can show that $E\left(v_t^{(1)}v_{t-1}^{(1)}\right) = E\left(u_t^3\right)E(u_t^2) > 0$ given that $E\left(u_t^3\right) > 0$ for the Gamma parameters we use. The sign change in the WN-GAM2 DGP generates negative autocorrelation[3] in $v_t^{(1)}$ because $E\left(v_t^{(1)}v_{t-1}^{(1)}\right) = -E\left(u_t^3\right)E(u_t^2) < 0$.

Figure 1.1 depicts null rejection probabilities for the IID DGP ($y_t = \epsilon_t$ is i.i.d.). There are slight over-rejections for $\widetilde{t}^{(1)}$ (null-not-imposed kernel HAR statistic) with fixed-b critical values (red squares dash-dotted line) for $T = 100$ because for this method there is variability in $\widetilde{v}_t^{(1)}$ from estimating ρ_1 that matters when T is relatively small. Imposing the null for the kernel HAR approach reduces over-rejections as illustrated by $\widetilde{t}^{(1)*}$ using fixed-b critical values (purple up-arrow dotted line). Using normal critical values for $\widetilde{t}^{(1)*}$ (gold circle dash-dotted line) shows some over-rejections and illustrates the benefits of using fixed-b critical values. The null rejections of $\widetilde{t}_{OS}^{(1)}$ (null-not-imposed, orange x's solid line) and $\widetilde{t}_{OS}^{(1)*}$ (null-imposed, light green down-arrow dashed line) are similar to the rejections of $\widetilde{t}^{(1)}$ and $\widetilde{t}^{(1)*}$. Rejections are close to 0.05 for all traditional methods (Bartlett formula (blue dot solid lines 'Bartlett(IID)'), generalized Bartlett (light blue star dashed line 'GB-WN'), Taylor (yellow rhombus dotted line 'Taylor') and White standard errors (green right-arrow dotted line 'White')). It is surprising to see that the i.i.d. bootstrap (black down-arrow dashed line 'IID-bootstrap') does not work for the i.i.d. DGP. Null rejections for the i.i.d. bootstrap are about 0.33 even when T increases to 2000. Interestingly, the moving block bootstrap (black circle dotted line 'MBB') performs better than the i.i.d. bootstrap even though the data has no

[3]Notice that the WN-1 and WN-GAM1,WN-GAM2 DGPs take the same form. The reason that $v_t^{(1)}$ is uncorrelated for WN-1 is because u_t is normally distributed. Normality implies that $E(u_t^3) = 0$ and it follows that $E\left(v_t^{(1)}v_{t-1}^{(1)}\right) = 0$.

dependence. Even so, rejections with the moving block bootstrap range from 0.15 with $T = 100$ to about 0.07 with $T = 2000$ whereas all non-bootstrap tests have rejections close to 0.05 when $T = 100$ and very close to 0.05 when $T = 2000$.

Figures 1.2-1.7 relax the i.i.d. assumption and give results for y_t being an MDS, GARCH and the various white noise series that satisfy, with the exception of the original Bartlett variance, the conditions of the traditional approaches. We see similar patterns as in the i.i.d. case, however more size distortions occur for $\widetilde{t}^{(1)}$ and $\widetilde{t}_{OS}^{(1)}$ (null-not-imposed) for smaller sample sizes. In contrast, $\widetilde{t}^{(1)*}$ and $\widetilde{t}_{OS}^{(1)*}$ (null-imposed) have rejections close to 0.05. This indicates potential size improvements by imposing the null, consistent with the findings in Lazarus et al. (2018) and Vogelsang (2018) in stationary regression settings. The traditional Bartlett formula shows over-rejections which is expected with the i.i.d. assumption violated. The moving block bootstrap continues to have substantial over-rejections especially for small sample sizes for all DGPs. The other traditional methods work reasonably well as expected given that y_t satisfies the required assumptions for those methods.

Figures 1.8 and 1.9 give results for the white noise case with Gamma distributed innovations. For the WN-GAM1 DGP (Figure 1.8) all tests show some over-rejections with $\widetilde{t}^{(1)*}$ and $\widetilde{t}_{OS}^{(1)*}$ (null-imposed) having rejections closest to 0.05. The null-not-imposed tests, $\widetilde{t}^{(1)}$ and $\widetilde{t}_{OS}^{(1)}$, have substantial over-rejections for small T but rejections approach 0.05 as T increases. All of the traditional methods have over-rejections even when T is large because this DGP violates the assumptions for those methods. In particular Taylor and White are designed for the case where $v_t^{(1)}$ is uncorrelated and that fails here. The generalized Bartlett formula uses a symmetry assumption for cross fourth moments of ϵ_t that is violated in the Gamma distribution case. Figure 1.9 shows that if we flip the sign on $u_{t-1}u_{t-2}$, rejections change dramatically with all tests under-rejecting. Under-rejections make sense because flipping the sign generates negative autocorrelation in $v_t^{(1)}$ for the WN-GAM2 DGP. The traditional methods can have very low rejections close to zero. As T increases the rejections using the estimating equation approach tends towards 0.05 but the traditional methods do not. The moving block bootstrap continues to over-reject and does not perform as well as non-bootstrap methods.

It is a common misconception that $y_t = \epsilon_t$ being uncorrelated implies that $v_t^{(1)} = \epsilon_{t-1}\epsilon_t$ will be uncorrelated. However, because it is possible for $\epsilon_{t-1}\epsilon_t$ to have serial correlation when ϵ_t is uncorrelated, the generalized-Bartlett, White, and Taylor approaches are not necessarily valid when y_t is uncorrelated. One benefit of the estimating equation approach is that it automatically handles white noise innovations including the case where $v_t^{(1)}$ has serial correlation.

Finally, our simulation results for the bootstrap are puzzling especially in the i.i.d. case given the relatively simple form of $\widehat{\rho}_1$. An analytical analysis of why the bootstrap is not performing as expected is part of an ongoing research project that we will report in a follow-up paper.

4.2 Null Rejections for Serially Correlated Time Series

Next we focus on cases of serially correlated time series where $\rho_k \neq 0$. We continue to focus on tests of the first order autocorrelation ($k = 1$) and consider the null hypothesis

$$H_0 : \rho_1 = \rho_1^{(0)},$$

where $\rho_1^{(0)}$ is the true value of ρ_1, and $\rho_1^{(0)}$ depends on the serial correlation structure of y_t. We exclude the Taylor and White approaches because they are no longer valid when $\rho_k \neq 0$. We do not report bootstrap results because of the bootstrap's relatively poor performance with uncorrelated data.

Two versions of the generalized Bartlett approach are included. One assumes that y_t is white noise (GB-WN) and the other assumes y_t follows an MA(1) process (GB-MA)[4]. The formula for GB-MA is given by

$$v_{1,1}^{B*} = \frac{\gamma_{\epsilon^2}(0)}{[\gamma_\epsilon(0)]^2}\left[\rho_{\epsilon^2}(1)(1 - 4\rho_1 + 4\rho_1^4) + \rho_{\epsilon^2}(2)\rho_1^2\right].$$

We derived this formula using the general expression in Francq and Zakoïan (2009). The corresponding estimator is obtained by plugging in estimators of the parameters. We estimate ρ_1 using (2). We estimate $\gamma_\epsilon(0)$ using the sample variance of $\widehat{\epsilon}_t$ where $\widehat{\epsilon}_t$ are the residuals from fitting an MA(1) model to $y_t - \overline{y}$. The parameters $\gamma_{\epsilon^2}(0)$, $\rho_{\epsilon^2}(1)$, $\rho_{\epsilon^2}(2)$ are estimated using sample analogs computed with $\widehat{\epsilon}_t^2$.

[4]We do not implement versions of the generalized Bartlett approach designed for the case when y_t has the AR(1) component because the form of the generalized Bartlett variance formula for the AR(1) case is complicated and is very difficult to implement.

Results are given for the MA(1) case in Figures 2-6 and the AR(1) case in Figures 6-11. Results for ARMA(1,1) specifications are similar and are omitted. We exclude DGPs WN-2,WN-NLMA and WN-BILIN for ϵ_t given the similarity in patterns to WN-1. We also exclude WN-GAM2. We continue to use two-tailed tests with 0.05 nominal level. Each figure has four panels corresponding to the sample sizes $T = 100$, $200, 500$, and 1000. The x-axis indicates the value of either θ or ϕ. For the MA(1) case, $\rho_1^{(0)} = \theta / \left(1 + \theta^2\right)$ and for the AR(1) case $\rho_1^{(0)} = \phi$.

Figure 2 gives results for MA(1) case with ϵ_t i.i.d. Not surprisingly, all approaches work reasonably well except for GB-WN which under-rejects unless $\theta = 0$. This is expected given that GB-WN is invalid except when $\theta = 0$. Figure 3 gives MA(1) results where ϵ_t follows the MDS DGP. The traditional Bartlett approach (MA(1)) over-rejects because ϵ_t is not i.i.d. For $T = 100$, GB-MA (green star dashed line) tends to over-reject. Rejections become closer to 0.05 as T increases. The small sample distortions are likely caused by the need to estimate θ. Similar to MA(1) with ϵ_t i.i.d, GB-WN continues to under-reject. The null-imposed kernel HAR test, $\widetilde{t}^{(1)*}$, works well whether normal critical values (N(0,1)-H_0) or fixed-b critical values (Fixed-b-H_0) are used. The null-imposed orthonormal series test, $\widetilde{t}_{OS}^{(1)*}$ (OS-H_0) has similar performance to $\widetilde{t}^{(1)*}$. Not imposing the null leads to over-rejections for both $\widetilde{t}^{(1)}$ (Fixed-b) and $\widetilde{t}_{OS}^{(1)}$ (OS) when T is relatively small. This again illustrates that more reliable inference under the null is obtained by imposing the null on the kernel and orthonormal series variance estimators. When ϵ_t is a GARCH process, Figure 4 shows that all methods works well except for the traditional Bartlett and GB-WN as one would expect. Figure 5 gives results for the case of ϵ_t being white noise (WN-1 DGP) and we see that patterns are similar to the MDS case. In contrast, patterns are clearly different when ϵ_t is the white noise driven by Gamma errors (WN-GAM1 DGP) as seen in Figure 6. None of the Bartlett approaches are valid in this case and rejections are either well above or well below 0.05. The null-imposed HAR approaches, $\widetilde{t}^{(1)*}$ and $\widetilde{t}_{OS}^{(1)*}$, perform best especially with fixed-b critical values. Not imposing the null can lead to nontrivial over-rejections. While rejections of the HAR tests get closer to 0.05 with larger sample sizes, there are still some size distortions even with $T = 2000$. Our conjecture is that the CLT and FCLT 'kick in' more slowly as T increases in the Gamma distribution case.

We now turn to Figures 7-11 for the AR(1) results. Keep in mind that both GB-MA and GB-WN use formulas based on a misspecified model and are not expected to perform well. Figure 7 gives results for ϵ_t i.i.d. We can see that the misspecified GB approaches have size distortions that persist with larger T. The Bartlett (AR(1)) and HAR tests perform reasonably well with small T with some slight over-rejections. Rejections are close to 0.05 with $T = 2000$. Figures 8, 9 and 10 give AR(1) results for ϵ_t MDS, GARCH and WN-1 respectively. When the errors are MDS and GARCH (Figures 8 and 9), we can see that the null-imposed HAR tests $\widetilde{t}^{(1)*}$ (Fixed-b-H_0) and $\widetilde{t}^{(1)*}_{OS}$ (OS-H_0) perform well with null rejections reasonably close to 0.05. When ϵ_t is white noise (Figures 10 and 11), all approaches exhibit over-rejections when $\phi > 0$ especially as ϕ approaches 1. Increasing T improves the performance of the HAR approaches.

4.3 Power Analysis

In this subsection we study finite sample power of the test statistics. We use size-adjusted power to account for the size distortions of the tests. This allows power comparisons with the same null rejections. Because we use size-adjusted finite sample critical values, there is no need to distinguish between using $N(0,1)$ and fixed-b critical values for $\widetilde{t}^{(1)*}$. In the power figures we label $\widetilde{t}^{(1)}$ as 'Kernel' and $\widetilde{t}^{(1)*}$ as 'Kernel-H_0'. We report three sets of power results for the case where y_t is AR(1) and ϵ_t is IID (DGP 1), WN-NLMA (DGP 6) and WN-GAM1 (DGP 8). The null hypothesis in all cases is $H_0 : \rho_1 = 0$ with the alternative given by $H_1 : \rho_1 = \phi$. We use a 0.1 grid for ϕ on the interval $[-0.5, 0.5]$. Results are reported for $T = 100$, 200, 300, and 500.

Figure 12 gives results for ϵ_t IID. Size-adjusted power is essentially the same across all tests. Figure 13 gives results for ϵ_t WN-NLMA. Size-adjusted power is similar across tests although one can see that the null-imposed HAR tests have slightly lower power for negative values of ρ_1. This is more apparent in Figure 14 where results for ϵ_t WN-GAM1 are given. With $T = 100$, power is lower for the null-imposed tests for negative values of ρ_1. Interestingly, these power differences disappear when $T = 500$. There are also some asymmetries in power around ρ_1 in the white noise cases, especially WN-GAM1, that do not occur with ϵ_t IID.

While the null-imposed HAR tests can have lower power than the null-not-imposed HAR tests, the power differences are relatively small and disappear as T increases. Given the superior null rejections of the null-imposed-tests and their respectable power, we can recommended them in practice.

4.4 Null Rejection Probabilities Across Lags

The finite sample results to this point have focused on the case of $k = 1$. In this subsection we provide results for other values of k. We report results for the $AR(1)$ case for $\phi = 0$ and $\phi = 0.5$ with k ranging from 1 to 10. The null hypothesis is

$$H_0 : \rho_k = \phi^k,$$

given the AR(1) structure. Results are reported for the HAR tests and the recursive MA approach used by the software Stata given by equation (1). We report results for $T = 50$, 100, 250 and 1000. We continue to focus on two-sided tests with a nominal level of 0.05. Results for $\phi = 0$ are given in Figures 15-19 for ϵ_t IID, MDS, GARCH, WN-1 and WN-GAM1. For ϵ_t IID (Figure 15) the HAR tests, especially the null-imposed versions, work well for all k with rejections very close to 0.05 as T increases. The Stata procedure (blue dot solid line labeled 'Software') works well with large T but under-rejects for small T and larger values of k. This makes sense because the estimated variance used by Stata increases mechanically as k increases. Figure 16 shows that when ϵ_t is an MDS, the null-imposed HAR tests continue to perform well but the null-not-imposed HAR tests have some over-rejections with small values of T. The Stata procedure relies on the i.i.d. assumption for ϵ_t and breaks down for $k = 1$. In the case of GARCH innovations, Figure 17 shows that the HAR tests perform well, again imposing the null works best. The Stata procedure completely breaks down. When ϵ_t is white noise, Figures 18 and 19 show that the null-imposed HAR tests continue to work well for all k including the $T = 50$ case. Not imposing the null results in HAR tests that can have substantial over-rejections for small values of k especially when T is not large. The Stata procedure breaks down for $k = 1, 2$ but works reasonably well for $k \geq 3$. These results show that when y_t is uncorrelated, the Stata procedure only works when y_t is i.i.d. In contrast, the HAR tests with the null-imposed work quite well including the case of y_t being white noise.

The results with $\phi = 0.5$ are given in Figures 20-24 for the same cases for ϵ_t. The Stata procedure is not valid for any of these cases given the AR(1) structure. The null-imposed HAR tests work well overall but do have some relatively minor size distortions when $T = 50$. The null-not-imposed HAR tests can have substantial over-rejections with small values of T and small values of k. An interesting contrast can also be seen in these figures for $\widetilde{t}^{(k)*}$ and $\widetilde{t}_{OS}^{(k)*}$. When these two tests have some over-rejections, they are less pronounced for $\widetilde{t}^{(k)*}$ than for $\widetilde{t}_{OS}^{(k)*}$. This is not because $\widetilde{t}^{(k)*}$ uses a kernel and $\widetilde{t}_{OS}^{(k)*}$ and uses series to estimate the long run variance. The reason is that the MSE criteria for smoothing parameters of long run variances leads to less smoothing than the test based criteria. Less smoothing (e.g. smaller bandwidths for kernel estimators) is well known to lead to tests with a greater tendency to over-reject in finite samples when fixed-smoothing critical values are used (see the simulations in Kiefer and Vogelsang (2005) for the kernel case). The reason that $\widetilde{t}_{OS}^{(k)*}$ tends to over-reject more than $\widetilde{t}^{(k)*}$ is because $\widetilde{K}$ leads to less smoothing than $\widetilde{M}$.

4.5 Shape of Confidence Intervals

In section 3.3, we showed that confidence intervals computed with the null-imposed HAR statistics can take three forms. In this section we investigate the likelihood of the forms for some representative DGPs from our simulation design. We provide results for confidence intervals using $\widetilde{t}^{(k)*}$ with fixed-b critical values. Results with $\widetilde{t}_{OS}^{(k)*}$ using t_K critical values are similar and are not reported. Tables 2 and 3 give results for the AR(1) case with ϵ_t IID and ϵ_t WN-NLMA (DGP 6). These results nicely show the range of possibilities. Results are given for $T = 50$, 100, 250, 500 and AR(1) values $\phi = 0, 0.25, 0.7, -0.7$. We use 10,000 replications.

Tables 2 and 3 are organized as follows. For each pair of values for ϕ and T, we report the empirical probabilities of each confidence interval type (Prob), the empirical coverage probability of the confidence interval (ECP), and the average confidence interval length ($\overline{CI}$) conditional on the confidence interval type and overall. The AR-IID results in Table 2 serve as a benchmark. The first panel of the table ($\phi = 0$) gives result for when y_t is i.i.d. We can see that for all sample sizes the probability of obtaining the typical

$[r_1, r_2]$ confidence interval is 1.0. As ϕ moves away from 0 and for smaller values of T, there are very small, but non-zero, probabilities of obtaining the confidence intervals $(-1, r_1] \cup [r_2, 1)$ and $(-1, 1)$.

Table 3 shows very different patterns from Table 2. With no autocorrelation or relatively weak autocorrelation ($\phi = 0.25$), there is about a 50% chance of shapes $(-1, r_1] \cup [r_2, 1)$ and $(-1, 1)$ with T small. In these cases, the empirical coverages and confidence lengths are larger than the $[r_1, r_2]$ case (this is obviously true by construction when the confidence interval is $(-1, 1)$). As T increases or ϕ moves farther away from zero, the probability of $[r_1, r_2]$ confidence interval shape increases. As one expect, average confidence interval lengths shrink as T increases.

These results show that for smaller sample sizes and more complex dependence in y_t and its innovations, ϵ_t, disjoint and possibly very wide confidence intervals can occur. While some empirical practitioners may be bothered by disjoint or wide confidence intervals, we view these cases as providing the practitioner with a signal that y_t has potentially complex serial correlation structure with innovations that have complex dependence in higher moments that matter for inference about the autocorrelations of y_t. In other words, disjoint or wide confidence intervals are signals of data that has limited information about autocorrelation structure.

5 Empirical Application

The autocorrelation function is widely used as a preliminary step in analyzing financial time series. The Bartlett formula is commonly used as part of graphical evidence of autocorrelation structure. For example, Bollerslev and Mikkelsen (1996) provides a figure of sample autocorrelations for absolute daily returns of the S&P 500 index with the 95% confidence bands[5] implied by the Bartlett formula for i.i.d. data to illustrate volatility clustering and its long-term dependence. Andersen, Bollerslev, Diebold and Labys (2003) provides figures of sample autocorrelations for daily exchange rate realized volatilities before and after fractional differencing along with the i.i.d. Bartlett confidence bands to graphically confirm evidence of long memory.

While the i.i.d. Bartlett confidence bands are routinely reported in practice, it is important to keep

[5] A confidence band is used to test the null hypothesis of zero autocorrelation and is not a confidence interval. An estimated value outside the band is a rejection of the null.

in mind the limitations of these confidence bands. First, the confidence bands are only valid if the data is i.i.d. If the data is uncorrelated but not i.i.d. (martingale difference, white noise), then the bands are no longer valid. Second, the bands can only be used to test the null hypothesis that the series is i.i.d. Once it is determined that the series has dependence, the bands **cannot** be used to assess significance of autocorrelations at specific lags because the bands are not generally valid when there is serial correlation.

A more informative approach is to report confidence intervals using $\widetilde{t}^{(k)*}$ or $\widetilde{t}_{OS}^{(k)*}$ allowing inference about autocorrelations that is valid for general serial correlation structures and innovations that are not necessarily i.i.d. As an illustration we provide some empirical results for S&P 500 index returns and absolute returns for two sets of time periods (before Covid and during/after Covid) that have the same number of observations (913 observations for each) but exhibit different estimated autocorrelation patterns and confidence intervals. Figure 25 provides plots of the returns and the absolute returns for the full time span of the observations from June 28, 2016 to September 28, 2023. Figure 26 plots estimated autocorrelations for S&P 500 returns for daily data from June 28, 2016 to February 12, 2020 (Panel (a)) and February 13, 2020 to September 28, 2023 (Panel (b)). Red circles are the sample autocorrelations given by (2) and blue dots are autocorrelations estimated by OLS using (4). The dashed red lines are i.i.d. Bartlett confidence bands. The gray area is the Stata confidence bands using equation (1). The black lines with bars are 95% confidence intervals computed using $\widetilde{t}^{(k)*}$ with fixed-b critical values. The dash-dot green lines are 95% confidence bands using $\widetilde{t}^{(k)*}$ that can be used to test a given autocorrelation is zero. One can equivalently test an autocorrelation is zero by checking that the confidence interval contains zero.

Figure 26 gives results for returns which provides information about market efficiency. Panel (a) shows that estimated autocorrelations of returns are close to zero and, in nearly every case, not statistically significant. If one used the Bartlett or Stata confidence bands, one would conclude there is no evidence to reject the null that returns are uncorrelated (equity market is efficient). However, that conclusion is subject to the caveat that the bands are only valid if the innovations are i.i.d. In contrast, the confidence intervals using $\widetilde{t}^{(k)*}$ allow more robust inference. Because nearly all the confidence intervals contain zero, we cannot reject the null returns are uncorrelated whether or not innovations are i.i.d. or are simply uncorrelated.

Panel (b) of Figure 26 is distinctly different and interesting because conclusions depend critically on the method used and its assumptions. Using the Bartlett or Stata confidence bands, one would conclude there is evidence to reject the null hypothesis that returns are uncorrelated in the Covid/Post-Covid period given that many sample autocorrelations are outside the bands. This conclusion is only valid if innovations are i.i.d. Furthermore, these bands *cannot* be used to conclude anything further about the autocorrelation structure because the confidence bands are *not* confidence intervals. In contrast, the $\widetilde{t}^{(k)*}$ confidence intervals tell a different story. While the estimated autocorrelations are larger in magnitude compared to the pre-Covid period, nearly all the confidence intervals contain zero. Therefore, using robust confidence intervals, one cannot reject that returns are uncorrelated in the Covid/Post-Covid period. The fact that confidence intervals are wider in this period is an indication that the innovations have potentially more complex higher order dependence and/or GARCH effects than the pre-Covid period.

Figure 27 gives results for absolute returns which provides information about volatility clustering and the dependence structure of volatility (Bollerslev and Mikkelsen (1996)). In Panel (a) we see positive estimated autocorrelations with tight confidence intervals. While the estimated autocorrelations are not large in magnitude, they persistent at long lags and are statistically significant (all confidence intervals do not contain zero). This evidence implies volatility clustering during the pre-Covid period. Panel (b) is an interesting contrast. While estimated autocorrelations are larger, confidence intervals are substantially wider. Notice that we cannot reject that the first six lags have zero autocorrelation. While it may be tempting to argue that there is stronger evidence for volatility clustering and higher persistence during the Covid/Post-Covid period, the wide confidence intervals suggest something else may be happening in this period that warrants further investigation. Here, if one only looked at the Bartlett or Stata confidence bands, a potentially misleading conclusion might be reached.

6 Conclusion

This paper develops an estimating equation approach for robust confidence intervals for the autocorrelation function of a stationary time series. Our approach is applicable to general stationary time series with uncorrelated innovations that can have dependence in higher order moments (innovations do not have to be i.i.d.). Except for narrow exceptions, the asymptotic variance of estimated autocorrelations take a sandwich form. The asymptotic variance can be directly estimated by well known HAR variance estimators allowing t-statistics and confidence intervals to be easily constructed. We consider HAR variance estimators that impose the null leading to more reliable inference. We provide conditions under which fixed-smoothing critical values can be used for t-tests and confidence intervals and recommend those critical values be used in practice.

Our extensive simulation study shows that the tests based on the null-imposed variance estimator in conjunction with fixed-smoothing critical values leads to inference about the autocorrelation function that works well in practice both in terms of controlling null rejection probabilities and having good power. Our approach can be used to report generally valid confidence intervals for covariance stationary time series under weak assumptions for the innovations. In contrast existing software packages typically report confidence bands based on strong assumptions that can only be used to test narrow hypotheses (and are often misused in practice). Our approach is an improvement and allows the testing of significantly broader hypothesis about the autocorrelation function in a highly robust manner.

Our simulation results also reveal a puzzle regarding the use of the bootstrap for inference about the autocorrelation function. For the case of uncorrelated data (including the case of i.i.d. data) we find that the block bootstrap and related bootstrap approaches do not perform as well as expected even in the case where the data is i.i.d. and the i.i.d. bootstrap is used. An analysis of the bootstrap applied to inference about the autocorrelation function is a topic of ongoing research that we plan to report in a follow-up paper.

References

Andersen, T., Bollerslev, T., Diebold, F. and Labys, P.: (2003), Modeling and forecasting realized volatility, *Econometrica* **71**(2), 579–625.

Andrews, D. W. K.: (1991), Heteroskedasticity and autocorrelation consistent covariance matrix estimation, *Econometrica* **59**, 817–854.

Bartlett, M.: (1946), On the theoretical specification and sampling properties of autocorrelated time-series, *Supplement to the Journal of the Royal Statistical Society* **8**(1), 27–41.

Bollerslev, T. and Mikkelsen, H.: (1996), Modeling and pricing long memory in stock market volatility, *Journal of Econometrics* **73**(1), 151–184.

Dalla, V., Giraitis, L. and Phillips, P.: (2022), Robust tests for white noise and cross-correlation, *Econometric Theory* **38**(5), 913–941.

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

de Jong, R. M. and Davidson, J.: (2000), Consistency of kernel estimators of heteroskedastic and autocorrelated covariance matrices, *Econometrica* **68**, 407–424.

Francq, C. and Zakoïan, J.: (2009), Bartlett's formula for a general class of nonlinear processes, *Journal of Time Series Analysis* **30**(4), 449–465.

Gonçalves, S. and Vogelsang, T. J.: (2011), Block bootstrap HAC robust tests: The sophistication of the naive bootstrap, *Econometric Theory* **27**(4), 745–791.

Hansen, B.: (2022), *Econometrics*, Princeton University Press.

Kiefer, N. M. and Vogelsang, T. J.: (2005), A new asymptotic theory for heteroskedasticity-autocorrelation robust tests, *Econometric Theory* **21**, 1130–1164.

Lazarus, E., Lewis, D. J. and Stock, J. H.: (2021), The size-power tradeoff in HAR inference, *Econometrica* **89**(5), 2497–2516.

Lazarus, E., Lewis, D. J., Stock, J. H. and Watson, M. W.: (2018), HAR inference: Recommendations for practice, *Journal of Business & Economic Statistics* **36**(4), 541–559.

Lobato, I.: (2001), Testing that a dependent process is uncorrelated, *Journal of the American Statistical Association* **96**(455), 1066–1076.

Müller, U. K.: (2007), A theory of robust long-run variance estimation, *Journal of Econometrics* **141**(2), 1331–1352.

Newey, W. K. and West, K. D.: (1987), A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* **55**, 703–708.

Phillips, P.: (2005), HAC estimation by automated regression, *Econometric Theory* **21**(1), 116–142.

Priestley, M. B.: (1981), *Spectral Analysis and Time Series*, Vol. 1, Academic Press, New York.

Romano, J. and Thombs, L.: (1996), Inference for autocorrelations under weak assumptions, *Journal of the American Statistical Association* **91**(434), 590–600.

Sun, Y.: (2013), A heteroskedasticity and autocorrelation robust F test using an orthonormal series variance estimator, *The Econometrics Journal* **16**(1), 1–26.

Sun, Y., Phillips, P. C. B. and Jin, S.: (2008), Optimal bandwidth selection in heteroskedasticity-autocorrelation robust testing, *Econometrica* **76**, 175–194.

Taylor, S.: (1984), Estimating the variances of autocorrelations calculated from financial time series, *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **33**(3), 300–308.

Vogelsang, T. J.: (2018), Comment on "HAR inference: Recommendations for practice", *Journal of Business & Economic Statistics* **36**(4), 569–573.

Vogelsang, T. and Nawaz, N.: (2017), Estimation and inference of linear trend slope ratios with an application to global temperature data, *Journal of Time Series Analysis* **38**(5), 640–667.

Wang, X. and Sun, Y.: (2020), An asymptotic F test for uncorrelatedness in the presence of time series dependence, *Journal of Time Series Analysis* **41**(4), 536–550.

Zhang, X. and Shao, X.: (2013), Fixed-smoothing asymptotics for time series, *The Annals of Statistics* **41**(3), 1329–1349.

Table 1: $cv_{\alpha/2}(b)$ Polynomial Coefficients, Parzen Kernel

λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8	λ_9
.4375	.1191	.0863	.4962	-.5787	.4326	.0254	-.0237	-.0237

Table 2: Shape of Confidence Intervals using $\widetilde{t}^{(k)*}$, AR-IID

	$\phi = 0.0$											
	T=50			T=100			T=250			T=1000		
Case	$\widehat{Prob}$	ECP	$\overline{CI}$	$\widehat{Prob}$	ECP	$\overline{CI}$	$\widehat{Prob}$	ECP	$\overline{CI}$	$\widehat{Prob}$	ECP	$\overline{CI}$
$[r_1, r_2]$	1.000	0.955	0.618	1.000	0.951	0.408	1.000	0.953	0.251	1.000	0.948	0.125
$(-1, r_1] \cup [r_2, 1)$	0.000	-	-	0.000	-	-	0.000	-	-	0.000	-	-
$(-1, 1)$	0.000	-	-	0.000	-	-	0.000	-	-	0.000	-	-
Total	1.000	0.955	0.618	1.000	0.951	0.408	1.000	0.953	0.251	1.000	0.948	0.125
	$\phi = 0.25$											
	T=50			T=100			T=250			T=1000		
Case	$\widehat{Prob}$	ECP	$\overline{CI}$	$\widehat{Prob}$	ECP	$\overline{CI}$	$\widehat{Prob}$	ECP	$\overline{CI}$	$\widehat{Prob}$	ECP	$\overline{CI}$
$[r_1, r_2]$	1.000	0.942	0.607	1.000	0.941	0.399	1.000	0.948	0.245	1.000	0.952	0.121
$(-1, r_1] \cup [r_2, 1)$	0.000	-	-	0.000	-	-	0.000	-	-	0.000	-	-
$(-1, 1)$	0.000	-	-	0.000	-	-	0.000	-	-	0.000	-	-
Total	1.000	0.942	0.607	1.000	0.941	0.399	1.000	0.948	0.245	1.000	0.952	0.121
	$\phi = 0.7$											
	T=50			T=100			T=250			T=1000		
Case	$\widehat{Prob}$	ECP	$\overline{CI}$	$\widehat{Prob}$	ECP	$\overline{CI}$	$\widehat{Prob}$	ECP	$\overline{CI}$	$\widehat{Prob}$	ECP	$\overline{CI}$
$[r_1, r_2]$	0.996	0.898	0.532	1.000	0.909	0.326	1.000	0.930	0.188	1.000	0.943	0.090
$(-1, r_1] \cup [r_2, 1)$	0.004	0.895	1.875	<0.001	1.000	1.885	0.000	-	-	0.000	-	-
$(-1, 1)$	<0.001	1.000	2.000	0.000	-	-	0.000	-	-	0.000	-	-
Total	1.000	0.898	0.538	1.000	0.909	0.327	1.000	0.930	0.188	1.000	0.943	0.090
	$\phi = -0.7$											
	T=50			T=100			T=250			T=1000		
Case	$\widehat{Prob}$	ECP	$\overline{CI}$	$\widehat{Prob}$	ECP	$\overline{CI}$	$\widehat{Prob}$	ECP	$\overline{CI}$	$\widehat{Prob}$	ECP	$\overline{CI}$
$[r_1, r_2]$	0.987	0.947	0.536	1.000	0.939	0.329	1.000	0.941	0.187	1.000	0.944	0.090
$(-1, r_1] \cup [r_2, 1)$	0.012	1.000	1.935	<0.001	1.000	1.945	0.000	-	-	0.000	-	-
$(-1, 1)$	0.002	1.000	2.000	0.000	-	-	0.000	-	-	0.000	-	-
Total	1.000	0.948	0.555	1.000	0.939	0.330	1.000	0.941	0.187	1.000	0.944	0.090

Table 3: Shape of Confidence Intervals using $\widetilde{t}^{(k)*}$, AR-WN-NLMA

$\phi = 0.0$	T=50			T=100			T=250			T=1000		
Case	$\widehat{Prob}$	ECP	$\overline{CI}$	$\widehat{Prob}$	ECP	$\overline{CI}$	$\widehat{Prob}$	ECP	$\overline{CI}$	$\widehat{Prob}$	ECP	$\overline{CI}$
$[r_1, r_2]$	0.488	0.968	1.125	0.747	0.966	0.883	0.944	0.955	0.587	1.000	0.958	0.304
$(-1, r_1] \cup [r_2, 1)$	0.134	0.985	1.592	0.109	0.987	1.549	0.036	1.000	1.487	<0.001	1.000	1.572
$(-1, 1)$	0.378	1.000	2.000	0.144	1.000	2.000	0.019	1.000	2.000	<0.001	1.000	2.000
Total	1.000	0.983	1.519	1.000	0.973	1.117	1.000	0.958	0.648	1.000	0.958	0.304
$\phi = 0.25$	T=50			T=100			T=250			T=1000		
Case	$\widehat{Prob}$	ECP	$\overline{CI}$	$\widehat{Prob}$	ECP	$\overline{CI}$	$\widehat{Prob}$	ECP	$\overline{CI}$	$\widehat{Prob}$	ECP	$\overline{CI}$
$[r_1, r_2]$	0.510	0.961	1.098	0.750	0.957	0.864	0.939	0.946	0.572	0.999	0.945	0.293
$(-1, r_1] \cup [r_2, 1)$	0.121	0.977	1.610	0.107	0.983	1.595	0.042	1.000	1.562	0.001	1.000	1.559
$(-1, 1)$	0.369	1.000	2.000	0.142	1.000	2.000	0.019	1.000	2.000	0.000	-	-
Total	1.000	0.977	1.493	1.000	0.966	1.104	1.000	0.950	0.640	1.000	0.945	0.295
$\phi = 0.7$	T=50			T=100			T=250			T=1000		
Case	$\widehat{Prob}$	ECP	$\overline{CI}$	$\widehat{Prob}$	ECP	$\overline{CI}$	$\widehat{Prob}$	ECP	$\overline{CI}$	$\widehat{Prob}$	ECP	$\overline{CI}$
$[r_1, r_2]$	0.662	0.957	0.791	0.812	0.942	0.608	0.930	0.925	0.388	0.996	0.929	0.193
$(-1, r_1] \cup [r_2, 1)$	0.093	0.972	1.671	0.076	0.984	1.769	0.049	0.992	1.854	0.004	1.000	1.859
$(-1, 1)$	0.245	1.000	2.000	0.112	1.000	2.000	0.021	1.000	2.000	0.000	-	-
Total	1.000	0.969	1.169	1.000	0.952	0.852	1.000	0.930	0.494	1.000	0.929	0.199
$\phi = -0.7$	T=50			T=100			T=250			T=1000		
Case	$\widehat{Prob}$	ECP	$\overline{CI}$	$\widehat{Prob}$	ECP	$\overline{CI}$	$\widehat{Prob}$	ECP	$\overline{CI}$	$\widehat{Prob}$	ECP	$\overline{CI}$
$[r_1, r_2]$	0.611	0.976	0.769	0.796	0.975	0.610	0.948	0.968	0.393	0.998	0.966	0.184
$(-1, r_1] \cup [r_2, 1)$	0.109	0.987	1.711	0.080	0.995	1.771	0.034	0.988	1.801	0.002	1.000	1.868
$(-1, 1)$	0.280	1.000	2.000	0.125	1.000	2.000	0.018	1.000	2.000	0.001	1.000	2.000
Total	1.000	0.984	1.216	1.000	0.980	0.876	1.000	0.969	0.470	1.000	0.966	0.188

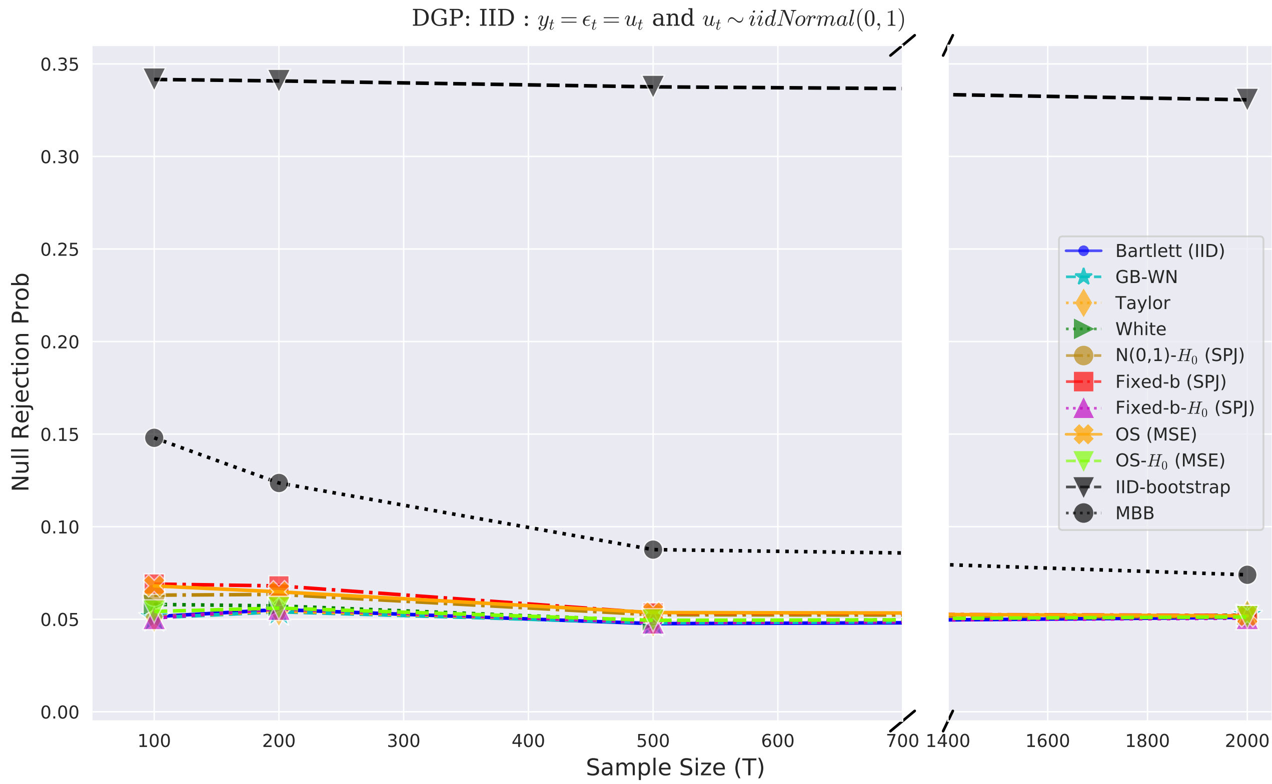

Figure 1.1: Graphs of null rejection probabilities, $H_0 : \rho_1 = 0$

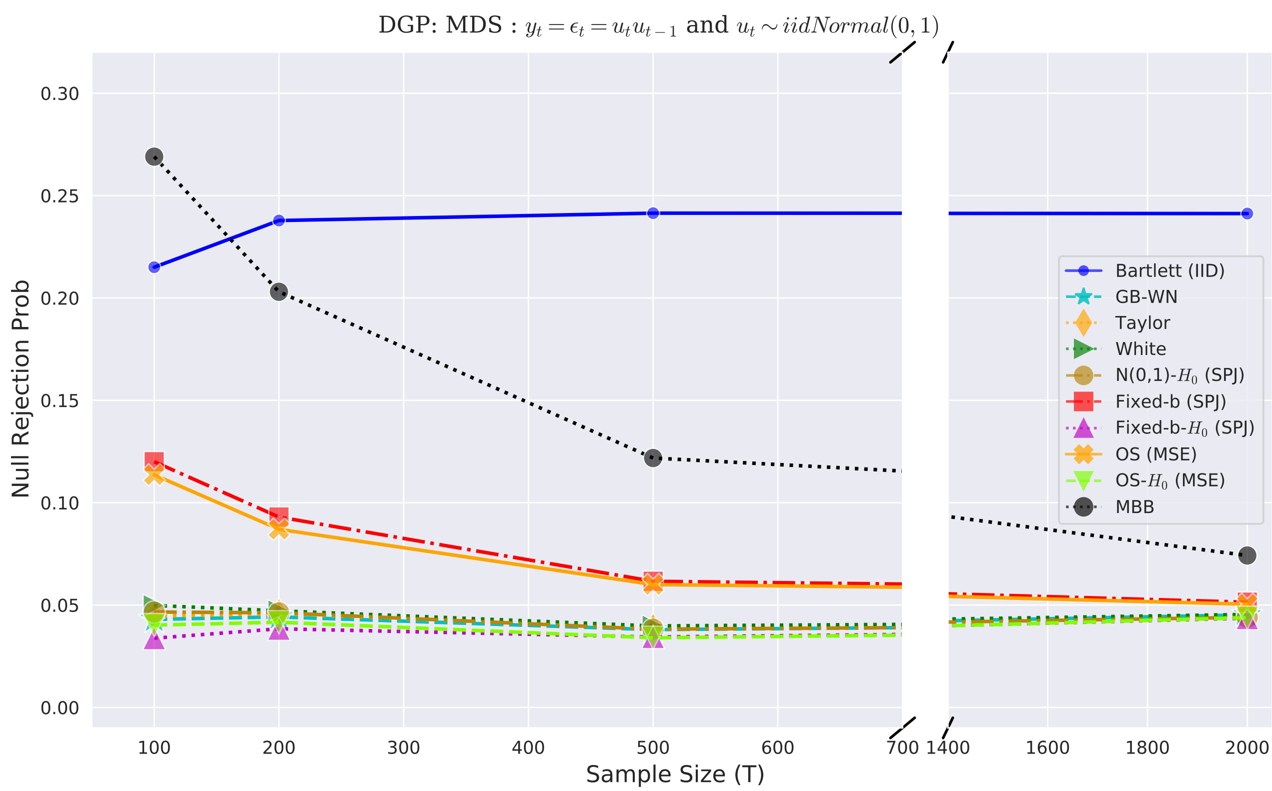

Figure 1.2: Null rejection probabilities, $H_0 : \rho_1 = 0$

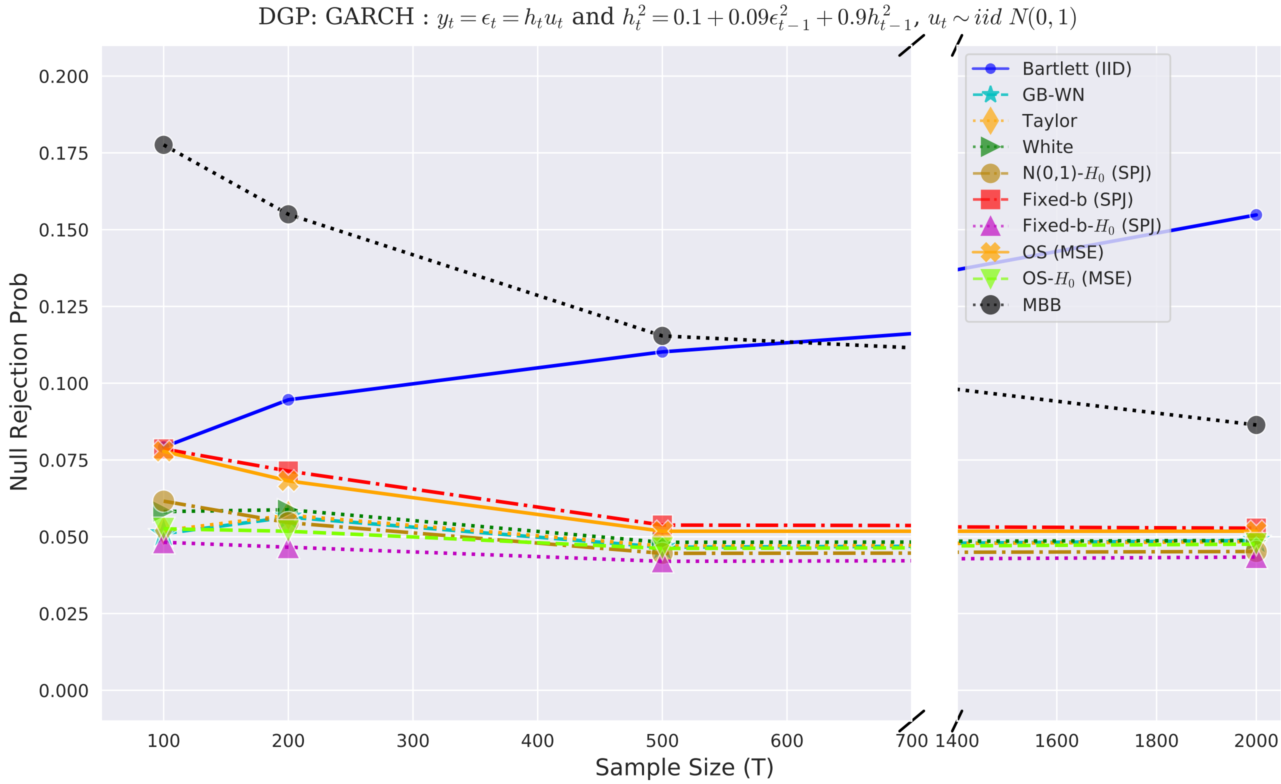

Figure 1.3: Null rejection probabilities, $H_0 : \rho_1 = 0$

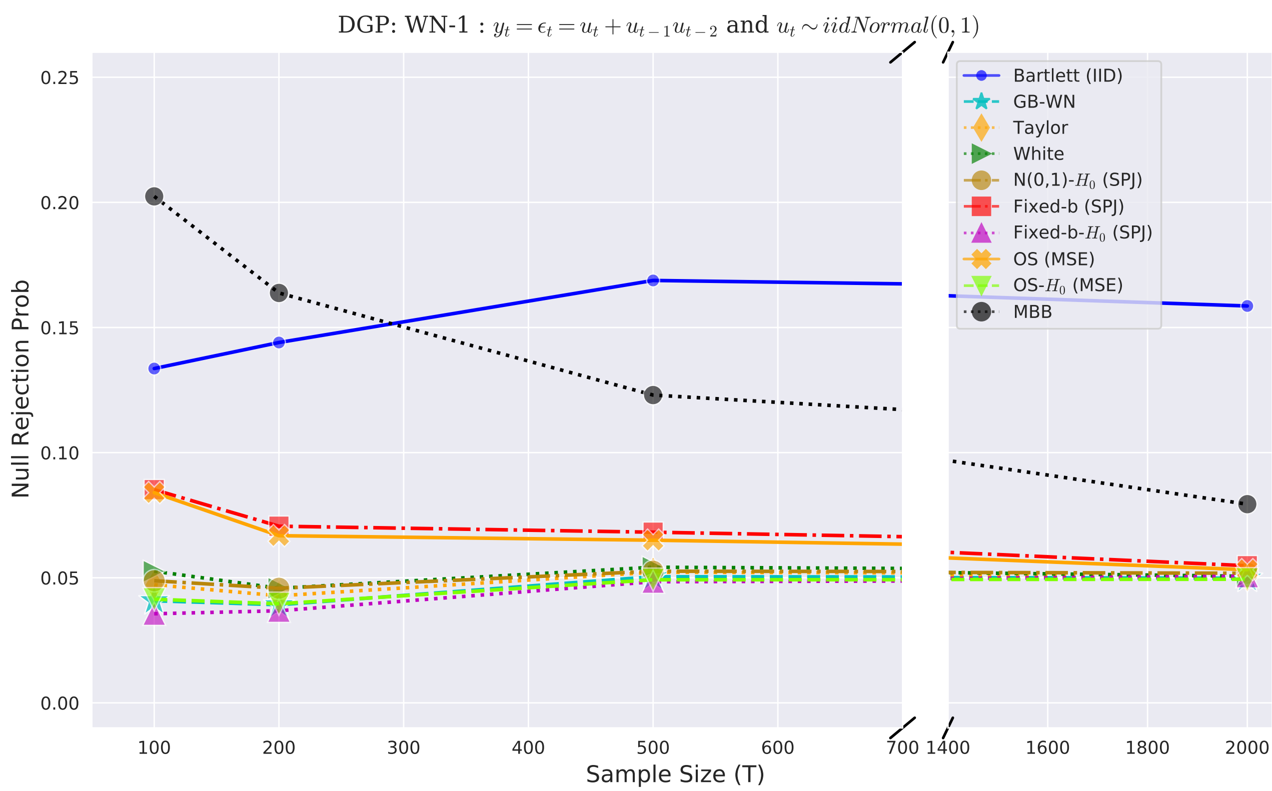

Figure 1.4: Null rejection probabilities, $H_0 : \rho_1 = 0$

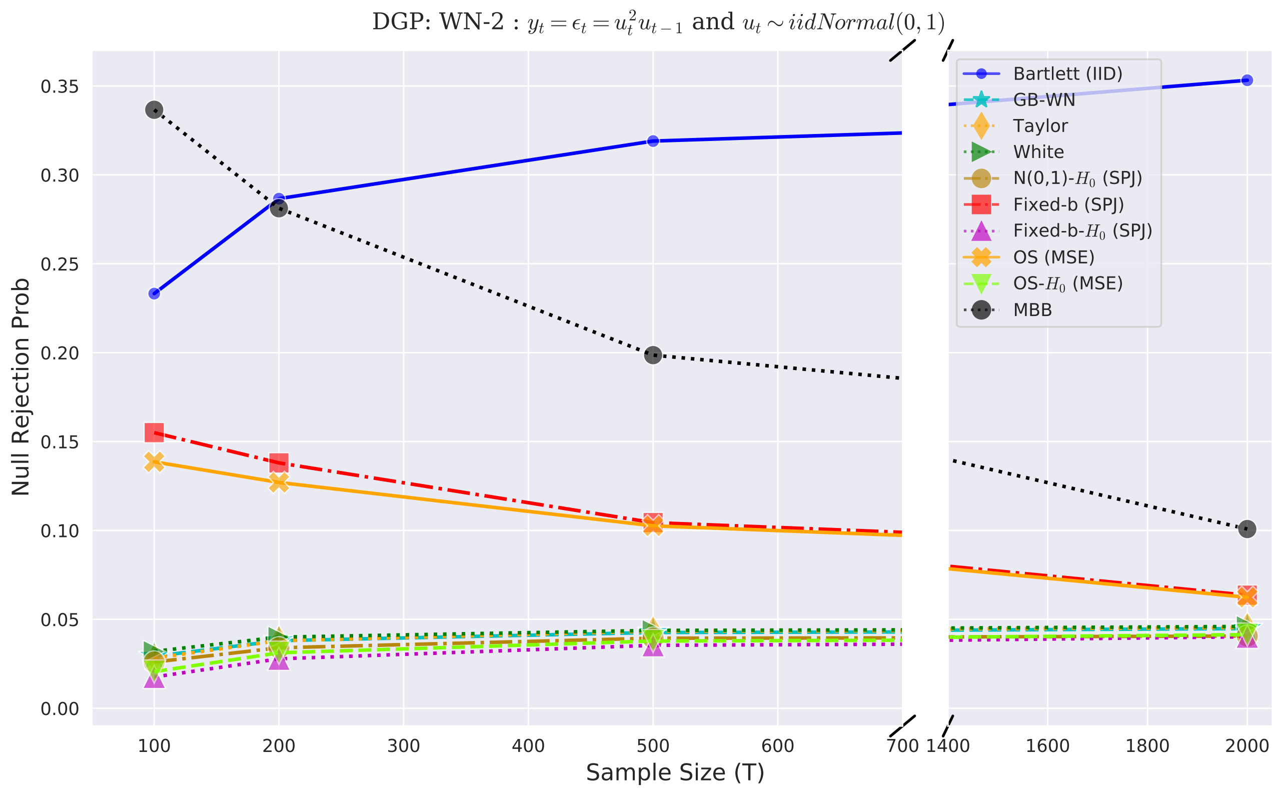

Figure 1.5: Null rejection probabilities, $H_0 : \rho_1 = 0$

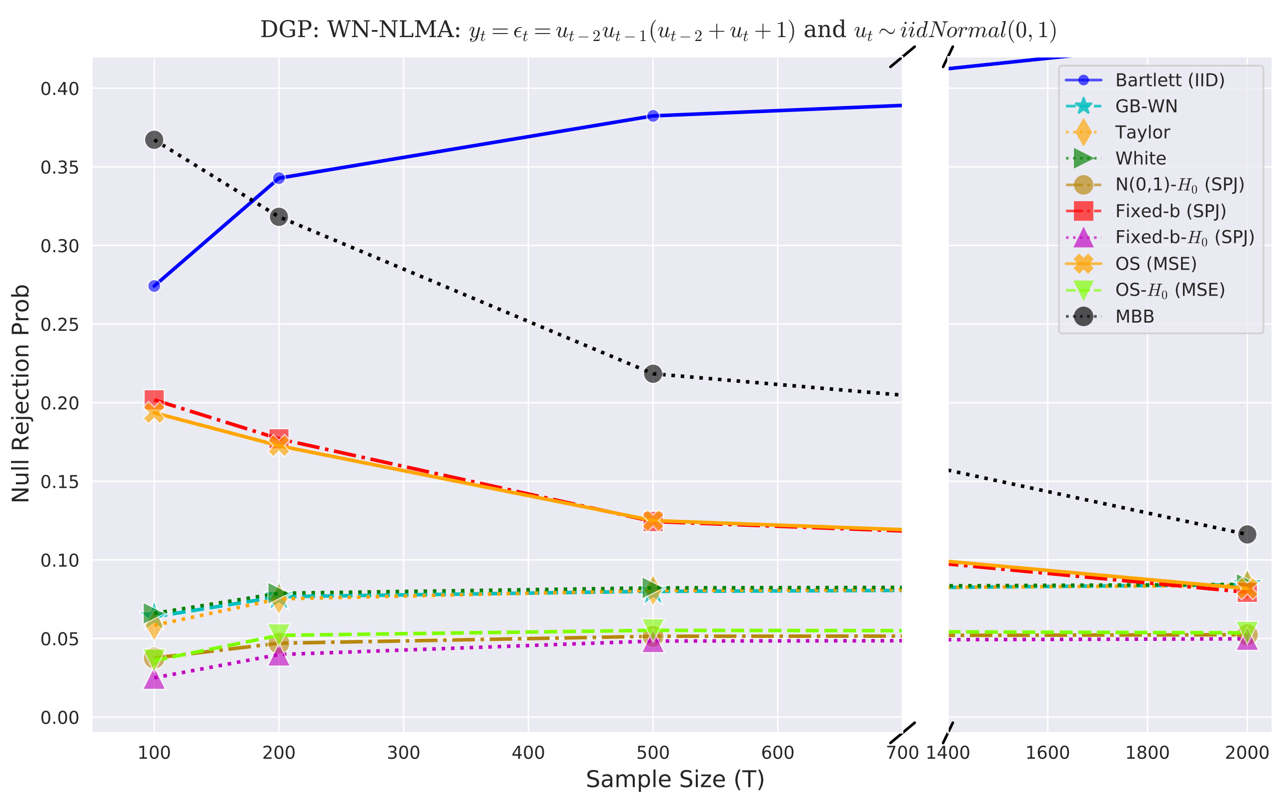

Figure 1.6: Null rejection probabilities, $H_0 : \rho_1 = 0$

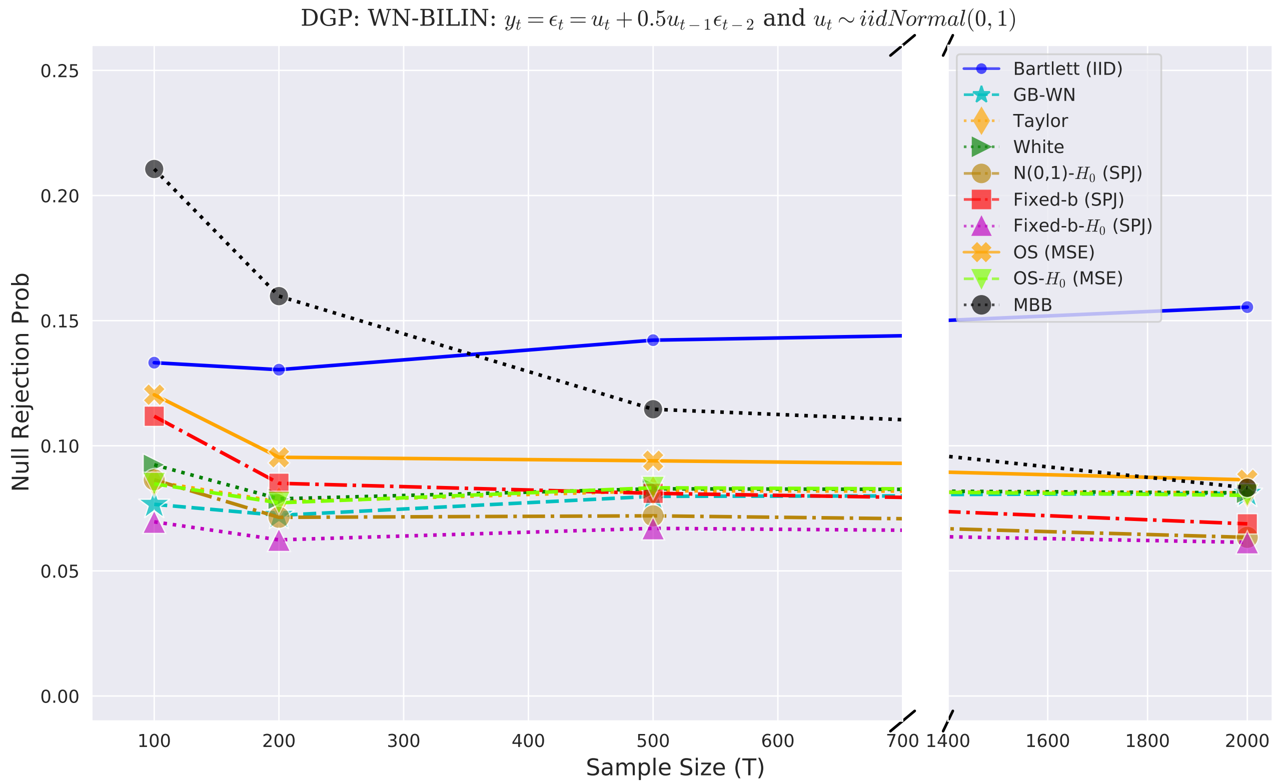

Figure 1.7: Null rejection probabilities, $H_0 : \rho_1 = 0$

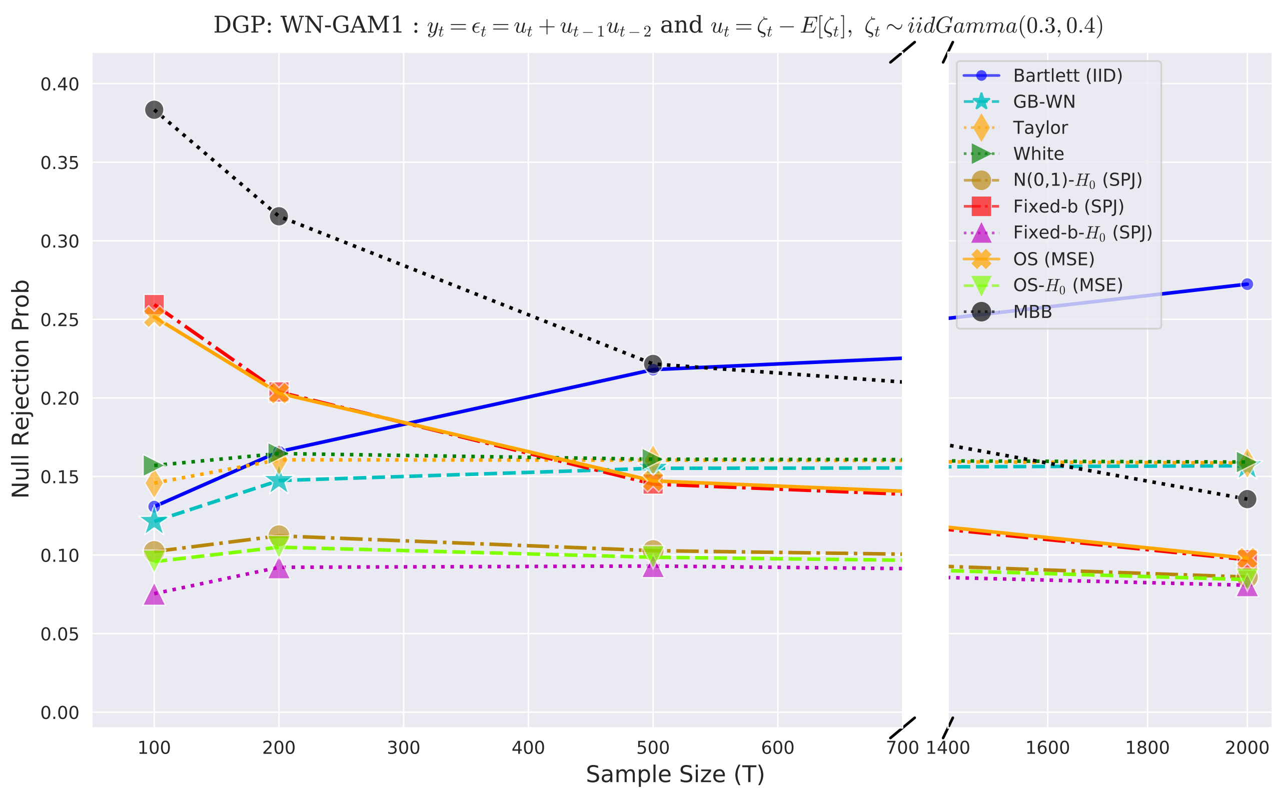

Figure 1.8: Null rejection probabilities, $H_0 : \rho_1 = 0$

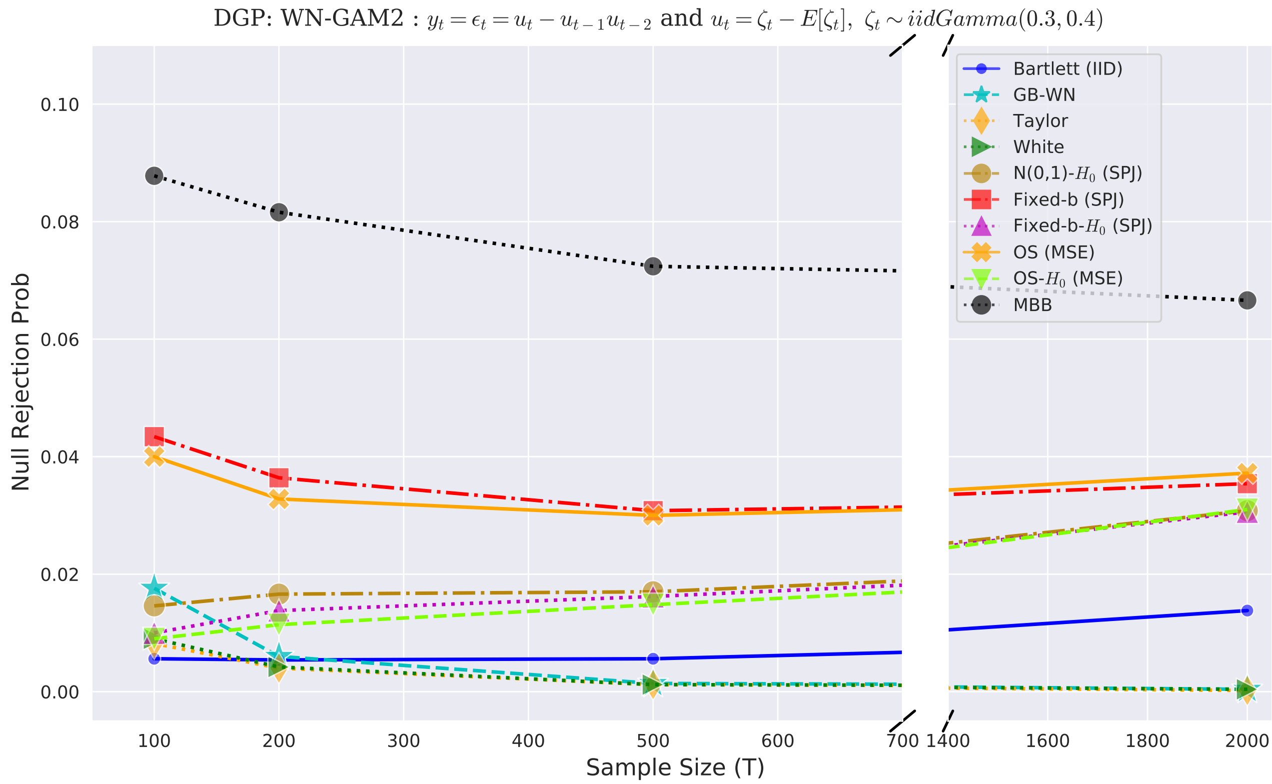

Figure 1.9: Null rejection probabilities, $H_0 : \rho_1 = 0$

DGP: MA-IID : $y_t = \epsilon_t + \theta\epsilon_{t-1}$, where $\epsilon_t \sim iidNormal(0,1)$

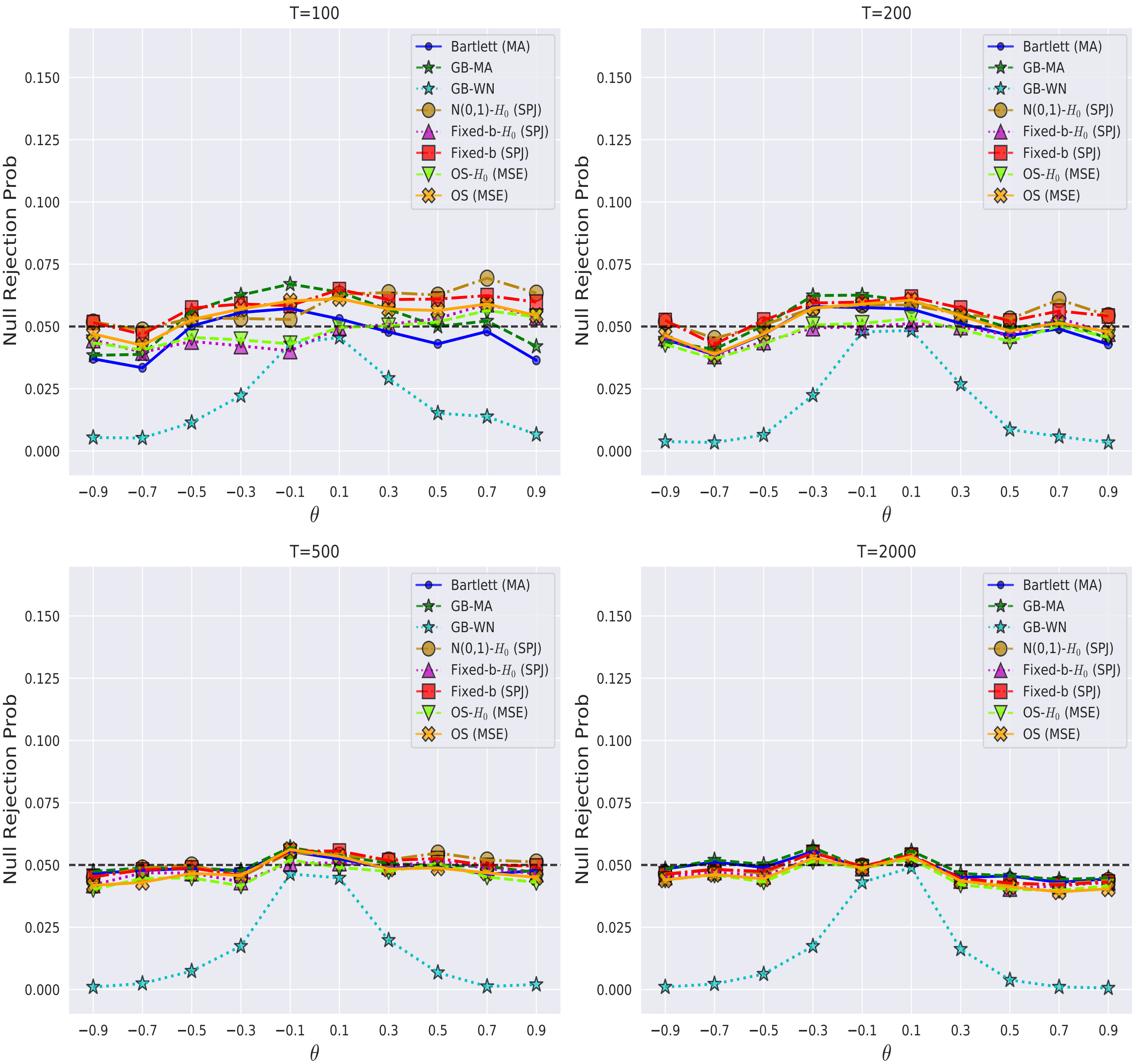

Figure 2: Null rejection probabilities, $H_0 : \rho_1 = \frac{\theta}{(1+\theta^2)}$, MA-IID

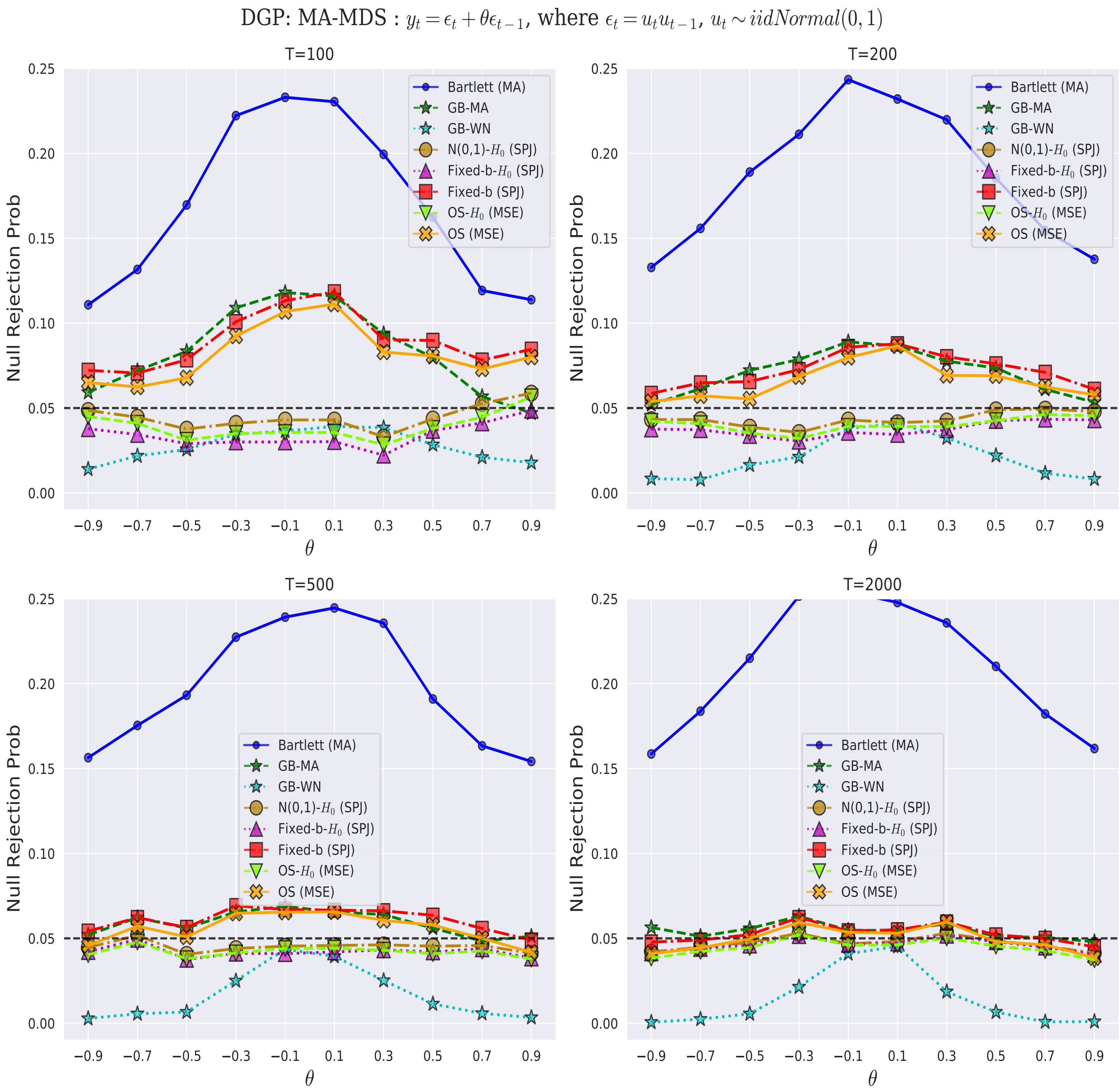

Figure 3: Null rejection probabilities, $H_0 : \rho_1 = \frac{\theta}{(1+\theta^2)}$, MA-MDS

DGP: MA-GARCH : $y_t = \epsilon_t + \theta\epsilon_{t-1}$, where $\epsilon_t = h_t u_t$ and $h_t^2 = 0.1 + 0.09\epsilon_{t-1}^2 + 0.9h_{t-1}^2$, $u_t \sim iidN(0,1)$

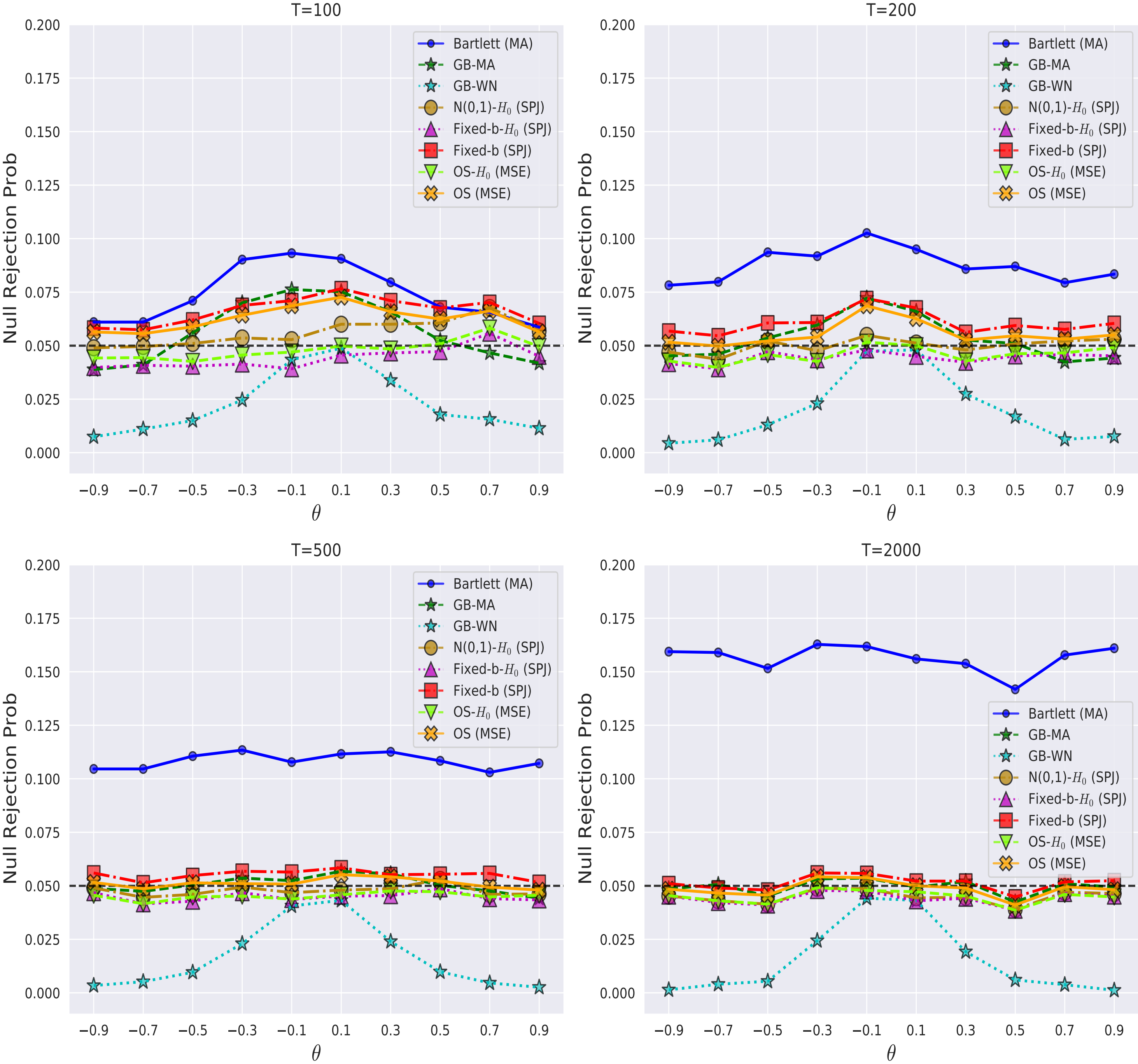

Figure 4: Null rejection probabilities, $H_0 : \rho_1 = \frac{\theta}{(1+\theta^2)}$, MA-GRACH

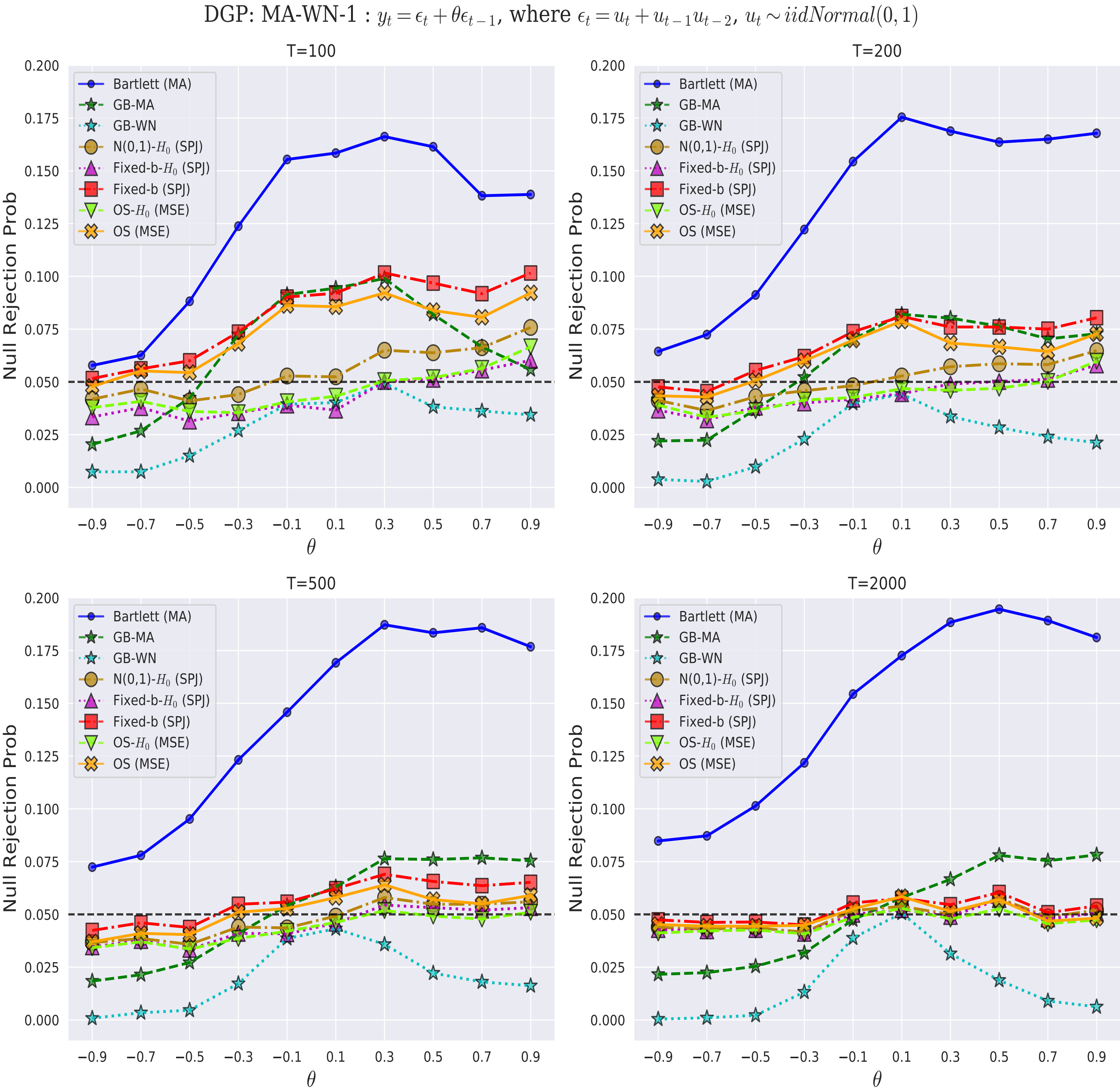

Figure 5: Null rejection probabilities, $H_0 : \rho_1 = \frac{\theta}{(1+\theta^2)}$, MA-WN-1

DGP: MA-WN-GAM1 : $y_t = \epsilon_t + \theta\epsilon_{t-1}$, where $\epsilon_t = u_t + u_{t-1}u_{t-2}$ and $u_t = \zeta_t - \mathrm{E}[\zeta_t]$, $\zeta_t \sim iidGamma(0.3, 0.4)$

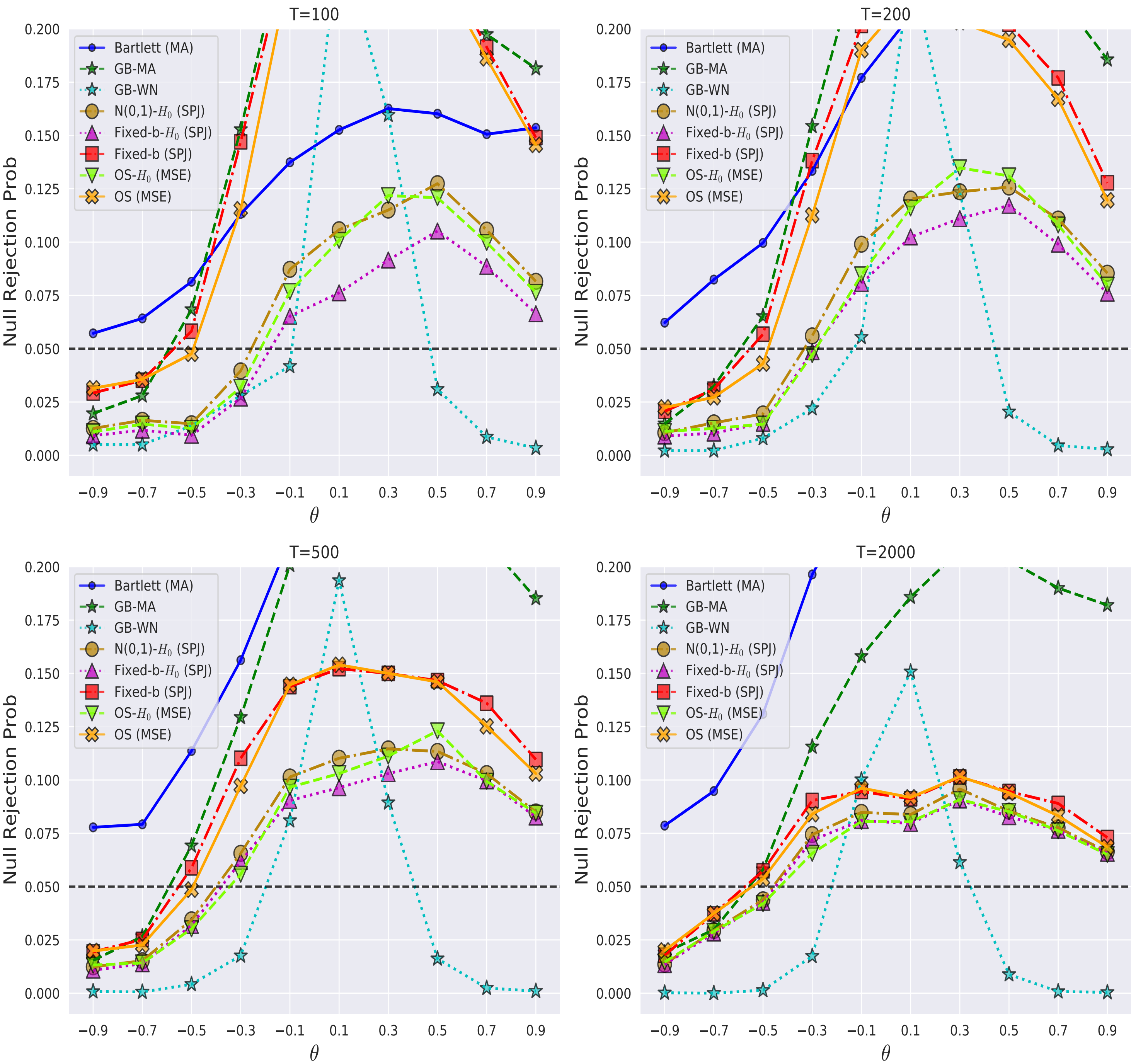

Figure 6: Null rejection probabilities, $H_0 : \rho_1 = \frac{\theta}{(1+\theta^2)}$, MA-WN-Gamma

DGP: AR-IID : $y_t = \phi y_{t-1} + \epsilon_t$, where $\epsilon_t \sim iidNormal(0,1)$

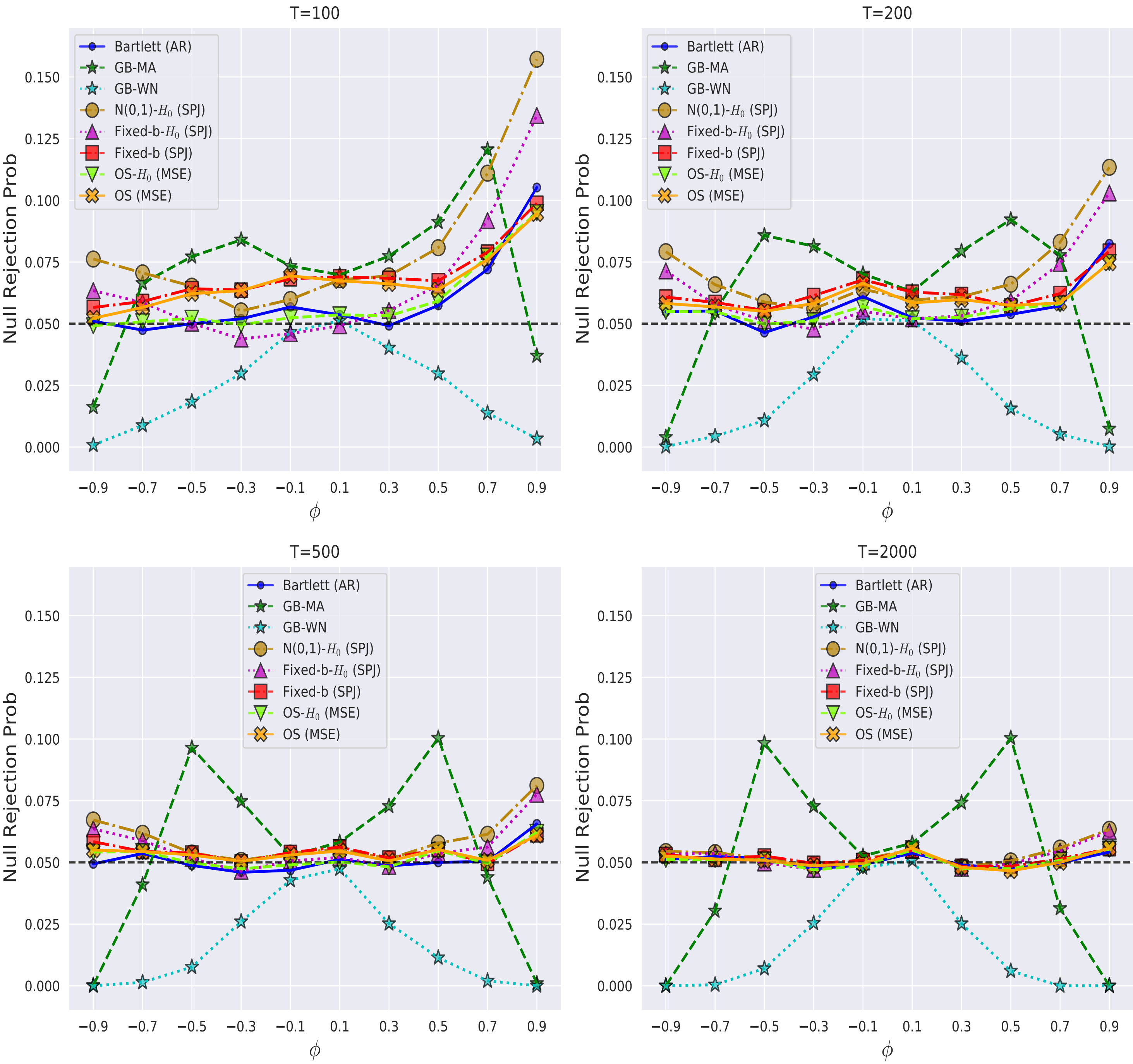

Figure 7: Null rejection probabilities, $H_0 : \rho_1 = \phi$, AR-IID

DGP: AR-MDS : $y_t = \phi y_{t-1} + \epsilon_t$, where $\epsilon_t = u_t u_{t-1}$, $u_t \sim iidNormal(0,1)$

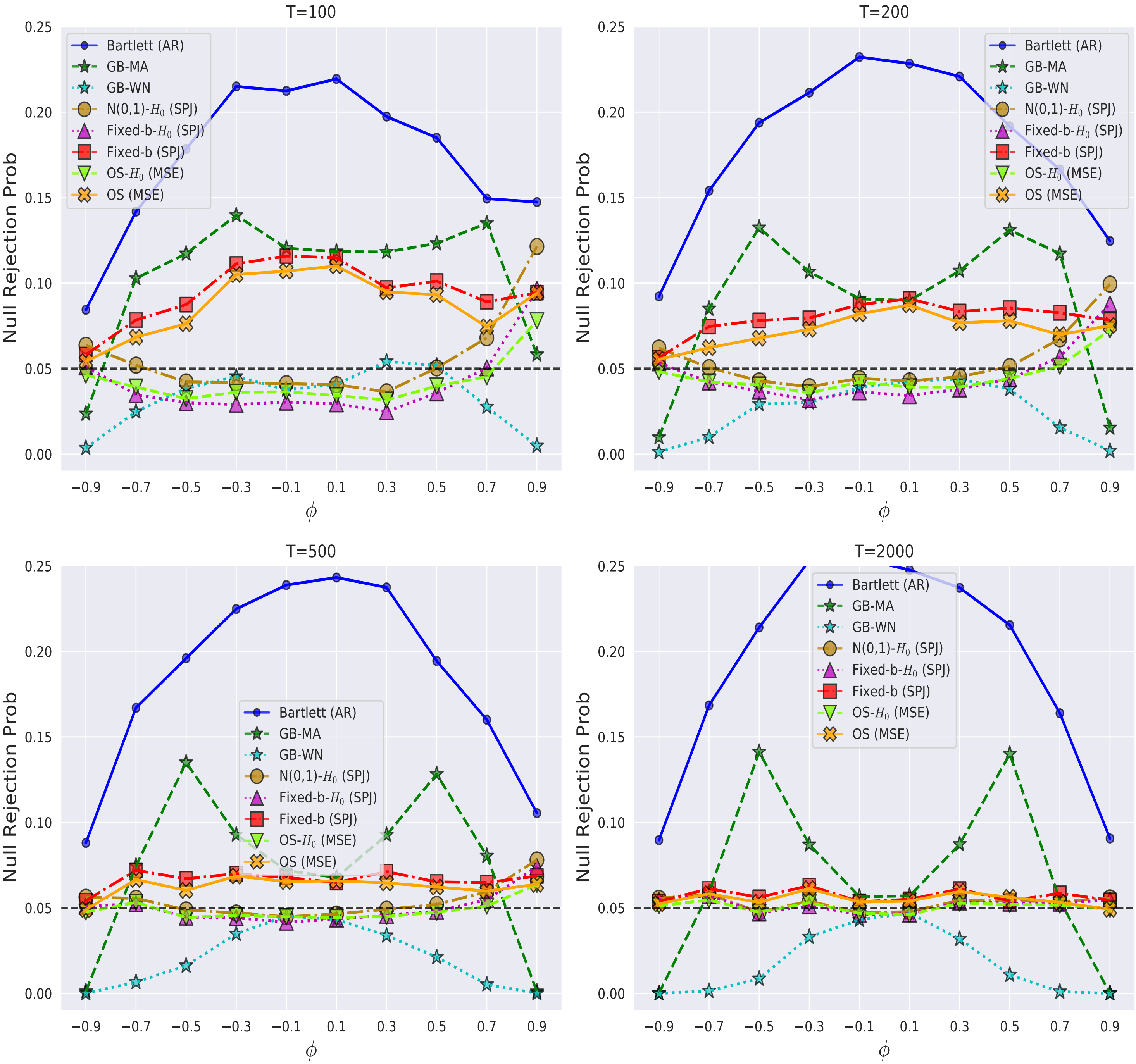

Figure 8: Null rejection probabilities, $H_0 : \rho_1 = \phi$, AR-MDS

DGP: AR-GARCH : $y_t = \phi y_{t-1} + \epsilon_t$, where $\epsilon_t = h_t u_t$ and $h_t^2 = 0.1 + 0.09\epsilon_{t-1}^2 + 0.9h_{t-1}^2$, $u_t \sim iidN(0,1)$

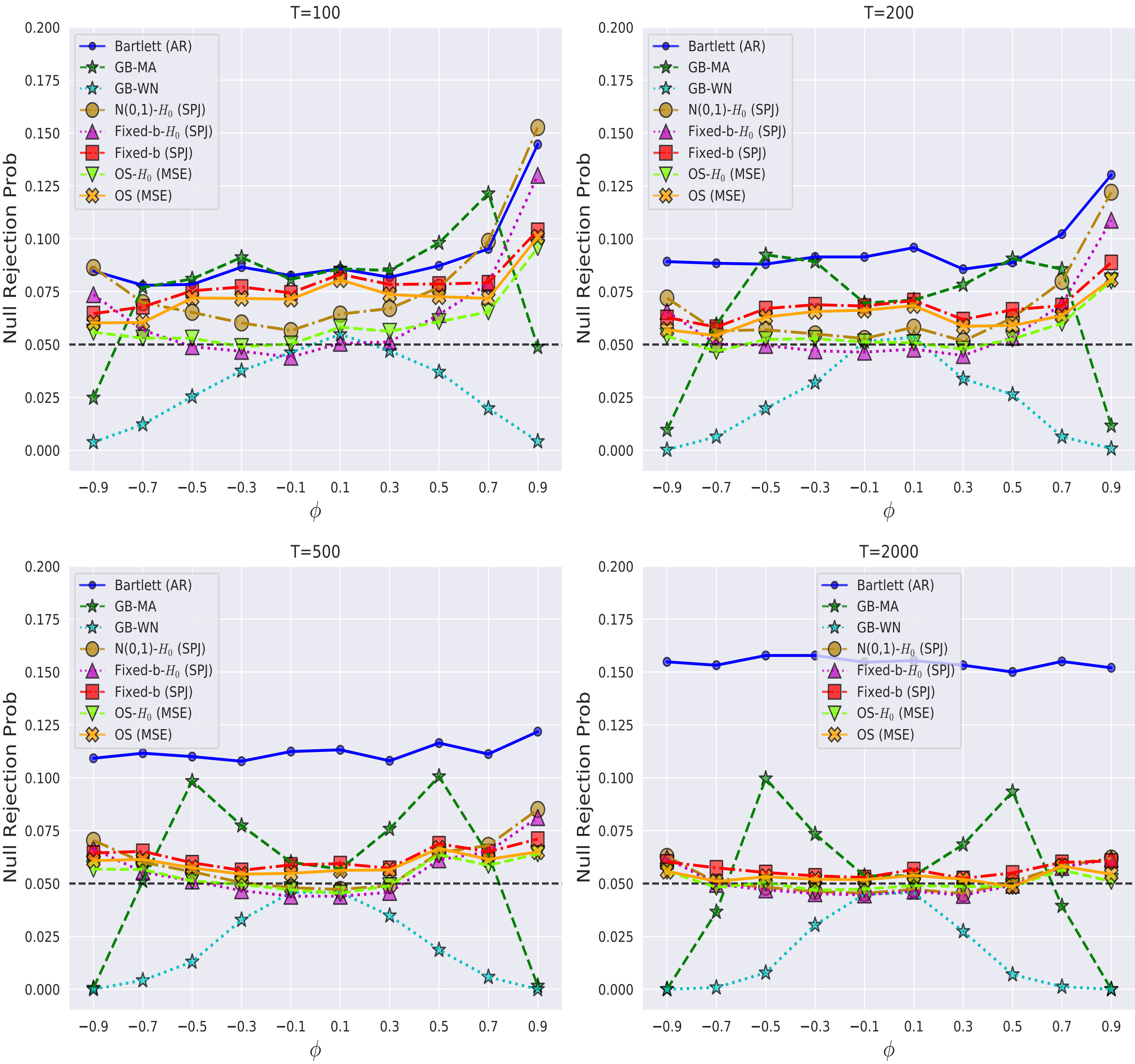

Figure 9: Null rejection probabilities, $H_0 : \rho_1 = \phi$, AR-GRACH

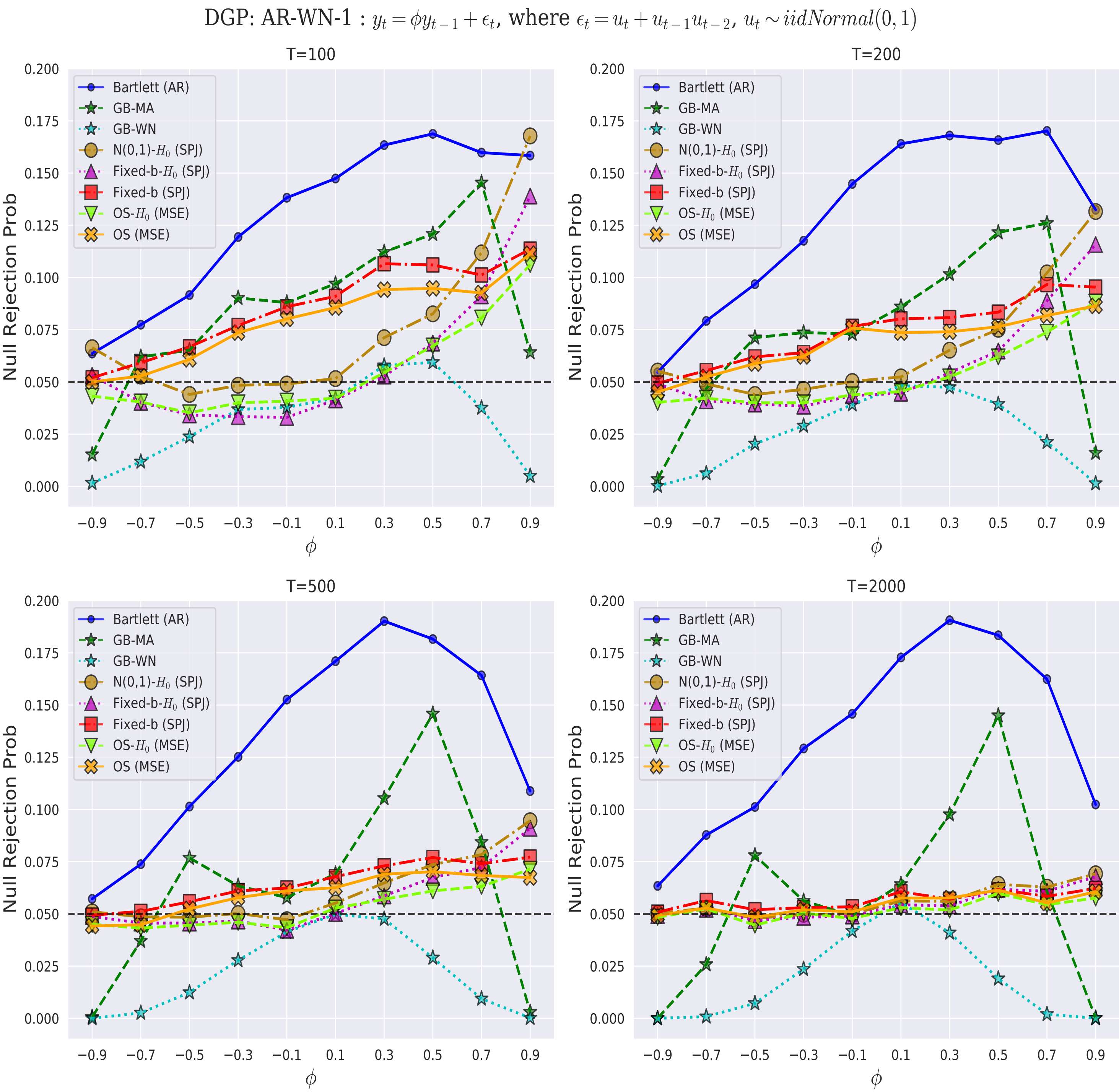

Figure 10: Null rejection probabilities, $H_0 : \rho_1 = \phi$, AR-WN-1

DGP: AR-WN-GAM1 : $y_t = \phi y_{t-1} + \epsilon_t$, where $\epsilon_t = u_t + u_{t-1}u_{t-2}$ and $u_t = \zeta_t - \mathrm{E}[\zeta_t]$, $\zeta_t \sim iidGamma(0.3, 0.4)$

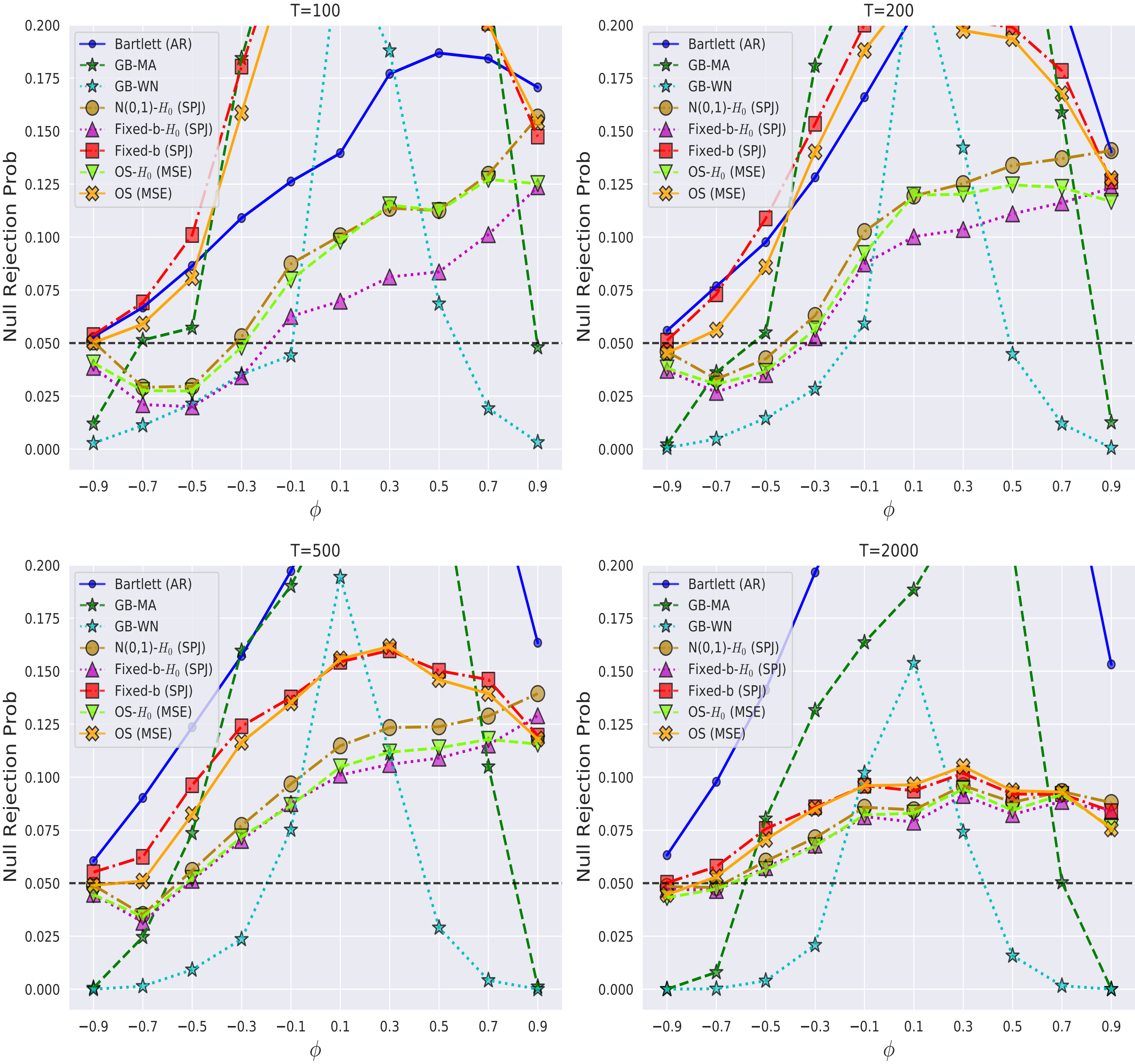

Figure 11: Null rejection probabilities, $H_0 : \rho_1 = \phi$, AR-WN-Gamma

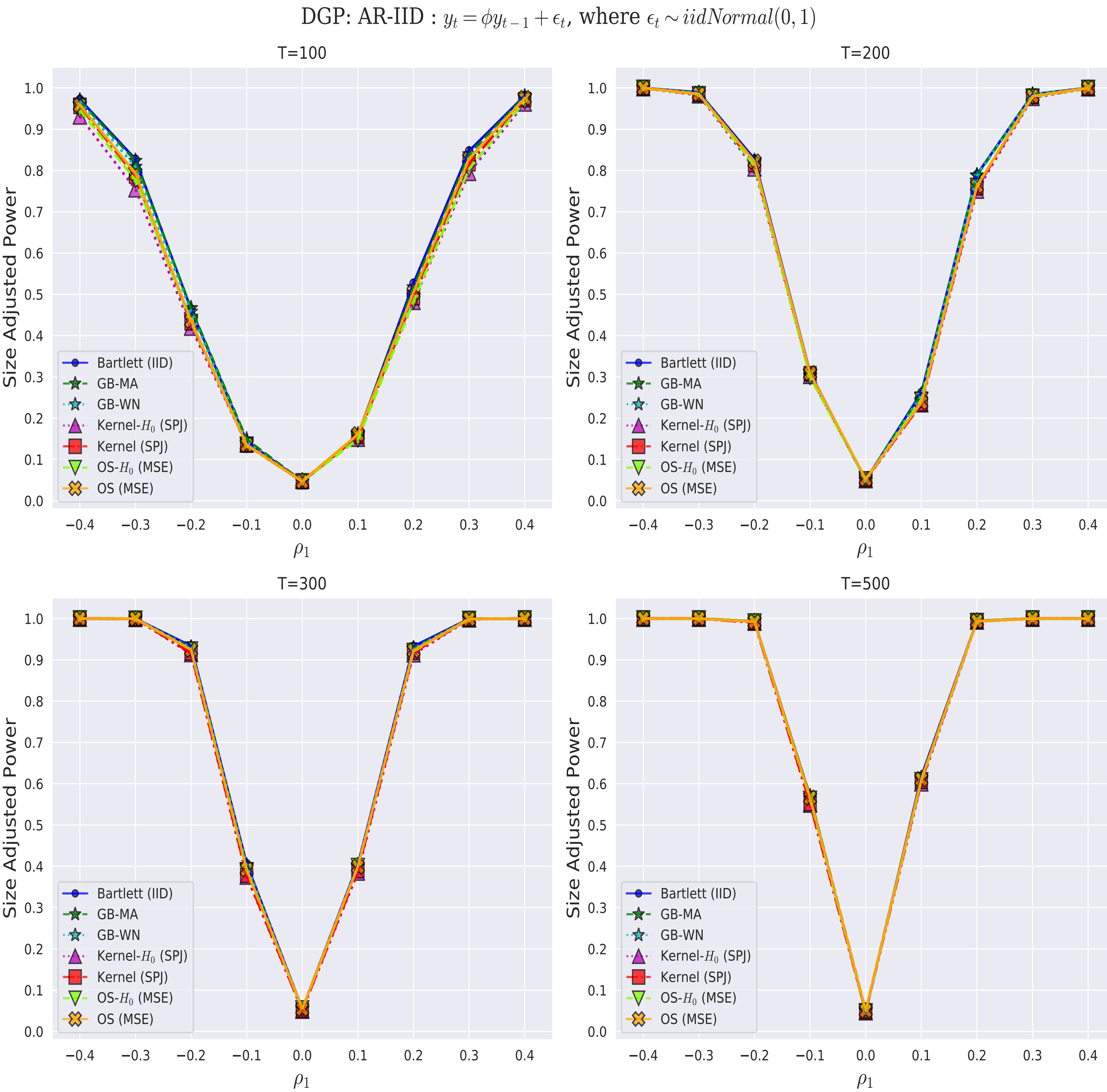

Figure 12: Size adjusted power, $H_0 : \rho_1 = 0$, $H_1 : \rho_1 = \phi$, AR-IID

DGP: AR-WN-NLMA : $y_t = \phi y_{t-1} + \epsilon_t$, where $\epsilon_t = u_{t-2}u_{t-1}(u_{t-2} + u_t + 1)$, $u_t \sim iidNormal(0, 1)$

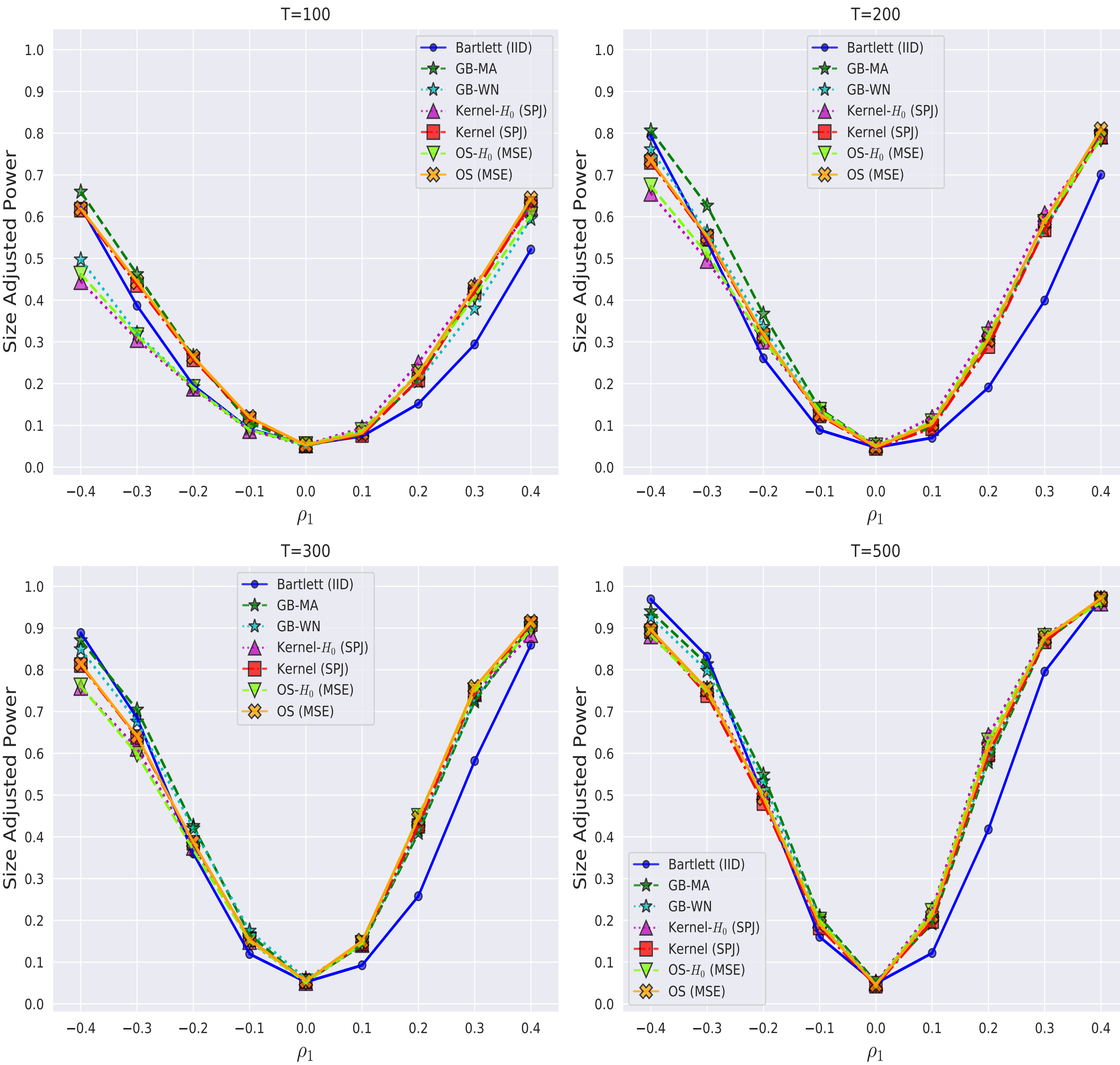

Figure 13: Size adjusted power, $H_0 : \rho_1 = 0$, $H_1 : \rho_1 = \phi$, AR-WN-NLMA

DGP: AR-WN-GAM1 : $y_t = \phi y_{t-1} + \epsilon_t$, where $\epsilon_t = u_t + u_{t-1}u_{t-2}$ and $u_t = \zeta_t - \mathrm{E}[\zeta_t]$, $\zeta_t \sim iidGamma(0.3, 0.4)$

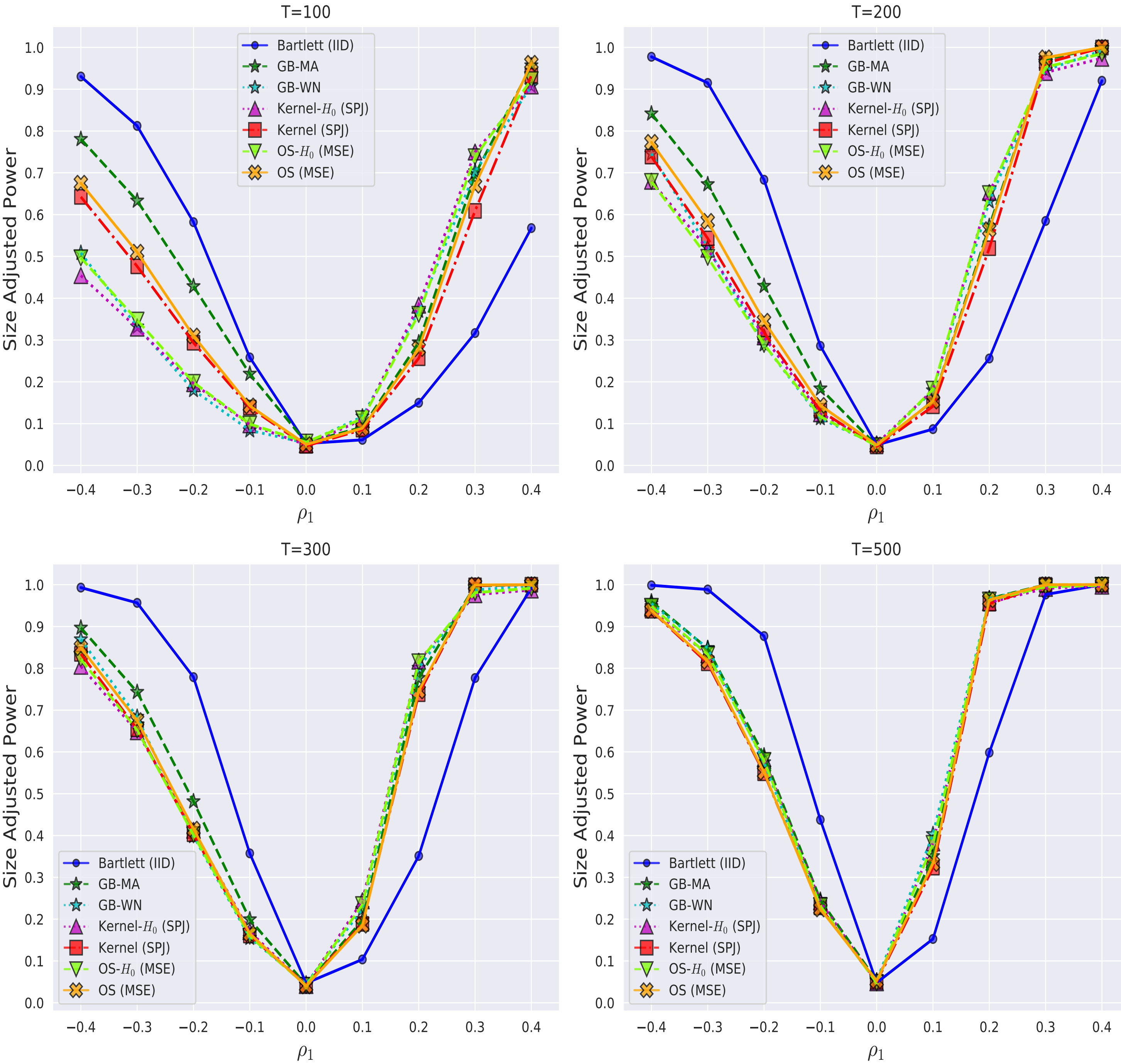

Figure 14: Size adjusted power, $H_0 : \rho_1 = 0$, $H_1 : \rho_1 = \phi$, AR-WN-Gamma

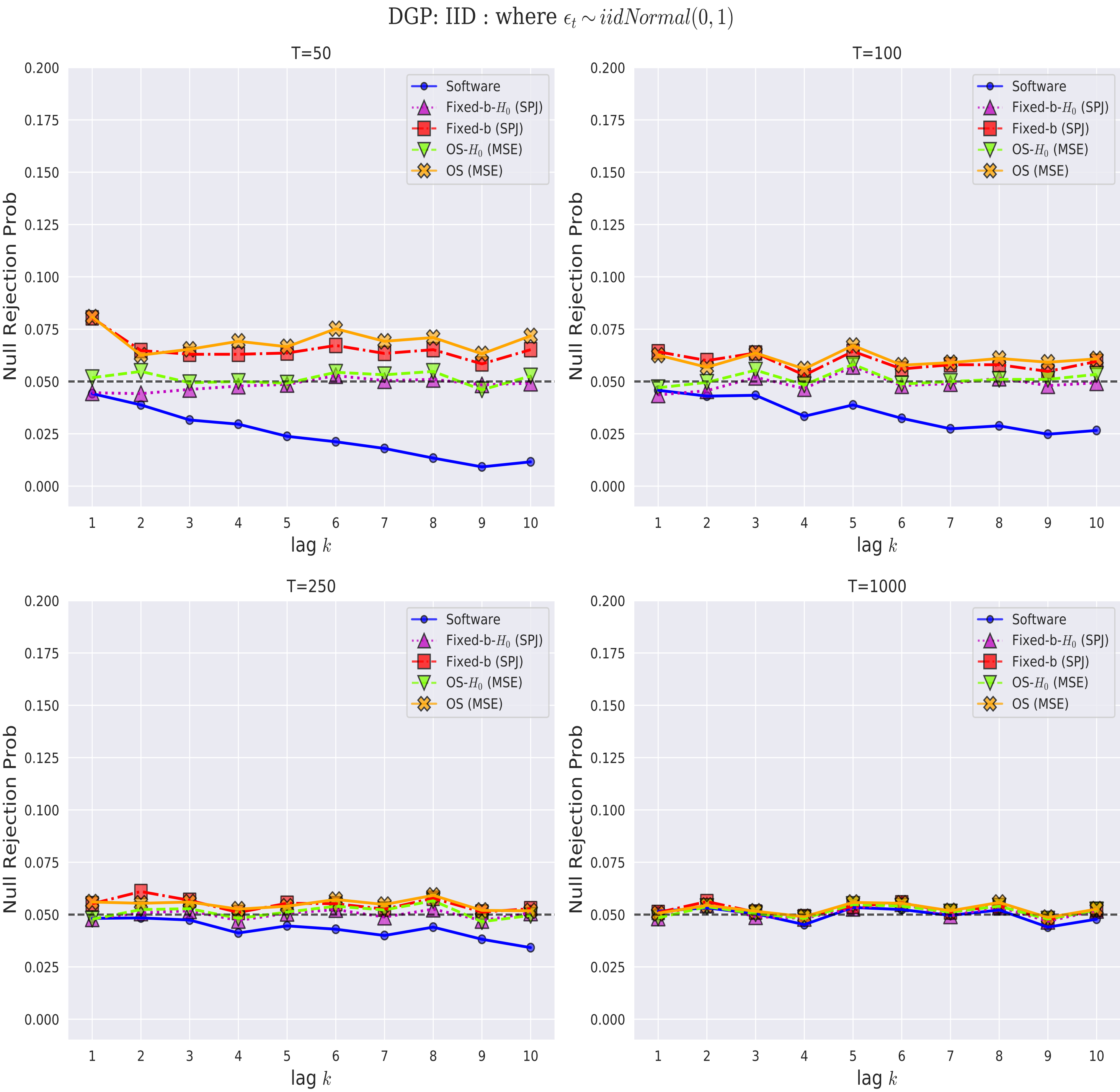

Figure 15: Null rejection probabilities, $H_0 : \rho_k = 0$, IID

DGP: MDS : where $\epsilon_t = u_t u_{t-1}$, $u_t \sim iidNormal(0,1)$

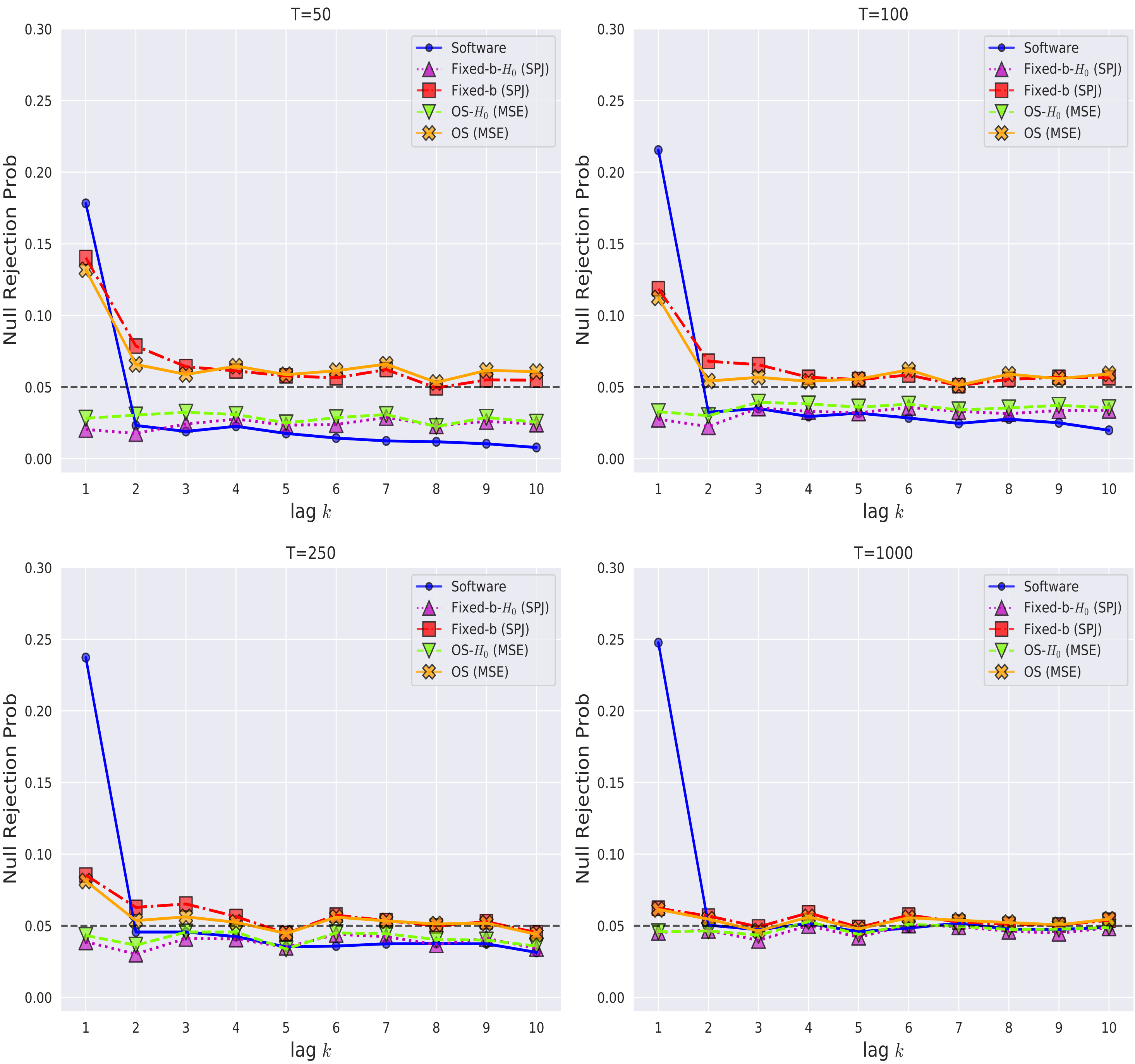

Figure 16: Null rejection probabilities, $H_0 : \rho_k = 0$, MDS

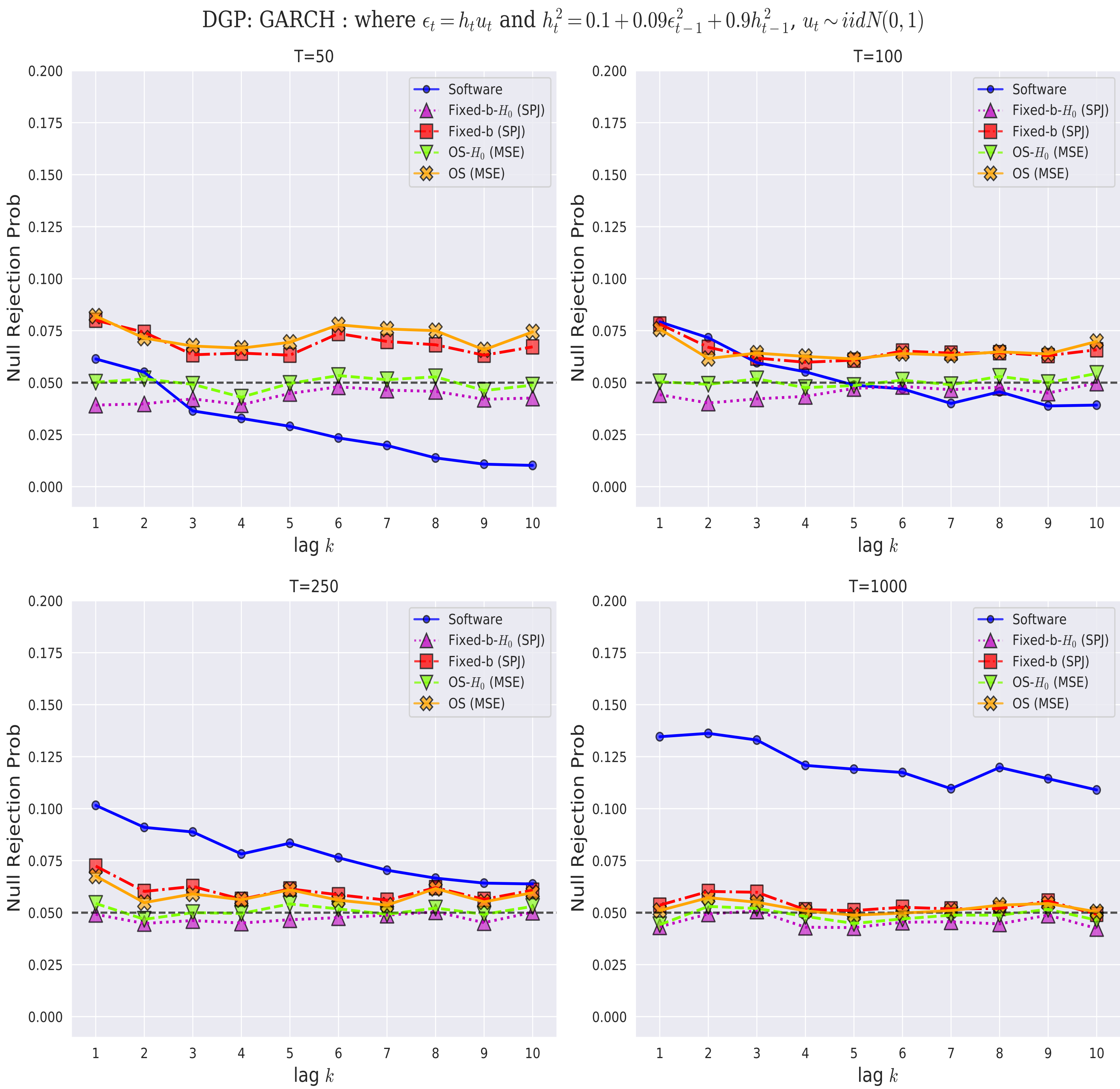

Figure 17: Null rejection probabilities, $H_0 : \rho_k = 0$, GRACH

DGP: WN-1 : where $\epsilon_t = u_t + u_{t-1}u_{t-2}$, $u_t \sim iidNormal(0,1)$

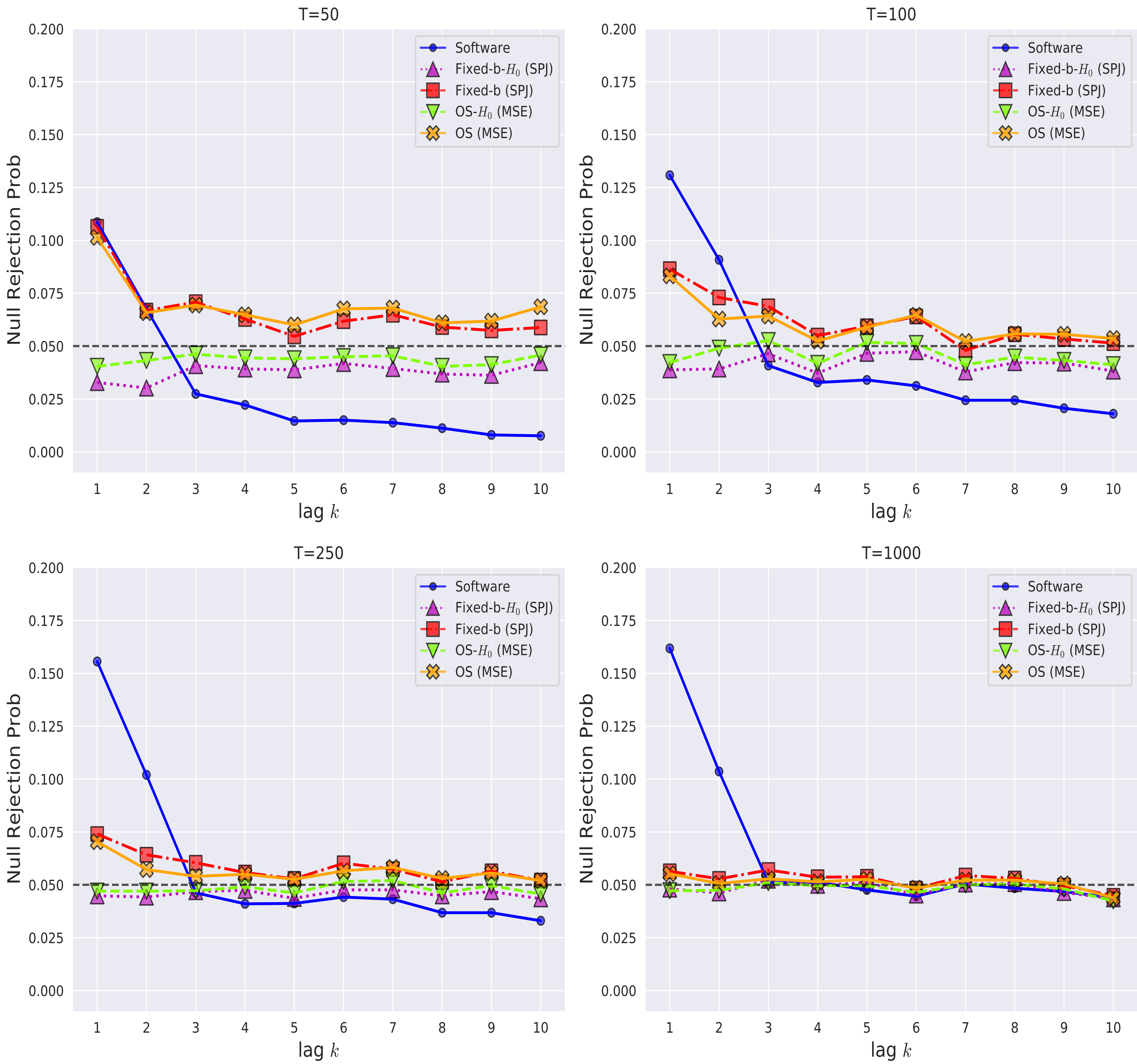

Figure 18: Null rejection probabilities, $H_0 : \rho_k = 0$, WN-1

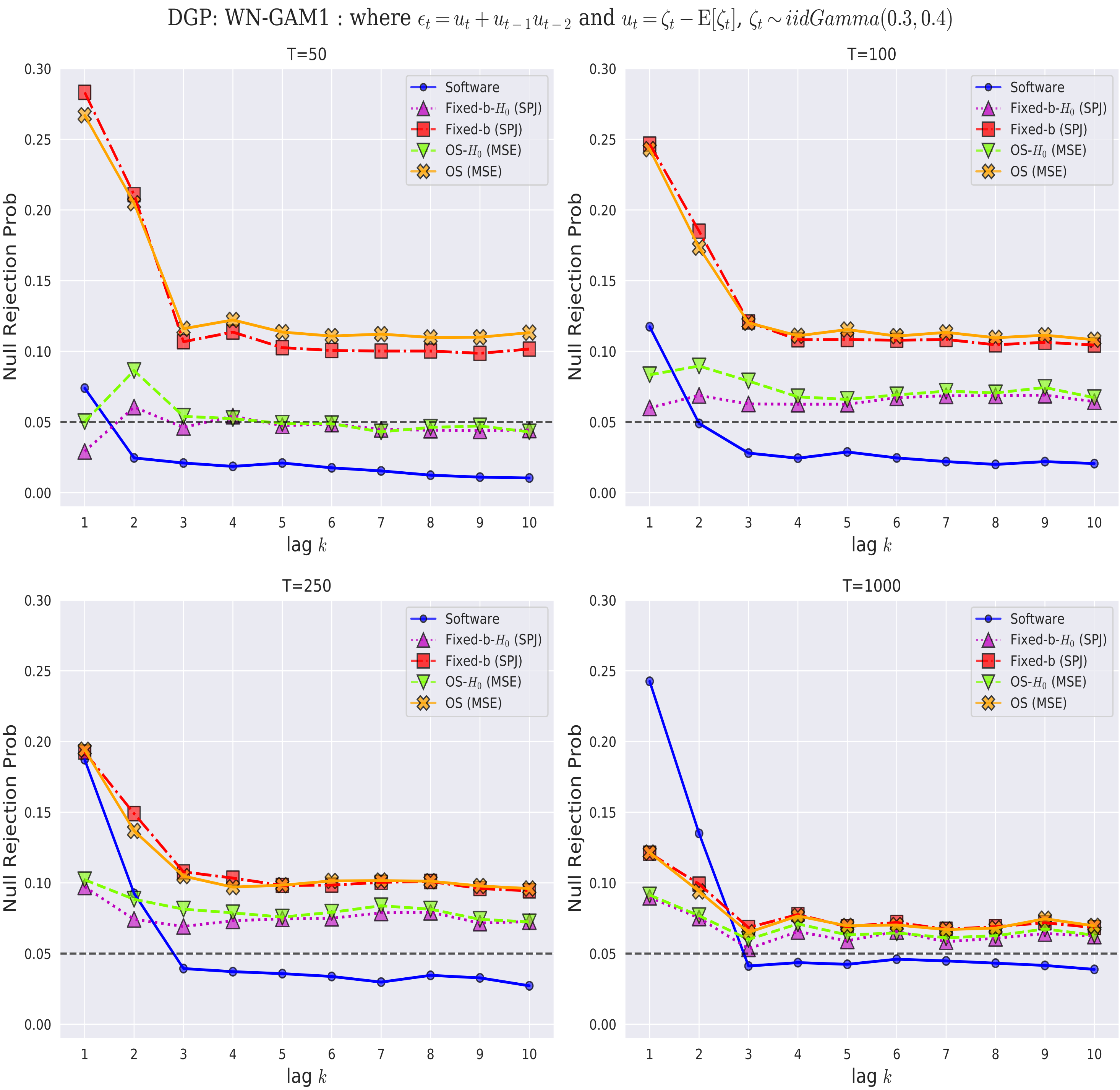

Figure 19: Null rejection probabilities, $H_0 : \rho_k = 0$, WN-Gamma

DGP: AR-IID : $y_t = \phi y_{t-1} + \epsilon_t$, where $\epsilon_t \sim iidNormal(0,1)$

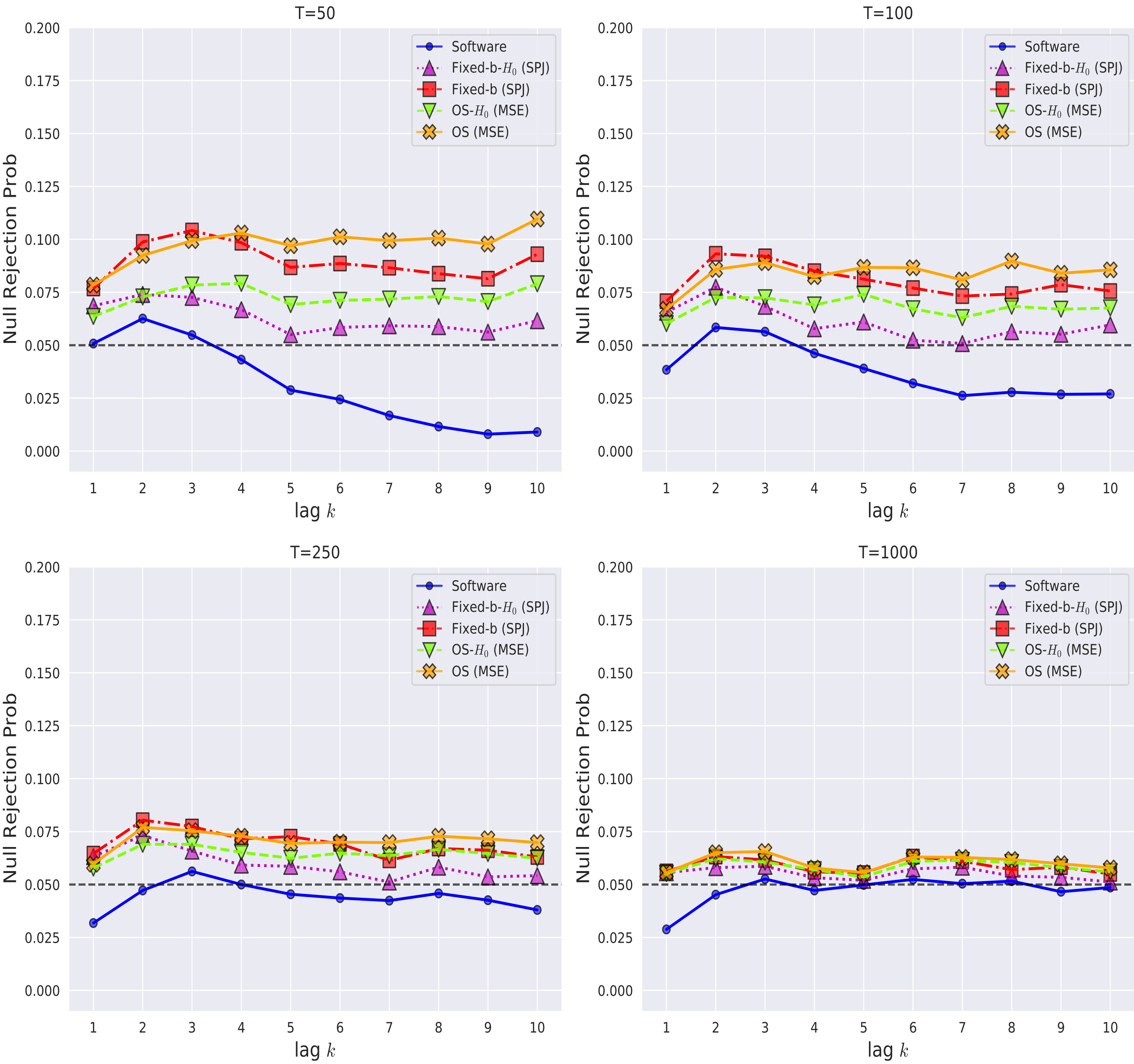

Figure 20: Null rejection probabilities, $H_0 : \rho_k = \phi^k$, $\phi = 0.5$, AR-IID

DGP: AR-MDS : $y_t = \phi y_{t-1} + \epsilon_t$, where $\epsilon_t = u_t u_{t-1}$, $u_t \sim iidNormal(0,1)$

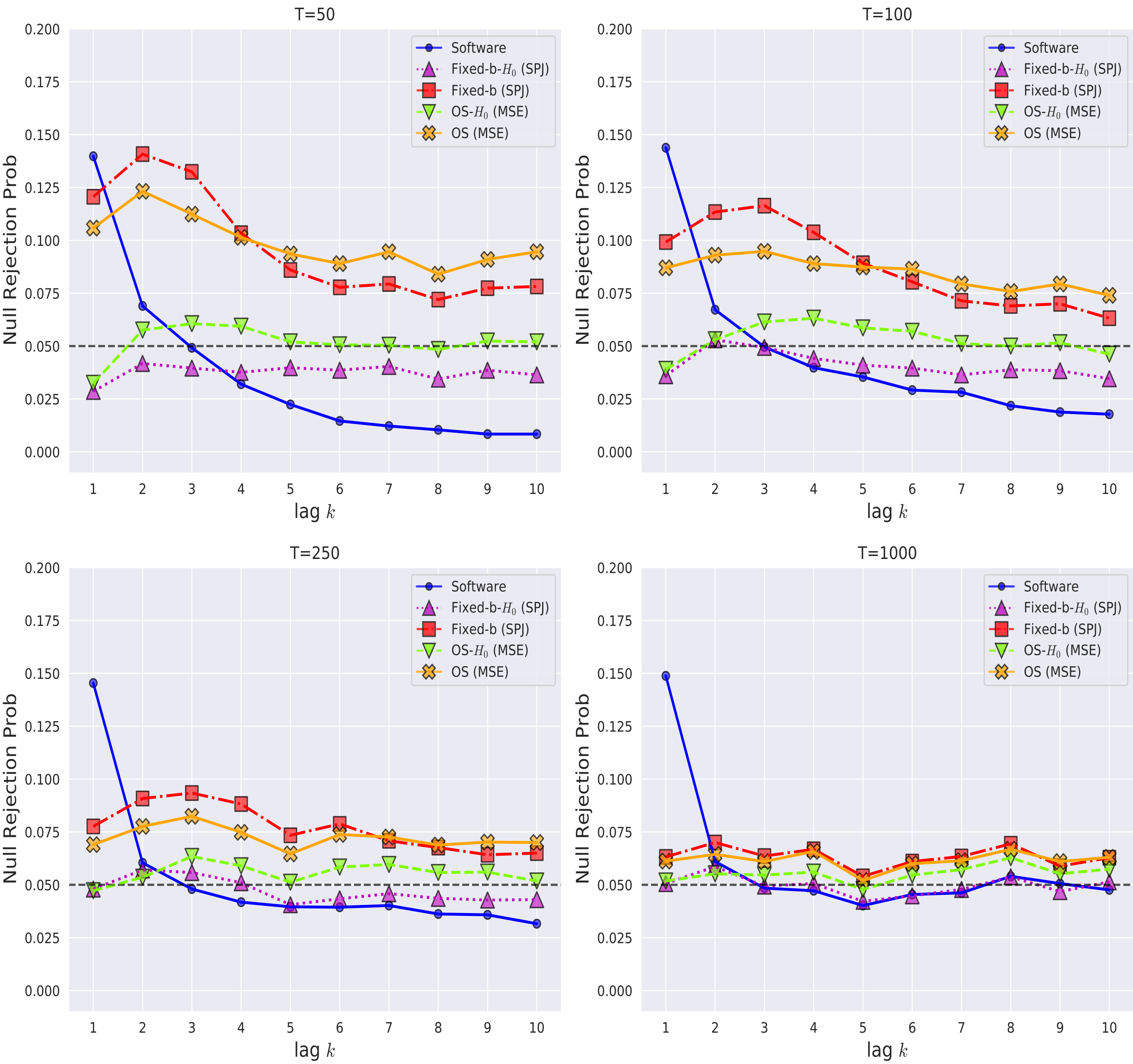

Figure 21: Null rejection probabilities, $H_0 : \rho_k = \phi^k$, $\phi = 0.5$, AR-MDS

DGP: AR-GARCH : $y_t = \phi y_{t-1} + \epsilon_t$, where $\epsilon_t = h_t u_t$ and $h_t^2 = 0.1 + 0.09\epsilon_{t-1}^2 + 0.9h_{t-1}^2$, $u_t \sim iidN(0,1)$

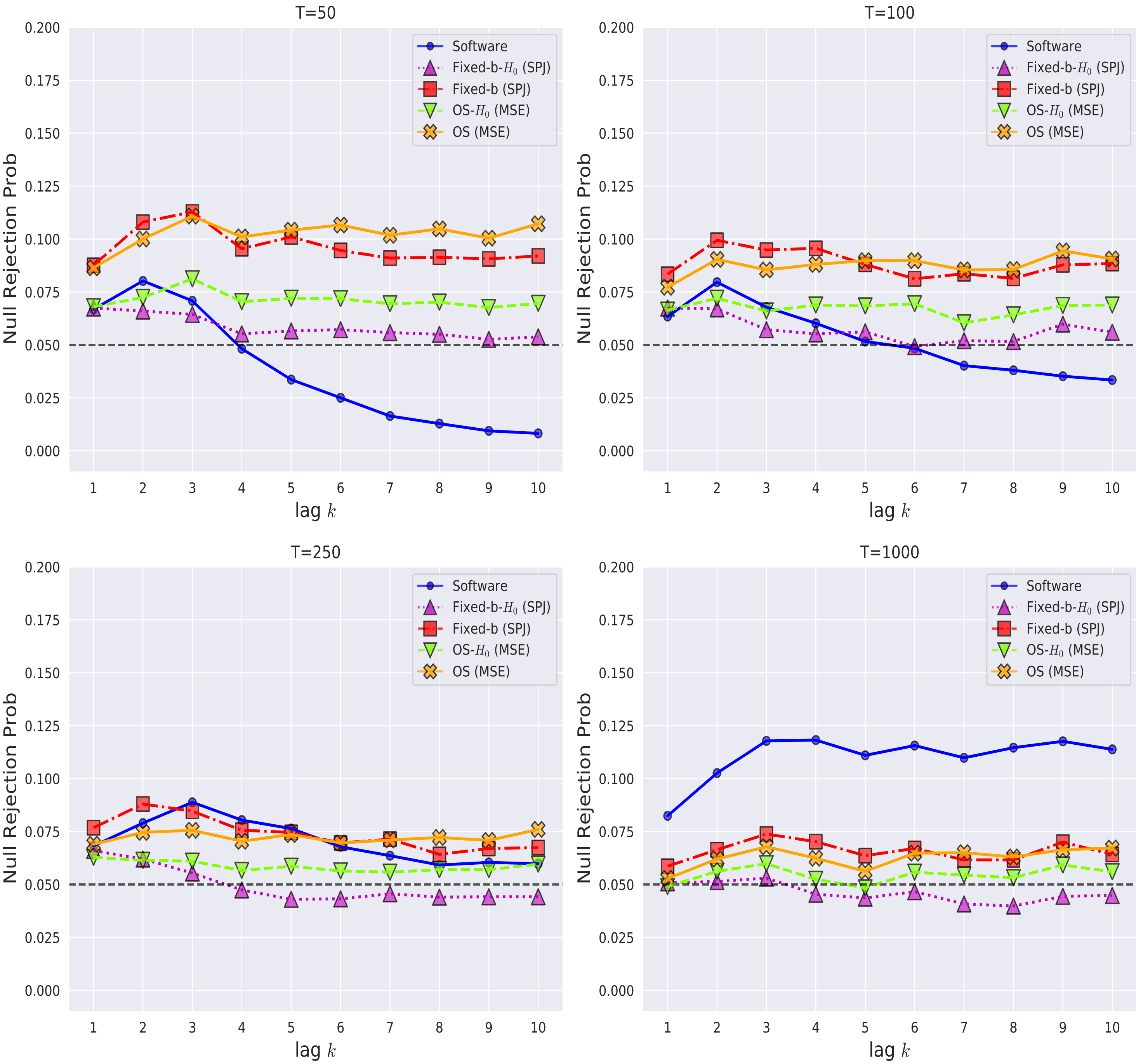

Figure 22: Null rejection probabilities, $H_0 : \rho_k = \phi^k$, $\phi = 0.5$, AR-GRACH

DGP: AR-WN-1 : $y_t = \phi y_{t-1} + \epsilon_t$, where $\epsilon_t = u_t + u_{t-1}u_{t-2}$, $u_t \sim iidNormal(0,1)$

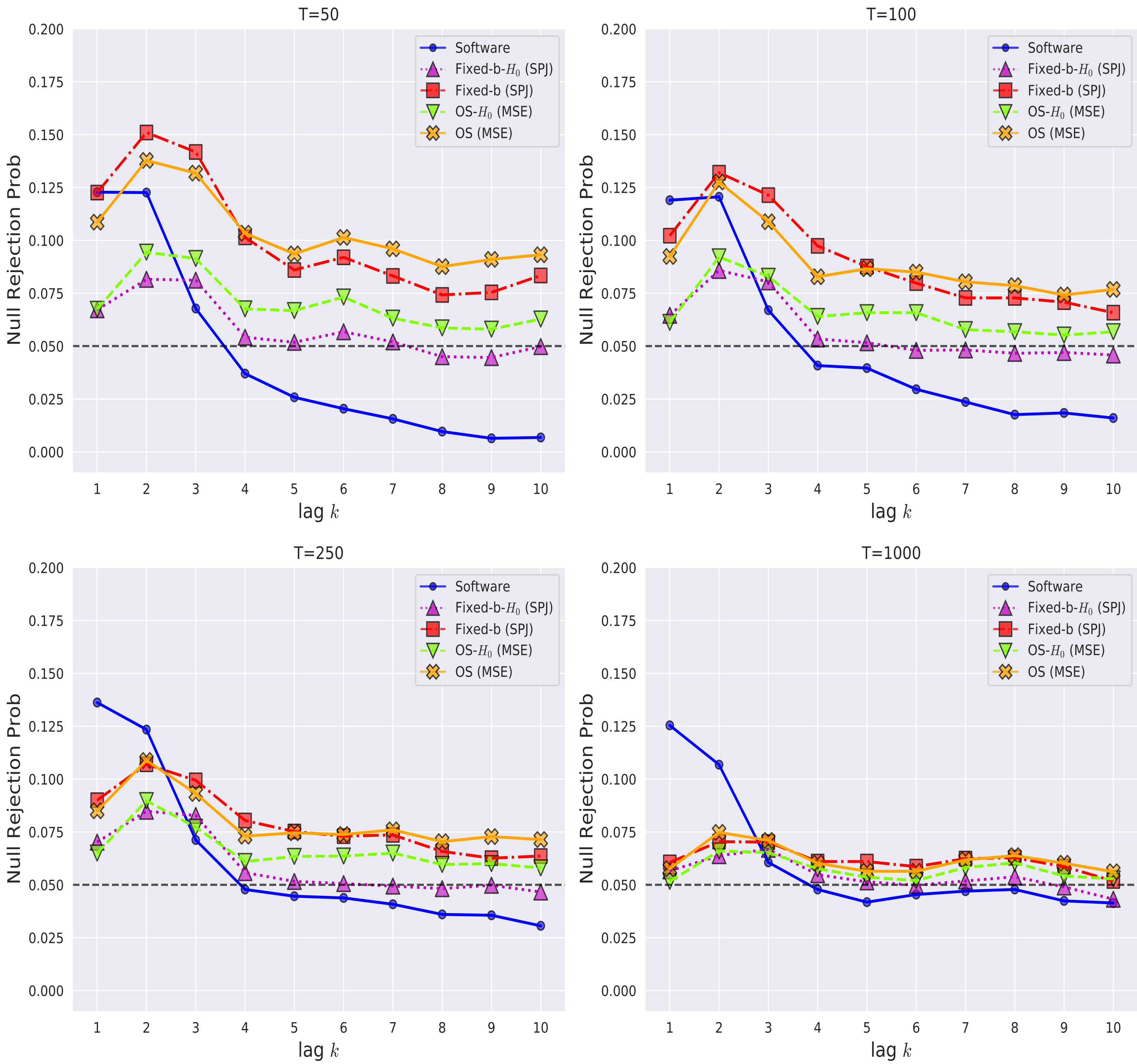

Figure 23: Null rejection probabilities, $H_0 : \rho_k = \phi^k$, $\phi = 0.5$, AR-WN-1

DGP: AR-WN-GAM1 : $y_t = \phi y_{t-1} + \epsilon_t$, where $\epsilon_t = u_t + u_{t-1}u_{t-2}$ and $u_t = \zeta_t - \mathrm{E}[\zeta_t]$, $\zeta_t \sim iidGamma(0.3, 0.4)$

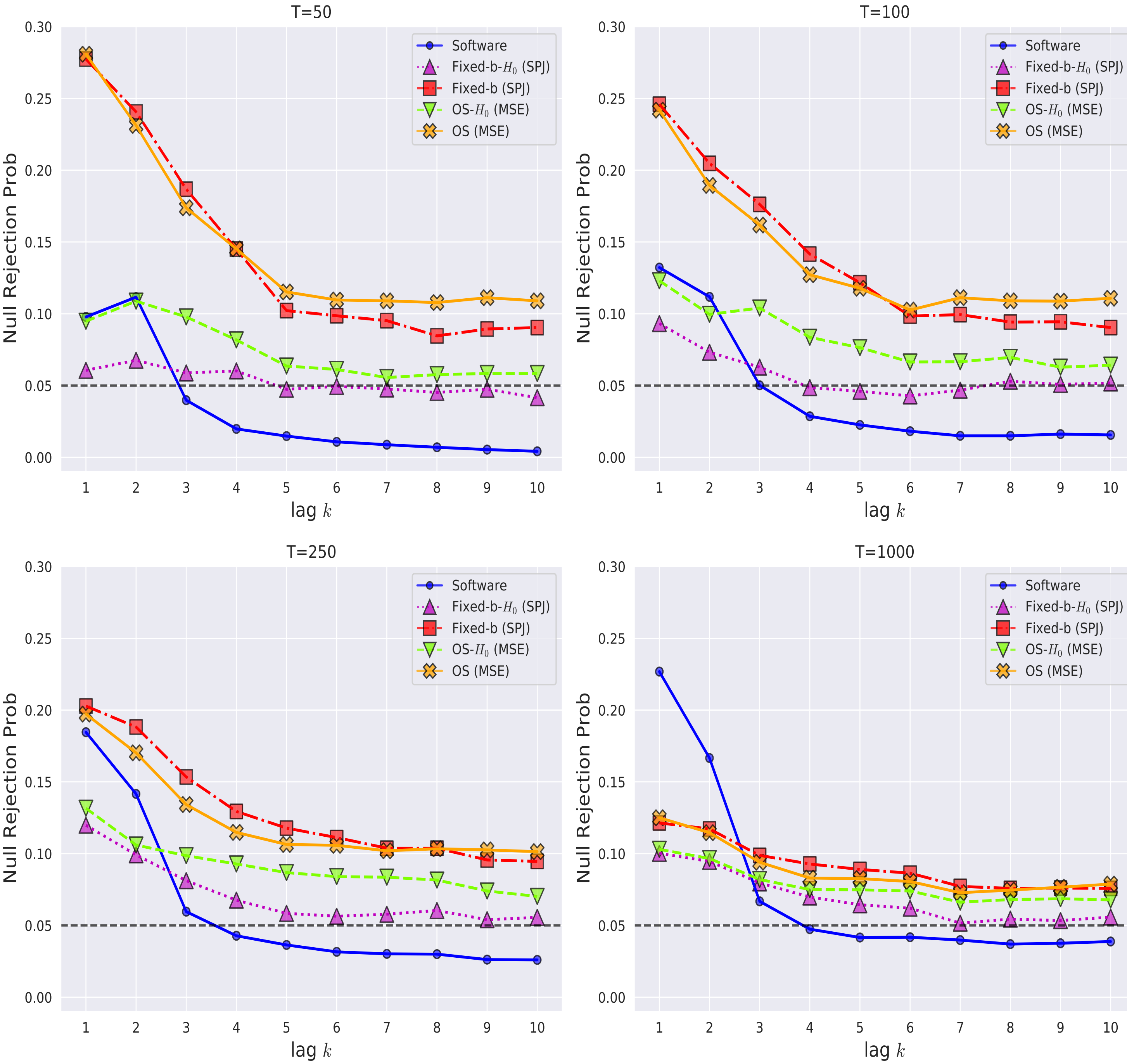

Figure 24: Null rejection probabilities, $H_0 : \rho_k = \phi^k$, $\phi = 0.5$, AR-WN-Gamma

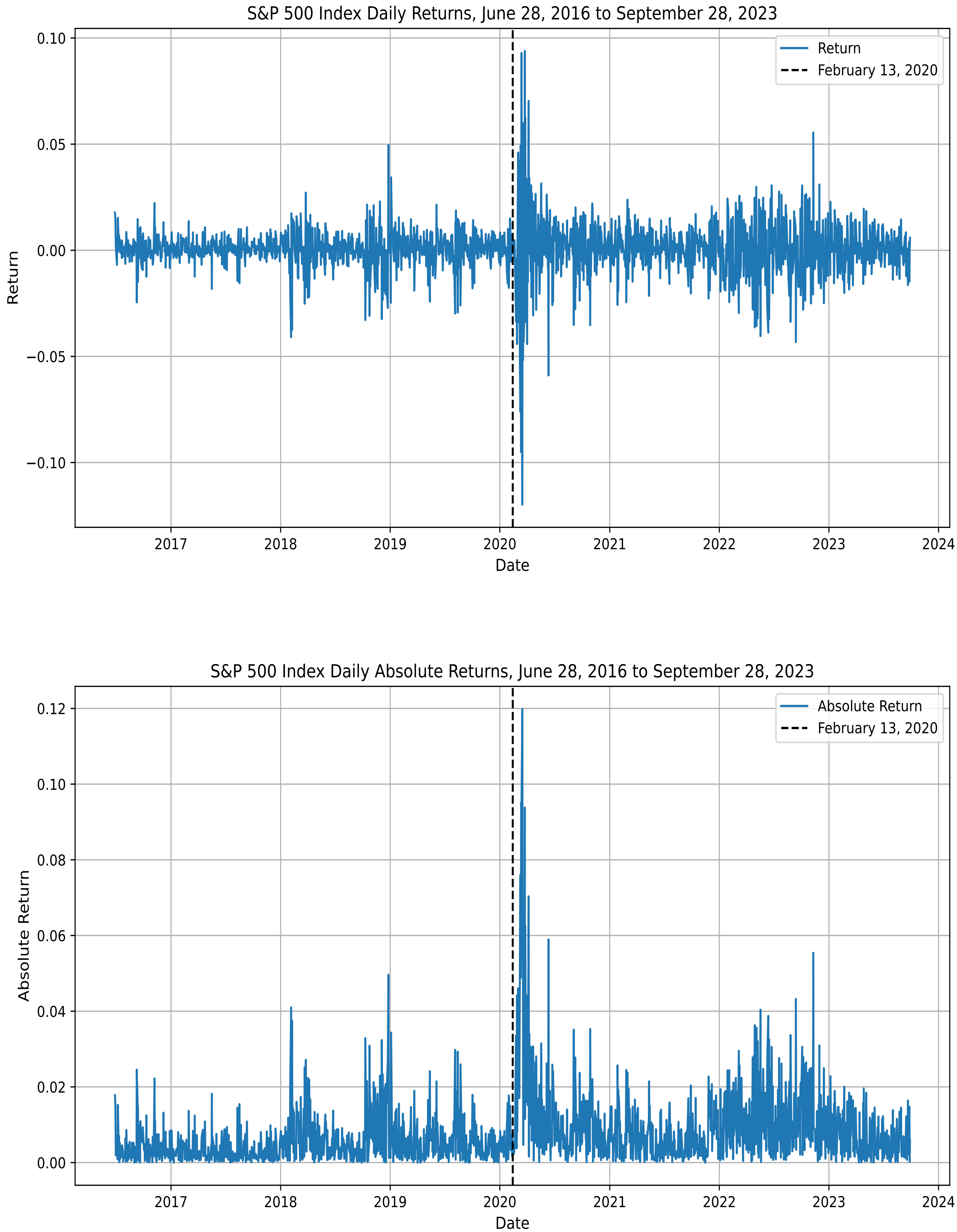

Figure 25: Graphs of S&P 500 index daily returns and absolute returns

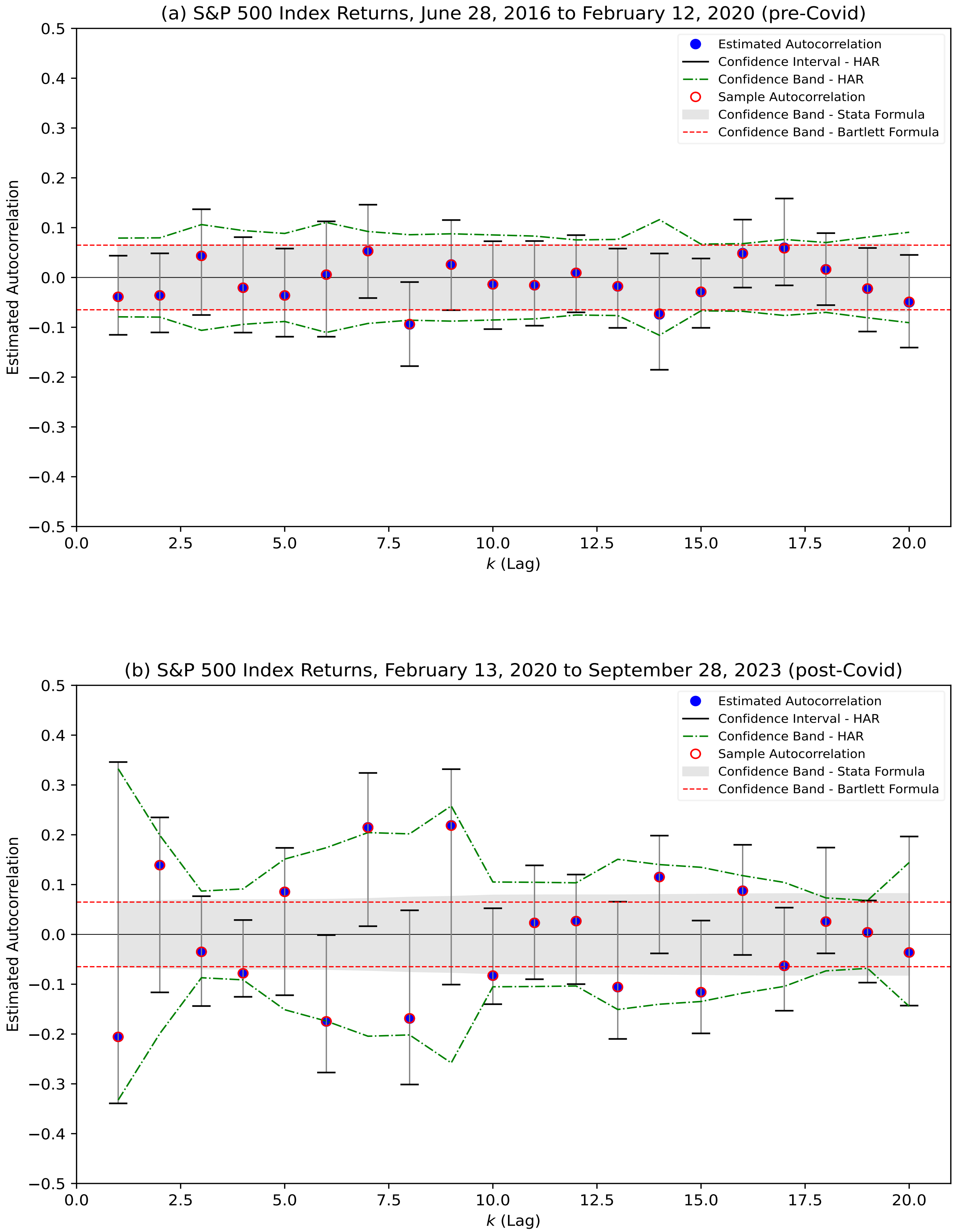

Figure 26: Estimated autocorrelations for S&P 500 index returns during pre- and post-Covid

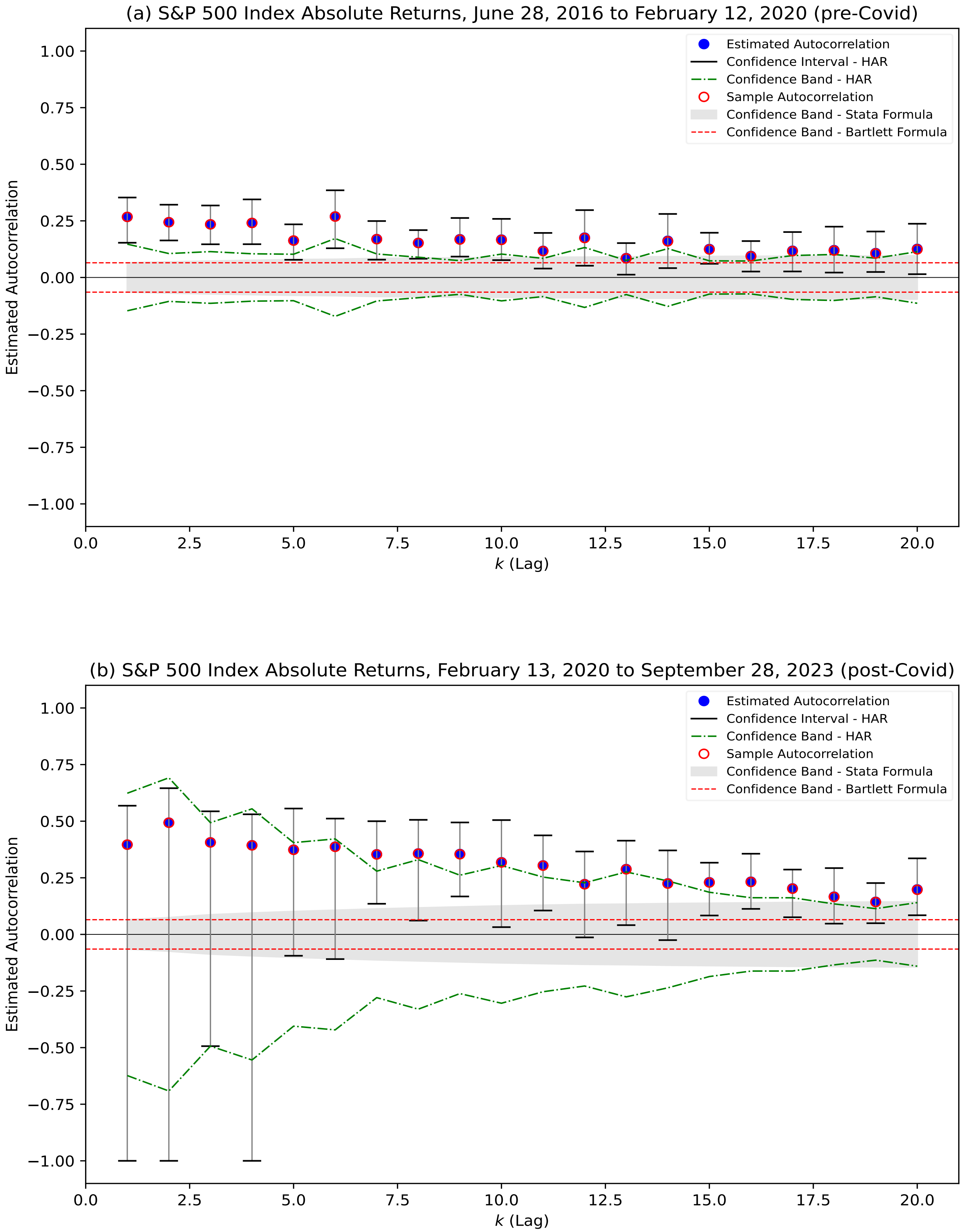

Figure 27: Estimated autocorrelations for S&P 500 index absolute returns during pre- and post-Covid