ESSAYS ON TIME SERIES ECONOMETRICS WITH FINANCIAL APPLICATIONS

By

Taeyoon Hwang

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Economics—Doctor of Philosophy

2024

ABSTRACT

This dissertation provides the developments and extensions of methodologies in time series econometrics and their financial applications. Chapter 1 of the disseration introduces and summarizes the following chapters. Chapter 2 develops an estimating equation approach to construct confidence intervals for autocorrelation functions for time series with general stationary serial correlation structures. Its empirical application using S&P 500 index returns shows that conclusions about market efficiency and volatility clustering during pre and post-Covid periods using the estimating equation approach contrast with conclusions using traditional (and often incorrectly used) methods. Chapter 3 develops fixed-b asymptotics results for heteroskedasticity autocorrelation robust (HAR) Wald tests for regressions for high frequency data using an existing continuous time framework. Its empirical application suggests that the validity of the uncovered interest parity hypothesis depends on whether normal or fixed-b critical values are used. Chapter 4 investigates the distribution of realized US corporate bond return volatility using a compound poisson process setting and realized volatility.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest gratitude to God for guiding me through my Ph.D. journey in good health. I firmly believe that it was with His presence that I have reached this point.

I am profoundly grateful to my advisor, Professor Tim Vogelsang, the person I am most thankful to during my Ph.D. journey. His guidance and support enabled me to complete this dissertation and reach the end of my Ph.D. journey. He was not only the best advisor but also a great friend. I will never forget the valuable research activities and moments of daily life we shared at MSU.

I would also like to extend my sincere thanks to the members of my dissertation committee, Professor Kyoo il Kim, Professor Antonio Galvao, and Professor Hao Jiang, for their constructive feedback and support. I am also grateful to Michigan State University for allowing me to start my Ph.D. and for the support provided throughout my studies. Additionally, I would like to thank the staff of the MSU Econ department for their support, and my colleagues, the Econ graduate students, for their camaraderie and shared experiences.

Special thanks go to Professor Sunku Hahn and Professor Jay Pil Choi, who guided me into the field of economics. I would also like to dedicate this dissertation to the late President Chang Young Jung, who is now with God and made great contributions to the academic field of Korean economics.

I would like to express my gratitude to everyone who prayed for me, especially the members of the Korean Church of Lansing. I am also thankful to my friends Hyunki Kim, Kijoo Song, and Doyeon Kim, who brought laughter into my life, and to Jaeho Choi and Jaehyeong Lim, who discussed the future with me and provided support. I would also like to dedicate this dissertation to my dearly beloved cousin, Mansoo Hwang, who was loved by God and went to be with Him far too early.

Lastly, I am deeply indebted to my parents and partner for their unwavering support and encouragement. To my parents, and my partner, Mira Choi, thank you for your constant love and understanding.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

This dissertation aims to develop and extend time series econometric methodologies with providing their financial applications. Therefore, the following chapters in this dissertation include econometric methods and their empirical application using financial time series data.

Chapter 2 of this dissertation develops an estimating equation approach to obtain valid confidence intervals for autocorrelations for stationary time series. The autocorrelation function is a fundamental statistical analysis tool, widely used in empirical research across various scientific domains. Yet, it is surprising to note that limited work has been done on providing easily implementable methods for inference about the autocorrelation function of time series data under empirically realistic assumptions (e.g., the relaxation of independent identically distributed (i.i.d.) assumption for underlying innovations). Romano and Thombs (1996) pointed out that the Bartlett formula, a primary approach for inference for the autocorrelation function, becomes invalid when the assumption of i.i.d innovations is relaxed. This issue, coupled with the lack of a robust and easily implementable method for such inference, poses challenges in the research domain: 1) The previous literature might yield potentially misleading economic implications when utilizing the Bartlett formula for inference, if the i.i.d assumption is violated. For example, Bollerslev and Mikkelsen (1996) and Andersen et al. (2003) present figures with the confidence bands based on the Bartlett formula to illustrate dependence properties within series of volatilities. 2) From a practical standpoint, many statistical packages commonly used by researchers rely on the Bartlett formula for inference.

Chapter 2 addresses the issue by developing a simple and easy to implement estimating equation approach for robust inference for the autocorrelation function. The estimating equation is estimated by ordinary least squares and inference is heteroskedasticity and autocorrelation robust (HAR). The approach is robust in three ways: innovations can be weak white noise, innovations can have asymmetric distributions, and inference does not require a specific model of serial correlation. Extensive Monte Carlo simulations in Chapter 2 highlight the robustness of the approach. An

1

empirical application using S&P 500 index returns shows that, in the post-Covid period, conclusions about market efficiency and volatility clustering using the approach contrast with conclusions using the traditional approaches.

Chapter 3 of this dissertation develops fixed-b asymptotics results for HAR Wald tests for high frequency data using the continuous time framework of Chang et al. (2023). In the fields of finance and macroeconomics, high frequency data is increasingly being adopted for research. Distinct asymptotics, rooted in a continuous time framework, have been established for regressions involving such high frequency data. These asymptotics consider the time between observations $\delta \rightarrow 0$ and sample span $T \rightarrow \infty$ jointly, different than standard asymptotics of discrete time, as suggested in Chang et al. (2023). In this context, Chapter 3 develops fixed-b asymptotic results for HAR Wald tests for high frequency stationary regression and cointegrating regression under the continuous time framework. Fixed-b asymptotics[1] captures the impact of kernel and bandwidth choices on the sampling distributions of HAR test statistics and typically provides more accurate inference than traditional asymptotics. Chapter 3 shows that fixed-b limits of HAR Wald tests for high frequency stationary regressions in the continuous time setting are the same as the standard discrete time fixed-b limits. The simulation study in Chapter 3 shows that fixed-b critical values provide rejection probabilities closer to nominal levels than traditional chi-square critical values under data generated by Ornstein-Uhlenbeck processes, which are continuous-time analogues of autoregregressive lag 1 (AR(1)) processes. As an empirical application, Chapter 3 provide some basic results on the uncovered interest parity (UIP) puzzle by using Yen/US dollar exchange rate returns and 2-year/10-year government bond yields of the US and Japan from 1991 to 2022, providing evidence that validity of the UIP hypothesis depends on whether normal or fixed-b critical values are used.

Chapter 4 of this dissertation focuses on empirical research on the volatility of a financial instrument, US corporate bonds. Given that corporate bond prices are illiquid and display irregular trading patterns that differ from other assets like stocks, I model the price dynamics of bond prices

---

[1]For more about fixed-b asymptotic theory, please see Kiefer and Vogelsang (2005).

with discrete jumps using a compound Poisson process (CPP). Then, I investigate the distribution of realized US corporate bond return volatility using realized volatility (RV) introduced in Andersen et al. (2001). Monte Carlo simulations are designed to examine finite sample properties of RV under CPP. These simulations consider various structures for the variance of the price jump including the Heston model and allow a different mean value for the number of daily transactions (number of price jumps) for the processes. The simulation results indicate that RV is a solid approximation for integrated volatility when the mean of daily transactions is set to about 146, showing that the mean absolute percentage error is around 10.2% for the case where the variance of the jump follows the Heston model. For the empirical analysis, I build series of daily realized volatilities for US corporate bonds from 2013 to 2018 by using high frequency corporate bond transaction data (recorded every second) from the Financial Industry Regulatory Authority's Trade Reporting and Compliance Engine and link them with corporate bond characteristics (such as credit ratings, issued amounts and yield) to examine the conditional distributions of the volatilities based on each bond characteristic.

BIBLIOGRAPHY

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Ebens, H. (2001). The distribution of realized stock return volatility. *Journal of financial economics*, 61(1):43–76.

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

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

Chang, Y., Lu, Y., and Park, J. Y. (2023). Understanding regressions with observations collected at high frequency over long span. Working paper, School of Economics, University of Sydney.

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

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

CHAPTER 2

AN ESTIMATING EQUATION APPROACH FOR ROBUST CONFIDENCE INTERVALS FOR AUTOCORRELATIONS OF STATIONARY TIME SERIES (CO-AUTHORED WITH TIM VOGELSANG)

2.1 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 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

---

The co-author has approved that the co-authored chapter is included. The co-author's contact: Tim Vogelsang, Department of Economics, 486 W. Circle Drive, 110 Marshall-Adams Hall, Michigan State University East Lansing, MI 48824-1038. email: tjv@msu.edu

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

$$\widehat{Var}\left(\widehat{\rho}_k\right) = \begin{cases} 1/T & k = 1 \\ \frac{1}{T}\left\{1 + 2\sum_{i=1}^{k-1}\widehat{\rho}_i^2\right\} & k > 1, \end{cases} \tag{2.1}$$

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.

2.2 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

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

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

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

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

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

$$\widehat{\rho}_k = \widehat{\gamma}_k/\widehat{\gamma}_0. \tag{2.2}$$

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, $\widehat{\rho} = (\widehat{\rho}_1, \ldots, \widehat{\rho}_m)'$, asymptotically follows a normal distribution with mean ρ, the vector of corresponding population autocorrelations up to lag m. The asymptotic variance-covariance matrix of $\widehat{\rho}$ 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:

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

8

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 $\sqrt{T}(\widehat{\rho} - \rho)$ 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 $c_{i+1,j+1} = \sum_{d=-\infty}^{\infty} \text{cov}\left(y_0 y_i, y_d y_{d+j} \right)$. 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 $\sqrt{T}(\widehat{\gamma} - \gamma)$ where $\gamma = (\gamma_0, ..., \gamma_m)'$ and $\widehat{\gamma} = (\widehat{\gamma}_0, ..., \widehat{\gamma}_m)'$. 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

$$v_{i,i}^{B} = 1, \ v_{i,i}^{B^*} = \frac{\gamma_{\epsilon^2}(i)}{[\gamma_\epsilon(0)]^2}, \tag{2.3}$$

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.

2.3 Theory

2.3.1 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:

$$y_t = c + \rho_k y_{t-k} + \eta_t^{(k)}, \tag{2.4}$$

where $c = \mu(1 - \rho_k)$ and $t = k + 1, k + 2, \ldots, T$. Regression (2.4) allows consistent estimation of c and ρ_k because

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

These conditions are easy to establish as follows. Taking the mean of both sides of (2.4) gives

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

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

10

Replacing $E(y_t)$ and $E(y_{t-k})$ with μ and because $c = \mu(1 - \rho_k)$, it follows that

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

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

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

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

$$\eta_t^{(k)} = y_t - c - \rho_k y_{t-k} = (y_t - \mu) - \rho_k(y_{t-k} - \mu). \tag{2.5}$$

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 (2.5) gives

$$\phi(L)\eta_t^{(k)} = \phi(L)(y_t - \mu) - \phi(L)\rho_k(y_{t-k} - \mu) = \phi(L)(y_t - \mu) - \rho_k\phi(L)L^k(y_t - \mu)$$
$$= \phi(L)(y_t - \mu) - \rho_k L^k \phi(L)(y_t - \mu) = \theta(L)\epsilon_t - \rho_k L^k \theta(L)\epsilon_t$$
$$= \left(1 - \rho_k L^k\right)\theta(L)\epsilon_t. \tag{2.6}$$

We see from (2.6) 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 (2.6) 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

11

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 rather than the rule, and inference based on estimation of (2.4) should be made robust to serial correlation (and conditional heteroskedasticity).

It is convenient to rewrite the estimation equation (2.4) as

$$y_t = \mathbf{x}'_{t-k}\beta + \eta_t^{(k)} \tag{2.7}$$

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 (2.7) 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,

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

$$= \left(T^{-1} \sum_{t=k+1}^{T} \mathbf{x}_{t-k}\mathbf{x}'_{t-k} \right)^{-1} T^{-1/2} \sum_{t=k+1}^{T} \mathbf{v}_t^{(k)} = \left(T^{-1} \sum_{t=k+1}^{T} \mathbf{x}_{t-k}\mathbf{x}'_{t-k} \right)^{-1} T^{-1/2}\mathbf{S}_T^{(k)}.$$

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 2.1 *$T^{-1/2} \sum_{t=k+1}^{[rT]} \mathbf{v}_t^{(k)} = T^{-1/2} \mathbf{S}_{[rT]}^{(k)} \Rightarrow \mathbf{\Lambda}^{(k)} \mathbf{W}_2(r)$, 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 2.2 *$T^{-1} \sum_{t=k+1}^{[rT]} \mathbf{x}_{t-k} \mathbf{x}'_{t-k} \xrightarrow{p} r\mathbf{Q} = r \begin{bmatrix} 1 & \mu \\ \mu & \gamma_0 + \mu^2 \end{bmatrix}$, where $r \in [0, 1]$.*

Assumption 2.1 is a functional central limit theorem (FCLT) for the scaled partial sums of $\mathbf{v}_t^{(k)}$. Assumption 2.1 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 2.1 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 2.1; see Davidson (1994). Assumption 2.2 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

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

13

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 (2.4) 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

$$\widetilde{\mathbf{v}}_t^{(k)} = \mathbf{x}_{t-k}\widetilde{\eta}_t^{(k)}, \quad \widetilde{\eta}_t^{(k)} = y_t - \mathbf{x}_{t-k}'\widetilde{\beta} = y_t - \widetilde{c} - \widetilde{\rho}_k y_{t-k}, \tag{2.8}$$

$k(x)$ 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 (2.4) 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.

2.3.2 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

15

t-statistics:

$$\widetilde{t}^{(k)} = \frac{(\widetilde{\rho}_k - a)}{\sqrt{\frac{1}{T-k}\widetilde{V}_{22}^{(k)}}}, \quad \widetilde{t}^{(k)*} = \frac{(\widetilde{\rho}_k - a)}{\sqrt{\frac{1}{T-k}\widetilde{V}_{22}^{(k)*}}} \tag{2.9}$$

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 et al. (2008), Gonçalves and Vogelsang (2011), Zhang and Shao (2013), Lazarus et al. (2018) and Lazarus et al. (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 2.1 *Let $M = bT$ where $b \in (0, 1]$ is fixed. Under Assumptions 1 and 2, as $T \to \infty$, the fixed-b limits of $\widetilde{\boldsymbol{\Omega}}^{(k)}$, and $\widetilde{\boldsymbol{\Omega}}^{(k)*}$ are given by*

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

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 $\widetilde{\boldsymbol{\Omega}}^{(k)}$ and $\widetilde{\boldsymbol{\Omega}}^{(k)*}$ 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

---

[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.

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

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

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

Using Theorem 2.1, the fixed-b limits of the t-statistics immediately follow from arguments in Kiefer and Vogelsang (2005) and are given by

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

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 2A.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

---

[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.

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 2.3 *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

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

giving the variance estimators

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

The corresponding t-statistics are given by

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

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

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

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 $\widetilde{t}_{OS}^{(k)*}$.

2.3.3 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 (2.9). 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 (2.4) 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

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

19

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

$$\ddot{\Omega}^{(k)*} = (T-k)^{-1}\sum_{t=k+1}^{T}\sum_{s=k+1}^{T}\ddot{v}_t^{(k)*}k_{ts}\ddot{v}_s^{(k)*}$$

$$= (T-k)^{-1}\sum_{t=k+1}^{T}\sum_{s=k+1}^{T}(\ddot{y}_{t-k}\ddot{y}_t - a\ddot{y}_{t-k}^2)k_{ts}(\ddot{y}_{s-k}\ddot{y}_s - a\ddot{y}_{s-k}^2),$$

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

$$\ddot{\Omega}^{(k)*} = \ddot{\Omega}_{11}^{(k)*} - 2a\ddot{\Omega}_{12}^{(k)*} + a^2\ddot{\Omega}_{22}^{(k)*}, \tag{2.10}$$

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

$$\left(\frac{(\widetilde{\rho}_k - a)}{\sqrt{\frac{1}{T-k} \ddot{Q}^{-1} \left(\ddot{\Omega}_{11}^{(k)*} - 2a\ddot{\Omega}_{12}^{(k)*} + a^2\ddot{\Omega}_{22}^{(k)*} \right)}} \right)^2 \leq cv_{\alpha/2}^2. \tag{2.11}$$

The inequality (2.11) can be rewritten as

$$c_2 a^2 + 2c_1 a + c_0 \leq 0, \tag{2.12}$$

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 (2.12).

The values of a satisfying the inequality (2.12) 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 $\widetilde{t}^{(k)*} = 0$ in this case and a

21

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

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

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 (2.10) and is scaled by a positive quantity. With appropriate choice of kernel, 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 $\ddot{\Lambda}_{\ell,2}^* = \frac{1}{\sqrt{T-k}}\sum_{t=k+1}^{T}\Phi_\ell\left(\frac{t}{T}\right)\ddot{y}_{t-k}^2$.

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 **?**) that are functions of $\widetilde{\mathbf{v}}_t^{(k)}$, the null-not-imposed proxy for $\mathbf{v}_t^{(k)}$ given by equation (2.8).

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

$$y_t = \phi y_{t-1} + \epsilon_t + \theta \epsilon_{t-1}, \tag{2.13}$$

where $\mu = 0$ without loss of generality given that we include in a intercept in the estimating equation (2.4). 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: $\epsilon_t = u_t + 0.5u_{t-1}\epsilon_{t-2}$, $u_t \sim i.i.d.N(0, 1)$.

23

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.

2.4.1 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 (2.13). 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 $v_{1,1}^B = 1$ whether or not y_t is i.i.d. For the generalized Bartlett formula, we use the formula (2.3) 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 et al. (2022). The Taylor (1984) $\widetilde{\tau}_1$ t-statistic is given by

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

24

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 2A.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 $\widetilde{v}_t^{(1)}$

---

[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 $E\left(v_t^{(1)} v_{t-1}^{(1)}\right) = 0$.

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 $\widetilde{t}^{(1)}_{OS}$ (null-not-imposed, orange x's solid line) and $\widetilde{t}^{(1)*}_{OS}$ (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 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 $\widetilde{t}^{(1)}_{OS}$ (null-not-imposed) for smaller sample sizes. In contrast, $\widetilde{t}^{(1)*}$ and $\widetilde{t}^{(1)*}_{OS}$ (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 2A.1.8) all tests show some over-rejections with $\widetilde{t}^{(1)*}$ and $\widetilde{t}^{(1)*}_{OS}$ (null-imposed) having rejections closest to 0.05. The null-not-imposed tests, $\widetilde{t}^{(1)}$ and $\widetilde{t}^{(1)}_{OS}$, 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 2A.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.

2.4.2 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

27

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 (2.2). 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 $\widehat{\epsilon}_t^2$.

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, $\rho_1^{(0)} = \theta/(1 + \theta^2)$ and for the AR(1) case $\rho_1^{(0)} = \phi$.

Figure 2A.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 2A.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

---

[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.

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 2A.4 shows that all methods works well except for the traditional Bartlett and GB-WN as one would expect. Figure 2A.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 2A.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 2A.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 $\widetilde{t}_{OS}^{(1)*}$ (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.

2.4.3 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

29

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 $\tilde{t}^{(1)*}$. In the power figures we label $\tilde{t}^{(1)}$ as 'Kernel' and $\tilde{t}^{(1)*}$ 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 2A.12 gives results for ϵ_t IID. Size-adjusted power is essentially the same across all tests. Figure 2A.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 2A.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.

2.4.4 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 (2.1). 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 2A.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 2A.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 2A.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 $\widetilde{t}^{(k)*}_{OS}$. When these two tests have some over-rejections, they are less pronounced for $\widetilde{t}^{(k)*}$ than for $\widetilde{t}^{(k)*}_{OS}$. This is not because $\widetilde{t}^{(k)*}$ uses a kernel and $\widetilde{t}^{(k)*}_{OS}$ 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 $\widetilde{t}^{(k)*}_{OS}$ tends to over-reject more than $\widetilde{t}^{(k)*}$ is because $\widetilde{K}$ leads to less smoothing than $\widetilde{M}$.

2.4.5 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 2A.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 2A.3 shows very different patterns from Table 2A.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.

2.5 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 et al. (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 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 2A.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 2A.26 plots estimated autocorrelations for S&P 500 returns for daily data from June

---

[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.

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 (2.2) and blue dots are autocorrelations estimated by OLS using (2.4). The dashed red lines are i.i.d. Bartlett confidence bands. The gray area is the Stata confidence bands using equation (2.1). 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 2A.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 2A.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 2A.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.

2.6 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.

BIBLIOGRAPHY

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

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*, volume 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. 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.

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

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.

TABLES AND FIGURES

Table 2A.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 2A.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 2A.3 Shape of Confidence Intervals using $\widetilde{t}^{(k)*}$, AR-WN-NLMA

$\phi = 0.0$

Case	T=50			T=100			T=250			T=1000		
	$\widehat{Prob}$	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

$\phi = 0.25$

Case	T=50			T=100			T=250			T=1000		
	$\widehat{Prob}$	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

$\phi = 0.7$

Case	T=50			T=100			T=250			T=1000		
	$\widehat{Prob}$	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

$\phi = -0.7$

Case	T=50			T=100			T=250			T=1000		
	$\widehat{Prob}$	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 2A.1.1 Graphs of null rejection probabilities, $H_0 : \rho_1 = 0$
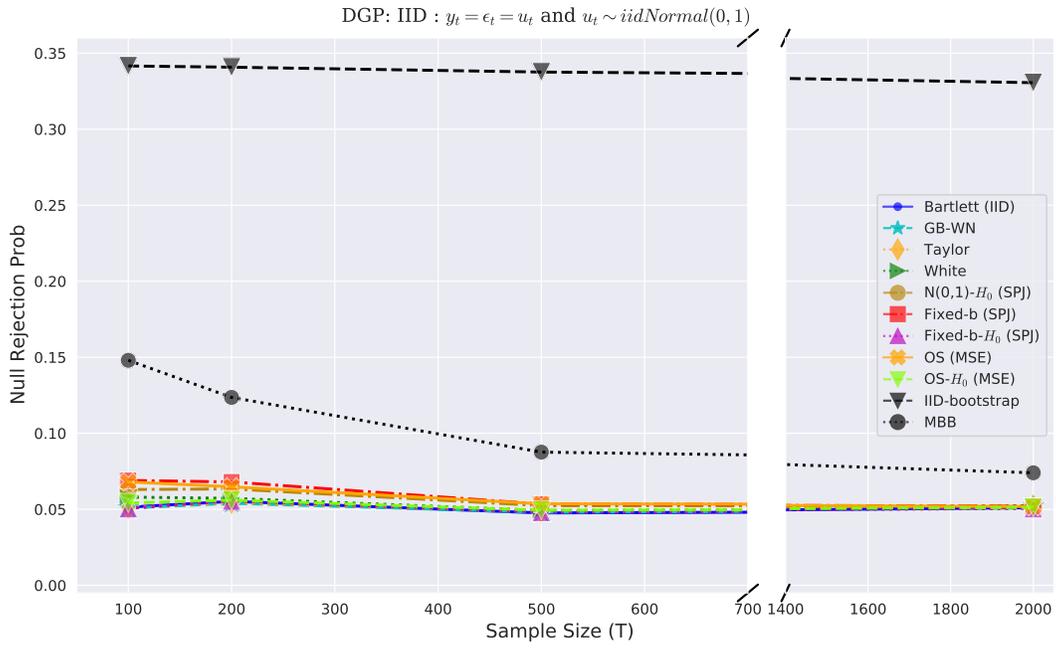
Figure 2A.1.2 Graphs of null rejection probabilities, $H_0 : \rho_1 = 0$
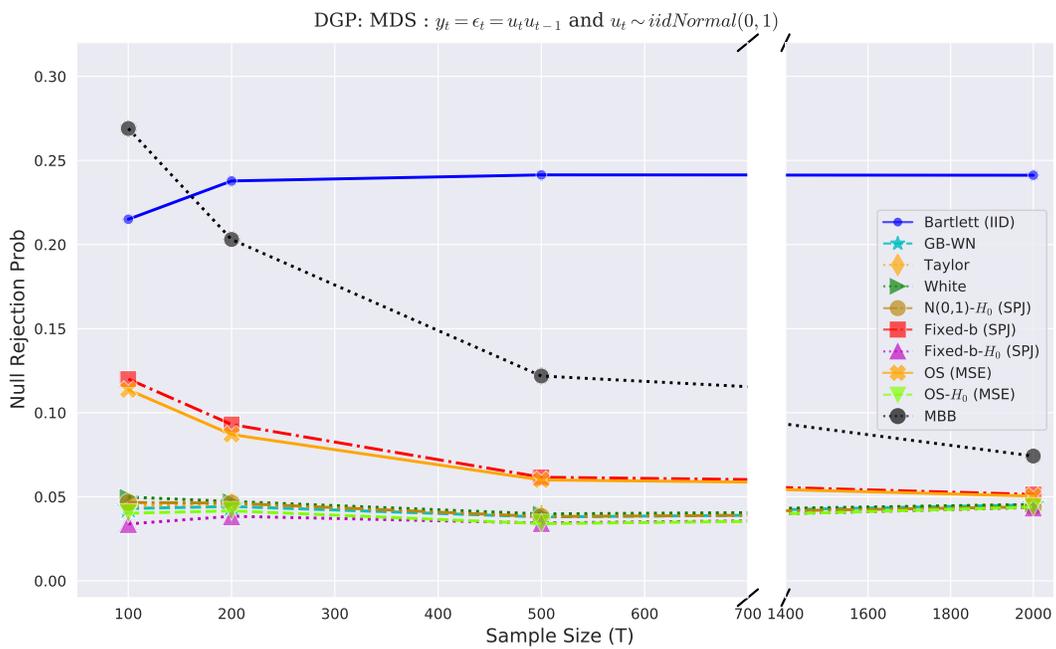
Figure 2A.1.3 Graphs of null rejection probabilities, $H_0 : \rho_1 = 0$

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

Figure 2A.1.3: Graphs of null rejection probabilities, $H_0 : \rho_1 = 0$

DGP: WN-1 : $y_t = \epsilon_t = u_t + u_{t-1}u_{t-2}$ and $u_t \sim iidNormal(0,1)$
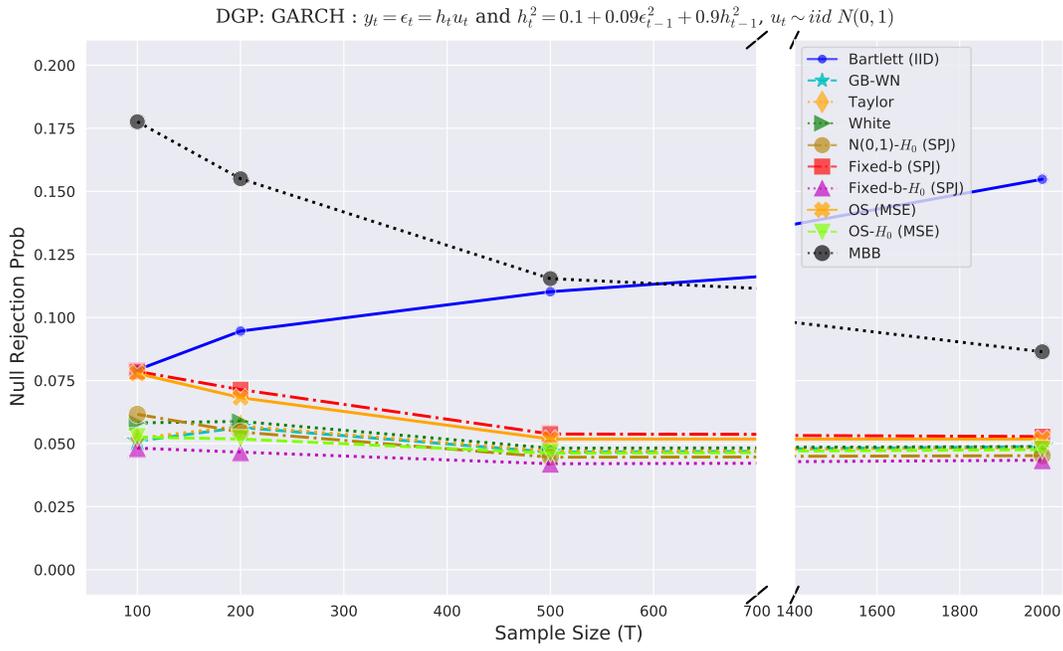
Figure 2A.1.5: Graphs of null rejection probabilities, $H_0 : \rho_1 = 0$

DGP: WN-2 : $y_t = \epsilon_t = u_t^2 u_{t-1}$ and $u_t \sim iidNormal(0, 1)$
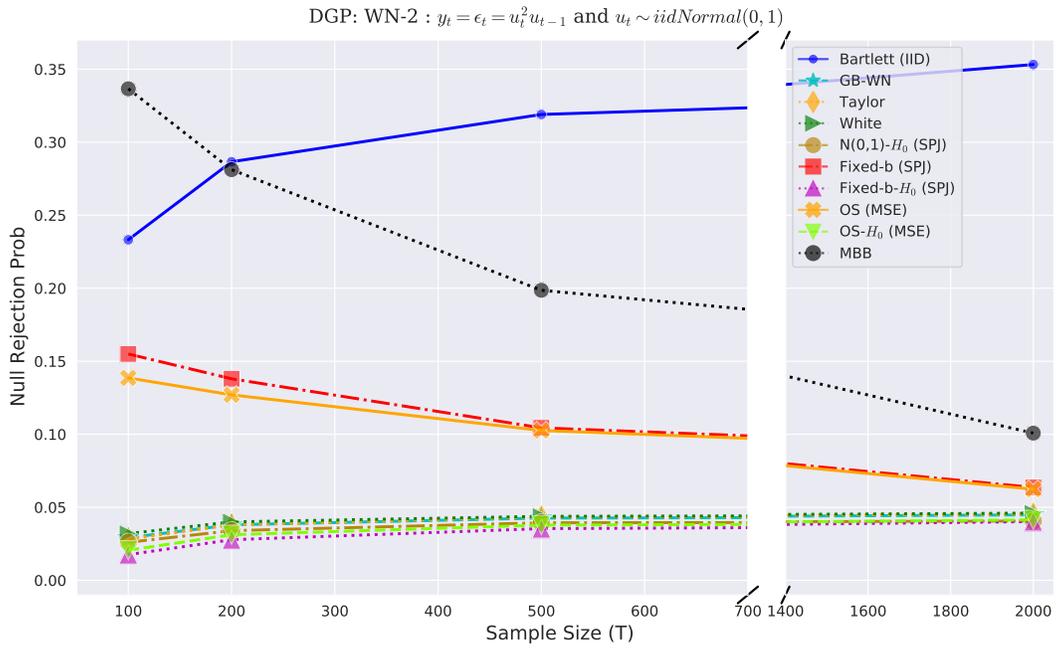
Figure 2A.1.6: Graphs of null rejection probabilities, $H_0 : \rho_1 = 0$

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

Figure 2A.1.7: Graphs of null rejection probabilities, $H_0 : \rho_1 = 0$

DGP: WN-BILIN: $y_t = \epsilon_t = u_t + 0.5u_{t-1}\epsilon_{t-2}$ and $u_t \sim iidNormal(0,1)$
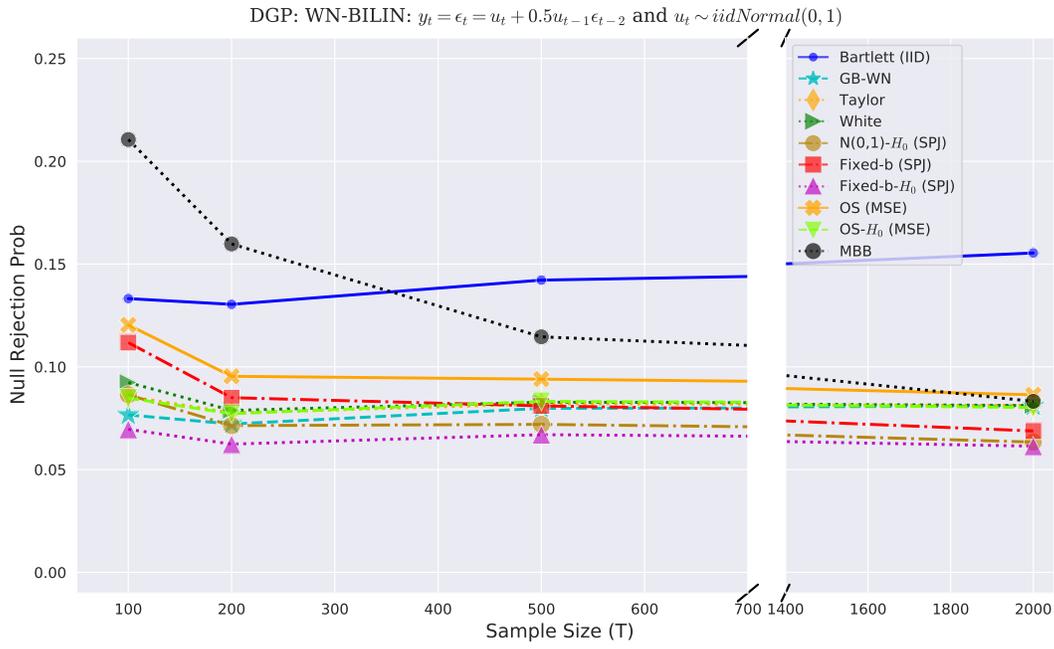
Figure 2A.1.8: Graphs of null rejection probabilities, $H_0 : \rho_1 = 0$

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

44
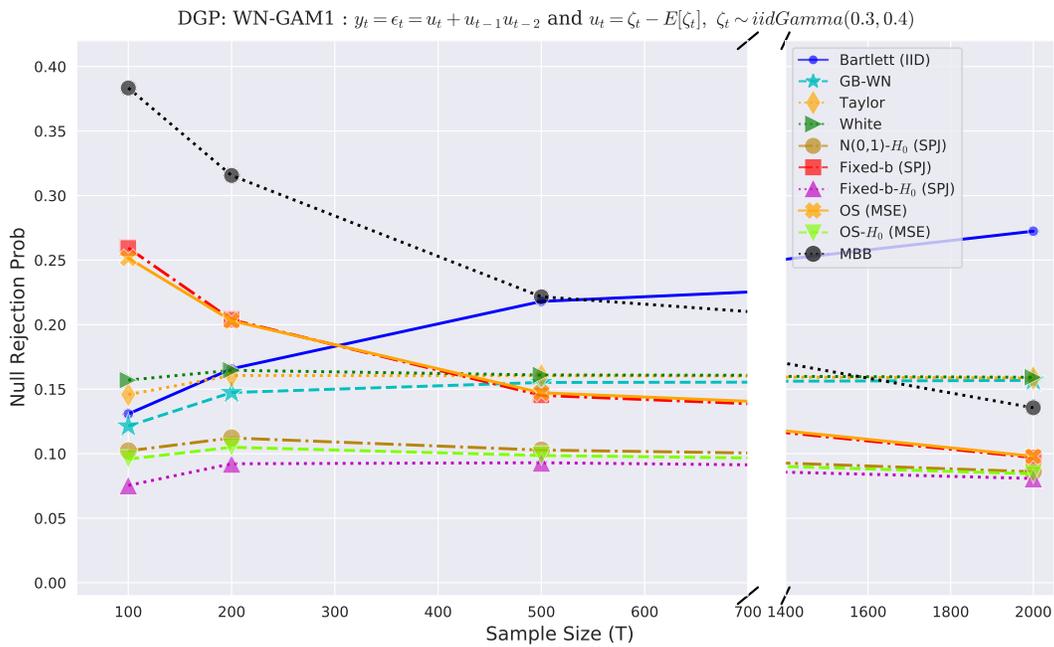
Figure 2A.1.9: Graphs of null rejection probabilities, $H_0 : \rho_1 = 0$

DGP: WN-GAM2 : $y_t = \epsilon_t = u_t - u_{t-1}u_{t-2}$ and $u_t = \zeta_t - E[\zeta_t]$, $\zeta_t \sim iidGamma(0.3, 0.4)$
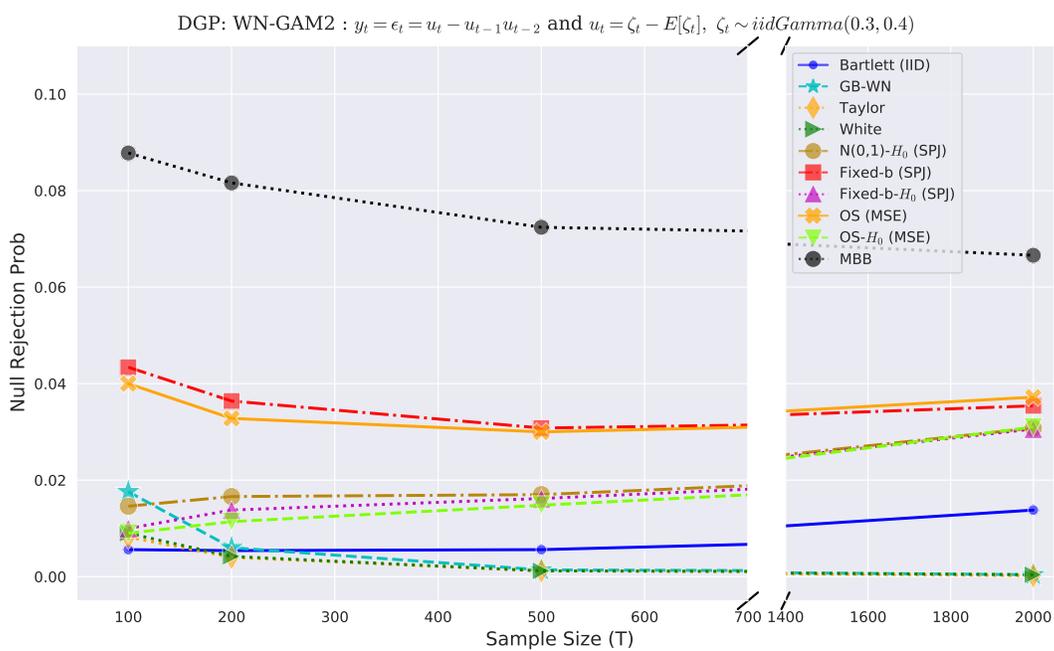
Figure 2A.2 Null rejection probabilities, $H_0 : \rho_1 = \frac{\theta}{(1+\theta^2)}$, MA-IID

DGP: MA-IID : $y_t = \epsilon_t + \theta\epsilon_{t-1}$, where $\epsilon_t \sim iidNormal(0,1)$
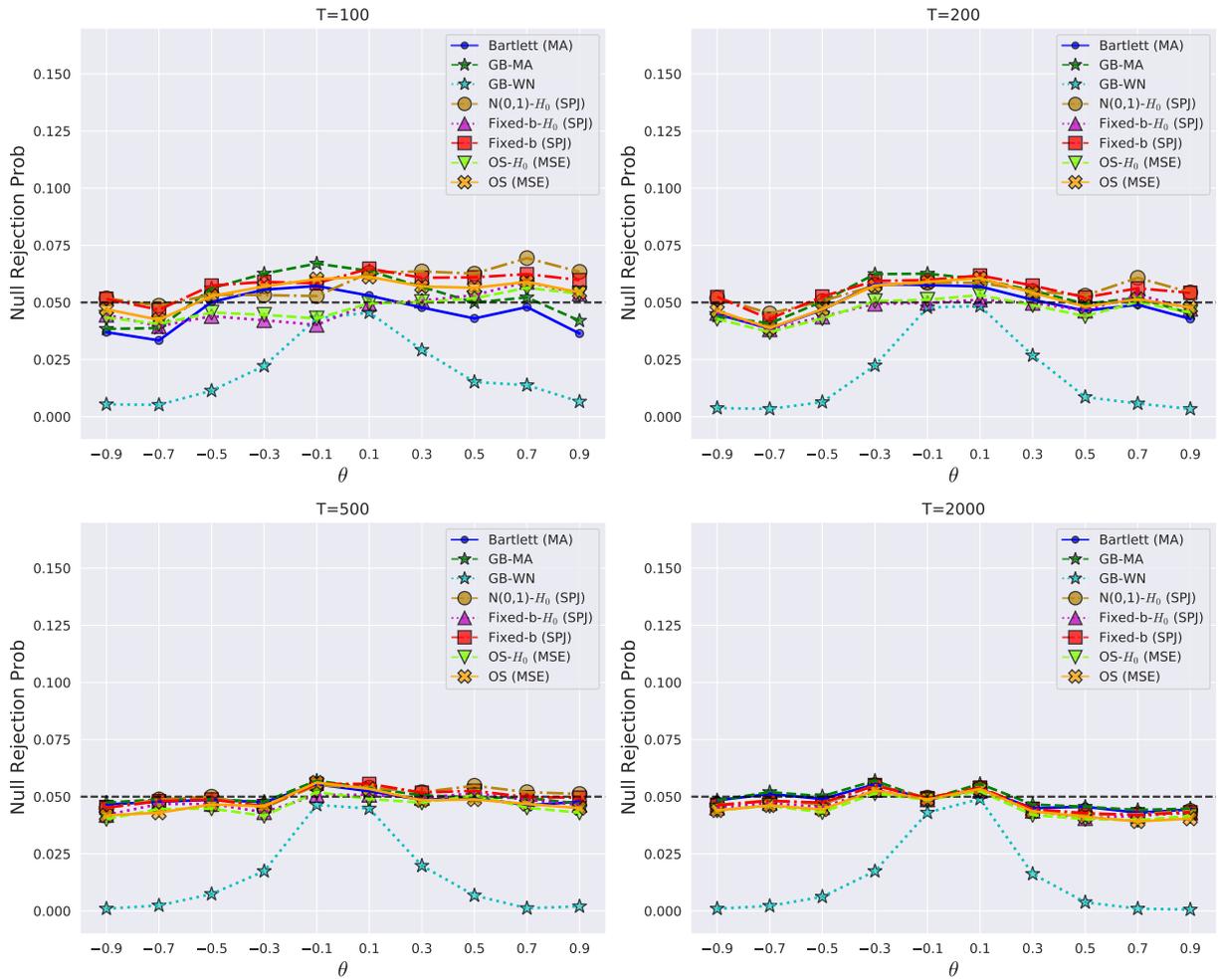
Figure 2A.3: Null rejection probabilities, $H_0 : \rho_1 = \frac{\theta}{(1+\theta^2)}$, MA-MDS

DGP: MA-MDS : $y_t = \epsilon_t + \theta\epsilon_{t-1}$, where $\epsilon_t = u_t u_{t-1}$, $u_t \sim iidNormal(0,1)$
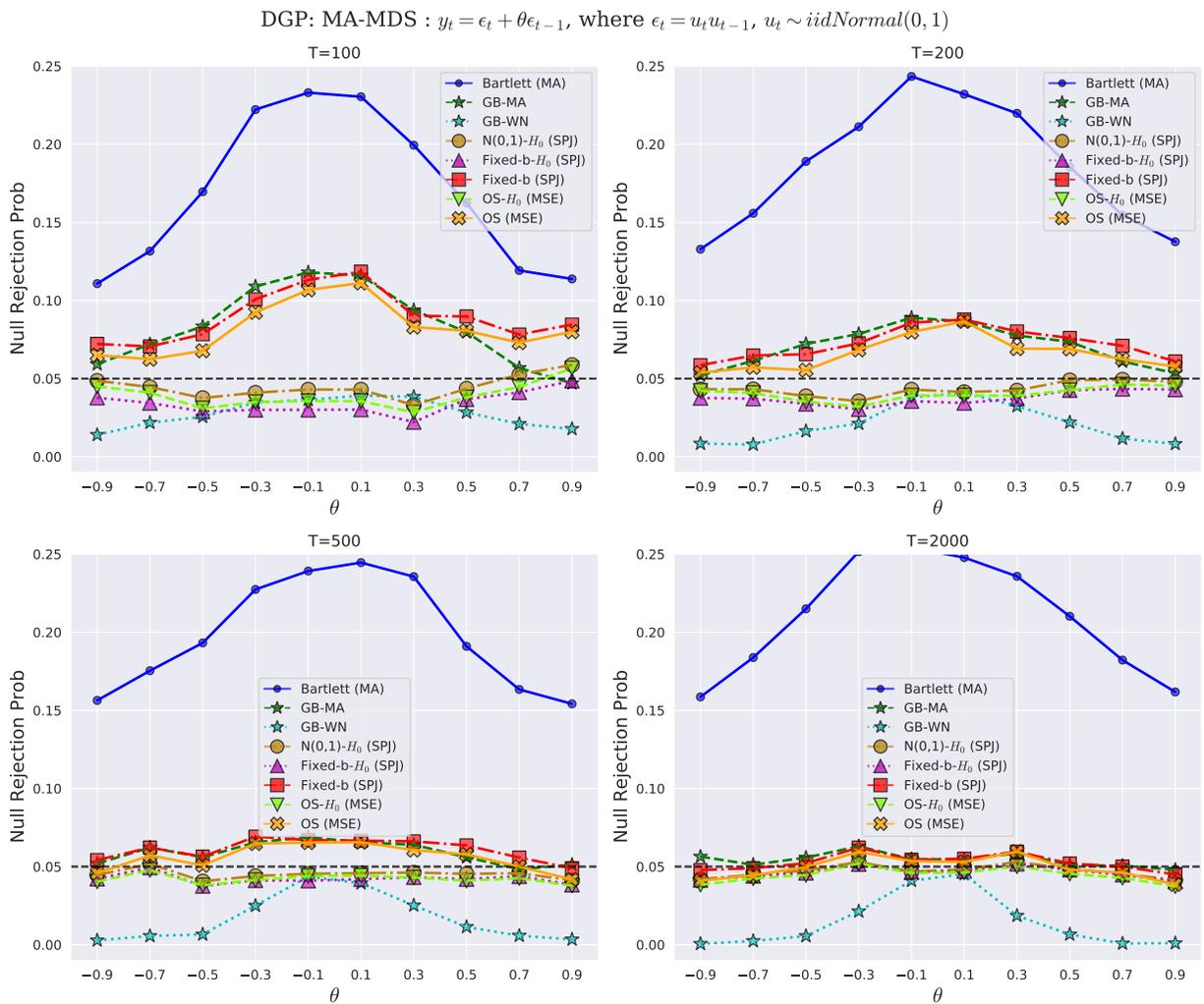
Figure 2A.4: Null rejection probabilities, $H_0 : \rho_1 = \frac{\theta}{(1+\theta^2)}$, MA-GRACH

DGP: MA-GARCH : $y_t = \epsilon_t + \theta \epsilon_{t-1}$, where $\epsilon_t = h_t u_t$ and $h_t^2 = 0.1 + 0.09 \epsilon_{t-1}^2 + 0.9 h_{t-1}^2$, $u_t \sim iid N(0,1)$
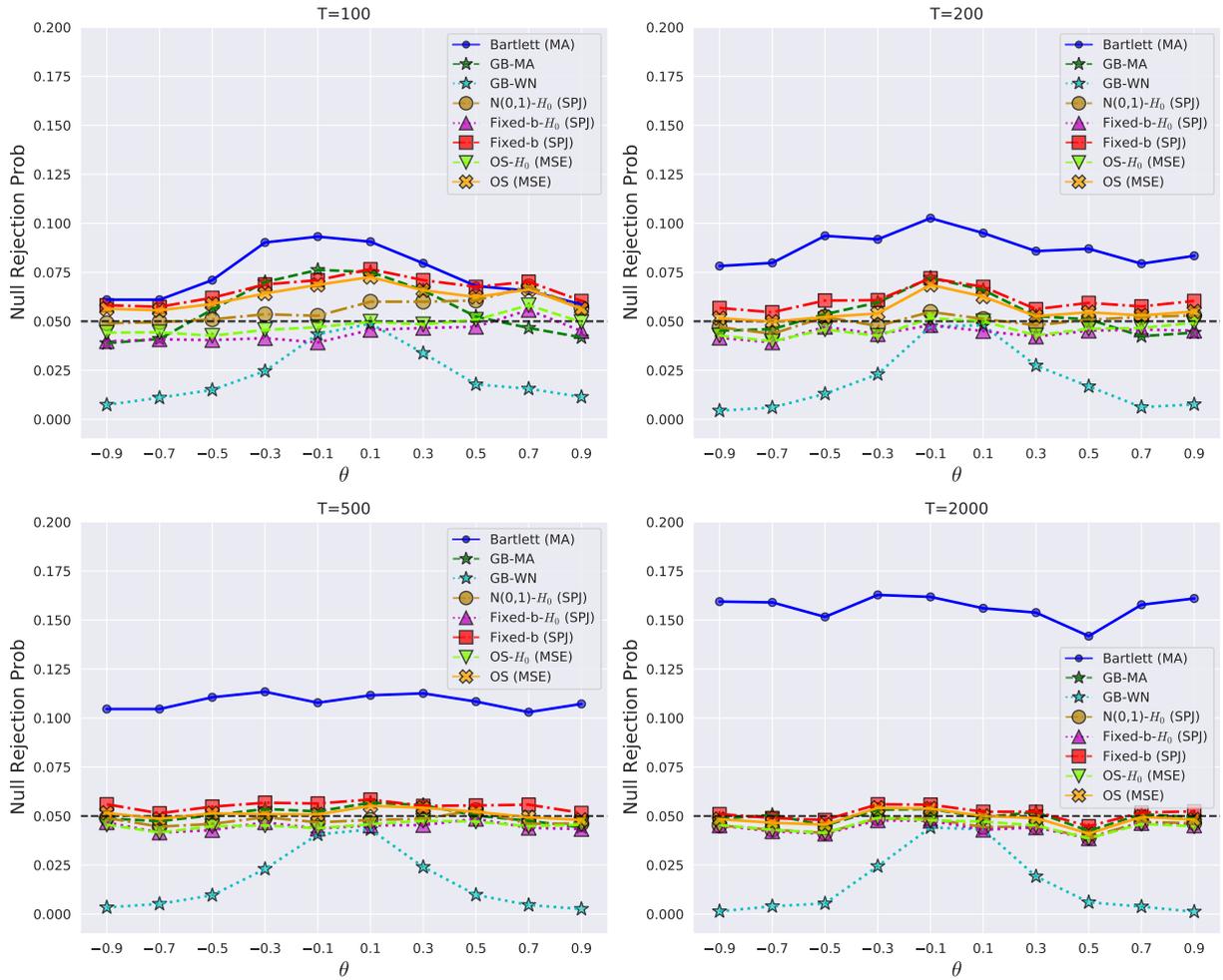
Figure 2A.5 Null rejection probabilities, $H_0 : \rho_1 = \frac{\theta}{(1+\theta^2)}$, MA-WN-1

DGP: MA-WN-1 : $y_t = \epsilon_t + \theta\epsilon_{t-1}$, where $\epsilon_t = u_t + u_{t-1}u_{t-2}$, $u_t \sim iidNormal(0,1)$
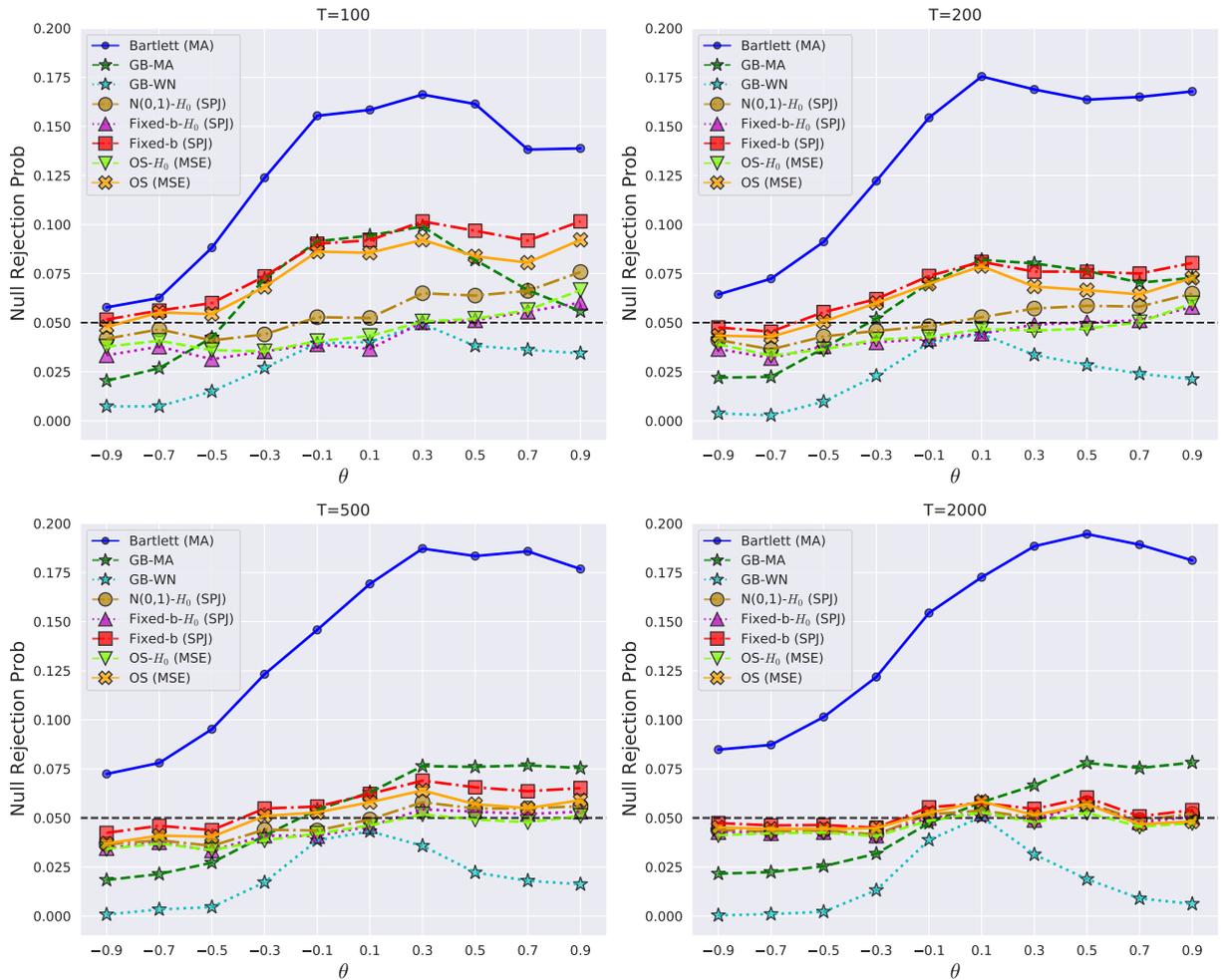
Figure 2A.6 Null rejection probabilities, $H_0 : \rho_1 = \frac{\theta}{(1+\theta^2)}$, MA-WN-Gamma

DGP: MA-WN-GAM1 : $y_t = \epsilon_t + \theta\epsilon_{t-1}$, where $\epsilon_t = u_t + u_{t-1}u_{t-2}$ and $u_t = \zeta_t - \mathrm{E}[\zeta_t]$, $\zeta_t \sim iidGamma(0.3, 0.4)$
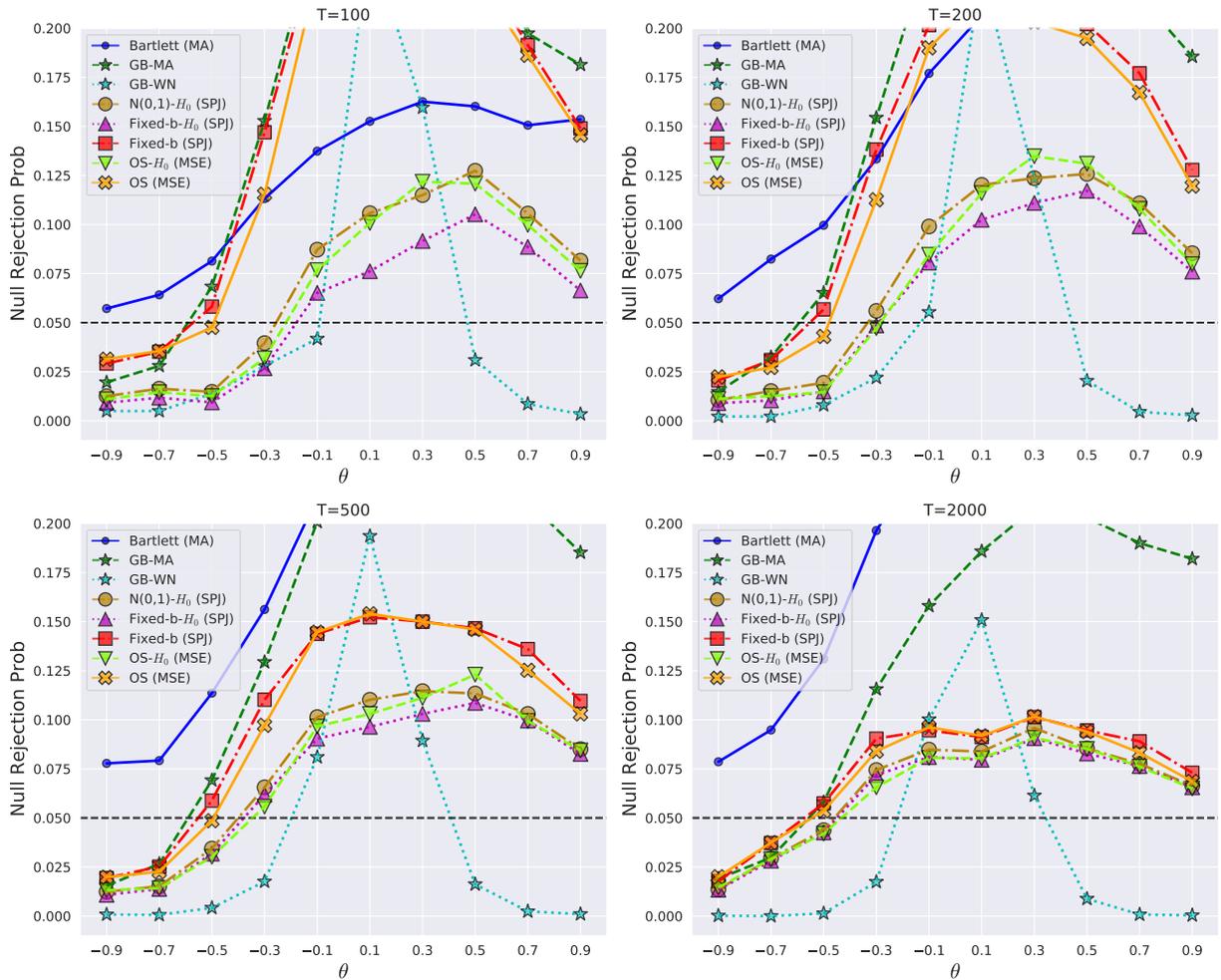
Figure 2A.7 Null rejection probabilities, $H_0 : \rho_1 = \phi$, AR-IID

DGP: AR-IID : $y_t = \phi y_{t-1} + \epsilon_t$, where $\epsilon_t \sim iidNormal(0, 1)$
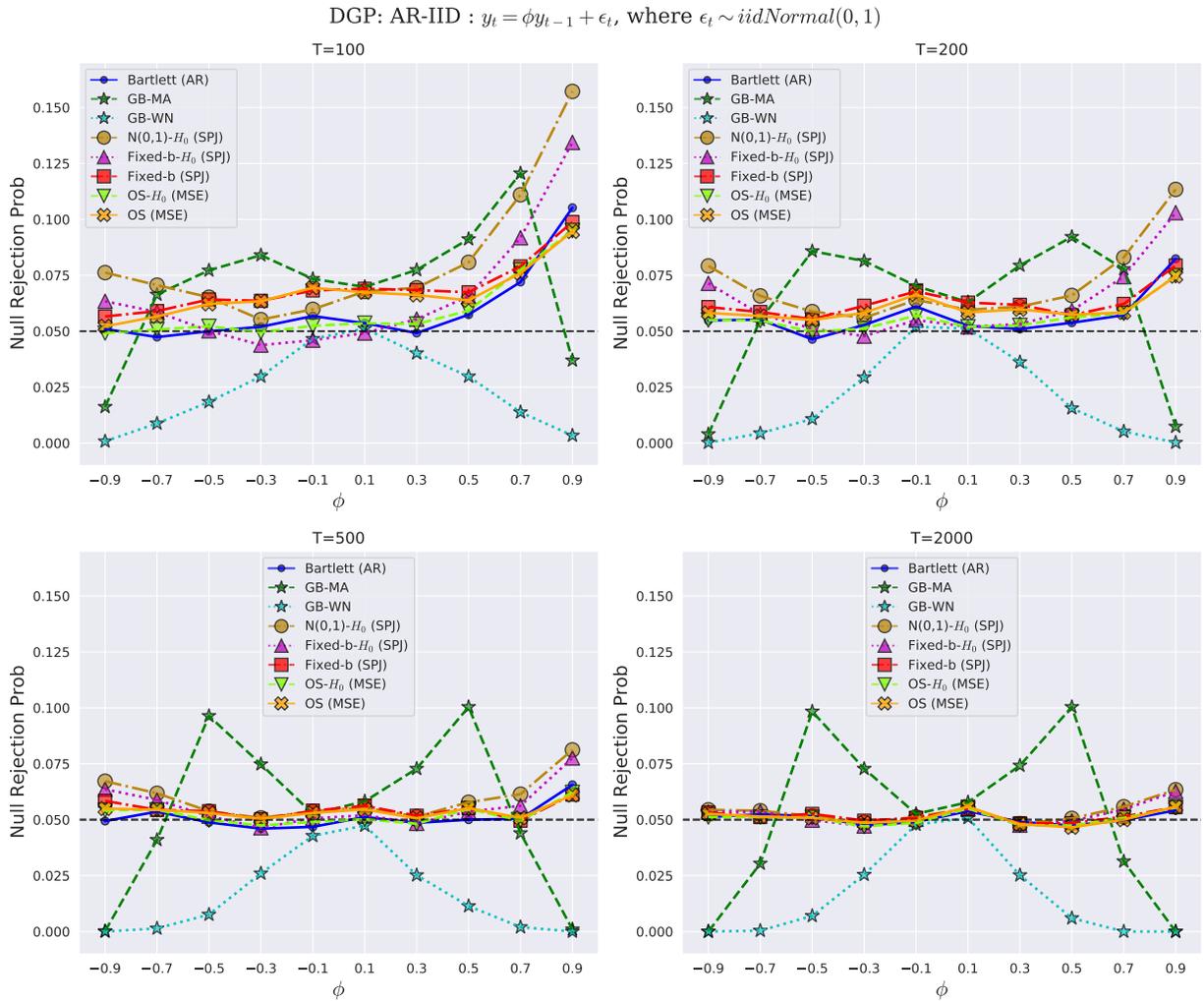
Figure 2A.8 Null rejection probabilities, $H_0 : \rho_1 = \phi$, AR-MDS

DGP: AR-MDS : $y_t = \phi y_{t-1} + \epsilon_t$, where $\epsilon_t = u_t u_{t-1}$, $u_t \sim iidNormal(0, 1)$
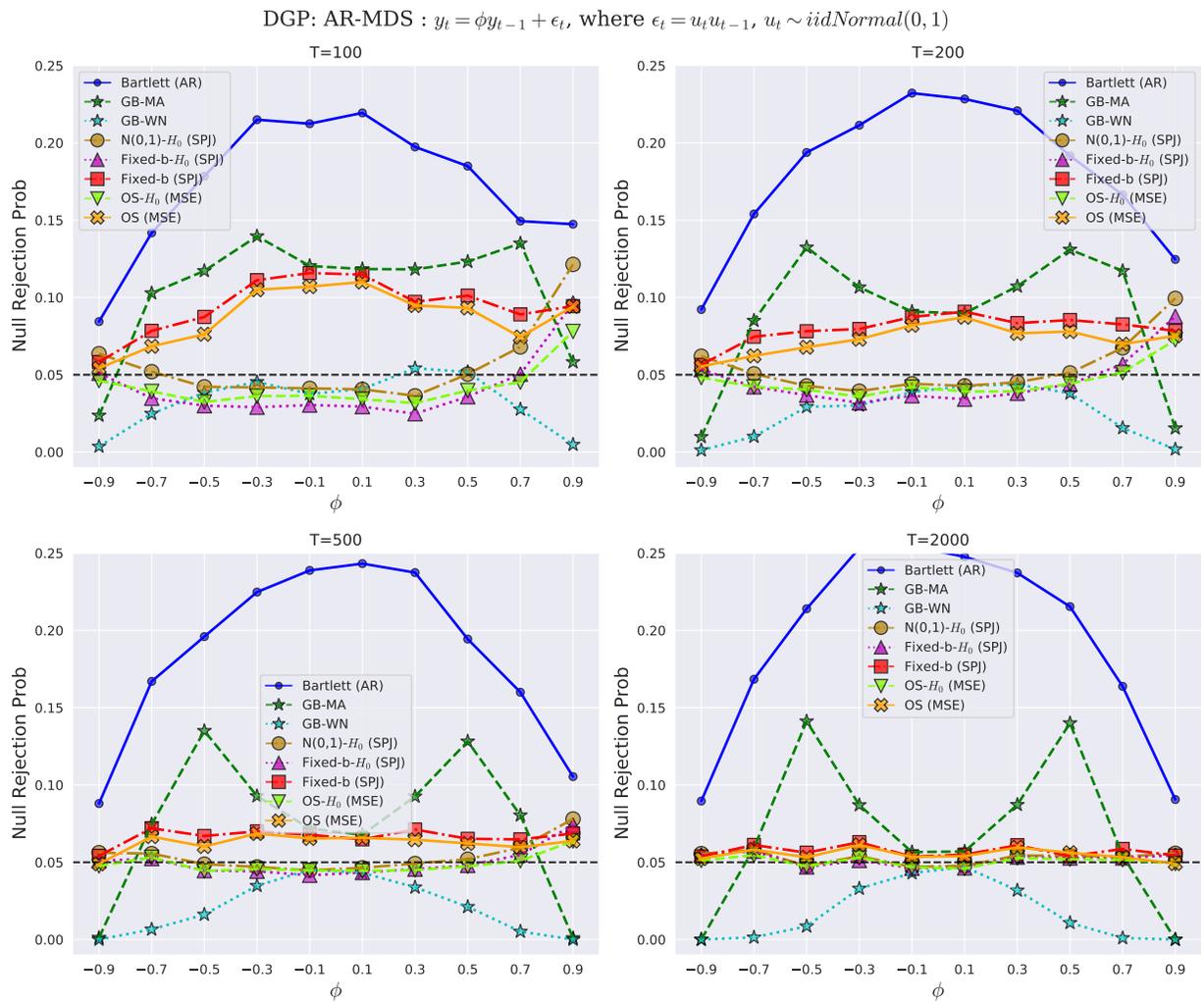
Figure 2A.9: Null rejection probabilities, $H_0 : \rho_1 = \phi$, AR-GRACH

DGP: AR-GARCH : $y_t = \phi y_{t-1} + \epsilon_t$, where $\epsilon_t = h_t u_t$ and $h_t^2 = 0.1 + 0.09\epsilon_{t-1}^2 + 0.9h_{t-1}^2$, $u_t \sim iidN(0,1)$
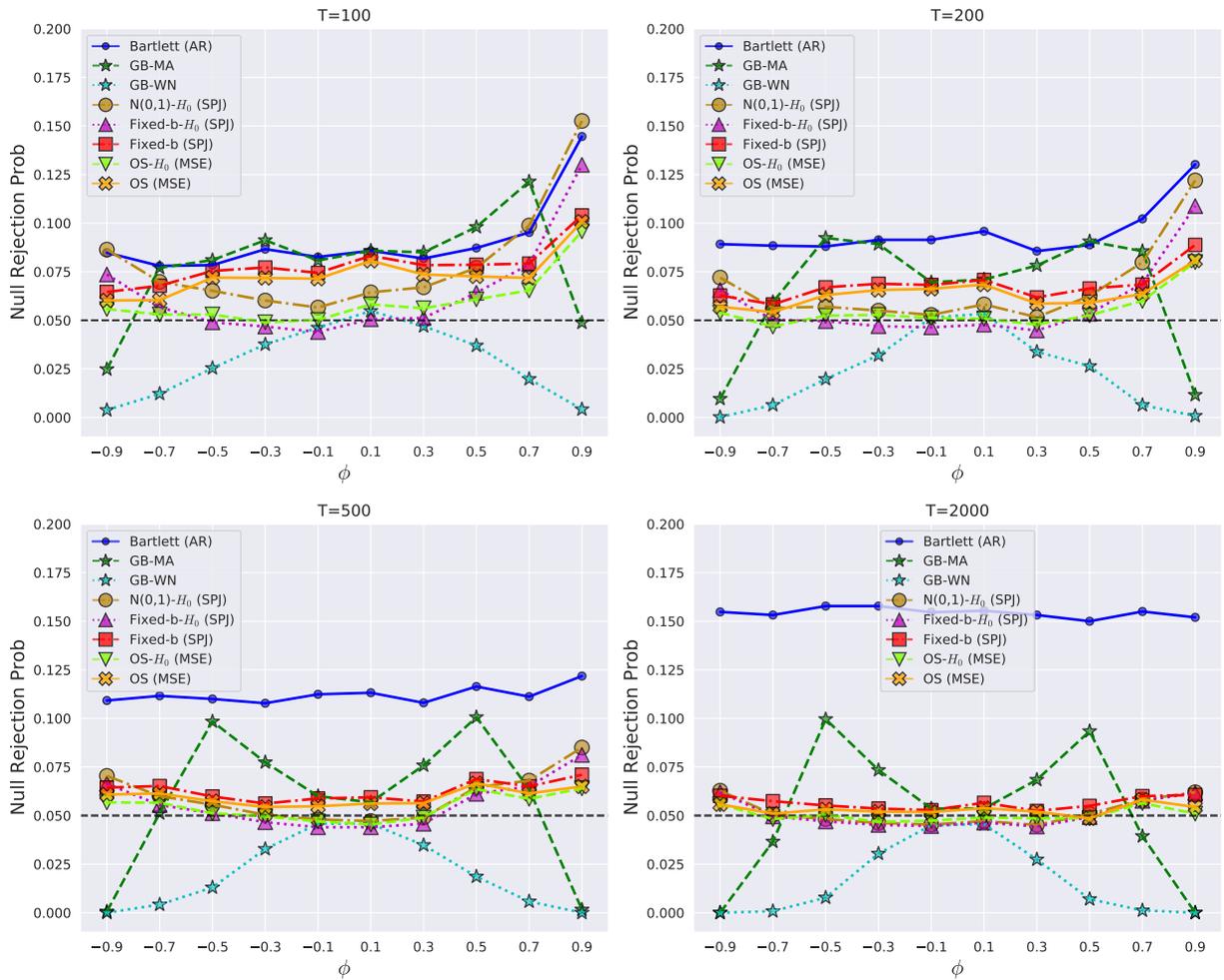
Figure 2A.10: Null rejection probabilities, $H_0 : \rho_1 = \phi$, AR-WN-1

DGP: AR-WN-1 : $y_t = \phi y_{t-1} + \epsilon_t$, where $\epsilon_t = u_t + u_{t-1}u_{t-2}$, $u_t \sim iidNormal(0,1)$
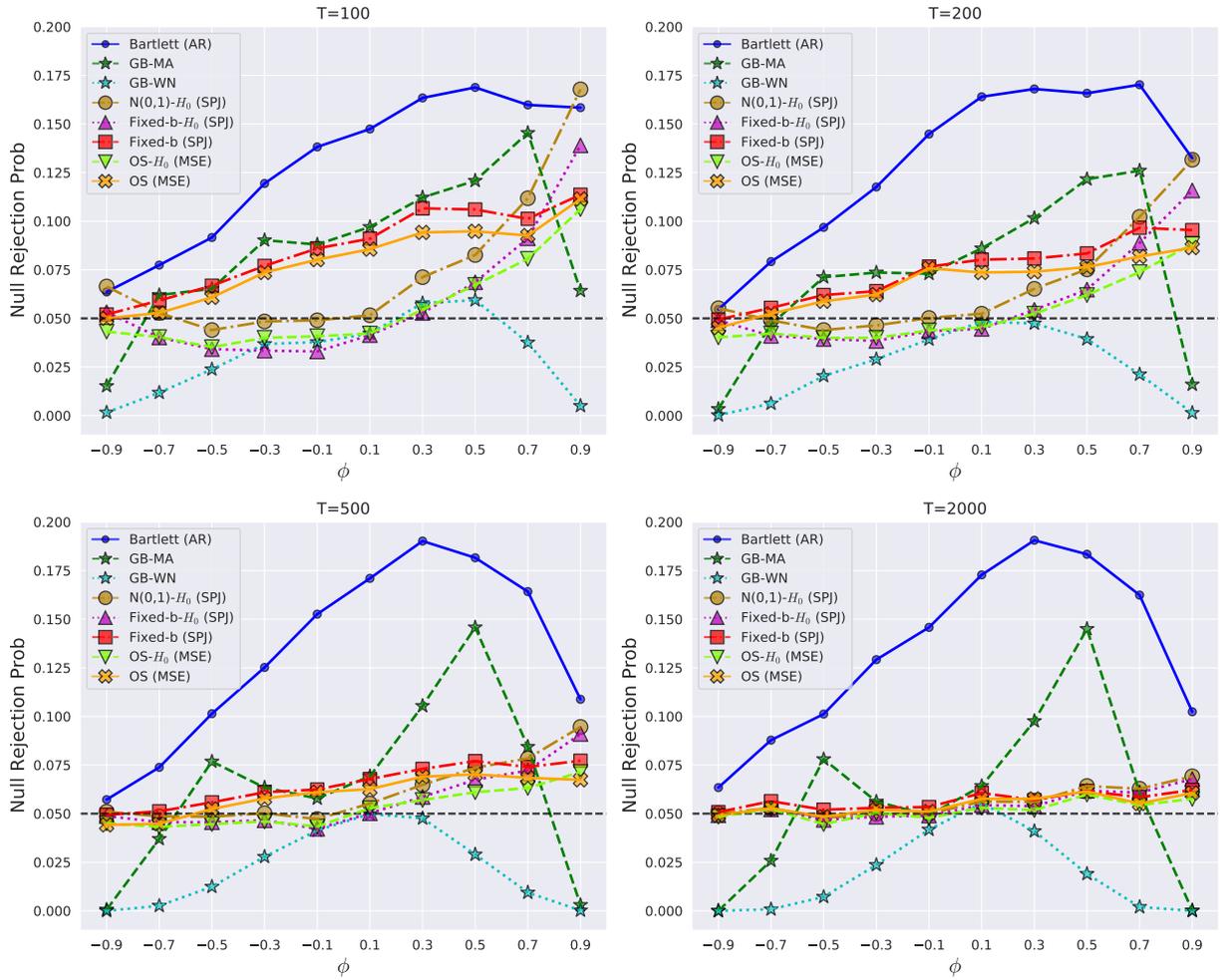
Figure 2A.11: Null rejection probabilities, $H_0 : \rho_1 = \phi$, AR-WN-Gamma

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

55
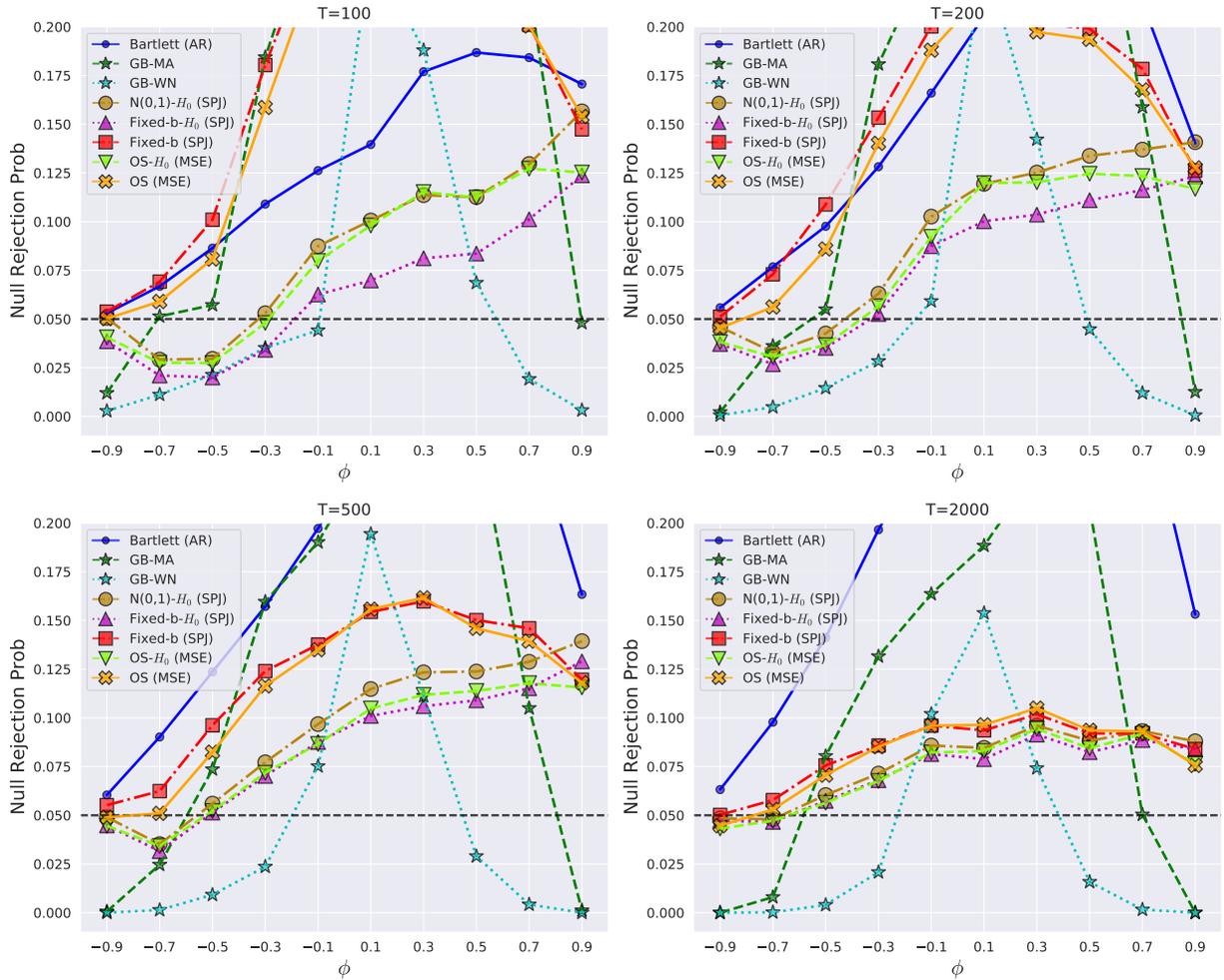
Figure 2A.12: Size adjusted power, $H_0 : \rho_1 = 0$, $H_1 : \rho_1 = \phi$, AR-IID

DGP: AR-IID : $y_t = \phi y_{t-1} + \epsilon_t$, where $\epsilon_t \sim iidNormal(0, 1)$
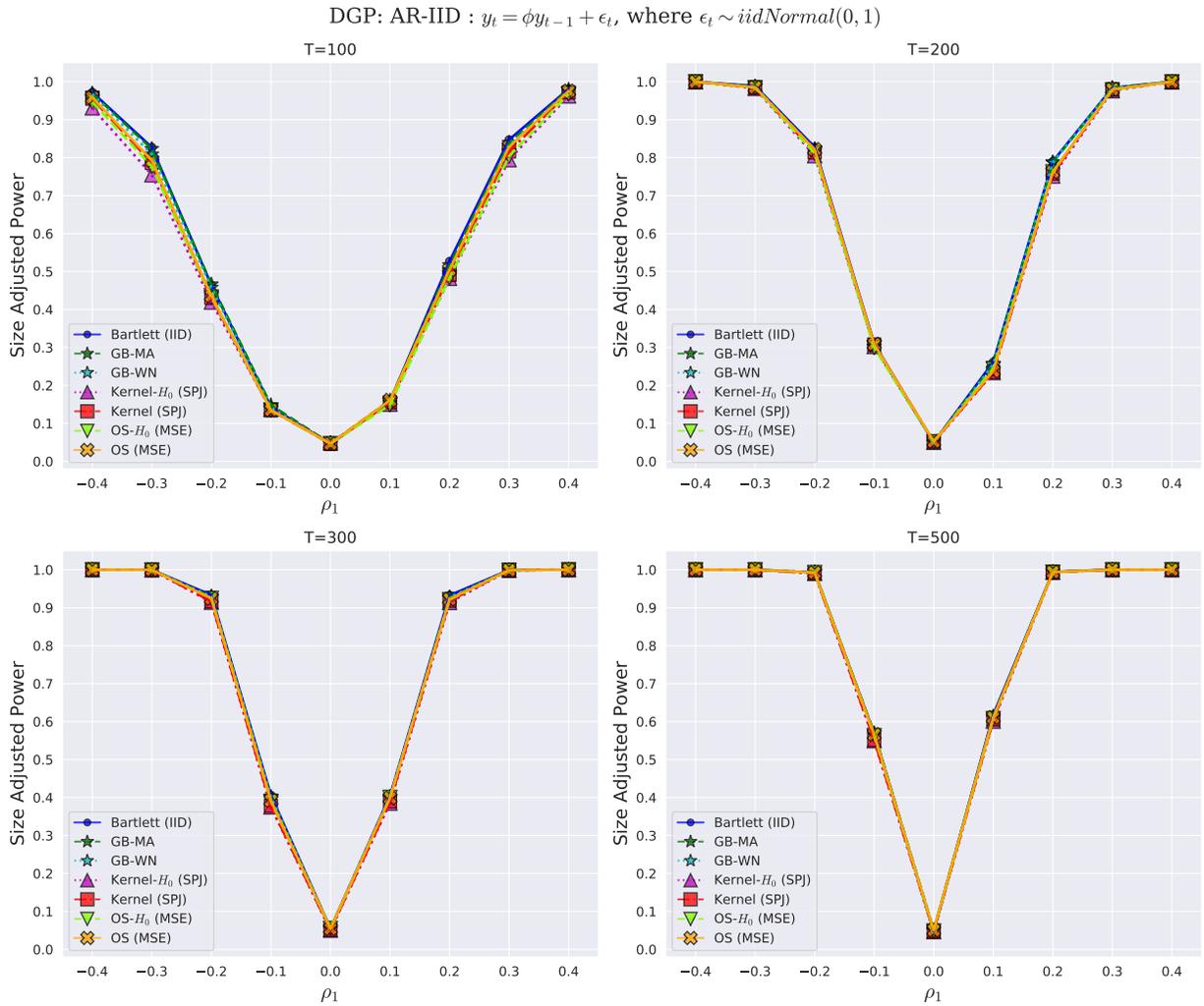
Figure 2A.13: Size adjusted power, $H_0 : \rho_1 = 0$, $H_1 : \rho_1 = \phi$, AR-WN-NLMA

DGP: AR-WN-NLMA : $y_t = \phi y_{t-1} + \epsilon_t$, where $\epsilon_t = u_{t-2} u_{t-1} (u_{t-2} + u_t + 1)$, $u_t \sim iidNormal(0, 1)$
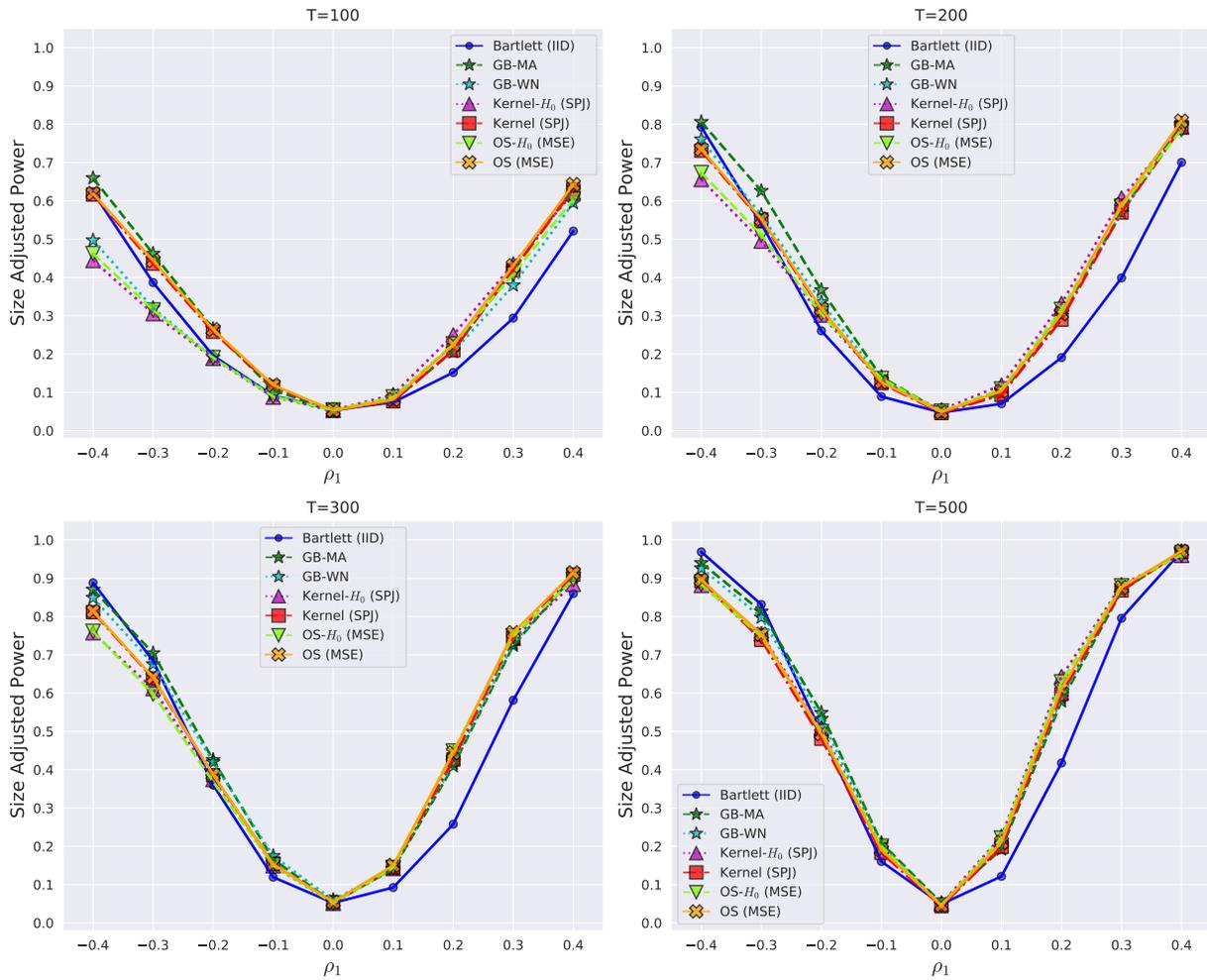
Figure 2A.14: Size adjusted power, $H_0 : \rho_1 = 0$, $H_1 : \rho_1 = \phi$, AR-WN-Gamma

DGP: AR-WN-GAM1 : $y_t = \phi y_{t-1} + \epsilon_t$, where $\epsilon_t = u_t + u_{t-1}u_{t-2}$ and $u_t = \zeta_t - \mathrm{E}[\zeta_t]$, $\zeta_t \sim iidGamma(0.3, 0.4)$
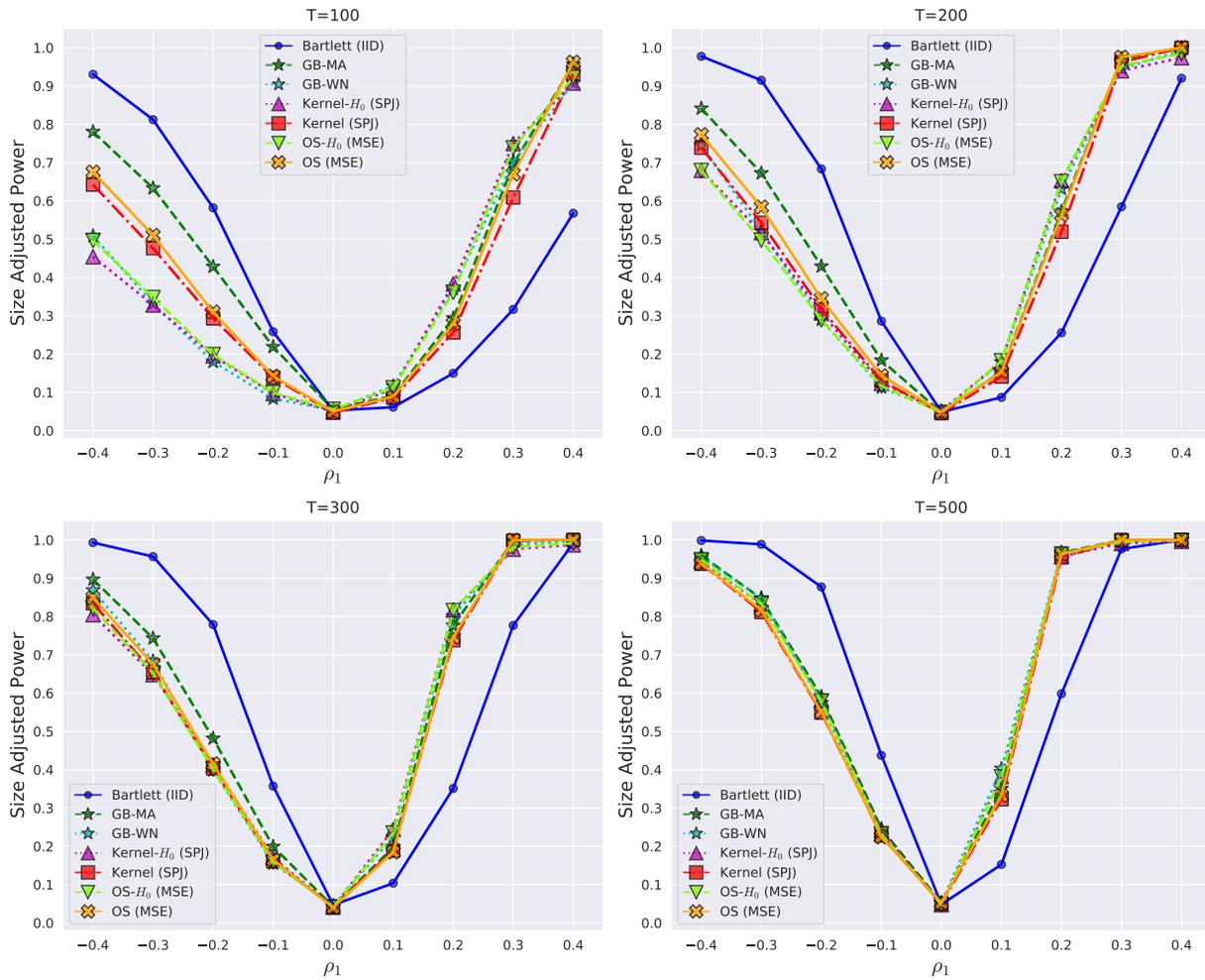
Figure 2A.15: Null rejection probabilities, $H_0 : \rho_k = 0$, IID

DGP: IID : where $\epsilon_t \sim iidNormal(0,1)$
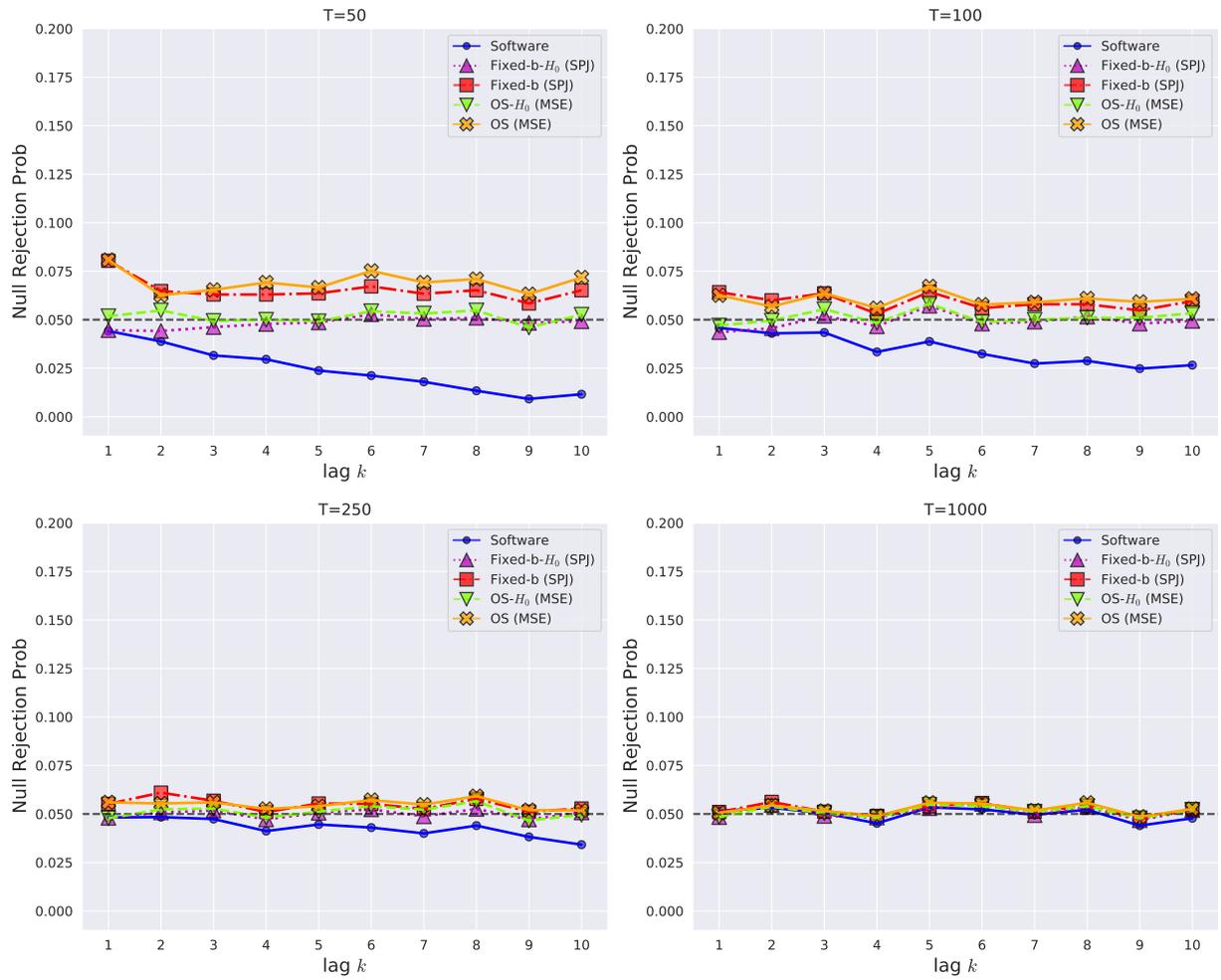
Figure 2A.16: Null rejection probabilities, $H_0 : \rho_k = 0$, MDS

DGP: MDS : where $\epsilon_t = u_t u_{t-1}$, $u_t \sim iidNormal(0,1)$
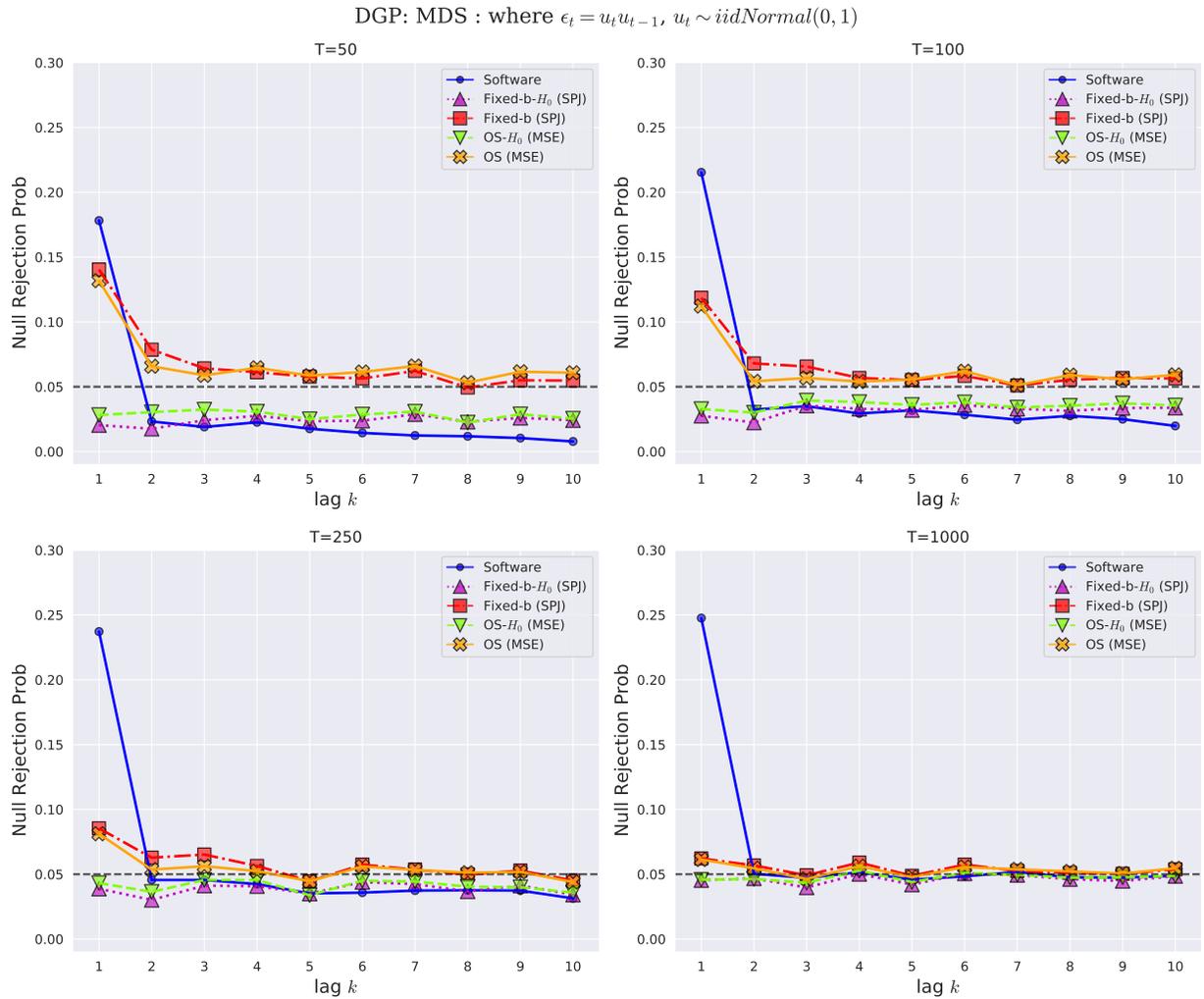
Figure 2A.17: Null rejection probabilities, $H_0 : \rho_k = 0$, GRACH

DGP: GARCH : where $\epsilon_t = h_t u_t$ and $h_t^2 = 0.1 + 0.09\epsilon_{t-1}^2 + 0.9h_{t-1}^2$, $u_t \sim iidN(0,1)$
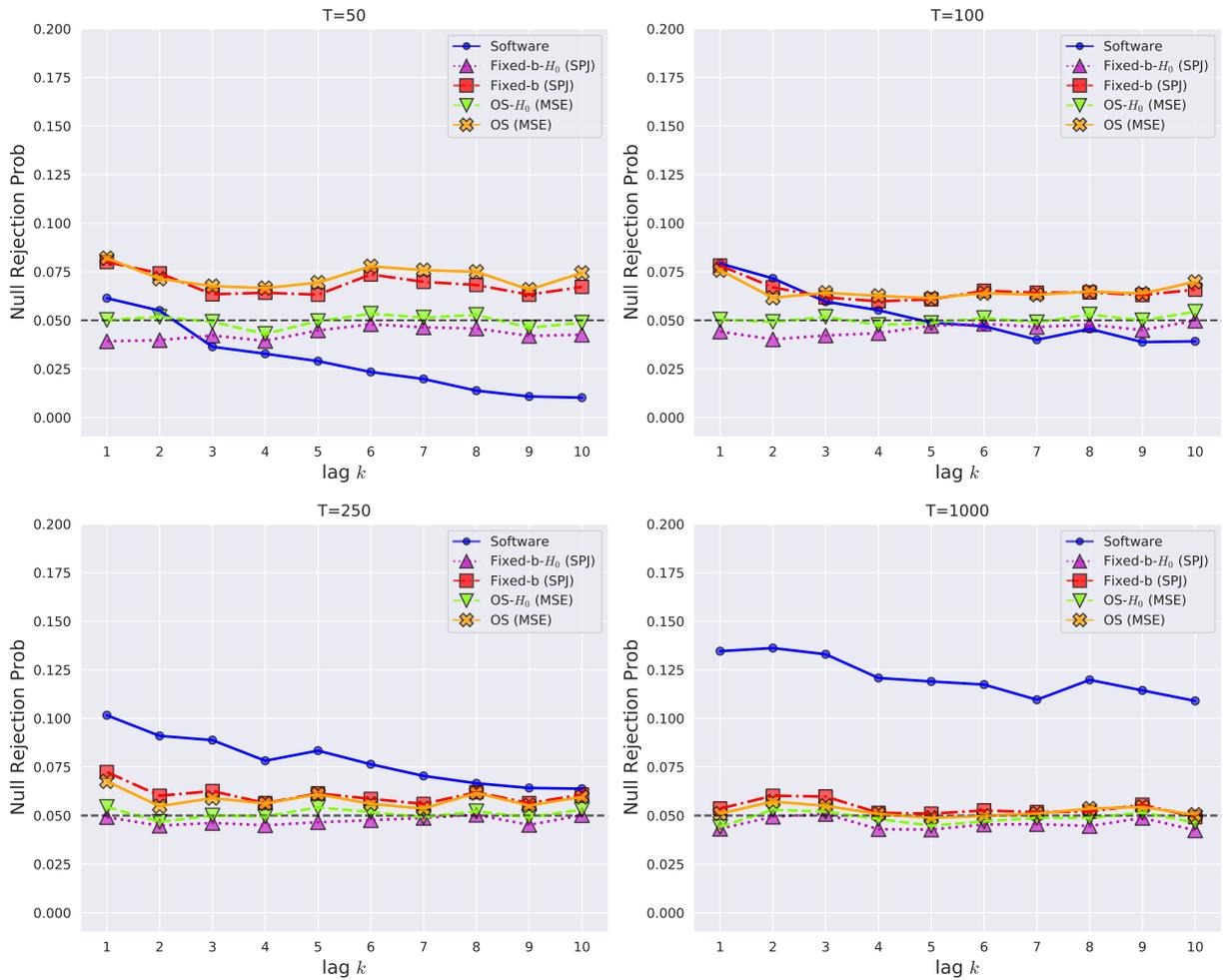
Figure 2A.18: Null rejection probabilities, $H_0 : \rho_k = 0$, WN-1

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

62
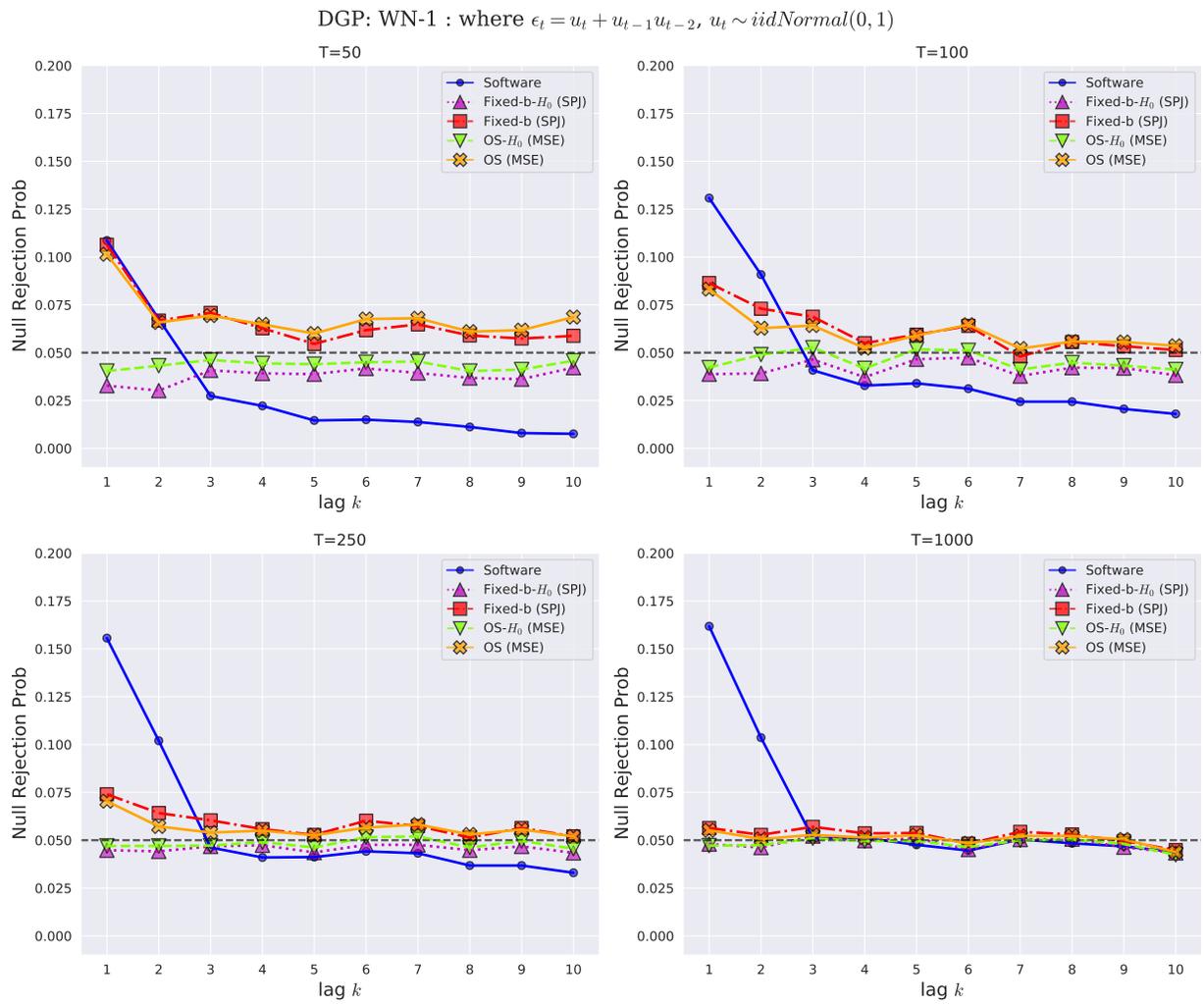
Figure 2A.19: Null rejection probabilities, $H_0 : \rho_k = 0$, WN-Gamma

DGP: WN-GAM1 : where $\epsilon_t = u_t + u_{t-1}u_{t-2}$ and $u_t = \zeta_t - \mathrm{E}[\zeta_t]$, $\zeta_t \sim iidGamma(0.3, 0.4)$
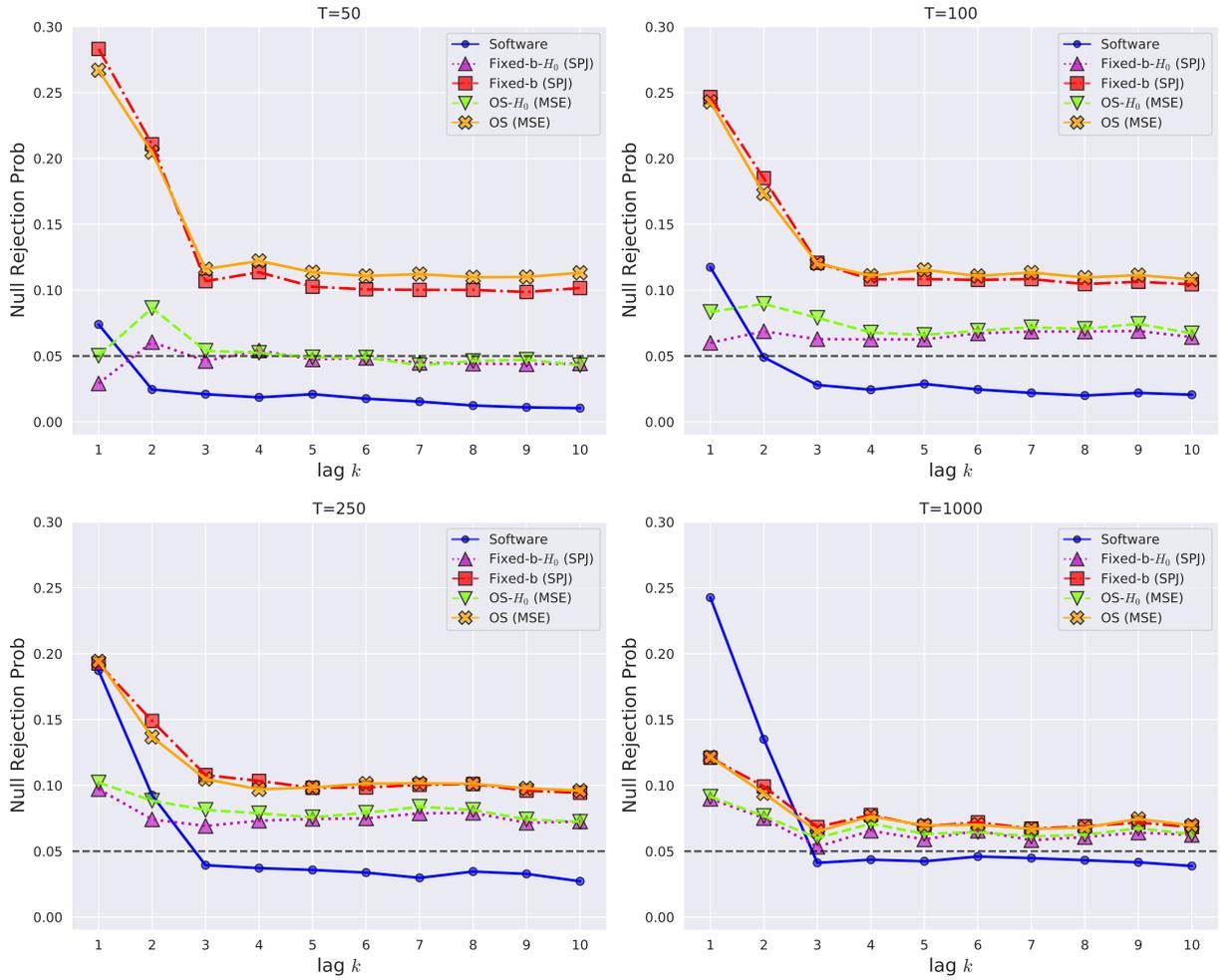
Figure 2A.20: Null rejection probabilities, $H_0 : \rho_k = \phi^k$, $\phi = 0.5$, AR-IID

DGP: AR-IID : $y_t = \phi y_{t-1} + \epsilon_t$, where $\epsilon_t \sim iidNormal(0,1)$
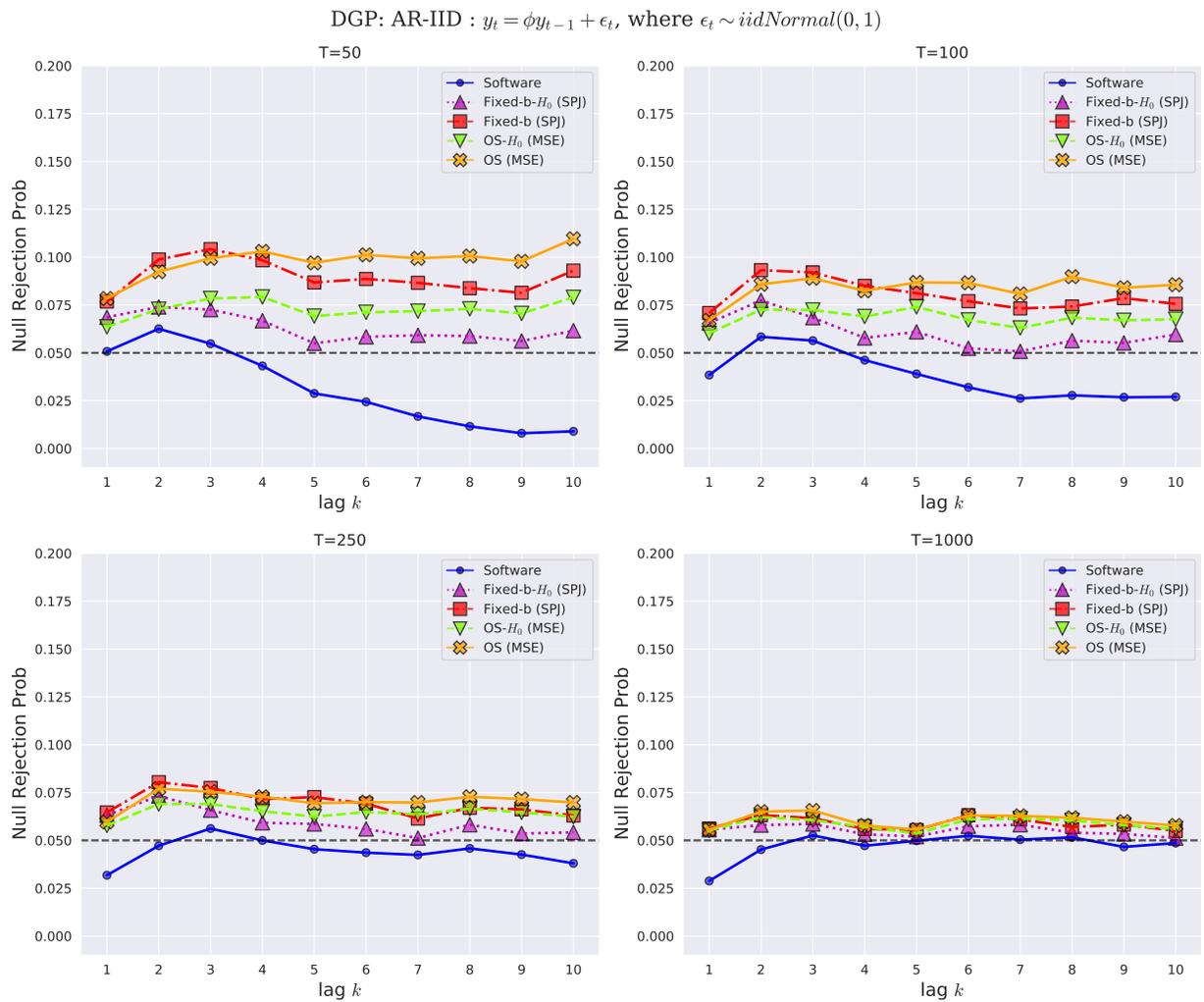
Figure 2A.21: Null rejection probabilities, $H_0 : \rho_k = \phi^k$, $\phi = 0.5$, AR-MDS

DGP: AR-MDS : $y_t = \phi y_{t-1} + \epsilon_t$, where $\epsilon_t = u_t u_{t-1}$, $u_t \sim iidNormal(0, 1)$
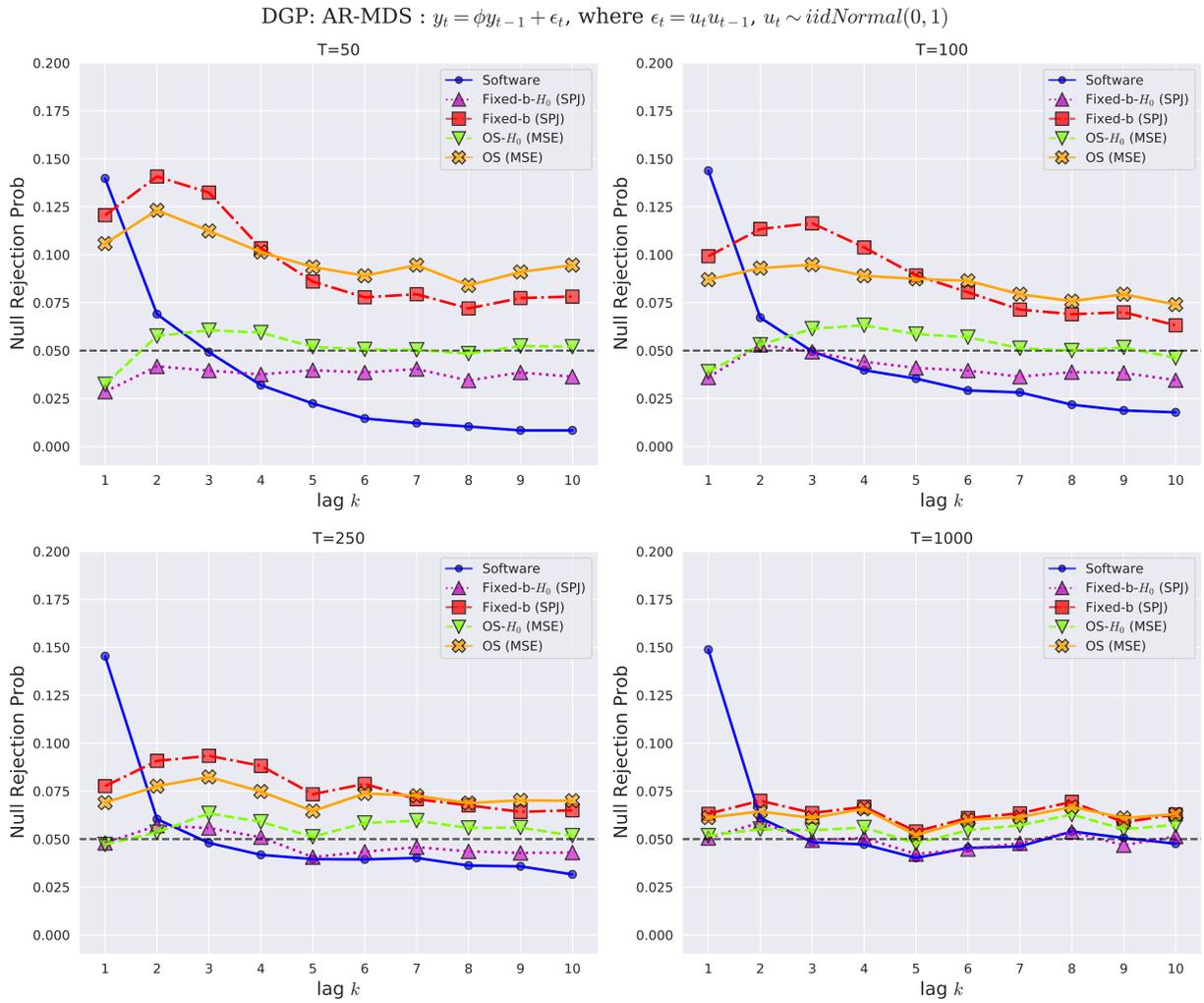
Figure 2A.22: Null rejection probabilities, $H_0 : \rho_k = \phi^k$, $\phi = 0.5$, AR-GRACH

DGP: AR-GARCH : $y_t = \phi y_{t-1} + \epsilon_t$, where $\epsilon_t = h_t u_t$ and $h_t^2 = 0.1 + 0.09\epsilon_{t-1}^2 + 0.9h_{t-1}^2$, $u_t \sim iidN(0,1)$
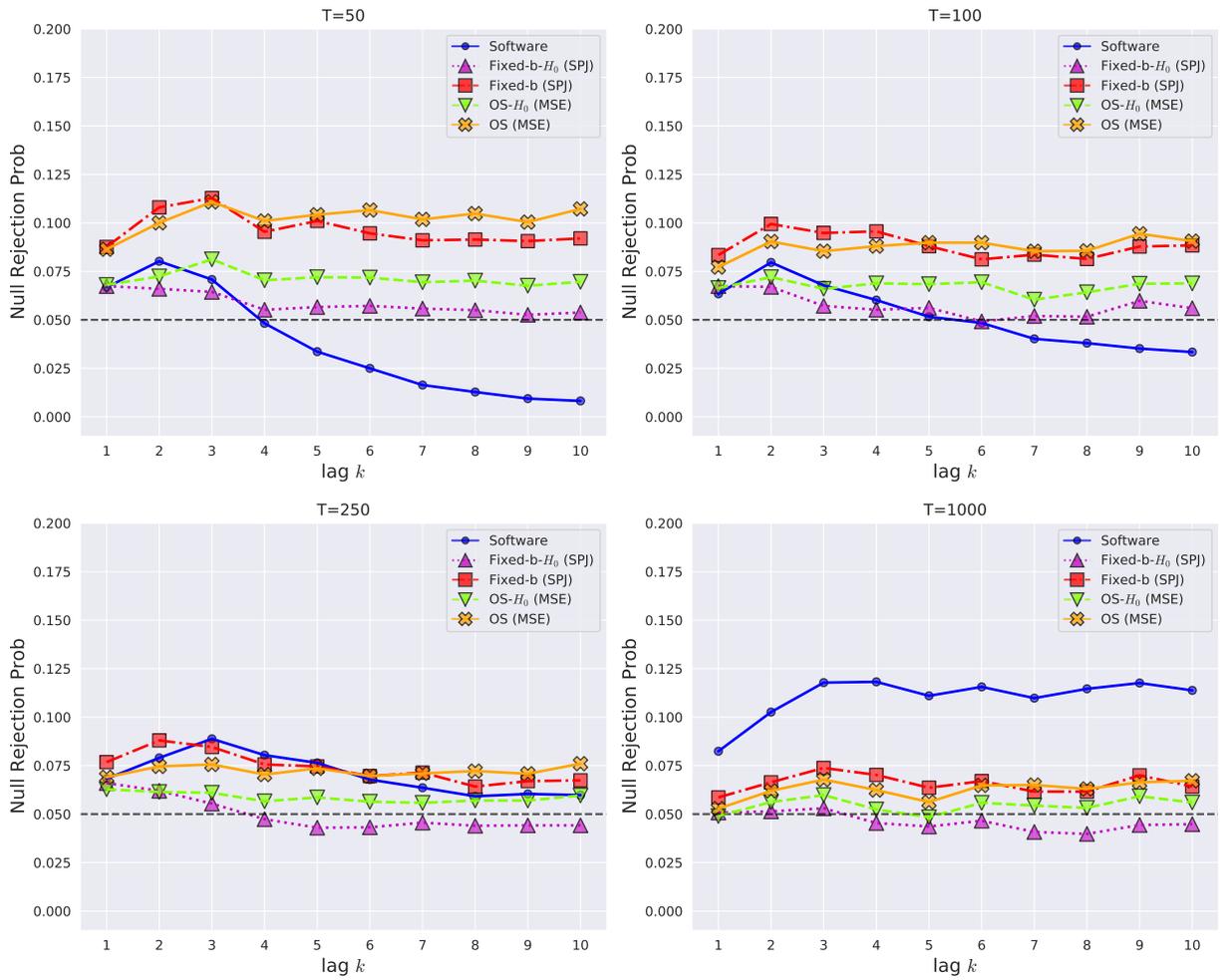
Figure 2A.23: Null rejection probabilities, $H_0 : \rho_k = \phi^k$, $\phi = 0.5$, AR-WN-1

DGP: AR-WN-1 : $y_t = \phi y_{t-1} + \epsilon_t$, where $\epsilon_t = u_t + u_{t-1} u_{t-2}$, $u_t \sim iidNormal(0,1)$
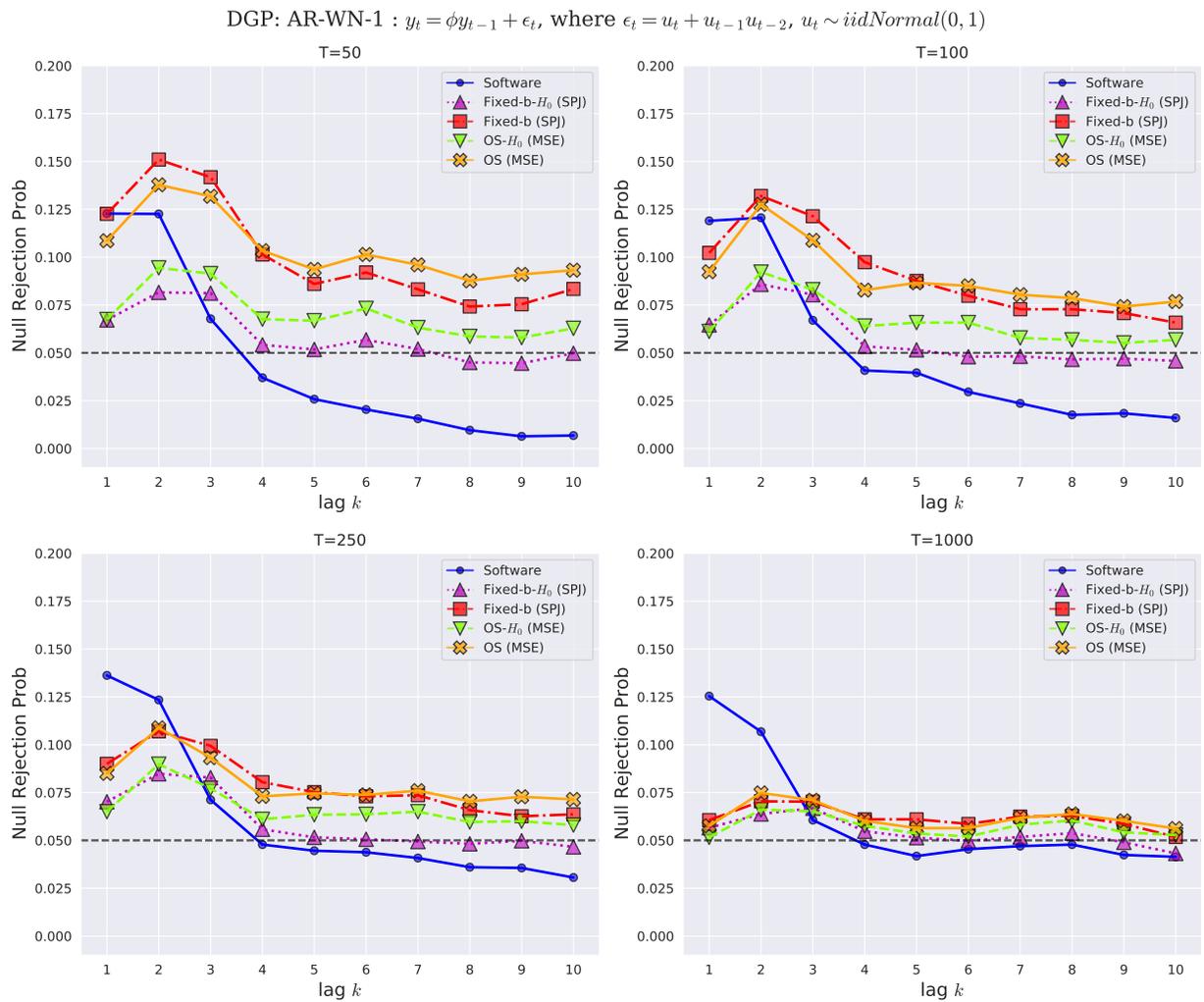
Figure 2A.24: Null rejection probabilities, $H_0 : \rho_k = \phi^k$, $\phi = 0.5$, AR-WN-Gamma

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

68
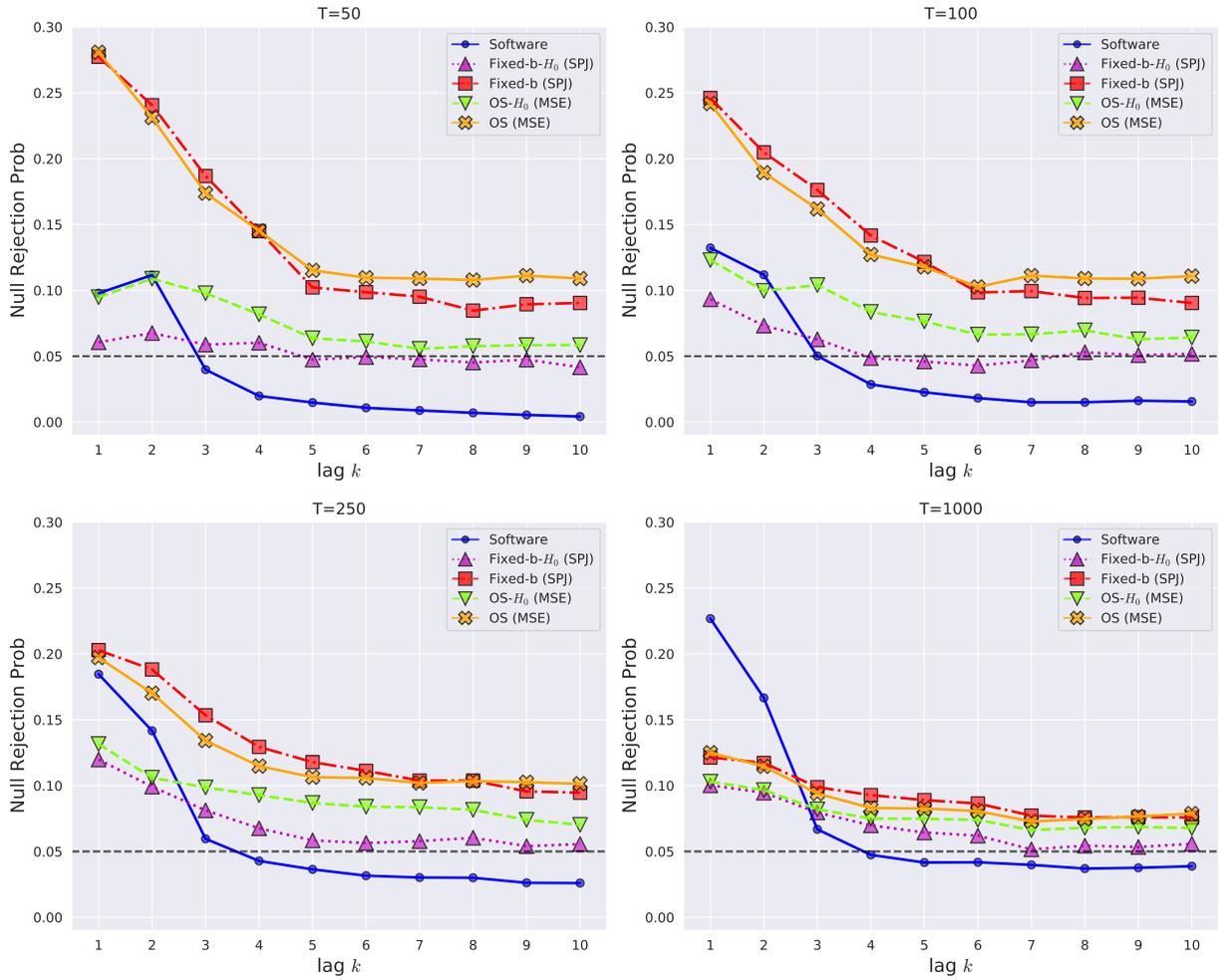
Figure 2A.25: Graphs of S&P 500 index daily returns and absolute returns
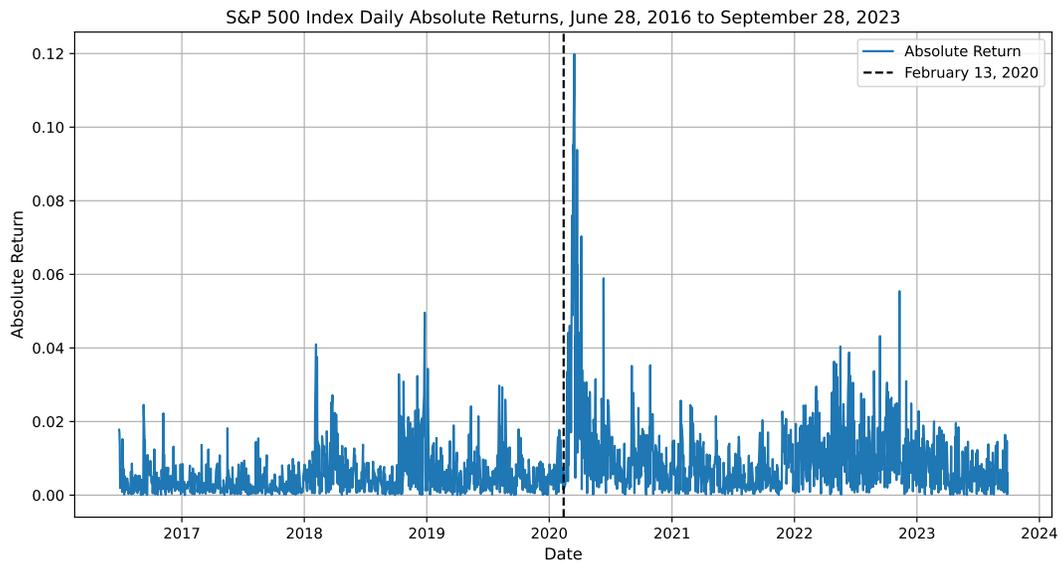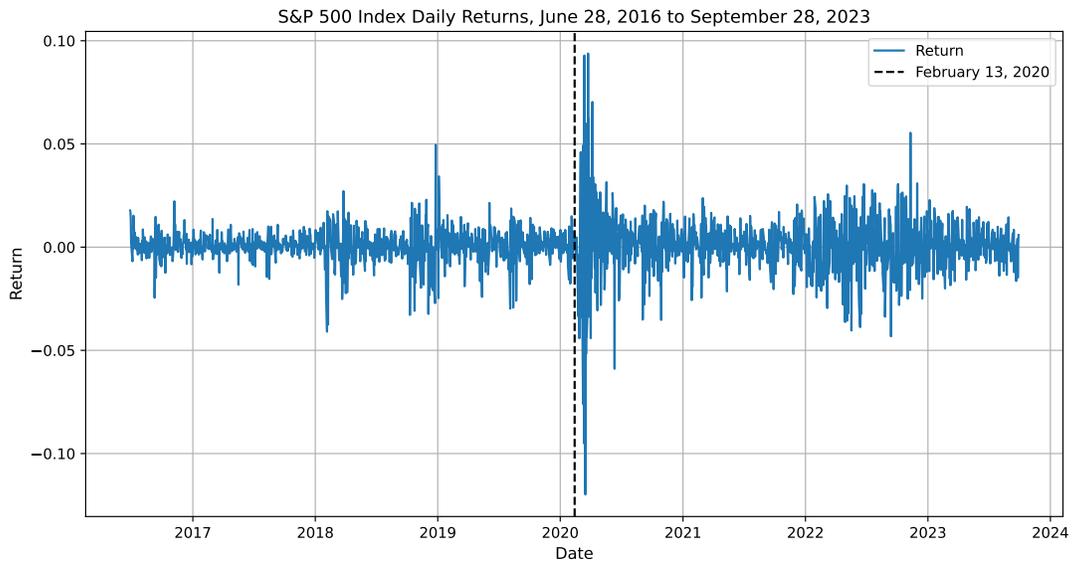
Figure 2A.26: Estimated autocorrelations for S&P 500 index returns during pre- and post-Covid
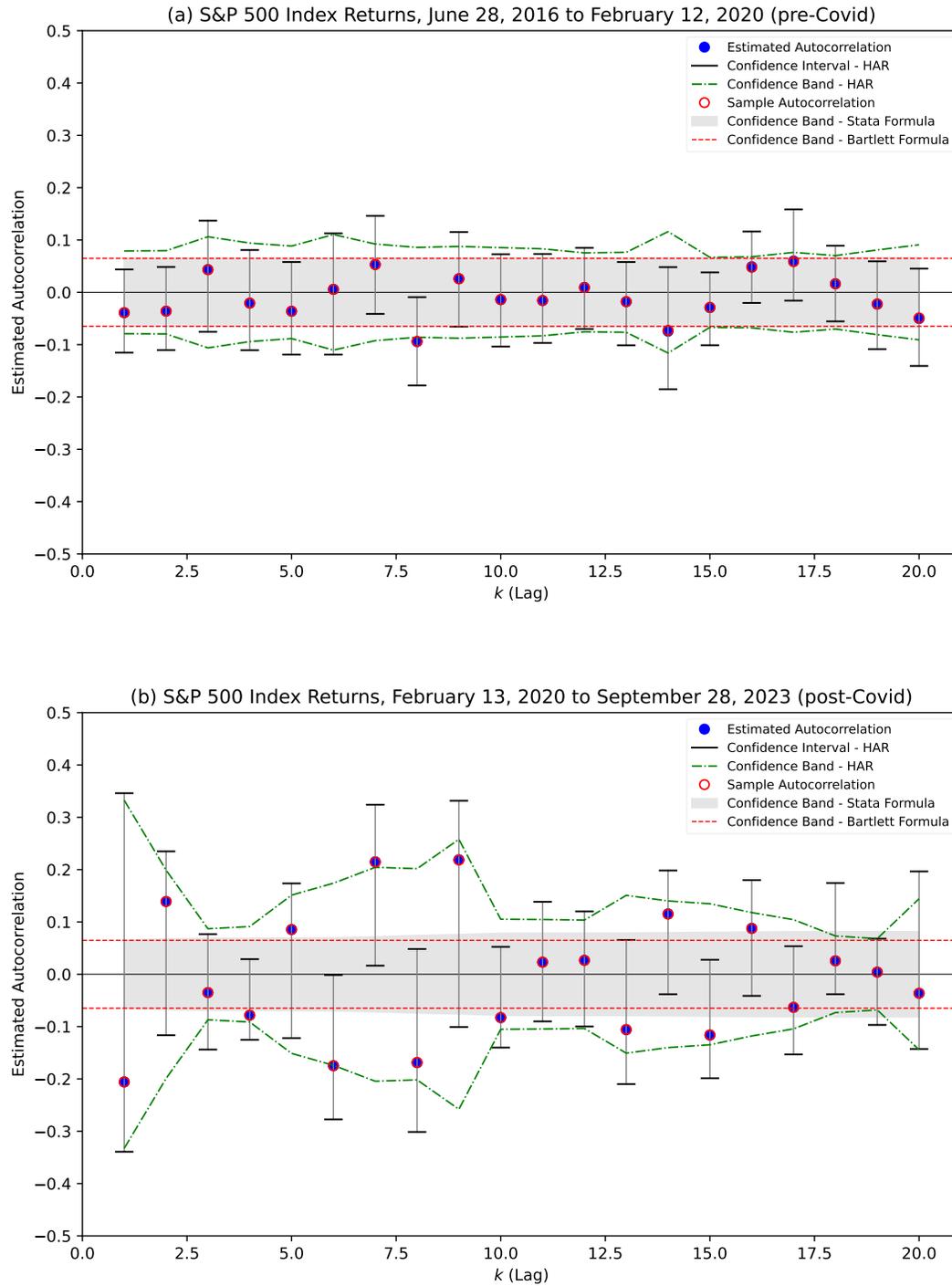
Figure 2A.27: Estimated autocorrelations for S&P 500 index absolute returns during pre- and post-Covid

(a) S&P 500 Index Absolute Returns, June 28, 2016 to February 12, 2020 (pre-Covid)

(b) S&P 500 Index Absolute Returns, February 13, 2020 to September 28, 2023 (post-Covid)
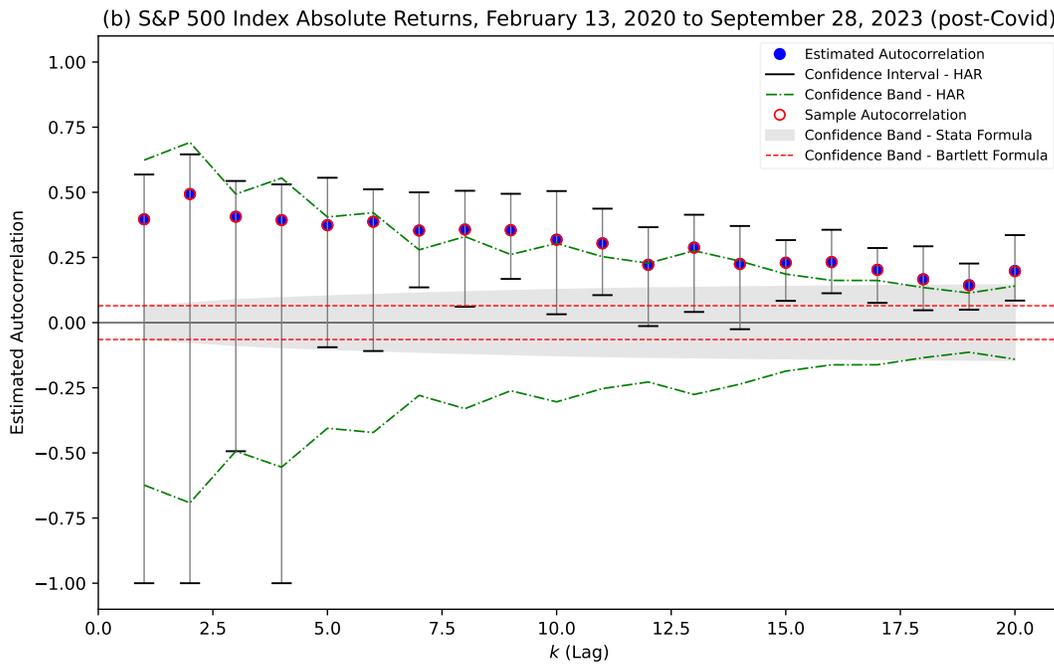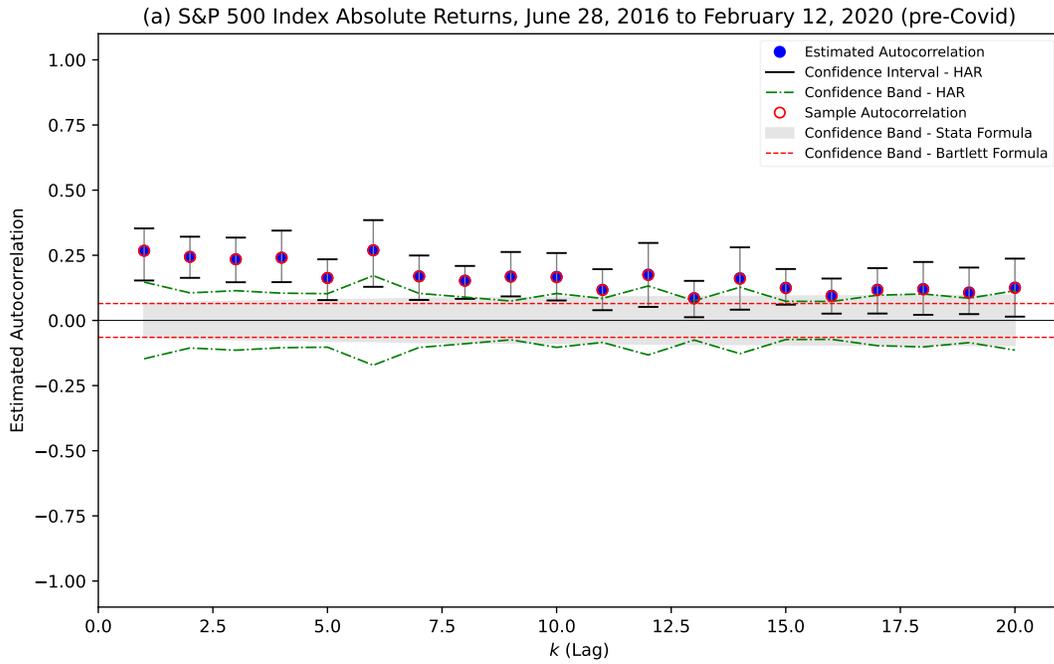
CHAPTER 3

SOME FIXED-*b* RESULTS FOR REGRESSIONS WITH HIGH FREQUENCY DATA OVER LONG SPANS
(CO-AUTHORED WITH TIM VOGELSANG)

3.1 Introduction

This paper develops fixed-*b* asymptotic results for heteroskedasticity autocorrelation robust (HAR) Wald statistics for regressions with high frequency data in a continuous time framework. Our results are obtained within the theoretical framework developed by Chang et al. (2023) (hereafter CLP). Our results complement and extend the analysis in CLP. Our results are related to, and complement, recent work by Pellatt and Sun (2023) who focus on orthonormal series estimators of long run variances and develop fixed-smoothing asymptotic results for corresponding HAR statistics.

Motivated by high frequency data, CLP investigate the asymptotic properties of HAR Wald tests in a regression model where the observed discrete time series data is generated by an underlying continuous time model. Focusing on consistency/inconsistency of kernel based long run variance estimators, CLP show that HAR Wald statistics can diverge to infinity under some high frequency conditions, but this spuriousness can disappear when using data-dependent bandwidth selection methods compatible to high frequency data. In particular, CLP conclude that the Andrews (1991) data dependent approach works more reliably with high frequency data than the Newey and West (1994) data dependent approach. While suggestive of finite sample properties, consistency/inconsistency of a long-run variance estimator only partially reflects the impact of the bandwidth/kernel on the sampling distribution of the HAR test statistic. In contrast, the fixed-*b* approach of Kiefer and Vogelsang (2005) more fully captures the impact of the bandwidth/kernel on the first order asymptotic distribution of the HAR test statistic.

In this paper we obtain fixed-b asymptotic results for the statistics analyzed by CLP using the same continuous time framework. We find that the fixed-b limits of the HAR Wald tests in stationary high frequency regressions estimated by ordinary least squares (OLS) are the same as the standard fixed-b limits in Kiefer and Vogelsang (2005). For cointegrating high frequency regressions the fixed-b limits generally have non-pivotal limits. However, for the special case where the stochastic processes in the continuous time regression follow Brownian motions and the regressors are independent of the errors, the fixed-b limits are pivotal and are the same as those obtained by Bunzel (2006) in discrete time settings. For the case of cointegration with endogeneity, we analyze the integrated modified OLS (IM-OLS) estimator of Vogelsang and Wagner (2014) and an associated test that is asymptotically pivotal under fixed-b asymptotics. We find that the fixed-b limit in the CLP high frequency setting is the same as that obtained by Vogelsang and Wagner (2014). Using the language of CLP, we can say that fixed-b critical values are high frequency compatible.

When fixed-b limits are pivotal with respect to serial correlation nuisance parameters but depend on the bandwidth and kernel, the use of fixed-b critical values rather than chi-square critical values is expected to improve inference regardless of the method used to obtain the bandwidth. See, for example, Kiefer and Vogelsang (2005) and Lazarus et al. (2018) for simulation evidence and Gonçalves and Vogelsang (2011), Lazarus et al. (2021), and Sun et al. (2008) for theoretical and simulation evidence.

We assess the performance of fixed-b critical values using a simulation study using the same data generating process (DGP) as CLP. Consistent with the existing fixed-b literature, we find the use of fixed-b critical values systematically performs better than chi-square critical values regardless of the method used to choose the bandwidth. We extend the simulation results of CLP by reporting results for additional persistence parameters for the Ornstein-Uhlenbeck process (OU process) used to generate the data. As in CLP we compare/contrast the Andrews (1991) and Newey and West (1994) data dependent methods. We also include the data dependent method proposed by Sun et al. (2008) where the bandwidth choice minimizes a weighted average of type I and type II errors.

73

Similar to CLP we find that the Andrews (1991) bandwidth performs reliably with respect to the frequency of observations especially when fixed-b critical values are used. The performance of the Newey and West (1994) bandwidth depends critically on the choice of pre-tuning parameters with some choices leading to severe over-rejections with high frequency data while other choices leading to better, but not fully satisfactory, performance. The Sun et al. (2008) bandwidth performance is similar to the Andrews (1991) bandwidth with less over-rejections when fixed-b critical values are used. The continuous rule of thumb (CRT) bandwidth rule suggested by CLP tends to over-reject more substantially than the Andrews (1991) and Sun et al. (2008) especially when the data has strong persistence.

While the Andrews (1991), Sun et al. (2008) and CRT bandwidths tend to perform well at high frequencies (assuming the persistence in the data is not too strong relative to the span), we find that null rejections are remarkably stable across sampling frequencies ranging from high to low. This stability holds for data with strong and mild persistence. We show that the source of this stability in null rejections is stability in bandwidth sample size ratios (b-values) across sampling frequencies. This stability holds by construction for the CRT bandwidth. It is more surprising this stability holds for the Andrews (1991) and Sun et al. (2008) bandwidths, and we provide a simple theoretical explanation.

We also report some power results in our simulations and find that for persistent series, power is stable across sampling frequencies. In contrast, power falls as sampling frequency decreases for mildly persistent series. Therefore, we can recommend that practitioners use data sampled at higher frequencies.

It is important to compare and contrast our analysis with Pellatt and Sun (2023). Both analyses obtain fixed-smoothing results for HAR tests in high frequency settings and provide reference distributions for critical values that improve finite sample inference relative to using standard critical values. In the stationary case we focus on kernel based tests whereas Pellatt and Sun (2023) focus on orthonormal series based tests. In cointegration settings we obtain results for kernel based tests for both OLS and IM-OLS estimators whereas Pellatt and Sun (2023) obtain results

for orthonormal series based tests for estimators based on the orthonormal series transformation proposed by Hwang and Sun (2018). Because we focus on kernel long run variance estimators, we are able to provide some useful and interesting results on bandwidth choice that further refine CLP's results on high-frequency compatible bandwidths. Taken together, our analysis and that of Pellatt and Sun (2023) provide a useful set of fixed-smoothing results for inference in the high frequency regression setting of CLP. That we and Pellatt and Sun (2023) find that methods originally proposed for discrete settings can be applied in the CLP high frequency setting in exactly the same way with existing fixed-smoothing reference distributions is a positive contribution for empirical practice and allows empirical researchers to use high or low frequency data for HAR inference with one set of methods.

The rest of the paper is organized as follows. In section 3.2, the model is given, and the continuous time framework of CLP is described. Section 3.3 reviews standard fixed-b asymptotic theory for HAR tests in stationary regressions and then provides fixed-b results for high frequency asymptotics using the continuous time framework of CLP. Results are provided for kernel based tests using OLS and, in the case of cointegration, IM-OLS. Because of the nonstandard form of fixed-b asymptotic distributions, Section 3.4 describes numerical methods based on simulations that are used to obtain critical value functions used for the finite sample simulations. Section 3.5 provides some finite sample simulation results. Section 3.6 has an illustrative empirical application for simple regressions used to test the uncovered interest parity condition. Section 3.7 gives some concluding remarks. Proofs are given in the Appendix.

3.2 Model

We focus on the model and setup used by CLP. Consider the continuous time regression model

$$Y_t = X_t'\beta + U_t, \tag{3.1}$$

where $0 \le t \le T$, T is the span (e.g. number of years), β is a $k \times 1$ vector of parameters, X_t is a $k \times 1$ vector of continuous time processes, and U_t is a scalar continuous time process.

Following CLP the continuous time model can be discretized as follows. Suppose data is sampled at discrete time periods with δ denoting the time interval between discrete observations.

Letting $i = 1, 2, \ldots, n$ index the discrete observations, the link between the discrete observations and the underlying continuous time processes is given by

$$y_i = Y_{i\delta} \text{ and } x_i = X_{i\delta},$$

where $i\delta$ is the time, t, at which discrete observation, i, is observed. Because $t \in [0, T]$, it follows that $n\delta = T$. Thus, y_i and x_i are discrete sample paths observed at time intervals δ from X_t and Y_t respectively. The sampling frequency is inversely related to δ.

We can write the discrete time regression analogous to (3.1) using x_i and y_i as

$$y_i = x_i'\beta + u_i. \tag{3.2}$$

where $u_i = U_{i\delta}$. Suppose (3.2) is estimated by ordinary least squares (OLS):

$$\widehat{\beta} = \left(\sum_{i=1}^{n} x_i x_i' \right)^{-1} \sum_{i=1}^{n} x_i y_i,$$

and we are interested in testing linear hypotheses about β of the form

$$H_0 : R\beta = r, \quad H_1 : R\beta \neq r,$$

where R is a known $q \times k$ matrix with rank q and r is a known $q \times 1$ vector. Following CLP we focus on two heteroskedasticity autocorrelation robust (HAR) Wald statistics. The first Wald statistic is appropriate for certain cointegration regressions and is given by

$$G(\widehat{\beta}) = \left(R\widehat{\beta} - r \right)' \left[\widehat{\omega}_n^2 R \left(\sum_{i=1}^{n} x_i x_i' \right)^{-1} R' \right]^{-1} \left(R\widehat{\beta} - r \right),$$

where $\widehat{\omega}_n^2$ is an estimator of the long run variance of u_i. The second Wald statistic is appropriate for stationary regressions and is given by

$$H(\widehat{\beta}) = (R\widehat{\beta} - r)' \left[R \left(\sum_{i=1}^{n} x_i x_i' \right)^{-1} n\widehat{\Omega}_n \left(\sum_{i=1}^{n} x_i x_i' \right)^{-1} R' \right]^{-1} (R\widehat{\beta} - r),$$

where $\widehat{\Omega}_n$ is an estimator of long run variance of $x_i u_i$. For the case of data sampled at a given frequency, the asymptotic properties (as $T, n \to \infty$, δ fixed) of these Wald statistics are well studied

76

in the literature. A key contribution of CLP is the analysis of these Wald statistics when $\widehat{\omega}_n^2$ and $\widehat{\Omega}_n$ are kernel long run variance estimators where the time interval between observations shrinks with the sample size, i.e. $\delta \to 0$ as $T, n \to \infty$. In this "high frequency" asymptotic setting, CLP establish conditions for the bandwidths of $\widehat{\omega}_n^2$ and $\widehat{\Omega}_n$ under which $\widehat{\omega}_n^2$ and $\widehat{\Omega}_n$ are consistent estimators leading to asymptotically valid inference using the Wald statistics. In particular, CLP show that the parametric plug-in data dependent bandwidth rule of Andrews (1991) ensures $\widehat{\omega}_n^2$ and $\widehat{\Omega}_n$ are consistent in the high frequency asymptotics case. In contrast, the non-parametric plug-in data dependent bandwidth rule of Newey and West (1994) results in $\widehat{\omega}_n^2$ and $\widehat{\Omega}_n$ being inconsistent because the bandwidths are too small in the high frequency asymptotics case.

Here we explore a related but different question. If the bandwidths are modeled as a fixed proportion of the sample size (i.e. the fixed-b asymptotics nesting is used for $\widehat{\omega}_n^2$ and $\widehat{\Omega}_n$), are the fixed-b limits in the CLP high frequency asymptotic setting the same as the well known limits for the fixed sampling frequency case (δ fixed)? As will be shown, the answer is yes if the assumptions in the CLP framework are slightly strengthened to be analogous to assumptions used in discrete (δ fixed) settings. This suggests that fixed-b critical values can be used to improve inference for sampling frequencies that range from low to high.

3.3 Theory

3.3.1 Fixed-b Theory for Discrete Stationary Regressions

Our starting point is a review of fixed-b theory for the $H(\widehat{\beta})$ statistic in stationary regressions for a given sampling frequency (δ fixed) as developed by Kiefer and Vogelsang (2005). Let Ω^v denote the long run variance of $v_i = x_i u_i$ defined as

$$\Omega^v = \Gamma^v(0) + \sum_{j=1}^{\infty} \left(\Gamma^v(j) + \Gamma^v(j)' \right),$$

where $\Gamma^v(j) = E(v_i v_{i-j}')$. Let Λ^v denote the matrix square root of Ω^v, i.e. $\Omega^v = \Lambda^v \Lambda^{v\prime}$. The kernel based nonparametric estimator of Ω^v is given by

$$\widehat{\Omega}_n^v = \widehat{\Gamma}_n^v(0) + \sum_{j=1}^{n-1} k\left(\frac{j}{M_n} \right) \left(\widehat{\Gamma}_n^v(j) + \widehat{\Gamma}_n^v(j)' \right),$$

77

where $\widehat{\Gamma}_n^v(j) = n^{-1} \sum_{i=j+1}^n \widehat{v}_i \widehat{v}'_{i-j}$ and $\widehat{v}_i = x_i \widehat{u}_i = x_i \left(y_i - x'_i \widehat{\beta} \right)$.

Fixed-b asymptotic results are obtained using an asymptotic nesting for the bandwidth, M_n, such that $M_n = bn$ where $b \in (0, 1]$ is held constant as the sample size, n, grows. With the frequency of observation held fixed (equivalent to δ held fixed), standard fixed-b theory applies under two sufficient assumptions. We use the symbol $\Rightarrow$ to denote weak convergence in distribution.

Assumption 3.1 *(a)* $n^{-1} \sum_{i=1}^{[rn]} x_i x'_i \xrightarrow{p} rQ$, *where* $r \in [0, 1]$ *and* Q^{-1} *exists, and (b)* $n^{-1/2} \sum_{t=1}^{[rn]} x_i u_i = n^{-1/2} \sum_{t=1}^{[rn]} v_i \Rightarrow \Lambda^v W_k(r)$, *where* $r \in [0, 1]$, *and* $W_k(r)$ *is a* $k \times 1$ *vector of independent Wiener processes,* $W_k(r) \sim N(0, r I_k)$.

Kiefer and Vogelsang (2005) show that under Assumption 3.1 the fixed-b limit of $\widehat{\Omega}_n^v$ is given by

$$\widehat{\Omega}_n^v \Rightarrow \Lambda^v P_k(b) \Lambda^{v\prime},$$

where $P_k(b)$ is a stochastic process that is a function of the Brownian bridge, $B_k(r) = W_k(r) - r W_k(1)$, where the form of $P_k(b)$ depends on the kernel, $k(x)$. Relevant to our simulations is the case where $k(x)$ is the Parzen kernel and $P_k(b)$ is given by

$$P_k(b) = -\iint_{|r-s|<b} \frac{1}{b^2} k'' \left(\frac{r-s}{b} \right) B_k(r) B_k(s)' dr ds, \tag{3.3}$$

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

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

For the case of the widely used Bartlett kernel

$$P_k(b) = \frac{2}{b} \int_0^1 B_k(r) B_k(r)' dr - \frac{1}{b} \int_0^{1-b} B_k(r) B_k(r+b)' dr - \frac{1}{b} \int_0^{1-b} B_k(r+b) B_k(r)' dr. \tag{3.4}$$

See Kiefer and Vogelsang (2005) for details. The fixed-b limit of $H(\widehat{\beta})$ is given by

$$H(\widehat{\beta}) \Rightarrow W_q(1)' P_q(b)^{-1} W_q(1).$$

78

This limiting random variable is a function of a $q \times 1$ vector of standard Wiener processes and depends on the kernel (through $P_q(b)$) and the bandwidth (through b) but is otherwise pivotal. Similar, although different, results are obtained for $G(\widehat{\beta})$ in cointegrated regressions. See Bunzel (2006) for details.

It is not obvious whether the fixed-b results for $G(\widehat{\beta})$ and $H(\widehat{\beta})$ in the fixed sampling frequency case continue to hold in the high frequency asymptotic framework of CLP. Using the theoretical tools of CLP, we obtain fixed-b results in the high frequency setting.

3.3.2 Fixed-b High Frequency Asymptotics for Stationary Regressions

In this section we obtain fixed-b asymptotic results for the $H(\widehat{\beta})$ statistic in the high frequency framework of CLP using slightly strengthened assumptions from CLP appropriate for stationary regressions. Following CLP, let Z (equivalently Z_t) denote a continuous time stochastic process and assume that $Z = Z^c + Z^d$ such that Z^c is the continuous component and Z^d is a jump component defined as $Z_t^d = \sum_{0 \le s \le t} \Delta Z_s$ where $\Delta Z_t = Z_t - Z_{t-}$. We assume a version of Assumption D1 from CLP holds, and we assume that a version of Lemma 3.1 from CLP holds for partial sums:

Assumption 3.2 *Defining $Z = XX'$ or XU and $z_i = Z_{i\delta}$ for $i = 1, \ldots, n$, suppose that for $r \in (0, 1]$*

$$\frac{1}{n} \sum_{i=1}^{[rn]} z_i = \frac{1}{T} \int_0^{rT} Z_t dt + O_p(\Delta_{\delta,rT}(\|Z\|)),$$

for all small δ and large T where $\Delta_{\delta,rT}(\|Z\|) = \sup_{0 \le s,t \le rT} \sup_{|t-s| \le \delta} \|Z_t^c - Z_s^c\|$ and $\|\cdot\|$ is the Euclidean norm.

Assumption 3.3 $\Delta_{\delta,T}(\|XX'\|) \to 0$ *and* $\sqrt{T}\Delta_{\delta,T}(\|XU\|) \to 0$ *as* $\delta \to 0$ *and* $T \to \infty$.

Assumptions 3.2 and 3.3 allow sample moments to approximate continuous time analogs. CLP argue that Assumption 3.3 is not particularly strong, nor is Assumption 3.6 given below (equivalent to Assumption D2 in CLP)[1]. The next assumption is sufficient to obtain continuous time fixed-b results and is equivalent to Assumption C1 of CLP strengthened to hold for partial sums.

---

[1]In the discussion after introducing Assumptions D1 and D2, CLP point out that these assumptions allow the continuous time processes to be Brownian motions but also allow the processes to have more local volatility and be more explosive globally. CLP argue that δ generally needs to go to 0 faster than T goes to ∞ and that the relative rate would depend on how locally volatile and explosive the processes are.

Assumption 3.4 *We assume that for $r \in (0, 1]$:* **(a)**

$$T^{-1} \int_0^{rT} X_t X_t' dt \xrightarrow{p} rQ,$$

as $T \to \infty$ for some nonrandom matrix $Q > 0$, and **(b)**

$$T^{-1/2} \int_0^{rT} X_t U_t dt \Rightarrow \Lambda W_k(r) \sim N(0, r\Omega),$$

as $T \to \infty$ where Ω is the long run variance of $X_t U_t$.

Assumption 3.4 is effectively a continuous time analog to Assumption 3.1 and is slightly stronger than Assumption C1 of CLP. It rules out certain nonstationary behavior for X_t and $X_t U_t$. Note that Assumption 3.4(b) is a continuous time functional central limit theorem (invariance principle) for $X_t U_t$ whereas Assumption C1(b) of CLP is a continuous time central limit theorem. An analogous condition to Assumption 3.4(b) was used by Lu and Park (2019) (their equation (2)) to obtain a continuous time fixed-*b* result for kernel long run variance estimators applied to a vector of continuous time processes *known* to be mean zero. The fixed-*b* results of Lu and Park (2019) cannot be directly applied to regression settings because of need to estimate β when constructing $x_i \widehat{u}_i$. This is equivalent to having to estimate an unknown mean before estimating a long run variance. This changes the fixed-*b* limit compared to the known mean case - see Hashimzade and Vogelsang (2008) for details.

Using Assumptions 3.2, 3.3, and 3.4, the following theorem holds for the $H(\widehat{\beta})$ statistic.

Theorem 3.1 *Let $M_n = bn$ where $b \in (0, 1]$ is fixed. Assume $H_0 : R\beta = r$ holds. Then, under Assumptions 3.2, 3.3, and 3.4, as $\delta \to \infty$ and $T \to \infty$, $H(\widehat{\beta}) \Rightarrow W_q(1)' P_q(b)^{-1} W_q(1)$.*

The proof is given in the appendix. Theorem 3.1 shows that the fixed-*b* limit of $H(\widehat{\beta})$ in the high frequency asymptotic framework of CLP is the same as in the fixed sampling frequency case as long as the assumptions used by CLP are strengthened to hold for partial sums. Critical values of the limiting distribution depends on the kernel and bandwidth sample size ratio, b, but otherwise are pivotal. Therefore, critical values are easily obtained using simulation methods.

Next we analyze both of the $G(\widehat{\beta})$ and $H(\widehat{\beta})$ statistics under assumptions in CLP suited for cointegrating regressions.

3.3.3 Fixed-b High Frequency Asymptotics for Cointegrating Regressions

As is well known in the literature, fixed-b limits of HAR statistics depend on the stationarity properties of X_t, U_t and X_tU_t. Cointegrating regression corresponds to the case where X_t is a Brownian motion (continuous time unit root process) and U_t is stationary. We consider the case where X_t includes an intercept and write $X_t = [1 \ \widetilde{X}_t]'$, and its discretized version x_i as $x_i = [1 \ \widetilde{x}_i]'$, where $\widetilde{X}_t$ and $\widetilde{x}_i$ are $(k-1) \times 1$ vectors. Two versions of assumptions used by CLP are sufficient to obtain fixed-b results for cointegrating regression. The space of cadlag functions is denoted by $D[0, 1]$.

Assumption 3.5 *(CLP Assumption C2) Assume that **(a)** for $X^T(r)$ defined as $X^T(r) = \Lambda_T^{-1}X_{rT}$ on [0,1] with an appropriate nonsingular normalizing sequence (Λ_T) of matrices, it follows that*

$$X^T(r) \Rightarrow X^\circ(r),$$

*in the product space of $D[0, 1]$ as $T \to \infty$ with linearly independent limit process $X^\circ(r)$, and **(b)** if we define $S^T(r)$ on $[0, 1]$ as $S^T(r) = T^{-1/2}\int_0^{rT} U_s ds$ then*

$$S^T(r) \Rightarrow U^\circ(r),$$

in $D[0, 1]$ jointly with $X^T(r) \Rightarrow X^\circ(r)$ in the product space of $D[0, 1]$ as $T \to \infty$, where $U^\circ(r)$ is a Brownian motion with $U^\circ(r) = \lambda_u w_u(r)$ where $\lambda_u^2 = \lim_{T\to\infty} T^{-1}E(\int_0^T U_t dt)^2 > 0$, which is assumed to exist. $w_u(r)$ is a standard Wiener process.

Assumption 3.6 *(CLP Assumption D2 (modified)) Assume **(a)** $\|\Lambda_T\|^2\Delta_{\delta,T}(\|XX'\|) \to 0$, $\sqrt{T}\|\Lambda_T\|\Delta_{\delta,T}(\|XU\|) \to 0$ and **(b)** $\sqrt{T}\Delta_{\delta,T}(\|U\|) \to 0$, $\|\Lambda_T\|\Delta_{\delta,T}(\|X\|) \to 0$ as $\delta \to 0$ and $T \to \infty$.*

Because we assume the first element of X_t is an intercept variable, the first element of X° is the identity function. As pointed out in CLP in the discussion of their Assumptions C1 and C2, the

random components of X° can be general diffusion processes; see also Kim and Park (2017). The classic cointegration case is obtained when the random components of X° are Brownian motions. Assumption 3.5(b) assumes a continuous time functional central limit theorem holds for the partial integrals of U_t. The limiting Brownian motion, U°, can be correlated with X°. A continuous time version of cointegrating regression without endogeneity between the regressors and error holds for the special case where X° is a Brownian motion that is independent of U°. We also extend the list of Z processes in Assumption 3.2 (labeled 3.2*) to include the processes in Assumption 3.6 (b).

The next theorem gives fixed-b results for the $G(\widehat{\beta})$ and $H(\widehat{\beta})$ statistics under Assumptions 3.5 and 3.6 (CLP Assumptions C2,D2). The limits depend on a $q \times k$ matrix R^* that depends on the form of Λ_T that is defined as follows. Suppose there exists a $q \times q$ nonsingular scaling matrix, Λ_T^R, such that $\lim_{T\to\infty} \Lambda_T^R T^{-1/2} R \Lambda_T^{-1}$ exists and is a matrix with rank equal to q. Then define

$$R^* = \lim_{T\to\infty} \Lambda_T^R T^{-1/2} R \Lambda_T^{-1}. \tag{3.5}$$

When the null hypothesis depends on estimated parameters that converge at the same rate, it will be the case that $R^* = R$. However, for a row of R that corresponds to a null hypothesis that is a linear combination of estimated parameters that converge at different rates, the corresponding row of R^* will have nonzero elements corresponding to the estimated parameters in that linear combination that converge the slowest.

Theorem 3.2 *Let $M_n = bn$ where $b \in (0, 1]$ is fixed. Assume $H_0 : R\beta = r$ holds. Under Assumptions 3.2*, 3.5 and 3.6, as $\delta \to 0$ and $T \to \infty$,*

$$G(\widehat{\beta}) \Rightarrow (R^*C)' \left[P_G(b)R^*Q_\circ^{-1}R^{*\prime}\right]^{-1} R^*C,$$

$$H(\widehat{\beta}) \Rightarrow (R^*C)' \left[R^*Q_\circ^{-1}P_H(b)Q_\circ^{-1}R^{*\prime}\right]^{-1} R^*C,$$

where

$$Q_\circ = \int_0^1 X^\circ(s)X^\circ(s)'ds, \quad C = Q_\circ^{-1}\int_0^1 X^\circ(s)\, dw_u(s),$$

$P_G(b)$ is a function of

$$B_G(r) = w_u(r) - \left(\int_0^r X^\circ(s)'ds\right)C,$$

82

and $P_H(b)$ is a function of

$$B_H(r) = \int_0^r X°(s)dw_u(s) - \left(\int_0^r X°(s)X°(s)'ds\right)C,$$

where the forms of $P_G(b)$ and $P_H(b)$ are the same as $P_k(b)$ with $B_G(r)$ and $B_H(r)$ in place of $B_k(r)$.

The proof is given in the appendix. In general, the fixed-b limits of $G(\widehat{\beta})$ and $H(\widehat{\beta})$ given by Theorem 3.2 are nonpivotal and depend on nuisance parameters related to the structure of $X°$ and dependence between $X°$ and $U°$. In the special case where $X°$ is a Brownian motion that is independent of $U°$, the limits in Theorem 3.2 simplify, and are identical to, the fixed-b limits obtained by Bunzel (2006) for cointegrating regressions. These limits depend on b and the kernel along with the number of stochastic regressors and the presence of the intercept regressor.

To be concrete, suppose that

$$X° (r) = \begin{bmatrix} 1 \\ \Lambda_{\widetilde{x}}W_{\widetilde{x}}(r) \end{bmatrix}, \tag{3.6}$$

where $W_{\widetilde{x}}(r)$ is a $(k-1) \times 1$ vector of independent Wiener processes that are independent of $w_u(r)$ $(U°)$. In this case the random part of $X°$ is a Brownian motion with long run variance $\Omega_{\widetilde{x}} = \Lambda_{\widetilde{x}}\Lambda_{\widetilde{x}}'$. Note that in this case the scaling matrix, Λ_T, has the form

$$\Lambda_T = \begin{bmatrix} 1 & 0_{1\times(k-1)} \\ 0_{(k-1)\times 1} & T^{1/2}I_{k-1} \end{bmatrix}.$$

The following Lemma gives the fixed-b limits for this special case.

Lemma 3.1 *Define*

$$g (r) = \begin{bmatrix} 1 \\ W_{\widetilde{x}}(r) \end{bmatrix}.$$

For the case where $X°$ is given by (3.6), the limits in Theorem 3.2 become

$$Q_\circ = \int_0^1 g(s)g(s)'ds, \quad C = Q_\circ^{-1}\int_0^1 g(s)dw_u(s),$$

$$B_G(r) = w_u(r) - \left(\int_0^r g(s)'ds\right)C, \quad B_H(r) = \int_0^r g(s)dw_u(s) - \left(\int_0^r g(s)g(s)'ds\right)C.$$

83

Careful examination of C, $B_G(r)$, and $B_H(r)$ in the Lemma reveals that these terms, and the limits of $G(\widehat{\beta})$ and $H(\widehat{\beta})$, are pivotal and only depend on $w_u(r)$ and $g(r)$ in addition to b and the kernel. For a given b and kernel, asymptotic critical values are easily simulated but depend on the dimension of $W_{\widetilde{x}}(r)$, the presence of the intercept in $g(r)$, and R^*.

3.3.4 Fixed-b High Frequency Asymptotics for Cointegrating Regressions with Endogenous Regressors

For the case of discrete time cointegrating regressions with endogenous regressors, there are many methods proposed in the literature to obtain asymptotically pivotal test statistics. In the continuous time framework, Pellatt and Sun (2023) analyze the discrete time approach of Hwang and Sun (2018) that uses a transformation of the regression using orthonormal basis functions. Pellatt and Sun (2023) show that the test statistics proposed by Hwang and Sun (2018) have the same asymptotic limits in the high frequency setting of CLP. Another approach that delivers asymptotic pivotal test statistics in discrete time is the integrated modified OLS (IM-OLS) approach of Vogelsang and Wagner (2014) where kernels are used to estimate relevant long run variance parameters. Therefore, it is natural to analyze IM-OLS tests in the CLP high frequency setting.

The IM-OLS approach is based on a simple transformation and augmentation of the discrete time regression (3.2) where we continue to focus on the case where $X_t = [1 \ \ \widetilde{X}_t]'$ and $x_i = [1 \ \ \widetilde{x}_i]'$. Partial summing both sides of (3.2) and including $\widetilde{x}_i$ after partial summing gives

$$S_i^y = S_i^{x\prime}\beta + \widetilde{x}_i'\gamma + S_i^u, \tag{3.7}$$

where $S_i^y = \sum_{j=1}^{i} y_j$, $S_i^x = \sum_{j=1}^{i} x_j$, $\widetilde{x}_i = \widetilde{x}_{i-1} + v_i^{\widetilde{x}}$, $S_i^u = \sum_{j=1}^{i} u_j$, and γ is a $(k-1) \times 1$ vector of parameters. It is convenient to stack S_i^x and $\widetilde{x}_i$ into a single vector x_i^* and write (3.7) more compactly as

$$S_i^y = x_i^{*\prime}\theta + S_i^u, \tag{3.8}$$

where

$$x_i^* = \begin{bmatrix} S_i^x \\ \widetilde{x}_i \end{bmatrix}, \quad \theta = \begin{bmatrix} \beta \\ \gamma \end{bmatrix}.$$

84

The IM-OLS estimator is given by OLS applied to (3.8):

$$\widetilde{\theta} = \left(\sum_{i=1}^{n} x_i^* x_i^{*\prime} \right)^{-1} \sum_{i=1}^{n} x_i^* S_i^y.$$

The corresponding continuous time regression is given by

$$S_t^Y = X_t^{*\prime} \theta + S_t^U,$$

where $S_t^Y = \int_0^t Y_s ds$, $S_t^U = \int_0^t U_s ds$, and $X_t^* = \begin{bmatrix} \int_0^t X_s ds \\ \widetilde{X}_t \end{bmatrix}$ where $\widetilde{X}_t = \int_0^t V_s^{\widetilde{x}} ds$. We focus on the

Wald statistic given by

$$\widetilde{W}^* = (R\widetilde{\theta} - r)' \left[\widetilde{\lambda}_{u \cdot \widetilde{x}}^{2*} R \left(\sum_{i=1}^{n} x_i^* x_i^{*\prime} \right)^{-1} \left(\sum_{i=1}^{n} c_i c_i' \right) \left(\sum_{i=1}^{n} x_i^* x_i^{*\prime} \right)^{-1} R' \right]^{-1} (R\widetilde{\theta} - r), \qquad (3.9)$$

where $c_i = \sum_{j=1}^{n} x_j^* - \sum_{j=1}^{i} x_j^*$. The dimension of the R matrix and r vector are adjusted with zeros to accommodate γ in the model but the restrictions being tested about β remain the same. The long run variance estimator is given by

$$\widetilde{\lambda}_{u \cdot \widetilde{x}}^{2*} = n^{-1} \sum_{i=2}^{n} \sum_{j=2}^{n} k\left(\frac{|i-j|}{M_n} \right) \Delta \widetilde{S}_i^{u*} \Delta \widetilde{S}_j^{u*},$$

where $\Delta \widetilde{S}_i^{u*} = \widetilde{S}_i^{u*} - \widetilde{S}_{i-1}^{u*}$ are the first differences of the OLS residuals, $\widetilde{S}_i^{u*}$, from the regression

$$S_i^y = S_i^{x\prime} \beta + \widetilde{x}_i' \gamma + z_i' \delta + S_i^{u*}, \qquad (3.10)$$

where $z_i = i \sum_{j=1}^{n} x_j^* - \sum_{j=1}^{i-1} \sum_{m=1}^{j} x_m^*$. The extra z_i regressors are included to ensure that the fixed-b limit of $\widetilde{W}^*$ is pivotal. See Vogelsang and Wagner (2014) for details.

To obtain the high frequency fixed-b limit of $\widetilde{W}^*$ we extend Assumptions 3.5 and 3.6 to accommodate the partials sums and augmented regressors. We focus on the case where $X^\circ(r)$ is given by (3.6) and we write the assumptions in terms of

$$X_t^{**} = \begin{bmatrix} X_t^* \\ t \int_0^T X_s^* ds - \int_0^t \left(\int_0^s X_v^* dv \right) ds \end{bmatrix},$$

which collects the continuous time variables corresponding to the regressors in (3.10).

Assumption 3.7 $\Lambda_T^{-1} X_{rT}^{**} \Rightarrow X_{**}^{\circ}(r)$ *on [0,1] with an appropriate nonsingular normalizing sequence* (Λ_T) *of matrices where*

$$X_{**}^{\circ}(r) = \begin{bmatrix} X_{*}^{o}(r) \\ r \int_0^1 X_{*}^{o}(s)ds - \int_0^r \left(\int_0^s X_{*}^{o}(v)dv \right) ds \end{bmatrix}, \quad X_{*}^{\circ}(r) = \Pi g_{*}(r),$$

$$\Pi = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \Lambda_{\widetilde{x}} & 0 \\ 0 & 0 & \Lambda_{\widetilde{x}} \end{bmatrix}, \quad g_{*}(r) = \begin{bmatrix} r \\ \int_0^r W_{\widetilde{x}}(s)ds \\ W_{\widetilde{x}}(r) \end{bmatrix},$$

and $S^T(r) = T^{-1/2} \int_0^{rT} U_s ds \Rightarrow U^{\circ}(r) = \lambda_{u \cdot \widetilde{x}} w_{u \cdot \widetilde{x}}(r) + \Omega_{u\widetilde{x}} \left(\Lambda_{\widetilde{x}}^{-1} \right)' W_{\widetilde{x}}(r) = B_u(r)$ *where* $w_{u \cdot \widetilde{x}}(r)$ *is a scalar standard Wiener process independent of* $W_{\widetilde{x}}(r)$, $\lambda_{u \cdot \widetilde{x}}^2 = \lambda_u^2 - \Omega_{u\widetilde{x}} \Omega_{\widetilde{x}}^{-1} \Omega_{u\widetilde{x}}'$, λ_u^2 *is the variance of* $U^{\circ}(r)$, *and* $\Omega_{u\widetilde{x}}$ *is the covariance between* $U^{\circ}(r)$ *and* $\Lambda_{\widetilde{x}} W_{\widetilde{x}}(r)$.

Assumption 3.8 $\sqrt{T}\Delta_{\delta,T}(\|U\|) \to 0, \Delta_{\delta,T}(\|\widetilde{X}\|) \to 0, \sqrt{T}\Delta_{\delta,T}(\|V^{\widetilde{x}}\|) \to 0, \|\Lambda_T\|^2 \Delta_{\delta,T}(\|X^{**}X^{**'}\|) \to 0$ *and* $\sqrt{T}\|\Lambda_T\|\Delta_{\delta,T}(\|X^{**}S^U\|) \to 0$ *as* $\delta \to 0$ *and* $T \to \infty$.

We also extend the list of Z processes in Assumption 3.2 (labeled 3.2**) to include the processes in Assumption 3.8. We can now state the fixed-b limiting result for the IM-OLS $\widetilde{W}^*$ statistic in the following theorem.

Theorem 3.3 *Let* $M_n = bn$ *where* $b \in (0,1]$ *is fixed. Assume* $H_0 : R\theta = r$ *holds. Under Assumptions* 3.2**, *3.7 and 3.8, as* $\delta \to 0$ *and* $T \to \infty$,

$$\widetilde{W}^* \Rightarrow \frac{\chi_q^2}{P_1^{**}(b)}$$

where χ_q^2 *is a chi-square random variable with q degrees freedom independent of* $P_1^{**}(b)$ *where* $P_1^{**}(b)$ *takes the same form as* $P_1(b)$ *in (3.3) or (3.4) with*

$$B_1^{**}(r) = \int_0^r dw_{u \cdot \widetilde{x}}(s) - g_{**}(r)' \left(\int_0^1 g_{**}(s) g_{**}(s)' \right)^{-1} \int_0^1 \left(\int_0^1 g_{**}(v)\,dv - \int_0^s g_{**}(v)\,dv \right) dw_{u \cdot \widetilde{x}}(s),$$

in place of $B_1(r)$ *with*

$$g_{**}(r) = \begin{bmatrix} g_{*}(r) \\ r \int_0^1 g_{*}(s)ds - \int_0^r \left(\int_0^s g_{*}(v)dv \right) ds \end{bmatrix}.$$

The limit given by Theorem 3.3 is the same as the limit in the discrete time case as obtained by Vogelsang and Wagner (2014).

3.4 Fixed-b Critical Values

Because fixed-b asymptotic limits are nonstandard, numerical methods are used to compute critical values. Here we focus on critical values using the Parzen kernel because the finite sample simulations use the Parzen kernel to make direct comparisons with CLP. Appendix B of Vogelsang (2011), the working paper version of Vogelsang (2012), provides a numerical method for computation of fixed-b critical values for the Bartlett kernel for stationary regressions. We use the same method here for the Parzen kernel for the following cases: **(i)** H-statistic stationary regression, **(ii)** G-statistic cointegration regression (without endogeneity) and **(iii)** H-statistic cointegration regression (without endogeneity). To align with our simulation results and empirical illustration, we report critical value functions for a simple regression with an intercept and one regressor for two hypotheses. The first is a test of the joint null hypothesis that the intercept parameter is zero ($\beta_1 = 0$) and the slope parameter is 1 ($\beta_2 = 1$). The second is a test that the slope parameter is 1. In the stationary case, the fixed-b critical values only depend on the number of restrictions, $q = 2$ and 1 respectively. In cointegration regressions, as shown by Propositions 1 and 2 of Bunzel (2006) and our Lemma 3.1, the fixed-b critical values also depend on the number of stochastic regressors in the model, the form of deterministic regressors (the intercept), and R^*.

Following Vogelsang (2011) let $cv_\alpha(b)$ denote the critical value for a given statistic for significance level α using a bandwidth sample size ratio b. Using 50,000 replications, $cv_\alpha(b)$ was simulated using normalized partial sums of independent, identically distributed (i.i.d.) $N(0,1)$ random variables using 1,000 steps to approximate the Wiener processes in the asymptotic distributions. These simulations were carried out for the values of $b = 0.02, 0.04, \ldots, 0.98, 1.0$. Using

the simulated critical values, we fit critical value functions of the form:

$$cv_\alpha(b) = z_{q,\alpha}^2 + \lambda_1(b \cdot z_{q,\alpha}^2) + \lambda_2(b \cdot \left(z_{q,\alpha}^2\right)^2) + \lambda_3(b \cdot \left(z_{q,\alpha}^2\right)^3) + \lambda_4(b^2 \cdot z_{q,\alpha}^2) + \lambda_5(b^2 \cdot \left(z_{q,\alpha}^2\right)^2)$$

(3.11)

$$+ \lambda_6(b^2 \cdot \left(z_{q,\alpha}^2\right)^3) + \lambda_7(b^3 \cdot z_{q,\alpha}^2) + \lambda_8(b^3 \cdot \left(z_{q,\alpha}^2\right)^2) + \lambda_9(b^3 \cdot \left(z_{q,\alpha}^2\right)^3),$$

where $z_{q,\alpha}^2$ is the critical value from a χ_q^2 (chi-square with q degrees of freedom) random variable. For $H_0 : \beta_1 = 0, \beta_2 = 1$ we have $q = 2$, and for $H_0 : \beta_2 = 1$ we have $q = 1$. Notice that, by construction, $cv_\alpha(0) = z_{q,\alpha}^2$ so that when $b = 0$ the critical values are chi-square. For a given statistic, the values of the λ_i coefficients were obtained using least squares. The fits, as measured by the regression R^2, are excellent in all cases (no smaller than 0.995). Table 3B.1 gives the λ_i coefficients.

As shown by Vogelsang and Wagner (2014) fixed-b critical values for $\widetilde{W}^*$ depend on the number of integrated regressors, the form of the deterministic regressors and the hypothesis being tested in addition to the kernel and bandwidth. To test the joint null hypothesis, $\beta_1 = 0$ and $\beta_2 = 1$, we simulated fixed-b critical values for $\widetilde{W}^*$ based on the Parzen kernel for testing the joint null hypothesis in a cointegrating regression with an intercept and one integrated regressor. Then using the fixed-b critical values, we fit a critical value function for $\alpha = 0.05$ for $\widetilde{W}^*$. The critical value function is given by

$$cv_{0.05}(b) = \begin{cases} 5.96 + 8.73 \cdot b + 551.46 \cdot b^2 - 1950.49 \cdot b^3 + 7145.52 \cdot b^4, \text{ when } b \leq 0.2 \\ \\ -770.1 + 16182.2 \cdot b - 144138 \cdot b^2 + 727975.7 \cdot b^3 - 2283734.7 \cdot b^4 \\ +4650336.4 \cdot b^5 - 6114285.0 \cdot b^6 + 4986435.8 \cdot b^7 - 2287834.8 \cdot b^8 + 450664.3 \cdot b^9, \\ \text{when } b > 0.2 \end{cases}$$

(3.12)

where the fit as measured by the regression R^2 was larger than 0.999.

For testing the slope parameter only, $\beta_2 = 1$, using simulated critical values for the case of an intercept and one integrated regressor provided by the supplementary material of Vogelsang and

Wagner (2014), we fit a critical value function for $\alpha = 0.05$ for $\widetilde{W}^*$. The critical value function is given by

$$cv_{0.05}(b) = \begin{cases} 3.84 - 6.61 \cdot b + 930.69 \cdot b^2 - 12634.81 \cdot b^3 + 102402.08 \cdot b^4 \\ -396004.20 \cdot b^5 + 605085.78 \cdot b^6, \text{ when } b \leq 0.2 \\ \\ -41.98 + 1353.9 \cdot b - 15398.9 \cdot b^2 + 96187.7 \cdot b^3 - 354315.7 \cdot b^4 \\ +821676.6 \cdot b^5 - 1182990.3 \cdot b^6 + 1018299.0 \cdot b^7 - 478903.5 \cdot b^8 + 94608.1 \cdot b^9, \\ \text{ when } b > 0.2 \end{cases}$$

(3.13)

where the fit as measured by the regression R^2 was larger than 0.999.

More generally, asymptotic critical values for $\widetilde{W}^*$ can be simulated using a simple Monte Carlo simulation procedure[2]. Suppose an empirical application uses regression (3.2), $y_i = x_i'\beta + u_i$, with $i = 1, 2, \ldots, n$ where $x_i = [1 \ \widetilde{x}_i]'$ is a $k \times 1$ vector and $\widetilde{x}_i$ is a $(k - 1) \times 1$ vector of unit root processes. Critical values for $\widetilde{W}^*$ for testing a given hypothesis about β using a given kernel and value of b can be computed as follows:

Step (1) Generate realizations of $\widetilde{x}_i$ and u_i using a large value of n such as $n = 1000$. The vector $\widetilde{x}_i$ is generated as $\widetilde{x}_i = \widetilde{x}_{i-1} + \epsilon_i^{\widetilde{x}}$ where $\{\epsilon_i^{\widetilde{x}}\}$ is a sequence of i.i.d. $N(0, I_{k-1})$ vectors with $\widetilde{x}_0$ equal to a zero vector. Generate a realization of $\{u_i\}$ as a sequence of i.i.d. $N(0, 1)$ random variables that are independent of $\{\epsilon_i^{\widetilde{x}}\}$. Let β_r denote the β vector with the null hypothesis imposed. Elements of β_r that do not involve the null hypothesis can be set to zero without loss of generality because $\widetilde{W}^*$ is exactly invariant to those elements. Using the realizations $x_i = [1 \ \widetilde{x}_i]'$ and u_i, a realization of y_i is calculated as $y_i = x_i'\beta_r + u_i$.

Step (2) Using the realized $\{y_i\}$ and $\{x_i\}$ data from Step 1, compute the IM-OLS statistic $\widetilde{W}^*$ using (3.9) with the given kernel and $M_n = b^\dagger n$ where $b^\dagger$ is the bandwidth sample size ratio from the empirical application.

---

[2]We thank a referee for suggesting that we include this algorithm in the paper.

Step (3) Repeat Steps (1) and (2) many times (for example 10,000 or more), and obtain the $(1 - \alpha)$ quantile of the realizations of $\widetilde{W}^*$. Use that quantile as the critical value for $\widetilde{W}^*$ computed from the empirical application.

As an example, recall the case of $k = 2$ for testing $\beta_1 = 0$ and $\beta_2 = 1$ as previously discussed. Here $\widetilde{x}_i$ is a scalar process generated as $\widetilde{x}_i = \widetilde{x}_{i-1} + \epsilon_i^{\widetilde{x}}$ where $\epsilon_i^{\widetilde{x}} \sim i.i.d.N(0, 1)$ and $\widetilde{x}_0 = 0$, and $u_i \sim i.i.d.N(0, 1)$ that is independent of $\epsilon_i^{\widetilde{x}}$. The vector β_r is given by $\beta_r = [0 \ 1]'$. For cases where more complicated deterministic regressors are included in x_i, those regressors would simply be included in the simulated x_i vector.

3.5 Simulations

3.5.1 Finite Sample Simulation Environment

In this section we use simulations to explore finite sample properties of $H(\widehat{\beta})$ for the stationary regression case and $\widetilde{W}^*$ for the cointegration case with and without endogeneity. Patterns are similar for the $G(\widehat{\beta})$ and $H(\widehat{\beta})$ statistics in the cointegration case without endogeneity and are not reported. We focus on performance of the kernel based tests across sampling frequencies to explore empirical null rejections, bandwidth behavior, and power across frequencies. We do not make comparisons across long run variance estimators given that Pellatt and Sun (2023) provide extensive comparisons between kernel long run variance estimators and orthonormal series long run variance estimators where overall performance between the two was found to be similar.

We use the same data generating process (DGP) as CLP to facilitate comparisons. The long run variances of $H(\widehat{\beta})$ and $\widetilde{W}^*$ are implemented with the Parzen kernel in all cases. We consider five bandwidth rules. These rules include three of the bandwidth rules used by CLP in their simulations: the Andrews (1991) AR(1) plug-in rule (AD), the Newey and West (1994) nonparametric plug-in rule (NW) using the pre-tuning parameters suggested by Newey and West (1994) and the CRT bandwidth rule proposed by CLP. The other two bandwidth rules are: the Sun et al. (2008) AR(1) plug-in rule that balances size distortions and power (SPJ), and a variant of the NW rule that uses different pre-tuning parameters which we label NW-Tune.

Following CLP, we focus on a continuous time regression model with an intercept and one

regressor given by

$$Y_t = \beta_1 + \beta_2 X_t + U_t,$$

where X_t now denotes a univariate stochastic regressor. Both X_t and U_t are Ornstein–Uhlenbeck processes given by

$$dX_t = -\kappa_x X_t dt + \sigma_x dV_t, \quad dU_t = -\kappa_u U_t dt + \sigma_u dW_t,$$

$$
\begin{bmatrix} V_t \\ W_t \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \pi & \sqrt{1-\pi^2} \end{bmatrix} \begin{bmatrix} \xi_{1t} \\ \xi_{2t} \end{bmatrix},
$$

where ξ_{1t} and ξ_{2t} are standard Brownian motions independent each other. The DGP for V_t and W_t is similar to the DGP used by Pellatt and Sun (2023). The variances of V_t and W_t are normalized to be one and π is the correlation between V_t and W_t. In the stationary simulations both κ_x and κ_u are strictly positive and $\pi = 0$. In the cointegration case $\kappa_x = 0$ and π can be non-zero. Using a span of $T = 30$ for a given replication, we generate 7,560 daily (252 weekday observations per year) sample paths for each OU process. Lower frequency series (such as monthly/quarterly) are constructed from the generated daily series. We focus on testing the joint null hypothesis, $H_0 : \beta_1 = 0, \beta_2 = 1$, using a nominal significance level of 0.05. When used, fixed-b critical values are computed using the critical value function given by (3.11) using coefficients from the H-stationary line of Table 3B.1 for the stationary regression case. The critical value function given by (3.12) is used for the cointegrating regression case. We used 2,000 replications in all cases.

We also carried out simulations for the case where X_t follows Feller's Square Root (SR) process as in CLP and Pellatt and Sun (2023) which is given by $dX_t = \kappa_x (\mu_x - X_t) dt + \sigma_x \sqrt{X_t} dV_t$. The patterns in the results are very similar to the OU process case and are not reported.

3.5.2 Bandwidth Formulas

To help with interpreting some of the finite sample patterns it is useful to examine some of the formulas for the bandwidth rules we used. The Andrews (1991) formula for the Parzen kernel is given by

$$\widehat{M}_n^{AD} = 2.6614 \left(\frac{4\widehat{\rho}^2}{(1-\widehat{\rho})^4} n \right)^{1/5},$$

where n is the sample size of the discretized series and $\widehat{\rho}$ is the estimated AR(1) parameter fit to $\widehat{v}_i = x_i\widehat{u}_i$ (stationary regression) or $\widehat{v}_i = \widehat{u}_i$ (cointegrating regression) where $\widehat{u}_i$ are the OLS residuals. For the stationary regression we follow Andrews (1991) and place zero weight on the intercept component of the long-run variance to ensure that the bandwidth rule is invariant to the scale of the data.

The Newey and West (1994) bandwidth takes the same form as $\widehat{M}_n^{AD}$ with the term that depends on $\widehat{\rho}$ estimated nonparametrically using kernel estimators that require bandwidths of their own (pre-tuning parameters). Formulas can be found in Newey and West (1994). Newey and West (1994) recommend the deterministic pre-tuning rule $4(n/100)^{4/25}$ and this gives the NW data dependent bandwidth. We also consider the pre-tuning rule $8.5(n/100)^{21/25}$ which gives the NW-Tune data dependent bandwidth. The formula for the Sun et al. (2008) bandwidth is

$$\widehat{M}_n^{SPJ} = \begin{cases} \left(\frac{2\widehat{\rho}c}{(1-\widehat{\rho})^2}n\right)^{1/3} & \text{if } \frac{2\widehat{\rho}c}{(1-\widehat{\rho})^2} > 0 \\ \log(n) & \text{otherwise} \end{cases}$$

where

$$c = \frac{12\left(wG'_{1,0}(z^2_{1,\alpha}) - G'_{1,\delta}(z^2_{1,\alpha})\right)}{0.539z^2_\alpha K_\tau(z^2_\alpha)}, \quad K_\tau(x) = \frac{\delta^2}{2x}G'_{3,\tau}(x),$$

and $\widehat{\rho}$ is the same AR(1) estimator used for $\widehat{M}_n^{AD}$, $G'_{j,\tau}(\cdot)$ is the probability density function of a (non)central chi-square random variable with j degrees of freedom and noncentrality parameter τ^2, and $z^2_{1,\alpha}$ is the critical value from a chi-square random variable with one degree of freedom. The parameters w and τ control the trade-off between size distortions and power. We use $w = 10$ and $\tau = 2$.

Finally, the CRT bandwidth of CLP is given by the formula

$$M_n^{CRT} = n^{1/5}\delta^{-4/5} = n^{1/5}\left(\frac{T}{n}\right)^{-4/5} = nT^{-4/5}.$$

3.5.3 Stationary Regression Results

For the stationary regressions we set $\pi = 0$ and use two pairs of persistence parameters: $(\kappa_x, \kappa_u) = (0.1, 6.9)$ and $(\kappa_x, \kappa_u) = (0.5, 0.5)$ where the first pair is from CLP and has relatively

low persistence in U_t. The second pair has similar persistence in X_t but much more persistence in U_t. We use the same standard deviation parameter values as in CLP: $(\sigma_x, \sigma_u) = (1.5514, 2.7566)$.

3.5.3.1 Empirical Null Rejections

Figure 3B.1 plots empirical null rejection probabilities for $H(\widehat{\beta})$. Panels (a) and (c) give results for $(\kappa_x, \kappa_u) = (0.1, 6.9)$ (the CLP case) whereas panels (b) and (d) give results for $(\kappa_x, \kappa_u) = (0.5, 0.5)$ (the more persistent case). The top panels give results using chi-square critical values whereas the bottom panels give results using fixed-b critical values. The x-axis is δ which ranges from 1/252 (daily frequency) to 1/4 (quarterly frequency). Figure 3B.1(a) essentially replicates some finite sample results from CLP although rejections using the AD bandwidth are just above 0.1 in contrast to the rejections in CLP where rejections were close to 0.07. Rejections using the SPJ bandwidth are slightly higher than those with AD, and rejections with CRT are between them. Rejections with NW are similar for low to medium frequencies and over-rejections occur at the daily (high) frequency. The alternative version of NW, NW-Tune, tends to over-reject across all sampling frequencies but does not show the big jump in over-rejection at the daily frequency. Figure 3B.1(c) shows that, except for NW, rejections are improved (closer to 0.05) when fixed-b critical values are used. Rejections are below 0.1 in all cases.

Figures 1(b,d) show what happens when the persistence is stronger for the given span. In Figure 3B.1(b) we see that all bandwidths lead to substantial over-rejections when the chi-square critical value is used. The NW bandwidth continues to over-reject more substantially at high frequencies consistent with the CLP finding that the NW bandwidth is not high frequency compatible. As Figure 3B.1(d) shows, using fixed-b critical values substantially reduces over-rejections for AD and especially SPJ. Modest improvements are seen for CRT and NW-Tune. Even with the improvements that the fixed-b critical values provide, over-rejections remain because the persistence is strong relative to the span (magnitude of T).

3.5.3.2 Bandwidth Patterns Across Sampling Frequencies

One pattern that is interesting in all four panels of Figure 3B.1 is that rejections for AD, SPJ and CRT are stable across frequency of observation especially when fixed-b critical values are

used. To help explain this stability, we computed average bandwidth to sample size ratios (across replications) denoted by

$$\widehat{b}_{AD} = \frac{\widehat{M}_n^{AD}}{n}, \quad \widehat{b}_{SPJ} = \frac{\widehat{M}_n^{SPJ}}{n}, \quad b_{CRT} = \frac{M_n^{CRT}}{n},$$

(and similarly for NW and NW-Tune). Plots of these ratios are given in the two panels of Figure 3B.2. We see that $\widehat{b}_{AD}$ and $\widehat{b}_{SPJ}$ are nearly flat across sampling frequencies with slight decreases at lower frequencies (larger δ). By construction, b_{CRT} is flat across sampling frequencies because

$$b_{CRT} = \frac{M_n^{CRT}}{n} = \frac{nT^{-4/5}}{n} = T^{-4/5}, \tag{3.14}$$

which is the same for all frequencies for a given value of T. The bandwidth ratio for NW decreases as the sampling frequency increases and NW-tune has a similar pattern that is shifted up. The fact that $\widehat{b}_{AD}$, $\widehat{b}_{SPJ}$ and b_{CRT} are stable across frequencies explains why rejections are similar across frequencies. It is well known from the fixed-smoothing literature that the extent to which over-rejections occur depends on the bandwidth sample size ratio and not the bandwidth itself.

Another interesting pattern in Figure 3B.1(c,d) is that rejections using SPJ are lower than AD or CRT when fixed-b critical values are used. This is obvious in Figure 3B.1(d). Why does SPJ give rejections closer to the nominal level? From Figure 3B.2(b) we see that $\widehat{b}_{SPJ} > \widehat{b}_{AD}$ on average and both are larger than b_{CRT}. This makes sense because the SPJ bandwidth rule is known to give larger bandwidths than AD given that SPJ balances size distortions and power rather than minimizing the mean square error of the variance estimator.

What is not obvious from Figure 3B.2(a,b) is why $\widehat{b}_{AD}$ and $\widehat{b}_{SPJ}$ are stable across sampling frequencies. In the next section we provide some simple theoretical arguments that can help explain this finite sample pattern.

3.5.3.3 Stability of Bandwidth Sample Size Ratios Across Sampling Frequencies

To understand Figure 3B.2(a,b), some simple theoretical calculations holding the span, T, fixed are useful. We hold T fixed because the patterns observed in those figures are for a given value of the span ($T = 30$). Recall that the AD and SPJ bandwidths are functions of the estimated AR(1) parameter of $x_i \widehat{u}_i$ which is a proxy for $x_i u_i$. Because $x_i u_i$ is the product of two independent discretized

OU processes, each of which are AR(1) processes with AR(1) parameters $\rho_x = \exp(-\kappa_x(T/n))$ and $\rho_u = \exp(-\kappa_u(T/n))$ respectively, $x_i u_i$ has an AR(1) structure with AR(1) parameter

$$\rho = \rho_x \rho_u = \exp\left(-(\kappa_x + \kappa_u)\frac{T}{n}\right). \tag{3.15}$$

Letting $\kappa_{xu} = \kappa_x + \kappa_u$ and using (3.15), we can write the AD and SPJ bandwidths as

$$M_n^{AD} = 2.6614\left(\frac{4\rho^2}{(1-\rho)^4}n\right)^{1/5} = 2.6614\left(\frac{4\exp\left(-2\kappa_{xu}\frac{T}{n}\right)}{(1-\exp\left(-\kappa_{xu}\frac{T}{n}\right))^4}n\right)^{1/5},$$

$$M_n^{SPJ} = \left(\frac{2\rho c}{(1-\rho)^2}n\right)^{1/3} = \left(\frac{2\exp\left(-\kappa_{xu}\frac{T}{n}\right)c}{(1-\exp\left(-\kappa_{xu}\frac{T}{n}\right))^2}n\right)^{1/3},$$

giving

$$b_{AD} = \frac{M_n^{AD}}{n} = 2.6614\left(\frac{4\exp\left(-2\kappa_{xu}\frac{T}{n}\right)}{(1-\exp\left(-\kappa_{xu}\frac{T}{n}\right))^4}\frac{1}{n^4}\right)^{1/5}, \tag{3.16}$$

$$b_{SPJ} = \frac{M_n^{SPJ}}{n} = \left(\frac{2\exp\left(-\kappa_{xu}\frac{T}{n}\right)c}{(1-\exp\left(-\kappa_{xu}\frac{T}{n}\right))^2}\frac{1}{n^2}\right)^{1/3}, \tag{3.17}$$

for the bandwidth ratios. Using the expansion

$$\exp\left(-\kappa_{xu}\frac{T}{n}\right) = 1 - \kappa_{xu}(\frac{T}{n}) + \frac{(\kappa_{xu}(\frac{T}{n}))^2}{2!} - \frac{(\kappa_{xu}(\frac{T}{n}))^3}{3!} + \cdots, \tag{3.18}$$

we can easily show (for T fixed) that

$$\lim_{n\to\infty} b_{AD} = 2.6614\left(\frac{4}{(\kappa_{xu}T)^4}\right)^{1/5}, \qquad \lim_{n\to\infty} b_{SPJ} = \left(\frac{c}{(\kappa_{xu}T)^2}\right)^{1/3}.$$

which suggests that, for data sampled at high frequencies, b_{AD} and b_{SPJ} are positive. The more persistent the data, i.e. the closer κ_{xu} is to zero, the larger b_{AD} and b_{SPJ} will be.

The stability of b_{CRT} across sampling frequencies in Figure 3B.2(a,b) is obvious and expected given (3.14). Because the formulas for (3.16) and (3.17) are not constant functions with respect to n, the large n limits of b_{AD} and b_{SPJ} are only useful in understanding what happens for very high frequency cases (large n). To provide a more complete picture, in Figure 3B.3 we plot the theoretical b_{AD} and b_{SPJ} functions (3.16) and (3.17) for the case of $T = 30$ for n ranging from high frequencies (daily, on the left) to low frequencies (yearly, on the right). For the high persistence

case $(\kappa_x, \kappa_u) = (0.5, 0.5)$, b_{AD} and b_{SPJ} are nearly flat across sampling frequencies especially in the daily $(\delta = 1/252)$ to quarterly $(\delta = 1/4)$ range. Interestingly, the stability of b_{AD} and b_{SPJ} also holds for lower frequency cases with a slight decline at the annual frequency $(\delta = 1)$. In the lower persistent case $(\kappa_x, \kappa_u) = (0.1, 6.9)$ we see that b_{AD} and b_{SPJ} continue to be flat in the daily to quarterly range but show noticeable decline at the annual frequency.

The surprising finding that b_{AD} and b_{SPJ} are stable in the simulations is explained by the relative flatness of b_{AD} and b_{SPJ} as functions of the sampling frequency at least for the AR(1) plug-in case. Because b_{AD} and b_{SPJ} do not converge to zero as the sampling frequency becomes very high (n becomes large), b_{AD} and b_{SPJ} remain "high frequency stable" - a complementary finding to the high frequency compatible finding of M_n^{AD} by CLP.

3.5.3.4   Finite Sample Power

In this section we report finite sample power (not size-adjusted) of $H(\widehat{\beta})$ for testing $H_0 : \beta_1 = 0$, $\beta_2 = 1$ using fixed-b critical values in all cases. Results are given in the four panels of Figure 3B.4. Panels (a) and (c) give results for $(\kappa_x, \kappa_u) = (0.1, 6.9)$ with alternatives $(\beta_1, \beta_2) = (0.02, 1.02)$, $(0.04, 1.06)$. Panels (b) and (d) give results for $(\kappa_x, \kappa_u) = (0.5, 0.5)$ with alternatives $(\beta_1, \beta_2) = (0.3, 1.3)$, $(0.95, 1.95)$. The format of the figures is the same as for the null simulations.

First, notice that power increases as the alternatives move farther away from the null (going from top panels to bottom panels). This is not surprising and is expected. The more interesting patterns are how power depends on the sampling frequency for a given alternative. In the high persistence case (Figure 3B.4(b,d)), power is nearly the same across sampling frequencies for the AD, SPJ and CRT bandwidths. For the less persistence case (Figure 3B.4(a,c)), power noticeably decreases as the sample frequency decreases (as δ increases). This suggests that using high frequency data when the data is not too persistent gives higher power. A similar finding was reported by Pellatt and Sun (2023).

This decline in power as the sampling frequency decreases cannot be explained by the bandwidths because the bandwidth sample size ratios are stable across sampling frequencies. To see why power decreases as the sampling frequency decreases, we can calculate the signal to noise

96

ratio for the slope estimator analogous to the calculations (3.16) and (3.17) for the bandwidths. The signal to noise ratio is simply the inverse of the approximate variance of $\widehat{\beta}_2$ given by

$$\left[var\left(\widehat{\beta}_2\right)\right]^{-1} \approx n \cdot var(x_i)\Omega^{-1}var(x_i),$$

where Ω is the long run variance of $x_i u_i$. Straightforward calculations give

$$n \cdot var(x_i)\Omega^{-1}var(x_i) = \frac{n\left(1 - \exp\left(-\kappa_{xu}\frac{T}{n}\right)\right)^2 \sigma_x^2 \kappa_u}{\left(1 - \exp\left(-2\kappa_{xu}\frac{T}{n}\right)\right)\sigma_u^2 \kappa_x} = \frac{\left[n\left(1 - \exp\left(-\kappa_{xu}\frac{T}{n}\right)\right)\right]^2 \sigma_x^2 \kappa_u}{n\left(1 - \exp\left(-2\kappa_{xu}\frac{T}{n}\right)\right)\sigma_u^2 \kappa_x}.$$

For T fixed it is easy to show, using (3.18), that as $n \to \infty$,

$$\lim_{n\to\infty} n \cdot var(x_i)\Omega^{-1}var(x_i) = \frac{\kappa_{xu}\sigma_x^2 \kappa_u}{2\sigma_u^2 \kappa_x}T.$$

In contrast, it is easy to see that as $n \to 0$,

$$\lim_{n\to 0} n \cdot var(x_i)\Omega^{-1}var(x_i) = 0.$$

We see that, for a given span, the signal to noise ratio is finite for high frequencies and decreases to zero as the sampling frequency decreases.

3.5.4 Cointegrating Regression Results

We report a set of simulation results for the IM-OLS statistic, $\widetilde{W}^*$, for cointegrating regressions in Figures 4-7 using a format that is analogous to Figures 1,2 and 4. The data dependent bandwidths for $\widetilde{\lambda}^{2*}_{u\cdot\widetilde{x}}$ were computed using the OLS residuals, not $\Delta\widetilde{S}^{u*}_i$, given that estimated values of ρ based on $\Delta\widetilde{S}^{u*}_i$ are severely biased. In all cases $\kappa_x = 0$. Following Pellatt and Sun (2023), results are reported for $\kappa_u = 0.393$ (highly persistent U_t) and $\kappa_u = 6.287$ (low persistent U_t). Results were obtained for $\kappa_u = 1.572$ but are not reported as those results consistently fall between results for the other two κ_u values. Following CLP we used $\sigma_u = 0.0097$ and $\sigma_x = 0.0998$. Results are given for $\pi = 0, 0.75$. Results for $\pi = 0.25, 0.5$ are very similar and are not reported. Fixed-b critical values based on (3.12) are used in all cases.

Overall, the patterns in Figures 5-7 are similar to the patterns in Figures 1-3 giving similar findings in cointegrating regressions, including the endogenous case, as in the stationary regression case. There can be substantial over-rejections in the high persistence case. As persistence decreases

97

(for the given value of T), empirical null rejections approach the nominal level. Figure 3B.6 shows that the AD, SPJ bandwidth ratios are stable across sampling frequencies for both levels of persistence. Figure 3B.7 shows that power in the high persistence case is not sensitive to the sampling frequency. Like for stationary regressions, power noticeably declines as the sampling frequency decreases in the lower persistence case.

3.5.5 Summary of Simulations

The simulation results for $H(\widehat{\beta})$ (stationary regression) and $\widetilde{W}^*$ (cointegrated regression) generated by OU processes can be summarized as follows. If the bandwidth sample size ratio is constant across sampling frequencies and fixed-b critical values are used, null rejections are similar. Therefore, with regard to size distortions, the sampling frequency does not matter. While higher frequency data has stronger autocorrelation for given values of κ_x and κ_u, it also has a larger sample size. Intuitively, the null rejections are stable across sampling frequencies when the same bandwidth sample size ratio is used because what matters is strength of autocorrelation relative to the number of observations. This balance between autocorrelation and sample size is stable across sampling frequencies.

Power, on the other hand, can depend on the sampling frequency if the persistence is not strong. Power increases as the sampling frequency increases. Because the b_{AD}, b_{SPJ}, and b_{CRT} are stable across sampling frequencies and provide stability in null rejections across sampling frequencies, using high frequency data can lead to higher power without sacrificing additional size distortions (over what is already present given the underlying persistence as measured by κ_x and κ_u relative to the span, T).

For given persistence in the data (given κ_x and κ_u), when fixed-b critical values are used, SPJ gives the least size distorted tests followed by AD followed by CRT. This happens because b_{SPJ} tends to be larger than b_{AD} which tends to be larger than b_{CRT}. Of course, given the well known trade-off between size distortions and power as b increases when fixed-b critical values are used, the power rankings are the opposite. Which bandwidth to use in practice depends on the implications of this trade-off to the practitioner.

3.6 Empirical Application

In this section we provide some basic empirical results on the uncovered interest parity (UIP) puzzle as an illustration. UIP is the well studied hypothesis that interest rate differentials between two countries should be equal to the expected return of the exchange rate. There is a large literature[3] that tests UIP using the following simple regression that fits within our analysis:

$$s_{i+\Delta} - s_i = \beta_1 + \beta_2(int_i - int_i^*) + u_i,$$

where s_i is the logarithm of the exchange rate at time i, int_i is the interest rate on a domestic bond of maturity Δ and int_i^* is the interest rate on a foreign country bond of the same maturity. The null hypothesis of UIP is $H_0 : \beta_1 = 0$ and $\beta_2 = 1$ and the null hypothesis of a milder version of UIP is $H_0 : \beta_2 = 1$ which is our focus here. Diez de los Rios and Sentana (2011) examine tests of (mild) UIP in a continuous time framework and focus on using high sampling frequencies in the data which we do here.

For our empirical illustration, we examine the UIP hypothesis for the US-Japan case, where the domestic country is the US and the foreign country is Japan. The data for the exchange rate and interest rates are obtained from Refinitiv Workplace (formerly Thomson-Reuters) and are daily observations. The sample period is from 1991/01/02 to 2022/11/01 giving a span of up to 30 years depending on the bond maturity horizon. For the interest rates int_t and int_t^*, we use the yields on the benchmark government bonds of the domestic (US) and the foreign country (Japan) with two different maturities. The first is a 2-year bond. The second is a 10-year bond following Chinn and Meredith (2004).

Because exchange rates and interest rates are highly persistent, we use a t-statistic based on $G(\widehat{\beta})$ for testing $H_0 : \beta_2 = 1$ which is appropriate for the case of a cointegration regression. We use fixed-b critical values calculated with (3.11) using coefficients from the G-cointegration line of Panel B of Table 3B.1. To accommodate possible endogeneity, we also use a t-statistics based on $\widetilde{W}^*$ of IM-OLS using critical values given by (3.13). We report results for four bandwidth selection

---

[3]See Engel et al. (2022) for a recent empirical paper and Engel (2014) for a broader survey of empirical work testing UIP.

rules (AD, SPJ, NW, CRT). For each bandwidth method, we calculate standard errors, the b ratio and the t-statistic for both the OLS and IM-OLS cases. We also provide 95% confidence intervals for β_2 using fixed-b critical values and $N(0, 1)$ critical values.

Table 3B.2 gives results for the 2-year bond maturity and Table 3B.3 gives results for the 10-year bond maturity. In each table results are given for daily, weekly, monthly and quarterly sampling frequencies. Both OLS and IM-OLS are considered. Looking at the daily frequency results in Table 3B.2, we see that the t-statistics based on $G(\widehat{\beta})$ using AD and SPJ are -2.189 and -2.473 respectively, rejecting the UIP hypothesis at the 5% level when using the normal critical value. However, the fixed-b critical values for the t-statistic using $\widehat{b}_{AD}$ and $\widehat{b}_{SPJ}$ are ±4.053 and ±5.213 respectively. Therefore, the UIP hypothesis is not rejected when using fixed-b critical values. The same finding is made with the t-statistics based on $\widetilde{W}^*$ of IM-OLS. Interestingly $\widehat{\beta}_2$ and $\widetilde{\beta}_2$ are stable across the frequencies but have opposite signs with $\widehat{\beta}_2$ negative and $\widetilde{\beta}_2$ close to 0.5.

Recall that $G(\widehat{\beta})$ and $\widetilde{W}^*$ use the same $\widehat{b}$ values for each bandwidth method. Notice that $\widehat{b}_{AD}$ and $\widehat{b}_{SPJ}$ yield wider confidence intervals with fixed-b critical values compared to normal critical values. These patterns hold across sampling frequencies. The values of the $\widehat{b}_{AD}$ and $\widehat{b}_{SPJ}$ are large (0.479 to 0.777) and each are roughly stable across sampling frequencies as expected given the finite sample simulations. Also, as expected, $\widehat{b}_{SPJ}$ is larger than $\widehat{b}_{AD}$. Large values of $\widehat{b}_{AD}$ and $\widehat{b}_{SPJ}$ are an indication of high persistence in the regression error. Using normal critical values with large bandwidth ratios would lead to misleading inference (type 1 error well above the nominal level). The NW bandwidths are substantially smaller than AD and SPJ. Table 3B.2 shows that $\widehat{b}_{NW}$ is very small for high sampling frequencies (0.007) and is increasing to 0.11 at the quarterly frequency, the same pattern that we observe in the finite sample simulations. This illustrates that the NW bandwidth is not high frequency compatible as argued by CLP, and confidence intervals based on NW can be misleadingly tight. By construction b_{CRT} is the same across sampling frequencies. However, the value of $b_{CRT} = 0.066$ is substantially smaller than $\widehat{b}_{AD}$ and $\widehat{b}_{SPJ}$ suggesting confidence intervals using b_{CRT} can be too tight relative to the persistence in the regression errors.

Table 3B.3 gives results for the 10-year benchmark government bond. While the point estimates

of β_2 are now larger than 1 (about 1.4) for both OLS and IM-OLS, bandwidth patterns and inference conclusions are similar to Table 3B.2. The $\widehat{b}$ values continue to be large and roughly stable across sampling frequencies. UIP cannot be rejected for either test.

3.7 Conclusion

In this paper we develop fixed-b asymptotic results for HAR Wald statistics for regressions with high frequency data in the continuous time framework of Chang et al. (2023). We find that the fixed-b limits of the HAR Wald tests in stationary high frequency regressions are the same as the standard fixed-b limits in Kiefer and Vogelsang (2005). For cointegrating high frequency regressions the fixed-b limits generally have non-pivotal limits. For the special case where the stochastic processes in the continuous time regression follow Brownian motions and the regressors are independent of the errors, the fixed-b limits are pivotal and are the same as those obtained by Bunzel (2006) in discrete time settings. We also analyzed a Wald statistic from Vogelsang and Wagner (2014) using their IM-OLS estimator and obtained fixed-b limits that are same as the limits in Vogelsang and Wagner (2014). Our results in conjunction with the results in Pellatt and Sun (2023) for orthonormal series approaches (including the cointegration estimator of Hwang and Sun (2018)) establish that fixed-b (more generally fixed-smoothing) critical values are high frequency compatible.

In a simulation study where data is generated according to OU processes we find that the Andrews (1991), Sun et al. (2008) and CLP's CRT bandwidths tend to perform well not only at high frequencies (assuming the persistence in the data is not too strong relative to the span), but give null rejections that are remarkably stable across sampling ranging from high to low frequencies. This stability holds for data with strong and mild persistence. We show that the source of this stability in null rejections is stability in bandwidth sample size ratios (b-values) across sampling frequencies. This stability holds by construction for the CRT bandwidth. It is more surprising this stability holds for the Andrews (1991), Sun et al. (2008) bandwidths and we provide a simple theoretical explanation.

We also report some power results in our simulations and find that for persistent series, power

is stable across sampling frequencies. In contrast, power falls as sampling frequency decreases for mildly persistent series. Therefore, we can recommend that practitioners use data sampled at higher frequencies.

BIBLIOGRAPHY

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

Bunzel, H. (2006). Fixed-b asymptotics in single-equation cointegration models with endogenous regressors. *Econometric Theory*, 22:743–755.

Chang, Y., Lu, Y., and Park, J. Y. (2023). Understanding regressions with observations collected at high frequency over long span. Working paper, School of Economics, University of Sydney.

Chinn, M. D. and Meredith, G. (2004). Monetary policy and long-horizon uncovered interest parity. *IMF staff papers*, 51(3):409–430.

Diez de los Rios, A. and Sentana, E. (2011). Testing uncovered interest parity: A continuous-time approach. *International Economic Review*, 52(4):1215–1251.

Engel, C. (2014). Exchange rates and interest parity. *Handbook of international economics*, 4:453–522.

Engel, C., Kazakova, K., Wang, M., and Xiang, N. (2022). A reconsideration of the failure of uncovered interest parity for the US dollar. *Journal of International Economics*, 136:103602.

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

Hashimzade, N. and Vogelsang, T. J. (2008). Fixed-*b* asymptotic approximation of the sampling behavior of nonparametric spectral density estimators. *Journal of Time Series Analysis*, 29:142–162.

Hwang, J. and Sun, Y. (2018). Simple, robust, and accurate F and t tests in cointegrated systems. *Econometric Theory*, 34(5):949–984.

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

Kim, J. and Park, J. ((2017)). Asymptotics for recurrent diffusions with application to high frequency regression. *Journal of Econometrics*, 196:37–54.

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.

Lu, Y. and Park, J. (2019). Estimation of longrun variance of continuous time stochastic process using descrete sample. *Journal of Econometrics*, 210:236–267.

Newey, W. K. and West, K. D. (1994). Automatic lag selection in covariance estimation. *Review of Economic Studies*, 61:631–654.

Pellatt, D. and Sun, Y. (2023). Asymptotic F test in regressions with observations collected at high frequency over long span. *Journal of Econometrics*, 235(2):1281–1309.

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

Vogelsang, T. and Wagner, M. (2014). Integrated modified OLS estimation and fixed-b inference for cointegrating regressions. *Journal of Econometrics*, 178(2):741–760.

Vogelsang, T. J. ((2011)). Heteroskedasticity, autocorrelation, and spatial correlation robust inference in linear panel models with fixed-effects. Working paper, Department of Economics, Michigan State University.

Vogelsang, T. J. ((2012)). Heteroskedasticity, autocorrelation, and spatial correlation robust inference in linear panel models with fixed-effects. *Journal of Econometrics*, 116:303–319.

APPENDIX 3A

PROOFS

We provide proofs for the case of the Bartlett kernel given the simple form of the corresponding long run variance estimators. Algebra is similar for other kernels. When writing formulas for $G(\widehat{\beta})$ and $H(\widehat{\beta})$ note that, when the null hypothesis $H_0 : R\beta = r$ is true, we can write

$$R\widehat{\beta} - r = R\widehat{\beta} - R\beta = R\left(\widehat{\beta} - \beta\right)$$

so that $R\widehat{\beta} - r$ can be replaced with $R\left(\widehat{\beta} - \beta\right)$ in the formulas for $G(\widehat{\beta})$ and $H(\widehat{\beta})$. Similarly for IM-OLS, we can replace $R\widehat{\theta} - r$ (with suitably augmented R and r) with $R\left(\widehat{\theta} - \theta\right)$.

Proof of Theorem 3.1: Under Assumptions 3.2, 3.3 and 3.4 as $\delta \to 0$ and $T \to \infty$, we have the following results.

$$\frac{1}{n} \sum_{i=1}^{[rn]} x_i x_i' = T^{-1} \int_0^{rT} X_t X_t' dt + o_p(1) \xrightarrow{p} rQ,$$

$$\sqrt{T} \frac{1}{n} \sum_{i=1}^{[rn]} x_i u_i = \frac{\sqrt{\delta}}{\sqrt{n}} \sum_{i=1}^{[rn]} x_i u_i = T^{-1/2} \int_0^{rT} X_t U_t dt + o_p(1) \Rightarrow \Lambda W_k(r).$$

Using these limits gives

$$\sqrt{T}\left(\widehat{\beta} - \beta\right) = \left(\frac{1}{n} \sum_{i=1}^{n} x_i x_i'\right)^{-1} \sqrt{T} \frac{1}{n} \sum_{i=1}^{n} x_i u_i \Rightarrow Q^{-1} \Lambda W_k(1).$$

Next, we derive the fixed-b limit of $\widehat{\Omega}$ upon appropriate scaling. For the Bartlett kernel, we can write

$$\widehat{\Omega}_n = \frac{2}{M_n n} \sum_{i=1}^{n-1} \widehat{S}_i^v (\widehat{S}_i^v)' - \frac{1}{M_n n} \sum_{i=1}^{n-M_n-1} \left(\widehat{S}_i^v (\widehat{S}_{i+M_n}^v)' + \widehat{S}_{i+M_n}^v (\widehat{S}_i^v)'\right),$$

where

$$\widehat{S}_{[rn]}^v = \sum_{j=1}^{[rn]} x_j \widehat{u}_j.$$

105

Scaling $\widehat{S}^v_{[rn]}$ by $T^{1/2}/n$ gives

$$T^{1/2}\frac{1}{n}\widehat{S}^v_{[rn]} = T^{1/2}\frac{1}{n}\sum_{j=1}^{[rn]}x_j\widehat{u}_j = T^{1/2}\frac{1}{n}\sum_{j=1}^{[rn]}x_j(y_j - x'_j\widehat{\beta})$$

$$= T^{1/2}\frac{1}{n}\sum_{j=1}^{[rn]}x_j(x'_j\beta + u_j - x'_j\widehat{\beta}) = T^{1/2}\frac{1}{n}\sum_{j=1}^{[rn]}x_j(u_j - x'_j(\widehat{\beta} - \beta))$$

$$= T^{1/2}\frac{1}{n}\sum_{j=1}^{[rn]}x_j u_j - T^{1/2}\frac{1}{n}\sum_{j=1}^{[rn]}x_j x'_j(\widehat{\beta} - \beta)$$

$$= T^{1/2}\frac{1}{n}\sum_{j=1}^{[rn]}x_j u_j - \frac{1}{n}\sum_{j=1}^{[rn]}x_j x'_j \sqrt{T}(\widehat{\beta} - \beta)$$

$$= T^{-1/2}\int_0^{rT} X_t U_t dt + o_p(1) - \left(T^{-1}\int_0^{rT}X_t X'_t dt + o_p(1)\right)\sqrt{T}(\widehat{\beta} - \beta)$$

$$\Rightarrow \Lambda W_k(r) - rQQ^{-1}\Lambda W_k(1) \equiv \Lambda B_k(r),$$

as $\delta \to 0$ and $T \to \infty$. Scaling $\widehat{\Omega}_n$ by $\delta = T/n$ gives

$$\delta\widehat{\Omega}_n = T\frac{1}{n}\widehat{\Omega}_n = \frac{2}{M_n n}T\frac{1}{n}\sum_{i=1}^{n-1}\widehat{S}^v_i(\widehat{S}^v_i)' - \frac{1}{M_n n}T\frac{1}{n}\sum_{i=1}^{n-M_n-1}[\widehat{S}^v_i(\widehat{S}^v_{i+M_n})' + \widehat{S}^v_{i+M_n}(\widehat{S}^v_i)']$$

$$= \frac{2}{M_n}\sum_{i=1}^{n-1}T^{1/2}\frac{1}{n}\widehat{S}^v_i(T^{1/2}\frac{1}{n}\widehat{S}^v_i)' - \frac{1}{M_n}\sum_{i=1}^{n-M_n-1}[T^{1/2}\frac{1}{n}\widehat{S}^v_i T^{1/2}\frac{1}{n}(\widehat{S}^v_{i+M_n})' + T^{1/2}\frac{1}{n}\widehat{S}^v_{i+M_n}(T^{1/2}\frac{1}{n}\widehat{S}^v_i)']$$

$$= \frac{2}{bn}\sum_{i=1}^{n-1}T^{1/2}\frac{1}{n}\widehat{S}^v_i(T^{1/2}\frac{1}{n}\widehat{S}^v_i)' - \frac{1}{bn}\sum_{i=1}^{n-bn-1}[T^{1/2}\frac{1}{n}\widehat{S}^v_i(T^{1/2}\frac{1}{n}\widehat{S}^v_{i+bn})' + T^{1/2}\frac{1}{n}\widehat{S}^v_{i+bn}(T^{1/2}\frac{1}{n}\widehat{S}^v_i)']$$

$$\Rightarrow \Lambda\left(\frac{2}{b}\int_0^1 B_k(r)B_k(r)'dr - \frac{1}{b}\int_0^{1-b}B_k(r)B_k(r+b)'dr - \frac{1}{b}\int_0^{1-b}B_k(r+b)B_k(r)'dr\right)\Lambda'$$

$$\equiv \Lambda P_k(b)\Lambda'.$$

The result for $H(\widehat{\beta})$ as $\delta \to 0$ and $T \to \infty$ is straightforward to obtain as follows:

$$H(\widehat{\beta}) = R\left(\widehat{\beta} - \beta\right)'\left[R\left(\sum_{i=1}^n x_i x'_i\right)^{-1}n\widehat{\Omega}_n\left(\sum_{i=1}^n x_i x'_i\right)^{-1}R'\right]^{-1}R\left(\widehat{\beta} - \beta\right)$$

$$= R\sqrt{T}\left(\widehat{\beta} - \beta\right)'\left[R\left(\frac{1}{n}\sum_{i=1}^n x_i x'_i\right)^{-1}T\frac{1}{n}\widehat{\Omega}_n\left(\frac{1}{n}\sum_{i=1}^n x_i x'_i\right)^{-1}R'\right]^{-1}R\sqrt{T}\left(\widehat{\beta} - \beta\right)$$

106

$$= R\sqrt{T}\left(\widehat{\beta} - \beta\right)'\left[R\left(T^{-1}\int_0^T X_t X_t'dt + o_p(1)\right)^{-1}T\frac{1}{n}\widehat{\Omega}_n\left(T^{-1}\int_0^T X_t X_t'dt + o_p(1)\right)^{-1}R'\right]^{-1}R\sqrt{T}\left(\widehat{\beta} - \beta\right)$$

$$\Rightarrow [RQ^{-1}\Lambda W_k(1)]'^{-1}[RQ^{-1}\Lambda P_k(b)\Lambda'Q^{-1}R']^{-1}[RQ^{-1}\Lambda W_k(1)]$$

$$= [\Lambda^* W_q(1)]'[\Lambda^* P_q(b)\Lambda^{*'}]^{-1}[\Lambda^* W_q(1)] = W_q(1)'P_q(b)^{-1}W_q(1).$$

Note that Λ^* is the square root matrix of $RQ^{-1}\Lambda\Lambda'Q^{-1}R'$.

Proof of Theorem 3.2: We first give the derivation for $H(\widehat{\beta})$ followed by the derivation for $G(\widehat{\beta})$. Recall that $X_t' = [1 \ \widetilde{X}_t]$, $x_i' = [1 \ \widetilde{x}_i]$ and $U^\circ(s) = \lambda_u w_u(s)$. Under Assumptions 3.2, 3.5 and 3.6 we have these results for the partial sums:

$$\frac{1}{n}\sum_{i=1}^{[rn]}\Lambda_T^{-1}x_i x_i'\Lambda_T^{-1'} = \frac{1}{T}\int_0^{rT}\Lambda_T^{-1}X_t X_t'\Lambda_T^{-1'}dt + o_p(1) \Rightarrow \int_0^r X^\circ(s)X^\circ(s)'ds,$$

$$T^{1/2}\frac{1}{n}\sum_{i=1}^{[rn]}\Lambda_T^{-1}x_i u_i = \frac{1}{\sqrt{T}}\int_0^{rT}\Lambda_T^{-1}X_t U_t dt + o_p(1) \Rightarrow \int_0^r X^\circ(s)dU^\circ(s) = \lambda_u\int_0^r X^\circ(s)dw_u(s).$$

Scaling $\left(\widehat{\beta} - \beta\right)$ by $T^{1/2}\Lambda_T'$ gives

$$T^{1/2}\Lambda_T'\left(\widehat{\beta} - \beta\right) = \left(\frac{1}{n}\sum_{i=1}^n\Lambda_T^{-1}x_i x_i'\Lambda_T^{-1'}\right)^{-1}\left(T^{1/2}\frac{1}{n}\sum_{i=1}^n\Lambda_T^{-1}x_i u_i\right)$$

$$\Rightarrow \left(\int_0^1 X^\circ(s)X^\circ(s)'ds\right)^{-1}\lambda_u\int_0^1 X^\circ(s)dw_u(s)$$

$$= \lambda_u\left(\int_0^1 X^\circ(s)X^\circ(s)'ds\right)^{-1}\int_0^1 X^\circ(s)dw_u(s) \equiv \lambda_u C.$$

Next, we need to determine the scaling for $\widehat{\Omega}_n$. First, scale $\widehat{S}_{[rn]}^v$ by $T^{1/2}\frac{1}{n}\Lambda_T^{-1}$ to give

$$T^{1/2}\frac{1}{n}\Lambda_T^{-1}\widehat{S}_{[rn]}^v = T^{1/2}\frac{1}{n}\Lambda_T^{-1}\sum_{i=1}^{[rn]}x_i\widehat{u}_i = T^{1/2}\frac{1}{n}\Lambda_T^{-1}\sum_{i=1}^{[rn]}x_i(y_i - x_i'\widehat{\beta})$$

$$= T^{1/2}\frac{1}{n}\Lambda_T^{-1}\sum_{i=1}^{[rn]}x_i(x_i'\beta + u_i - x_i'\widehat{\beta}) = T^{1/2}\frac{1}{n}\Lambda_T^{-1}\sum_{i=1}^{[rn]}x_i(u_i - x_i'(\widehat{\beta} - \beta))$$

$$= T^{1/2}\frac{1}{n}\Lambda_T^{-1}\sum_{i=1}^{[rn]}x_i u_i - T^{1/2}\frac{1}{n}\Lambda_T^{-1}\sum_{i=1}^{[rn]}x_i x_i'(\widehat{\beta} - \beta)$$

$$= T^{1/2} \frac{1}{n} \sum_{i=1}^{[rn]} \Lambda_T^{-1} x_i u_i - \frac{1}{n} \sum_{i=1}^{[rn]} \Lambda_T^{-1} x_i x_i' \Lambda_T^{-1\prime} T^{1/2} \Lambda_T'(\widehat{\beta} - \beta)$$

$$\Rightarrow \int_0^r X^\circ(s) \lambda_u dw_u(s) - \left(\int_0^r X^\circ(s) X^\circ(s)' ds \right) \left[\lambda_u \left(\int_0^1 X^\circ(s) X^\circ(s)' ds \right)^{-1} \int_0^1 X^\circ(s) dw_u(s) \right]$$

$$= \lambda_u \left(\int_0^r X^\circ(s) dw_u(s) - \int_0^r X^\circ(s) X^\circ(s)' ds \left(\int_0^1 X^\circ(s) X^\circ(s)' ds \right)^{-1} \int_0^1 X^\circ(s) dw_u(s) \right) \equiv \lambda_u B_H(r).$$

Next, scale $\widehat{\Omega}_n$ as $\delta \Lambda_T^{-1} \widehat{\Omega}_n \Lambda_T^{-1\prime} = (T/n) \Lambda_T^{-1} \widehat{\Omega}_n \Lambda_T^{-1\prime}$ to give

$$T \frac{1}{n} \Lambda_T^{-1} \widehat{\Omega}_n \Lambda_T^{-1\prime} = \frac{2}{M_n n} T \frac{1}{n} \Lambda_T^{-1} \sum_{i=1}^{n-1} \widehat{S}_i^v (\widehat{S}_i^v)' \Lambda_T^{-1\prime} - \frac{1}{M_n n} T \frac{1}{n} \Lambda_T^{-1} \sum_{i=1}^{n-M_n-1} [\widehat{S}_i^v (\widehat{S}_{i+M_n}^v)' + \widehat{S}_{i+M_n}^v (\widehat{S}_i^v)'] \Lambda_T^{-1\prime}$$

$$= \frac{2}{bn} \sum_{i=1}^{n-1} T^{1/2} \frac{1}{n} \Lambda_T^{-1} \widehat{S}_i^v T^{1/2} \frac{1}{n} (\widehat{S}_i^v)' \Lambda_T^{-1\prime} - \frac{1}{bn} \sum_{t=1}^{n-bn-1} [T^{1/2} \frac{1}{n} \Lambda_T^{-1} \widehat{S}_i^v T^{1/2} \frac{1}{n} (\widehat{S}_{i+bn}^v)' \Lambda_T^{-1\prime}$$

$$+ T^{1/2} \frac{1}{n} \Lambda_T^{-1} \widehat{S}_{i+bn}^v T^{1/2} \frac{1}{n} (\widehat{S}_i^v)' \Lambda_T^{-1\prime}]$$

$$= \frac{2}{bn} \sum_{i=1}^{n-1} T^{1/2} \frac{1}{n} \Lambda_T^{-1} \widehat{S}_i^v T^{1/2} \frac{1}{n} (\widehat{S}_i^v)' \Lambda_T^{-1\prime} - \frac{1}{bn} \sum_{t=1}^{n-bn-1} [T^{1/2} \frac{1}{n} \Lambda_T^{-1} \widehat{S}_i^v T^{1/2} \frac{1}{n} (\widehat{S}_{i+bn}^v)' \Lambda_T^{-1\prime}$$

$$+ T^{1/2} \frac{1}{n} \Lambda_T^{-1} \widehat{S}_{i+bn}^v T^{1/2} \frac{1}{n} (\widehat{S}_i^v)' \Lambda_T^{-1\prime}]$$

$$\Rightarrow \frac{2}{b} \int_0^1 \lambda_u B_H(r) \lambda_u B_H(r)' dr - \frac{1}{b} \int_0^{1-b} \lambda_u B_H(r) \lambda_u B_H(r+b)' dr$$

$$- \frac{1}{b} \int_0^{1-b} \lambda_u B_H(r+b) \lambda_u B_H(r)' dr$$

$$= \lambda_u^2 \left(\frac{2}{b} \int_0^1 B_H(r) B_H(r)' dr - \frac{1}{b} \int_0^{1-b} B_H(r) B_H(r+b)' dr - \frac{1}{b} \int_0^{1-b} B_H(r+b) B_H(r)' dr \right) \equiv \lambda_u^2 P_H(b).$$

The result for $H(\widehat{\beta})$ as $\delta \to 0$ and $T \to \infty$ is straightforward to obtain as follows:

$$H(\widehat{\beta}) = \left(R \left(\widehat{\beta} - \beta \right) \right)' \left[R \left(\sum_{i=1}^n x_i x_i' \right)^{-1} n \widehat{\Omega}_n \left(\sum_{i=1}^n x_i x_i' \right)^{-1} R' \right]^{-1} R(\widehat{\beta} - \beta)$$

$$= \left(R \Lambda_T^{-1\prime} \Lambda_T' \left(\widehat{\beta} - \beta \right) \right)' \left[R \Lambda_T^{-1\prime} \left(\sum_{i=1}^n \Lambda_T^{-1} x_i x_i' \Lambda_T^{-1\prime} \right)^{-1} n \Lambda_T^{-1} \widehat{\Omega}_n \Lambda_T^{-1\prime} \left(\sum_{i=1}^n \Lambda_T^{-1} x_i x_i' \Lambda_T^{-1\prime} \right)^{-1} \Lambda_T^{-1} R' \right]^{-1}$$

$$\times R \Lambda_T^{-1\prime} \Lambda_T'(\widehat{\beta} - \beta)$$

$$= \left(\Lambda_T^R R \Lambda_T^{-1\prime} \Lambda_T' \left(\widehat{\beta} - \beta \right) \right)'$$

$$\times \left[\Lambda_T^R R \Lambda_T^{-1\prime} \left(\sum_{i=1}^n \Lambda_T^{-1} x_i x_i' \Lambda_T^{-1\prime} \right)^{-1} n \Lambda_T^{-1} \widehat{\Omega}_n \Lambda_T^{-1\prime} \left(\sum_{i=1}^n \Lambda_T^{-1} x_i x_i' \Lambda_T^{-1\prime} \right)^{-1} \Lambda_T^{-1} R' \Lambda_T^{R\prime} \right]^{-1}$$

$$\times \Lambda_T^R R \Lambda_T^{-1\prime} \Lambda_T' \left(\widehat{\beta} - \beta \right)$$

$$= \left(\Lambda_T^R T^{-1/2} R \Lambda_T^{-1\prime} T^{1/2} \Lambda_T' \left(\widehat{\beta} - \beta \right) \right)'$$

$$\times \left[\Lambda_T^R T^{-1/2} R \Lambda_T^{-1\prime} \left(\frac{1}{n} \sum_{i=1}^n \Lambda_T^{-1} x_i x_i' \Lambda_T^{-1\prime} \right)^{-1} T \frac{1}{n} \Lambda_T^{-1} \widehat{\Omega}_n \Lambda_T^{-1\prime} \left(\frac{1}{n} \sum_{i=1}^n \Lambda_T^{-1} x_i x_i' \Lambda_T^{-1\prime} \right)^{-1} \Lambda_T^{-1} R' T^{-1/2} \Lambda_T^{R\prime} \right]^{-1}$$

$$\times \Lambda_T^R T^{-1/2} R \Lambda_T^{-1\prime} T^{1/2} \Lambda_T' (\widehat{\beta} - \beta)$$

$$\Rightarrow (R^* \lambda_u C)' \left[R^* \left(\int_0^1 X^\circ(s) X^\circ(s)' ds \right)^{-1} \lambda_u^2 P_H(b) \left(\int_0^1 X^\circ(s) X^\circ(s)' ds \right)^{-1} R^{*\prime} \right]^{-1} R^* \lambda_u C$$

$$= (R^* C)' \left[R^* Q_\circ^{-1} P_H(b) Q_\circ^{-1} R^{*\prime} \right]^{-1} R^* C.$$

Note that we use (3.5) for the limit of $\Lambda_T^R T^{-1/2} R \Lambda_T^{-1\prime}$.

The derivation for $G(\widehat{\beta})$ is similar and only requires the fixed-b limit of $\widehat{\omega}_n^2$. Similar to $\widehat{\Omega}_n$, for the case of the Bartlett kernel $\widehat{\omega}_n^2$ can be written as a function of $\widehat{S}_i^u = \sum_{j=1}^{[rn]} \widehat{u}_j$ as

$$\widehat{\omega}_n^2 = \frac{2}{M_n n} \sum_{i=1}^{n-1} \left(\widehat{S}_i^u \right)^2 - \frac{2}{M_n n} \sum_{i=1}^{n-M_n-1} \widehat{S}_i^u \widehat{S}_{i+M_n}^u.$$

To determine the scaling needed for $\widehat{\omega}_n^2$, we scale $\widehat{S}_i^u$ by $T^{1/2}/n$ to give

$$T^{1/2} \frac{1}{n} \widehat{S}_{[rT]}^u = T^{1/2} \frac{1}{n} \sum_{j=1}^{[rn]} \widehat{u}_j = T^{1/2} \frac{1}{n} \sum_{j=1}^{[rn]} (y_j - x_j' \widehat{\beta})$$

$$= T^{1/2} \frac{1}{n} \sum_{j=1}^{[rn]} (x_j' \beta + u_j - x_j' \widehat{\beta}) = T^{1/2} \frac{1}{n} \sum_{j=1}^{[rn]} (u_j - x_j'(\widehat{\beta} - \beta))$$

$$= T^{1/2} \frac{1}{n} \sum_{j=1}^{[rn]} u_j - T^{1/2} \frac{1}{n} \sum_{j=1}^{[rn]} x_j'(\widehat{\beta} - \beta) = T^{1/2} \frac{1}{n} \sum_{j=1}^{[rn]} u_j - \frac{1}{n} \sum_{j=1}^{[rn]} x_j' \Lambda_T^{-1\prime} T^{1/2} \Lambda_T'(\widehat{\beta} - \beta)$$

$$= T^{-1/2} \int_0^{rT} U_t dt + o_p(1) - \left(\frac{1}{T} \int_0^{rT} X_t' \Lambda_T^{-1'} dt + o_p(1) \right) T^{1/2} \Lambda_T'(\widehat{\beta} - \beta)$$

$$\Rightarrow \lambda_u w_u(r) - \lambda_u \int_0^r X^\circ(s)' ds \left(\int_0^1 X^\circ(s) X^\circ(s)' ds \right)^{-1} \int_0^1 X^\circ(s) dw_u(s)$$

$$= \lambda_u \left(w_u(r) - \int_0^r X^\circ(s)' ds \left(\int_0^1 X^\circ(s) X^\circ(s)' ds \right)^{-1} \int_0^1 X^\circ(s) dw_u(s) \right) \equiv \lambda_u B_G(r).$$

Next, scaling $\widehat{\omega}_n^2$ by $\delta = T/n$ gives

$$\delta \widehat{\omega}_n^2 = T \frac{1}{n} \widehat{\omega}_n^2 = \frac{2}{M_n n} T \frac{1}{n} \sum_{i=1}^{n-1} \left(\widehat{S}_i^u \right)^2 - \frac{2}{M_n n} T \frac{1}{n} \sum_{i=1}^{n-M_n-1} \widehat{S}_i^u \widehat{S}_{i+M_n}^u$$

$$= \frac{2}{bn} \sum_{i=1}^{n-1} \left(T^{1/2} \frac{1}{n} \widehat{S}_i^u \right) (T^{1/2} \frac{1}{n} \widehat{S}_i^u) - \frac{2}{bn} \sum_{i=1}^{n-bn-1} \left(T^{1/2} \frac{1}{n} \widehat{S}_i^u \right) \left(T^{1/2} \frac{1}{n} \widehat{S}_{i+bn}^u \right)$$

$$\Rightarrow \frac{2}{b} \int_0^1 \lambda_u B_G(r) \lambda_u B_G(r) dr - \frac{2}{b} \int_0^{1-b} \lambda_u B_G(r) \lambda_u B_G(r+b) dr$$

$$= \lambda_u^2 \left(\frac{2}{b} \int_0^1 B_G(r)^2 dr - \frac{2}{b} \int_0^{1-b} B_G(r) B_G(r+b) dr \right) \equiv \lambda_u^2 P_G(b).$$

The result for $G(\widehat{\beta})$ follows using similar arguments as for $H(\widehat{\beta})$ as $\delta \to 0$ and $T \to \infty$:

$$G(\widehat{\beta}) = \left(R(\widehat{\beta} - \beta) \right)' \left[\widehat{\omega}_n^2 R \left(\sum_{t=1}^T x_i x_i' \right)^{-1} R' \right]^{-1} R \left(\widehat{\beta} - \beta \right)$$

$$= \left(R \Lambda_T^{-1'} \Lambda_T' \left(\widehat{\beta} - \beta \right) \right)' \left[\widehat{\omega}_n^2 R \Lambda_T^{-1'} \left(\sum_{t=1}^T \Lambda_T^{-1} x_i x_i' \Lambda_T^{-1'} \right)^{-1} \Lambda_T^{-1} R' \right]^{-1} R \Lambda_T^{-1'} \Lambda_T' \left(\widehat{\beta} - \beta \right)$$

$$= \left(\Lambda_T^R R \Lambda_T^{-1'} \Lambda_T' \left(\widehat{\beta} - \beta \right) \right)' \left[\widehat{\omega}_n^2 \Lambda_T^R R \Lambda_T^{-1'} \left(\sum_{t=1}^T \Lambda_T^{-1} x_i x_i' \Lambda_T^{-1'} \right)^{-1} \Lambda_T^{-1} R' \Lambda_T^{R'} \right]^{-1} \Lambda_T^R R \Lambda_T^{-1'} \Lambda_T' \left(\widehat{\beta} - \beta \right)$$

$$= \left(\Lambda_T^R T^{-1/2} R \Lambda_T^{-1'} T^{1/2} \Lambda_T' \left(\widehat{\beta} - \beta \right) \right)' \left[T \frac{1}{n} \widehat{\omega}_n^2 \Lambda_T^R T^{-1/2} R \Lambda_T^{-1'} \left(\frac{1}{n} \sum_{t=1}^T \Lambda_T^{-1} x_i x_i' \Lambda_T^{-1'} \right)^{-1} \Lambda_T^{-1} R' T^{-1/2} \Lambda_T^{R'} \right]^{-1}$$

$$\times \Lambda_T^R T^{-1/2} R \Lambda_T^{-1'} T^{1/2} \Lambda_T' \left(\widehat{\beta} - \beta \right)$$

$$\Rightarrow (R^* \lambda_u C)' \left[\lambda_u^2 P_G(b) R^* \left(\int_0^1 X^\circ(s) X^\circ(s)' ds \right)^{-1} R^{*'} \right]^{-1} R^* \lambda_u C = (R^* C)' \left[P_G(b) R^* Q_\circ^{-1} R^{*'} \right]^{-1} R^* C.$$

Proof of Theorem 3.3: Define the scaling matrix

$$
A_{IM} = \begin{bmatrix} T^{-1/2} & 0 & 0 \\ 0 & T^{-1}I_{k-1} & 0 \\ 0 & 0 & I_{k-1} \end{bmatrix}.
$$

Following Vogelsang and Wagner (2014) let A_R be a $q \times q$ matrix such that A_R^{-1} exists and satisfies $\lim_{T\to\infty} R_T^* = R^*$ where $R_T^* = A_R^{-1}RA_{IM}$. Replacing $R\widetilde{\theta} - r$ with $R\left(\widetilde{\theta} - \theta\right)$ we rewrite $\widetilde{W}^*$ using A_{IM} and A_R^{-1}:

$$
\widetilde{W}^* = [A_R^{-1}RA_{IM}A_{IM}^{-1}(\widetilde{\theta} - \theta)]'
$$

$$
\times \left[\widetilde{\lambda}_{u\cdot\widetilde{x}}^{2*}A_R^{-1}RA_{IM}\left(A_{IM}\sum_{i=1}^{n}x_i^*x_i^{*\prime}A_{IM}\right)^{-1}\left(A_{IM}\sum_{i=1}^{n}c_ic_i'A_{IM}\right)\left(A_{IM}\sum_{i=1}^{n}x_i^*x_i^{*\prime}A_{IM}\right)^{-1}A_{IM}R'A_R^{-1\prime}\right]^{-1}
$$

$$
\times [A_R^{-1}RA_{IM}A_{IM}^{-1}(\widetilde{\theta} - \theta)].
$$

Noting that $A_R^{-1}RA_{IM}$ is R_T^* by definition and letting $B_{IM} = T^{1/2}A_{IM}$, we have upon scaling inside the middle inverse term:

$$
\widetilde{W}^* = [R_T^*A_{IM}^{-1}(\widetilde{\theta} - \theta)]'
$$

$$
\times \left[\frac{T}{n}\widetilde{\lambda}_{u\cdot\widetilde{x}}^{2*}R_T^*\left(\frac{1}{n}\sum_{i=1}^{n}B_{IM}\frac{1}{n}x_i^*x_i^{*\prime}\frac{1}{n}B_{IM}\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\left(B_{IM}\frac{1}{n}\right)\frac{1}{n}c_ic_i'\frac{1}{n}\left(\frac{1}{n}B_{IM}\right)\right)\right.
$$

$$
\left.\times\left(\frac{1}{n}\sum_{i=1}^{n}B_{IM}\frac{1}{n}x_i^*x_i^{*\prime}\frac{1}{n}B_{IM}\right)^{-1}R_T^{*\prime}\right]^{-1}[R_T^*A_{IM}^{-1}(\widetilde{\theta} - \theta)].
$$

Under Assumption 3.2^{**} and 3.8, we have the following results for partial sums:

$$
\frac{1}{n}\sum_{i=1}^{[rn]}u_i = \frac{1}{T}\int_0^{rT}U_t dt + o_p(T^{-1/2}), \tag{3A.1}
$$

$$
\frac{1}{n}\sum_{i=1}^{[rn]}\widetilde{x}_i = \frac{1}{T}\int_0^{rT}\widetilde{X}_t dt + o_p(1), \tag{3A.2}
$$

$$
\frac{1}{n}\widetilde{x}_{[rn]} = \frac{1}{n}\sum_{i=1}^{[rn]}v_i^{\widetilde{x}} = \frac{1}{T}\int_0^{rT}V_t^{\widetilde{x}}dt + o_p(T^{-1/2}) = \frac{1}{T}\widetilde{X}_{rT} + o_p(T^{-1/2}), \tag{3A.3}
$$

111

Recall that $x^*_{[rn]} = \begin{bmatrix} [rn] \\ \sum_{i=1}^{[rn]} \widetilde{x}_i \\ \widetilde{x}_{[rn]} \end{bmatrix}$. Scaling $x^*_{[rn]}$ by B_{IM}/n gives,

$$B_{IM} \cdot \frac{1}{n} x^*_{[rn]} = B_{IM} \begin{bmatrix} \frac{1}{n}[rn] \\ \frac{1}{n}\sum_{i=1}^{[rn]} \widetilde{x}_i \\ \frac{1}{n}\widetilde{x}_{[rn]} \end{bmatrix} = B_{IM} \begin{bmatrix} r + o(1) \\ \frac{1}{T}\int_0^{rT} \widetilde{X}_s ds + o_p(1) \\ \frac{1}{T}\widetilde{X}_{rT} + o_p(T^{-1/2}) \end{bmatrix} = B_{IM} \cdot \frac{1}{T} X^*_{rT} + o_p(1),$$

using (3A.2), (3A.3) and $X^*_{rT} = \begin{bmatrix} rT \\ \int_0^{rT} \widetilde{X}_s ds \\ \widetilde{X}_{rT} \end{bmatrix}$. The limit of $B_{IM} \cdot \frac{1}{T} X^*_{rT}$ follows as

$$B_{IM} \cdot \frac{1}{T} X^*_{rT} = B_{IM} \begin{bmatrix} r \\ \frac{1}{T}\int_0^{rT} \widetilde{X}_s ds \\ \frac{1}{T}\widetilde{X}_{rT} \end{bmatrix} = \begin{bmatrix} r \\ \frac{1}{T}\int_0^{rT} T^{-1/2}\widetilde{X}_s ds \\ T^{-1/2}\widetilde{X}_{rT} \end{bmatrix} \Rightarrow \begin{bmatrix} r \\ \Lambda_{\widetilde{x}} \int_0^{r} W_{\widetilde{x}}(s) ds \\ \Lambda_{\widetilde{x}} W_{\widetilde{x}}(r) \end{bmatrix} = \Pi g_*(r).$$

$$(3A.4)$$

Therefore, we have the following results:

$$\frac{1}{n}\sum_{i=1}^{n} B_{IM}\frac{1}{n}x^*_i x^{*\prime}_i \frac{1}{n}B_{IM} = \frac{1}{T}\int_0^T B_{IM}\frac{1}{T}X^*_t X^{*\prime}_t \frac{1}{T}B_{IM} dt + o_p(1) \Rightarrow \Pi \int_0^1 g_*(s)g_*(s)' ds \Pi',$$

$$\frac{1}{n}\sum_{i=1}^{n} B_{IM}\frac{1}{n}x^*_i T^{1/2}\frac{1}{n}S^u_i = \frac{1}{T}\int_0^T B_{IM}\frac{1}{T}X^*_t T^{-1/2}S^U_t dt + o_p(1) \Rightarrow \Pi \int_0^1 g_*(s)B_u(s) ds.$$

We now derive the limit of $A^{-1}_{IM}(\widetilde{\theta} - \theta)$ as $T \to \infty$ and $\delta \to 0$. Using algebra from Vogelsang and Wagner (2014) and straightforward scaling gives

$$A^{-1}_{IM}(\widetilde{\theta} - \theta) = \left(\frac{1}{n}\sum_{i=1}^{n} T^{1/2} A_{IM}\frac{1}{n}x^*_i x^{*\prime}_i \frac{1}{n}T^{1/2} A_{IM} \right)^{-1} \left(\frac{1}{n}\sum_{i=1}^{n} T^{1/2} A_{IM}\frac{1}{n}x^*_i T^{1/2}\frac{1}{n}S^u_i \right) - \left(0, 0, \Omega_{u\widetilde{x}}\Omega_{\widetilde{x}}^{-1} \right)'$$

$$= \left(\frac{1}{n}\sum_{i=1}^{n} B_{IM}\frac{1}{n}x^*_i x^{*\prime}_i \frac{1}{n}B_{IM} \right)^{-1} \left(\frac{1}{n}\sum_{i=1}^{n} B_{IM}\frac{1}{n}x^*_i T^{1/2}\frac{1}{n}S^u_i \right) - \left(0, 0, \Omega_{u\widetilde{x}}\Omega_{\widetilde{x}}^{-1} \right)'.$$

Together these results give

$$A^{-1}_{IM}(\widetilde{\theta} - \theta) \Rightarrow \left(\Pi \int_0^1 g_*(s)g_*(s)' ds \Pi' \right)^{-1} \left(\Pi \int_0^1 g_*(s)B_u(s) ds \right) - \left(0, 0, \Omega_{u\widetilde{x}}\Omega_{\widetilde{x}}^{-1} \right)'$$

$$= \lambda_{u\cdot\widetilde{x}} (\Pi')^{-1} \left(\int_0^1 g_*(s)g_*(s)' ds \right)^{-1} \int_0^1 g_*(s)w_{u\cdot\widetilde{x}}(s) ds,$$

112

where the last equality holds using arguments from the proof of Theorem 3.2 from Vogelsang and Wagner (2014) and the limit is identical to the limit obtained in Vogelsang and Wagner (2014).

Now consider the terms inside the inverse of $\widetilde{W}^*$. The argument for the limit of $(1/n)\sum_{i=1}^{n}(B_{IM}/n)(c_i/n)(c_i'/n)(B_{IM}/n)$ is similar to that of $(1/n)\sum_{i=1}^{n}(B_{IM}/n)x_i^*x_i^{*\prime}(B_{IM}/n)$ and is given by

$$\frac{1}{n}\sum_{i=1}^{n}\left(B_{IM}\frac{1}{n}\right)\frac{1}{n}c_ic_i'\frac{1}{n}\left(\frac{1}{n}B_{IM}\right)$$

$$=\frac{1}{n}\sum_{i=1}^{n}\left(B_{IM}\frac{1}{n}\right)\left(\frac{1}{n}\sum_{j=1}^{n}x_j^*-\frac{1}{n}\sum_{j=1}^{i}x_j^*\right)\left(\frac{1}{n}\sum_{j=1}^{n}x_j^*-\frac{1}{n}\sum_{j=1}^{i}x_j^*\right)'\left(\frac{1}{n}B_{IM}\right)$$

$$=\frac{1}{T}\int_0^T\left(\frac{1}{T}\int_0^T B_{IM}\frac{1}{T}X_s^*ds-\frac{1}{T}\int_0^t B_{IM}\frac{1}{T}X_s^*ds\right)\left(\frac{1}{T}\int_0^T B_{IM}\frac{1}{T}X_s^*ds-\frac{1}{T}\int_0^t B_{IM}\frac{1}{T}X_s^*ds\right)'dt+o_p(1)$$

$$\Rightarrow\Pi\int_0^1\left(\int_0^1 g_*(s)ds-\int_0^r g_*(s)ds\right)\left(\int_0^1 g_*(s)ds-\int_0^r g_*(s)ds\right)'dr\Pi'.$$

This limit together with the limits of $(1/n)\sum_{i=1}^{n}(B_{IM}/n)x_i^*x_i^{*\prime}(B_{IM}/n)$ and R_T^* give an expression identical to that obtained by Vogelsang and Wagner (2014) for the inverse term apart from $(T/n)\widetilde{\lambda}_{u\cdot\widetilde{x}}^{2*}$. Therefore, the limit of the parts of $\widetilde{W}^*$ that do not depend on $(T/n)\widetilde{\lambda}_{u\cdot\widetilde{x}}^{2*}$ follow a $\lambda_{u\cdot\widetilde{x}}^2\chi_q^2$ random variable using arguments in Vogelsang and Wagner (2014). The last step is to show that $(T/n)\widetilde{\lambda}_{u\cdot\widetilde{x}}^{2*}\Rightarrow\lambda_{u\cdot\widetilde{x}}^2 P_1^{**}(b)$. While more tedious, the derivation follows the same steps as the derivation of the limit of $(T/n)\widehat{\omega}_n^2$ in Theorem 3.2 using (3A.1) and (3A.4). Details are omitted.

APPENDIX 3B

TABLES AND FIGURES

Table 3B.1 Fixed-b Right Tail Critical Value Polynomial Coefficients, Parzen Kernel

	λ_1	λ_1	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8	λ_9	R^2
Panel A. Joint Null Hypothesis That Intercept is Zero ($\beta_0 = 0$) and Slope is One ($\beta_1 = 1$).										
G-cointegration	.7289	.6333	-.0421	2.6426	-.3118	.1336	.8846	-.1857	.1841	.9956
H-stationary	.2688	.9880	-.0490	5.1346	-2.3769	.1518	-2.1792	3.0649	.2219	.9998
H-cointegration	1.3250	1.4466	-.0802	6.0381	-3.2097	.3246	-.1194	2.8672	.3661	.9987

	λ_1	λ_1	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8	λ_9	R^2
Panel B. Null Hypothesis That Slope is One ($\beta_1 = 1$).										
G-cointegration	.8130	.2969	-.0079	.7675	-1.1113	.7940	-.1950	.4250	-.2409	.9258
H-stationary	.9475	.3250	-.0078	.7082	.5386	-.01370	1.4406	-.7790	.2669	.9993
H-cointegration	1.3752	.3003	.1324	.1909	-1.9013	.8501	-.0859	.9864	-.3429	.9388

Table 3B.2 UIP Tests (US-Japan, 2-year government bond, $\Delta = 504$ days)

		Bandwidth Rules			
		AD	SPJ	NW	CRT
		OLS $\widehat{\beta}_2 = -0.337$			
Daily, $\sqrt{G(\widehat{\beta})}$	S.E.	0.611	0.54	0.308	0.806
$n = 7678$	$t_{\beta_2=1}$	-2.189	-2.473	-4.341	-1.658
$\widehat{\beta}_1 = 0.011$	Fixed-b CV	±4.053	±5.213	±1.98	±2.15
	CI $N(0,1)$	(-1.53, 0.86)	(-1.4, 0.72)	(-0.94, 0.267)	(-1.92, 1.24)
	CI Fixed-b	(-2.81, 2.14)	(-3.15, 2.48)	(-0.95, 0.273)	(-2.07, 1.4)
		IM-OLS $\widetilde{\beta}_2 = 0.514$			
Daily, $\sqrt{\widetilde{W}^*}$	S.E.	0.363	0.179	0.462	0.959
$n = 7678$	$t_{\beta_2=1}$	-1.34	-2.711	-1.053	-0.507
$\widetilde{\beta}_1 = -0.004$	Fixed-b CV	±11.377	±16.557	±1.959	±2.305
	CI Fixed-b	(-3.62, 4.64)	(-2.46, 3.48)	(-0.39, 1.418)	(-1.7, 2.72)
	$\widehat{b}$-ratio	0.504	0.7	0.007	0.066
		OLS $\widehat{\beta}_2 = -0.322$			
Weekly, $\sqrt{G(\widehat{\beta})}$	S.E.	0.616	0.541	0.509	0.807
$n = 1545$	$t_{\beta_2=1}$	-2.147	-2.445	-2.596	-1.638
$\widehat{\beta}_1 = 0.011$	Fixed-b CV	±3.917	±5.03	±2.019	±2.15
	CI $N(0,1)$	(-1.53, 0.89)	(-1.38, 0.74)	(-1.32, 0.676)	(-1.9, 1.26)
	CI Fixed-b	(-2.73, 2.09)	(-3.04, 2.4)	(-1.35, 0.706)	(-2.06, 1.41)
		IM-OLS $\widetilde{\beta}_2 = 0.521$			
Weekly, $\sqrt{\widetilde{W}^*}$	S.E.	0.392	0.199	0.651	0.963
$n = 1545$	$t_{\beta_2=1}$	-1.222	-2.405	-0.735	-0.497
$\widetilde{\beta}_1 = -0.005$	Fixed-b CV	±10.628	±15.901	±2.005	±2.305
	CI Fixed-b	(-3.64, 4.68)	(-2.64, 3.69)	(-0.78, 1.827)	(-1.7, 2.74)
	$\widehat{b}$-ratio	0.479	0.671	0.021	0.066
		OLS $\widehat{\beta}_2 = -0.327$			
Monthly, $\sqrt{G(\widehat{\beta})}$	S.E.	0.601	0.541	0.733	0.811
$n = 364$	$t_{\beta_2=1}$	-2.209	-2.452	-1.812	-1.635
$\widehat{\beta}_1 = 0.01$	Fixed-b CV	±4.435	±5.717	±2.1	±2.15
	CI $N(0,1)$	(-1.5, 0.85)	(-1.39, 0.73)	(-1.76, 1.109)	(-1.92, 1.26)
	CI Fixed-b	(-2.99, 2.34)	(-3.42, 2.77)	(-1.87, 1.212)	(-2.07, 1.42)
		IM-OLS $\widetilde{\beta}_2 = 0.533$			
Monthly, $\sqrt{\widetilde{W}^*}$	S.E.	0.288	0.134	0.88	0.963
$n = 364$	$t_{\beta_2=1}$	-1.621	-3.492	-0.531	-0.485
$\widetilde{\beta}_1 = -0.005$	Fixed-b CV	±13.351	±18.12	±2.184	±2.305
	CI Fixed-b	(-3.31, 4.38)	(-1.89, 2.96)	(-1.39, 2.454)	(-1.69, 2.75)
	$\widehat{b}$-ratio	0.572	0.777	0.05	0.066
		OLS $\widehat{\beta}_2 = -0.367$			
Quarterly, $\sqrt{G(\widehat{\beta})}$	S.E.	0.618	0.539	0.924	0.821
$n = 124$	$t_{\beta_2=1}$	-2.214	-2.537	-1.48	-1.666
$\widehat{\beta}_1 = 0.013$	Fixed-b CV	±3.946	±5.069	±2.292	±2.15
	CI $N(0,1)$	(-1.58, 0.84)	(-1.42, 0.69)	(-2.18, 1.443)	(-1.98, 1.24)
	CI Fixed-b	(-2.8, 2.07)	(-3.1, 2.36)	(-2.48, 1.75)	(-2.13, 1.4)
		IM-OLS $\widetilde{\beta}_2 = 0.52$			
Quarterly, $\sqrt{\widetilde{W}^*}$	S.E.	0.378	0.19	1.056	0.977
$n = 124$	$t_{\beta_2=1}$	-1.271	-2.531	-0.455	-0.491
$\widetilde{\beta}_1 = -0.004$	Fixed-b CV	±10.789	±16.046	±2.692	±2.305
	CI Fixed-b	(-3.56, 4.6)	(-2.52, 3.56)	(-2.32, 3.361)	(-1.73, 2.77)
	$\widehat{b}$-ratio	0.484	0.677	0.11	0.066

Notes: The null hypothesis is $H_0 : \beta_2 = 1$. The rows S.E., $t_{\beta_2=1}$, Fixed-b CV, CI, and $\widehat{b}$-ratio are reported for $\widehat{\beta}_2$ and $\widetilde{\beta}_2$, the OLS and IM-OLS estimators respectively. The t-statistics are computed as $\sqrt{G(\widehat{\beta})}$ and $\sqrt{\widetilde{W}^*}$ with the signs equal to the signs of $\widehat{\beta}_2 - 1$ and $\widetilde{\beta}_2 - 1$ respectively. The value of $\widehat{b}$ is the same for both test statistics.

Table 3B.3 UIP Tests (US-Japan, 10-year government bond, $\Delta = 2520$ days)

		Bandwidth Rules			
		AD	SPJ	NW	CRT
		OLS $\widehat{\beta}_2 = 1.356$			
Daily, $\sqrt{G(\widehat{\beta})}$	S.E.	0.501	0.437	0.139	0.386
$n = 5706$	$t_{\beta_2=1}$	0.71	0.814	2.558	0.922
$\widehat{\beta}_1 = -0.036$	Fixed-b CV	±3.333	±4.217	±1.983	±2.199
	CI $N(0,1)$	(0.37, 2.34)	(0.5, 2.21)	(1.08, 1.628)	(0.6, 2.11)
	CI Fixed-b	(-0.31, 3.02)	(-0.49, 3.2)	(1.08, 1.631)	(0.51, 2.2)
		IM-OLS $\widetilde{\beta}_2 = 1.518$			
Daily, $\sqrt{\widetilde{W}^*}$	S.E.	0.247	0.186	0.241	0.27
$n = 5706$	$t_{\beta_2=1}$	2.098	2.789	2.146	1.92
$\widetilde{\beta}_1 = -0.038$	Fixed-b CV	±7.325	±12.254	±1.96	±2.431
	CI Fixed-b	(-0.29, 3.33)	(-0.76, 3.79)	(1.04, 1.991)	(0.86, 2.17)
	$\widehat{b}$-ratio	0.364	0.534	0.008	0.081
		OLS $\widehat{\beta}_2 = 1.365$			
Weekly, $\sqrt{G(\widehat{\beta})}$	S.E.	0.491	0.423	0.226	0.383
$n = 1148$	$t_{\beta_2=1}$	0.743	0.863	1.617	0.953
$\widehat{\beta}_1 = -0.036$	Fixed-b CV	±3.407	±4.324	±2.025	±2.199
	CI $N(0,1)$	(0.4, 2.33)	(0.54, 2.19)	(0.92, 1.807)	(0.61, 2.12)
	CI Fixed-b	(-0.31, 3.04)	(-0.46, 3.19)	(0.91, 1.822)	(0.52, 2.21)
		IM-OLS $\widetilde{\beta}_2 = 1.537$			
Weekly, $\sqrt{\widetilde{W}^*}$	S.E.	0.243	0.178	0.242	0.271
$n = 1148$	$t_{\beta_2=1}$	2.211	3.008	2.216	1.984
$\widetilde{\beta}_1 = -0.039$	Fixed-b CV	±7.741	±12.799	±2.016	±2.431
	CI Fixed-b	(-0.34, 3.42)	(-0.75, 3.82)	(1.05, 2.025)	(0.88, 2.19)
	$\widehat{b}$-ratio	0.38	0.552	0.024	0.081
		OLS $\widehat{\beta}_2 = 1.358$			
Monthly, $\sqrt{G(\widehat{\beta})}$	S.E.	0.456	0.363	0.349	0.386
$n = 271$	$t_{\beta_2=1}$	0.786	0.987	1.028	0.929
$\widehat{\beta}_1 = -0.036$	Fixed-b CV	±3.984	±5.12	±2.141	±2.199
	CI $N(0,1)$	(0.47, 2.25)	(0.65, 2.07)	(0.67, 2.042)	(0.6, 2.11)
	CI Fixed-b	(-0.46, 3.17)	(-0.5, 3.22)	(0.61, 2.105)	(0.51, 2.21)
		IM-OLS $\widetilde{\beta}_2 = 1.501$			
Monthly, $\sqrt{\widetilde{W}^*}$	S.E.	0.201	0.143	0.255	0.266
$n = 271$	$t_{\beta_2=1}$	2.493	3.517	1.968	1.885
$\widetilde{\beta}_1 = -0.038$	Fixed-b CV	±10.998	±16.231	±2.283	±2.431
	CI Fixed-b	(-0.71, 3.71)	(-0.81, 3.81)	(0.92, 2.083)	(0.85, 2.15)
	$\widehat{b}$-ratio	0.491	0.685	0.063	0.081
		OLS $\widehat{\beta}_2 = 1.331$			
Quarterly, $\sqrt{G(\widehat{\beta})}$	S.E.	0.448	0.355	0.447	0.394
$n = 92$	$t_{\beta_2=1}$	0.74	0.934	0.742	0.841
$\widehat{\beta}_1 = -0.035$	Fixed-b CV	±4.121	±5.303	±2.318	±2.199
	CI $N(0,1)$	(0.45, 2.21)	(0.64, 2.03)	(0.46, 2.207)	(0.56, 2.1)
	CI Fixed-b	(-0.51, 3.18)	(-0.55, 3.21)	(0.3, 2.367)	(0.46, 2.2)
		IM-OLS $\widetilde{\beta}_2 = 1.506$			
Quarterly, $\sqrt{\widetilde{W}^*}$	S.E.	0.199	0.142	0.295	0.282
$n = 92$	$t_{\beta_2=1}$	2.537	3.558	1.712	1.792
$\widetilde{\beta}_1 = -0.038$	Fixed-b CV	±11.744	±16.863	±2.769	±2.431
	CI Fixed-b	(-0.83, 3.85)	(-0.89, 3.9)	(0.69, 2.323)	(0.82, 2.19)
	$\widehat{b}$-ratio	0.516	0.714	0.118	0.081

See the notes to Table 2.

Figure 3B.1 Null rejection probabilities of $H\left(\widehat{\beta}\right)$ under OU processes, stationary regression

(a) $\kappa_u = 6.9$, $\kappa_x = 0.1$, chi-square cv

(b) $\kappa_u = 0.5$, $\kappa_x = 0.5$, chi-square cv

(c) $\kappa_u = 6.9$, $\kappa_x = 0.1$, fixed-b cv

(d) $\kappa_u = 0.5$, $\kappa_x = 0.5$, fixed-b cv
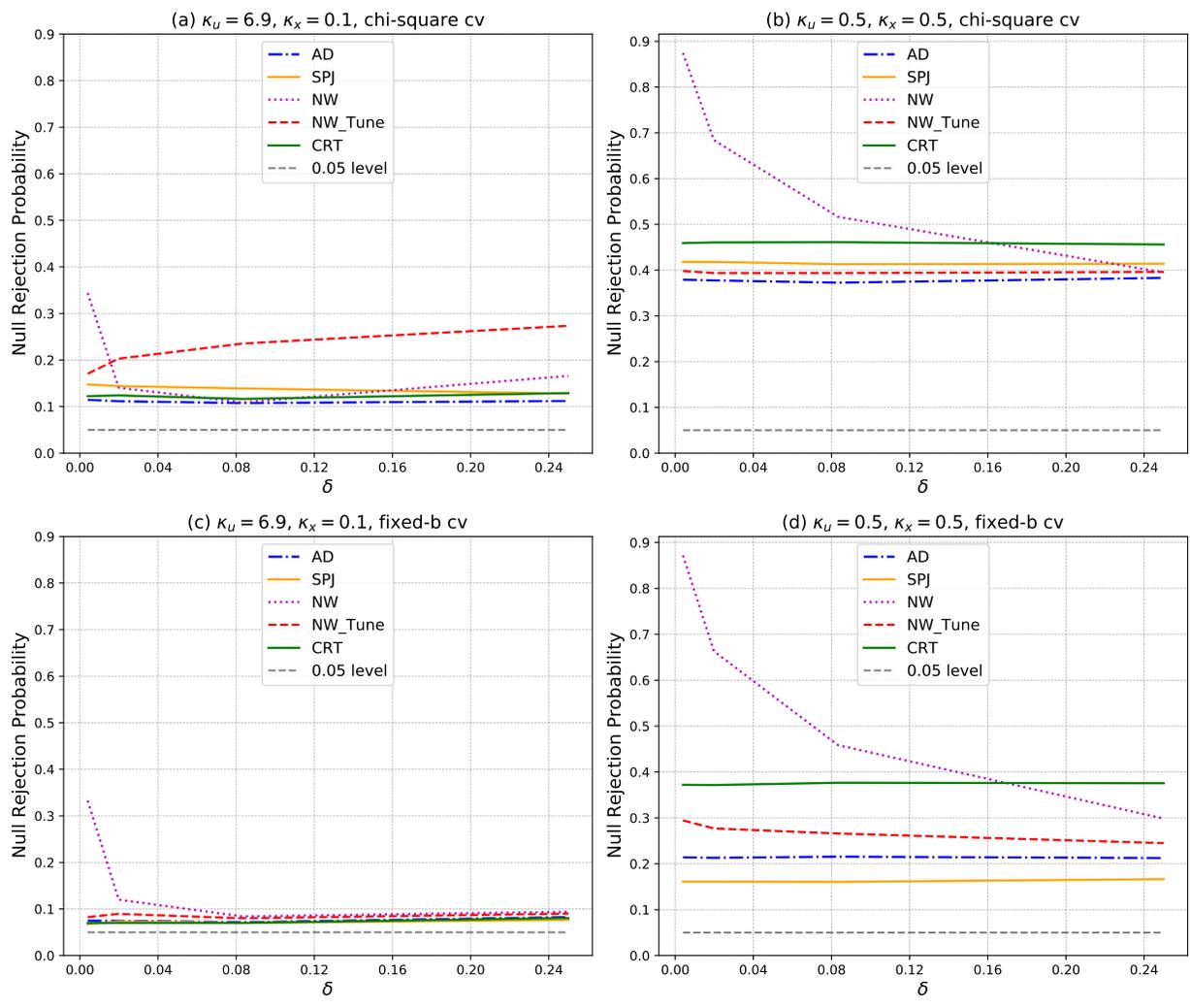
Figure 3B.2 Average bandwidth ratios under OU processes, stationary regression

(a) $\kappa_u = 6.9$, $\kappa_x = 0.1$ (b) $\kappa_u = 0.5$, $\kappa_x = 0.5$
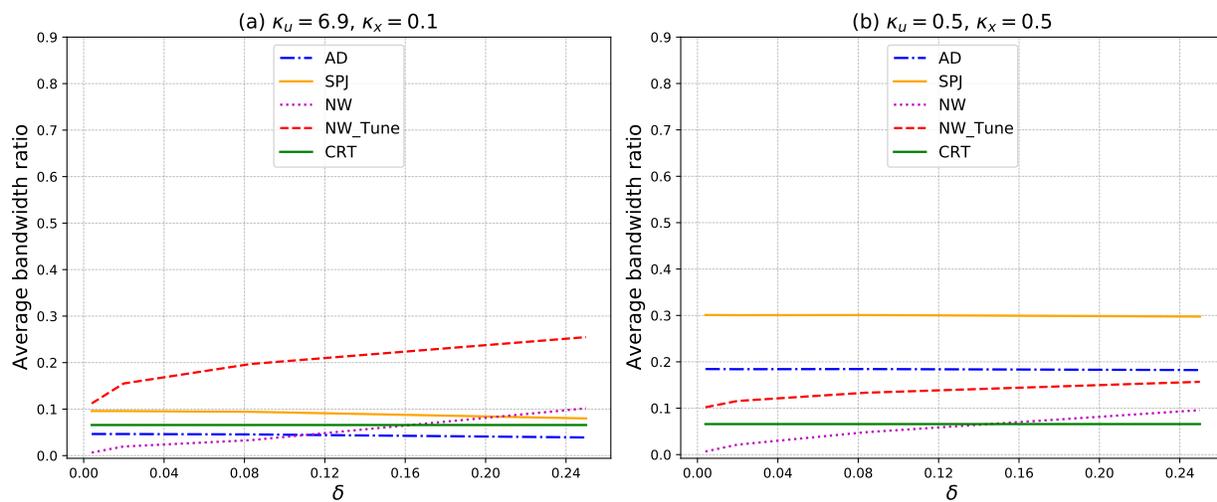
Figure 3B.3 Theoretical bandwidth ratios
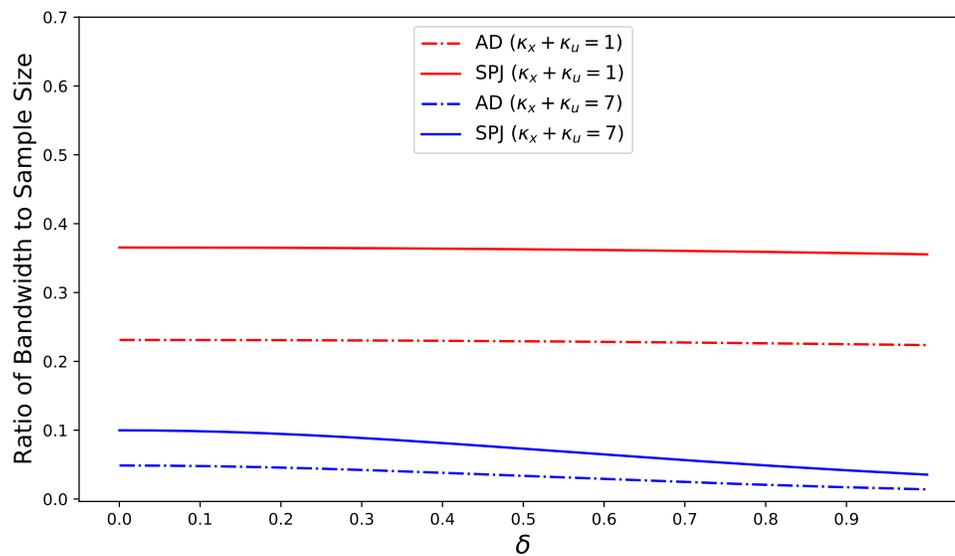
Figure 3B.4 Finite sample power of $H\left(\widehat{\beta}\right)$ under OU processes, stationary regression
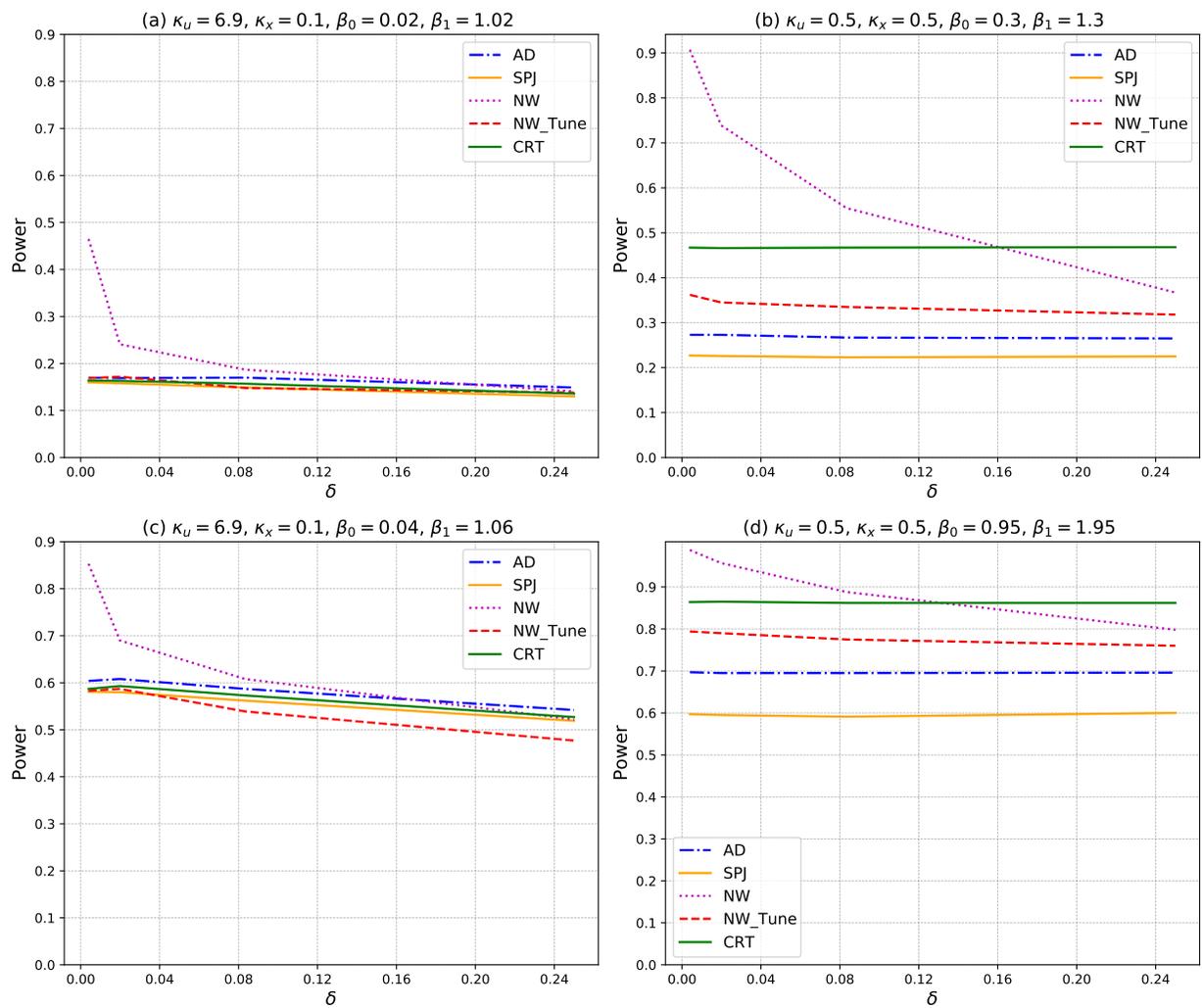
Figure 3B.5 Null rejection probabilities of $\widetilde{W}^*$ of IM-OLS under OU processes, cointegrating regression
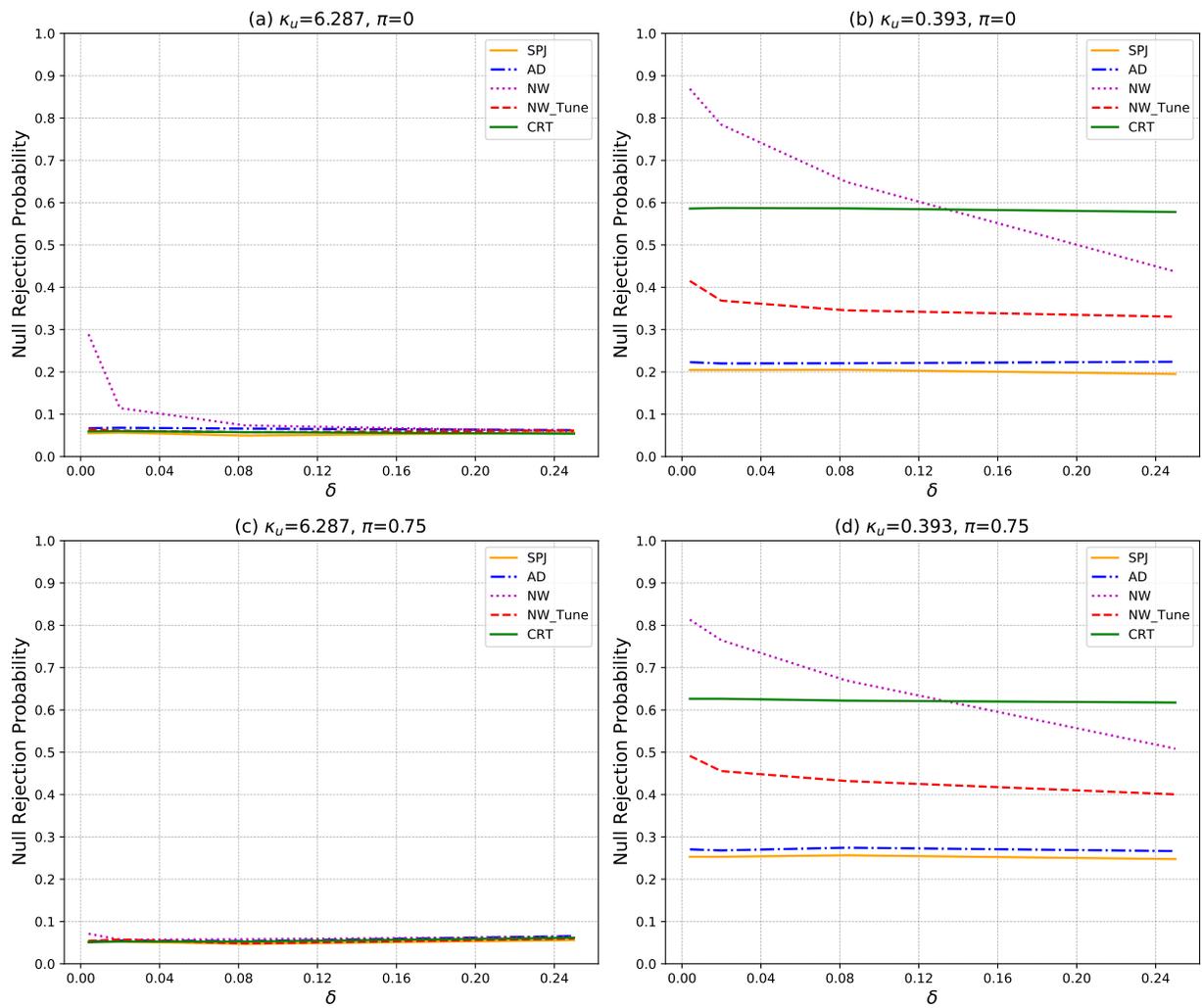
Figure 3B.6 Average bandwidth ratios under OU processes, cointegrating regression
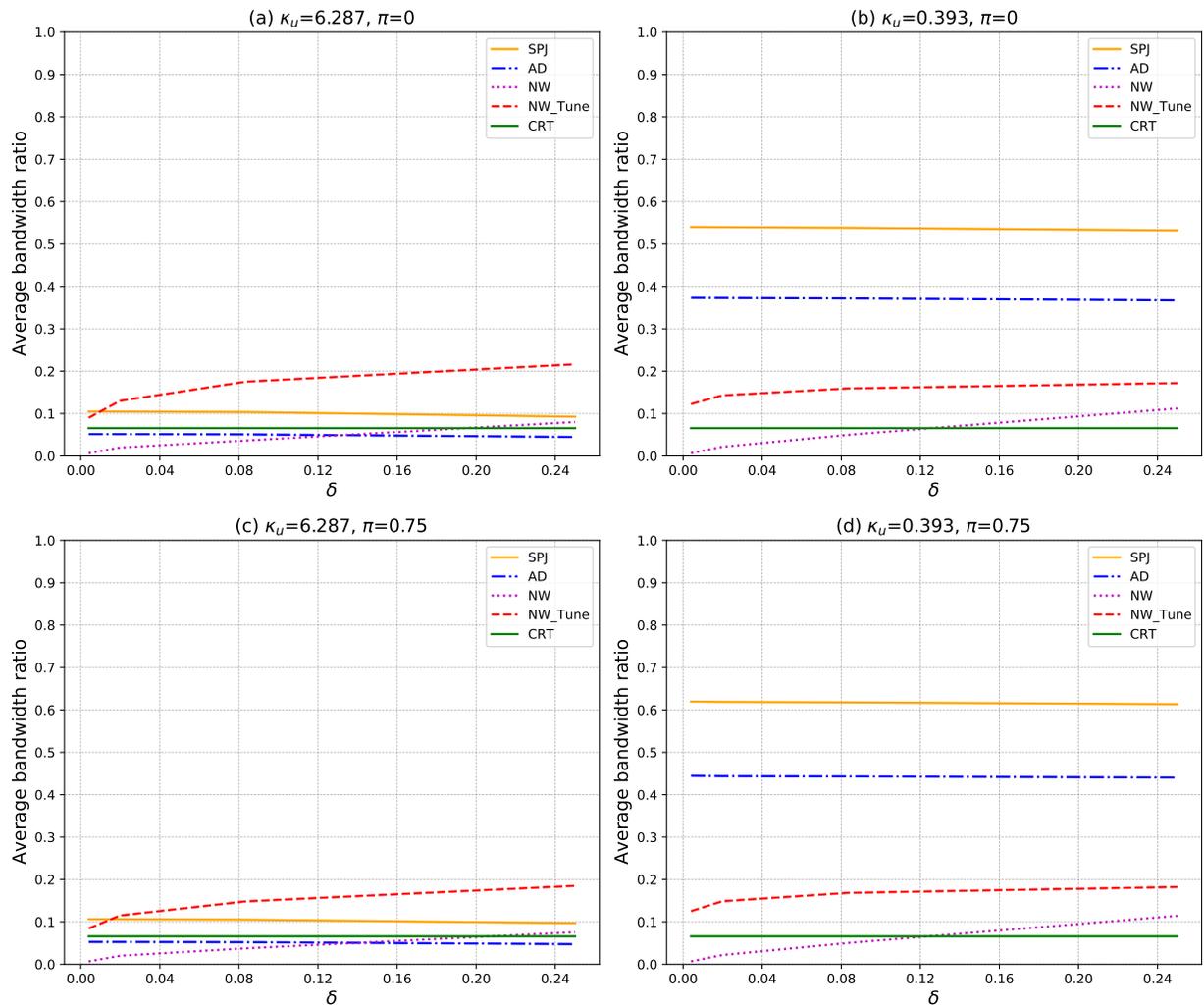
Figure 3B.7 Finite sample power of $\widetilde{W}^*$ of IM-OLS under OU processes, cointegrating regression
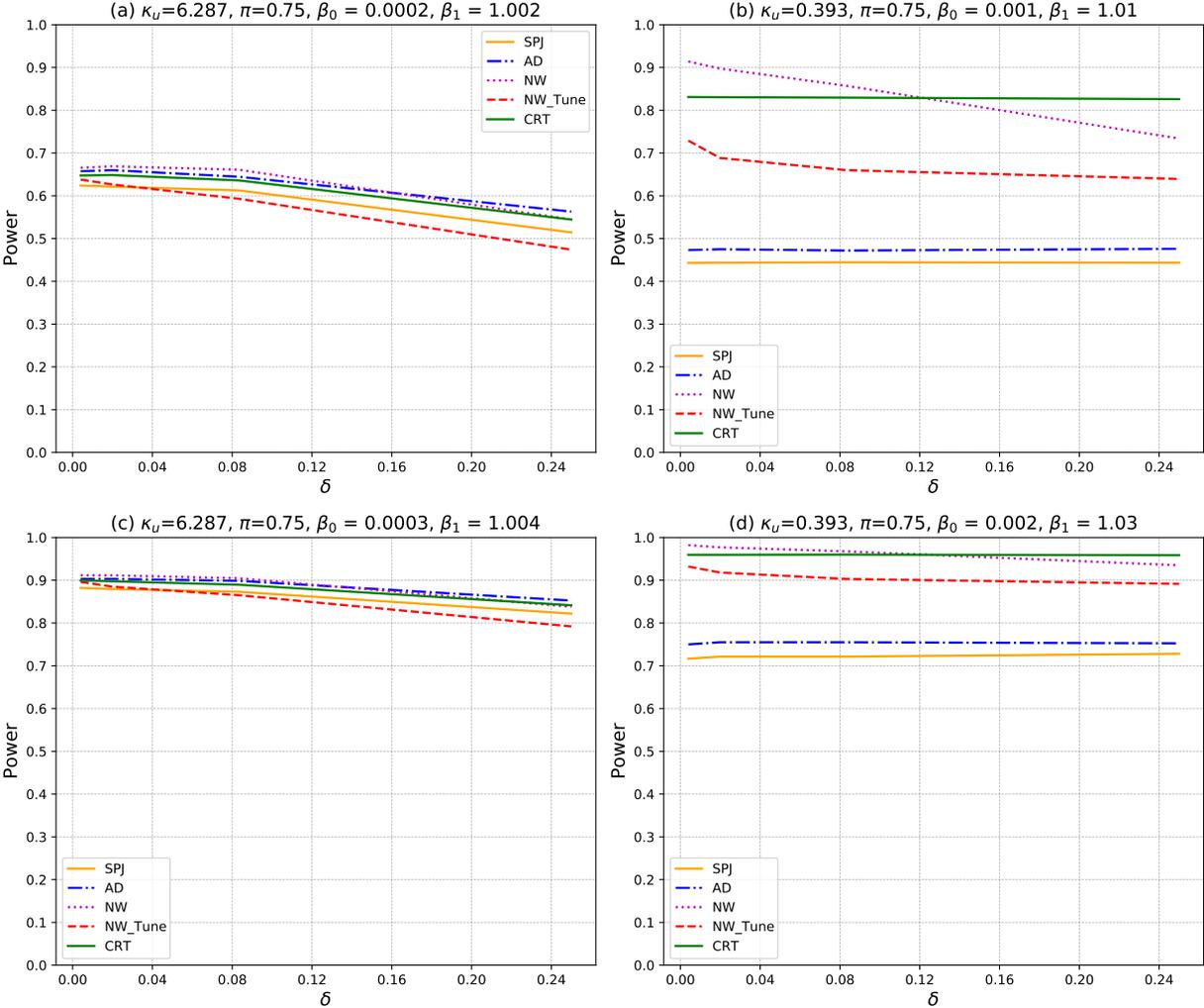
CHAPTER 4

THE DISTRIBUTION OF REALIZED US CORPORATE BOND RETURN VOLATILITY

4.1 Introduction

Volatility in the US corporate bond market (especially, daily volatility) has been less frequently studied despite of its huge importance for the financial sector, macroeconomic conditions, and portfolio management. The reasons could be that (1) US corporate bonds are illiquid, and (2) their transactions are irregularly spaced.

Due to these characteristics and the high-frequency nature (recorded every second) of corporate bond transaction data, the distribution of the volatility of corporate bonds has not been explored extensively. Campbell and Taksler (2003) studied the effect of equity volatility on corporate bond yields, but it is not about the volatility of corporate bonds.

To address the irregular price movements and illiquid transaction behavior of corporate bonds, I utilize a Compound Poisson Process (CPP) to model the price dynamics of US corporate bonds. CPP can well describe the irregular price movements and illiquid transaction behavior of corporate bonds. Although the current continuous time stochastic diffusion model can take the irregularity of US corporate bond transactions into account, it does not effectively describe the illiquidity of US corporate bonds. CPP is designed to describe the behavior of a continuous stochastic process which has random discrete jumps over time. Since its main focus is on the discrete change of the process by the arrival of the random jump, it is adequate for modeling the prices of illiquid assets. Unlike US stock trades, the US corporate bonds are not actively traded during a day, mostly showing random transactions less than hundreds during a day according to US corporate bond transaction data. Therefore, CPP is an appropriate process to describe the price dynamics of US corporate bonds.

To estimate the volatility of corporate bonds, I utilize realized (daily) volatility suggested by Andersen et al. (2001a). The realized volatility uses intraday returns measured at a fixed regular interval (e.g. 5 minutes, 10 minutes) to estimate daily volatility of an asset. I slightly modify the calculation of the realized volatility to adjust it to accommodate a CPP setting.

In Section 4.2, I review the RV and related literature. In Section 4.3, I explain the CPP setting for US corporate bonds and slightly modify the RV under the setting. Then, I conduct Monte Carlo simulations for the RV in Section 4.4.

For empirical analysis, I utilize the richness of Trade Reporting and Compliance Engine (TRACE) of Financial Industry Regulatory Authority (FINRA). Section 4.5 explains the data used for empirical analysis. The data allows one to obtain 99% of every bond transaction of US corporate bond market which is recorded every second. Additionally, bond characteristics data for TRACE can be obtained using Refinitiv Workspace. With the datasets, in Section 4.6, I analyze the distributions of the RV of US corporate bonds. I investigate not only the unconditional distribution of the RV of US corporate bonds, but also the conditional distribution of the RV using the bond characteristics. I also provide an analysis for linkages between the corporate bond volatility and returns of other financial instruments in Section 4.6. Then I conclude in Section 4.7.

4.2 Preliminaries

Andersen et al. (2001a,b, 2003) proposed realized volatility (RV), which is an ex-post, model-free volatility measure. The method starts from the continuous time diffusion model given by

$$dP_t = \mu dt + \sigma_t dW_t, \tag{4.1}$$

where P is a logarithmic price of the asset and W is a standard Brownian motion. Under this model, the integrated volatility (IV) is defined as

$$IV = \int_0^t \sigma_s^2 dW_s. \tag{4.2}$$

They proposed the realized volatility as an estimator for IV, given by

$$RV = \sum_{\tau=1}^m r_\tau^2, \tag{4.3}$$

where $r_\tau = p_\tau - p_{\tau-1}$ and $\{p_\tau\}_{\tau=1}^m$ is a sequence of intraday log-prices with a fixed interval, such as 5 minutes. In Andersen et al. (2003), they showed that RV is a consistent estimator for IV using semi-martingale theory in Protter (1990). Using their seminal work about RV, many subsequent

studies extended the continuous time diffusion model by adding a jump component giving

$$dP_t = \mu dt + \sigma_t dW_t + dJ_t, \tag{4.4}$$

where J is a pure jump Lèvy process. As shown in subsequent studies, this can be expressed as a Brownian semi-martingale with jumps given by

$$dP_t = \mu dt + \sigma_t dW_t + \kappa_t dq_t, \tag{4.5}$$

where κ_t is size of the jump and dq_t is 1 when there is a jump and 0 if there is no jump at time t. Thus, the return of an asset on interval $[0, t]$ is expressed as

$$r_{t,0} = \mu t + \int_0^t \sigma_s dW_s + \sum_{j=1}^{N_t} k_j, \tag{4.6}$$

where the return of an asset is computed as $r_{t,0} = P_t - P_0$. Here N_t represents a number of jumps on the interval $(0, t]$, following a poisson process with intensity λ with jump size of k_j. Recall that IV of the continuous time diffusion process is

$$IV = \int_0^t \sigma_s^2 dW_s. \tag{4.7}$$

Additionally, it has been pointed out by the previous literature that RV converges to the following expression under the model with a jump component,

$$RV \xrightarrow{p} \int_0^t \sigma_s^2 dW_s + \sum_{j=1}^{N_t} \kappa_j^2 \tag{4.8}$$

as $m \to \infty$. Thus, in order to estimate the integrated volatility, Barndorff-Nielsen and Shephard (2004) devise bipower variation (BV) as a consistent estimator for integrated volatility given by

$$BV = \frac{\pi}{2} \sum_{\tau=2}^{m} r_\tau r_{\tau-1} \xrightarrow{p} \int_0^t \sigma_s^2 dW_s \tag{4.9}$$

as $m \to \infty$. Hence, BV can be a consistent estimator when the jumps exist in the price dynamics of the asset.

4.3 Model

Stochastic volatility models for assets like stocks and exchange rate typically start with a continuous time diffusion process for the logarithmic price (P_t) of an asset given by

$$dP_t = \mu dt + \sigma_t dW_t. \tag{4.10}$$

As I discussed in the previous section, one can extend the process to include jumps given by

$$dP_t = \mu dt + \sigma_t dW_t + \kappa_t dq_t. \tag{4.11}$$

The processes are based on continuous time diffusion models valid for assets like stocks and exchange rates which are heavily traded within time intervals. In contrast, US corporate bonds are illiquid, and thus, their transactions are irregularly spaced. As they are illiquid, the price of a bond is constant for some time intervals and shows sudden discrete jumps at irregular time points. Therefore, the model for the price of a corporate bond should exclude continuous part for its diffusion model, including only the jump part in its diffusion model for the price. To model this illiquidity and irregularity of the corporate bond price, I adopt a Compound Poisson Process (CPP) for the bond price diffusion model. The process P_t is a Compound Poisson Process and is defined as

$$P_t = \sum_{j=1}^{N_t} \kappa_j, \tag{4.12}$$

where N_t follows a poisson process (which is the number of jumps) with an intensity parameter λ and each κ_j is an *i.i.d.* random variable (which represents a jump size). Hence, utilizing the CPP setting, my model for the (logarithmic) price of a US corporate bond at time t, P_t, is suggested as

$$P_t = P_0 + \sum_{j=1}^{N_t} \kappa_j, \tag{4.13}$$

where I can think of the discrete jump size κ_j as the percentage change (return) of the bond price at that time point. The return of a US corporate bond during the interval $(0, t]$ is expressed as

$$r_{t,0} = \sum_{j=1}^{N_t} \kappa_j. \tag{4.14}$$

Next, I slightly modify RV for the CPP setting to estimate the volatility of US corporate bonds. First, I define integrated volatility (IV) for US corporate bond price using the CPP setting. I start by modifying the CPP of equation 4.13 by adding some assumptions. For the poisson process N_t in equation 4.13, I assume an inhomogeneous poisson process with intensity $\lambda(t)$ which is depending on time, rather than a homogeneous poisson process with a constant intensity rate λ, since the transaction of financial assets shows the diurnal pattern that the frequency of the trade is dependent on time. For the size of the jump of the log price (discrete change), I assume that the jump size κ_j follows normal distribution, thus $\kappa_j \sim N(0, \phi_j^2(t))$, where $\phi_j^2(t)$ of the CPP model for the bond price corresponds with the instantaneous variance of the continuous time diffusion model. Therefore, the integrated volatility of US corporate bond price on interval $(0, t]$ is defined as

$$IV_{(0,t]}^{bond} = \sum_{j=1}^{N_t} \phi_j^2. \tag{4.15}$$

Then, I modify the RV to estimate the volatility of US corporate bonds under the CPP setting as follows.

$$RV = \sum_{j=1}^{N_t} \kappa_j^2. \tag{4.16}$$

I use this RV for Monte Carlo simulations and empirical analysis for US corporate bonds throughout the paper. Notice that the suggested RV can be interpreted as the sum of the squares of returns (the sum of the squares of discrete jumps in (logarithmic) bond price) at some time points during the time interval $(0, t]$.

4.4   Monte Carlo Simulation

To investigate the properties of the RV estimator under the CPP setting, I conduct Monte Carlo simulations. I generate data using the below CPP setting with three different models for volatility modeling. For the first model for the volatility (Case (1)), I assume that the variance of the jump size, $\phi_j^2(t)$, does not change over time within a day. Thus, the volatility is a constant within a day for Case(1). In the second model (Case (2)), I assume that $\phi_j^2(t)$ is randomly generated from uniform distribution, which means that the instantaneous volatility of US corporate bond changes every second within a day, following uniform distribution. In the third model (Case (3)), $\phi_j^2(t)$

changes every second similar to the second model, but it follows Heston model.

$$P_t = P_0 + \sum_{j=1}^{N_t} \kappa_j, \qquad (4.17)$$

$$N_t \text{ is a poisson process with intensity } \lambda(t), \qquad (4.18)$$

$$\kappa_j \sim N(0, \phi_j^2(t)), \qquad (4.19)$$

$$RV = \sum_{j=1}^{N_t} \kappa_j^2, \qquad (4.20)$$

$$IV = \sum_{j=1}^{N_t} \phi_j^2, \qquad (4.21)$$

where N_t is the number of jumps (trades) within a day. As discussed, I introduce three different models for volatility modeling for the simulations. Case (1) for volatility modeling is given by $\phi_j^2(t) = \phi_d^2$. Case (2) is modeled as $\phi_j^2(t) \sim unif[0, 1]$. Case (3) follows the Heston model and is given by $d\phi_t^2 = \kappa(\theta - \phi_t^2) + \eta\phi_t dW_t$, where W_t is standard Brownian motion.

I use the "Thinning Algorithm" for generating poisson process where t is measured in seconds. I conduct 5000 times (days) of simulations for each case. The closed form of $\lambda(t)$ will not affect the results, but it is set as a quadratic equation to accommodate the diurnal pattern of financial assets, called the "Volatility Smile." I evaluate the finite sample performance of the RV as an estimator of IV using mean absolute percentage error (MAPE) = $\frac{1}{n}\sum_{t=1}^{n}|\frac{RV-IV}{IV}|$. I also have three different cases for the mean of intraday observations (the mean number of jump occurrences in a day) for the simulations. I call it 'Illiquid Bond Market' when the mean of intraday observations is set to 26.6, and 'Liquid Bond Market' when the mean is set to about 60.3. When the mean is set to about 146.7, I call it 'Thick Bond Market'.

Table 4A.1 reports the results of the Monte Carlo simulations. Based on the simulations, the RV obtains comparably low MAPE under Case (1) than under other cases across the three different averages of intraday observations. The MAPEs for all cases decline as the average of the intraday observations grows, which is expected. The RV under Case (2) shows 19% of MAPE if the mean of the observations is about 60 for a day. The MAPE for Case (2) is around 12% when having 146

jumps. The RV under Case (3) represents 16% of MAPE when the average of the observations is about 60 for a day. The MAPE for Case (3) declines around 10% when having 146 jumps.

4.5 Data

The data for the empirical analysis of realized US corporate bond return volatility is obtained from the Trade Reporting and Compliance Engine (TRACE) of the Financial Industry Regulatory Authority (FINRA). I use two different databases to gather the data. The first database is the Wharton Research Data Services (WRDS) where I acquire high-frequency transaction data of US corporate bonds in TRACE. The transactions are recorded every second. The data includes the price, date, and time of over-the-counter secondary market corporate bond transactions, which covers more than 99% of US corporate bond transactions. The second database is Refinitiv Workspace, where I obtain the data for the corporate bond characteristics such as credit ratings and original issued amounts. The credit ratings data includes ratings from both Moody's and Fitch Ratings. The two datasets are easily linked together using the unique 9 digit number given to US corporate bonds called the Committee on Uniform Securities Identification Procedures (CUSIP). My analysis is from January 1, 2013, to December 31, 2018. For simple statistics of the characteristics of interest, the means of the yield rate and the issued amount for the bonds during the period are 6.41 and 2,120,885,360.13, respectively. The standard deviations of each variable are 6.31 and 2,035,185,824.37.

I clean the TRACE dataset following Dick-Nielsen (2009). I remove cancellation transactions and their original transaction, corrections and their original transaction, reversals and their original transaction, and lastly, agency transactions (double counting problem). About 20% of 100 million transactions in the data is removed by the steps.

To ensure more than the average daily observation of 60 for each bond (which could correspond to the liquid market in the Monte Carlo simulations), I select bonds with more than 15,000 transactions per year, as 15,000 divided by 250 trading days in a year equals 60. I also remove bonds with trading days less than 200 days.

4.6 Empirical Results

4.6.1 The Unconditional Distribution of Corporate Bond Volatility

I construct the unconditional distribution of the daily realized US corporate bond return volatility. Similar to the way Andersen et al. (2001b) constructed the unconditional distribution for 30 DJIA stocks, I report the distribution of the daily realized US corporate bond return volatility. First, I calculate the daily realized volatility from the chosen corporate bonds which satisfy the standards discussed in the previous section. Next, I calculate and report the mean, standard deviation, skewness, and kurtosis of the daily realized volatilities for each combination of corporate bond and year. Then, I construct the unconditional distribution of the mean of the daily RV. Table 4A.4 shows the numbers (percentiles) describing the mean, standard deviation, skewness and kurtosis of the daily realized volatilities. As shown in Table 4A.4, the median for the mean of the daily realized volatilities is about 0.0042 and the mean value of it is about 0.0052. I also report the percentiles for statistics from the logarithm of the standard deviation. The standard deviation is calculated as the root of realized volatility. Its median is about -2.931 and the mean is -3.052. Figure 4A.1 shows the unconditional distribution of the series of the mean of daily RVs for each combination of corporate bond and year. The distributions is left-skewed. In figure 4A.2, I compare the unconditional distribution of the mean of log standard deviation of RV with standard normal distribution with mean of -3.

4.6.2 The Conditional Distribution of Corporate Bond Volatility by Bond Characteristics

In this section, I construct the conditional distribution of the mean of daily RV by using bond characteristics from the bond characteristics dataset. The characteristics that I use are credit rating, (original) issued amount, and yield rate. For credit rating, I divide the volatility series into two groups, 'Investment Grade' bonds and 'High Yield' bonds. I classify a bond as a 'Investment Grade' bond, if either a credit rating from Moody's for the bond is above (or equal to) Baa3 or a credit rating from Fitch for the bond is above (or equal to) BBB-. If either a credit rating from Moody's for the bond is below Baa3 or a credit rating from Fitch for the bond is below BBB-, then I classify the bond as a 'High Yield' bond. If I have both credit ratings (from Moody's and Fitch)

for the bond and the credit ratings conflict, I follow the rating from Moody's for the classification. Issued amount is the size of bond issued. I divide the daily volatilities as two groups by the median value of the issued amount of the sample. Finally, I used yield rate as the one of bond characteristics to divide the volatilities as two group. The criteria for the grouping is 5%. The bonds with a yield rate higher than 5% are in one group, and the bonds with a yield rate less than 5% are in another group.

As one can see in Table 4A.5, the conditional distributions of the mean of daily RV by bond characteristics vary between the groups. The series of the mean of daily bond volatilities with credit rating of investment grade has less mean and median than the series of high yield, which implies that investment grade bonds are less volatile than high yield corporate bonds. Surprisingly, the group of average daily volatility for less issued bonds have 3 times higher median than that of the largely issued bonds group. The group of bonds with yield rate higher than 5% also have higher value for the median than the group with less than 5%.

I also compare the conditional distributions of different bond characteristics using the graphs. The graphs shows significant difference between the distributions with different characteristics as I confirmed in the table as well. In Figure 4A.3, I draw two different distributions by credit rating criteria, where the distribution with blue color represents the conditional distribution of mean of daily RV of investment grade bonds and orange colored distribution represents that of high yield bond. One can see that the conditional distribution of the high yield bonds is slightly shifted to the right relative to that of the investment grade bonds, which means that high yield bonds are more volatile than investment grade bonds. The difference of the distributions by bond characteristics seems much more obvious in Figure 4A.4 and Figure 4A.5. The conditional distributions of small size issued bonds in Figure 4A.4 and high yield rate bonds in Figure 4A.5 not only have bigger mean and median, but also exhibit the different shapes with fatter and longer tails compared to those of large-size issued bonds and low yield rate bonds. These figures can provide evidence that bond characteristics are the main factors determining the volatility of corporate bonds.

4.6.3 The Conditional Distribution of Corporate Bond Volatility by Linkages with Other Financial Markets

In this section, I examine the conditional distribution of the averaged daily realized volatility of corporate bonds on a given day, considering the market conditions of other financial markets on the same day. I consider three financial assets. The first one is the returns of the S&P 500 index, which is representative market index for the US stock market. Secondly, I choose the returns of the CBOE Volatility Index (VIX.), which is an implied volatility measure of the US stock market and is often called as a fear gauge of the financial markets. Lastly, I use returns of US 30 year treasury bond yield as a leading indicator for other types of bonds. The data for three financial instruments is from Center for Research in Security Prices (CRSP). The data period is the same as the period mentioned for the corporate bond data in Section 4.5.

I derive a time series index for the US corporate bond market volatility by averaging the RV of corporate bonds for each day, which represents the degree of intraday volatility of the US corporate bond market for that day. Then, I investigate how the (averaged) realized volatility index of the US corporate bond market behaves under different conditions in other financial markets. One can see that the volatility of the corporate bond market for that day shows different patterns depending on the conditions of the other financial markets. In Table 4A.6, when there are sizable price movements in the other financial instruments, the means for the index of the corporate bond market shows larger values compared to the means on days with relatively smaller price movements. When there is more than a 1% price change in the S&P 500 index, whether it is a negative or a positive shock, the mean for the volatility index of corporate bond market is larger compared to the mean on days when the S&P 500 price change is within 1%. For CBOE VIX and 30 year T-Bill, one observes similar patterns where sizable price changes over the particular thresholds (7% for VIX and 0.025% for T-Bill) in these instruments yield a higher volatility index for the corporate bond market. This implies that the conditional distributions of the index for the corporate bond market by the size of price movements in other financial markets can vary.

Figure 4A.6 shows the conditional distribution of the z-scores of (averaged) realized volatility

132

index of the US corporate bonds for the days when the absolute value of the (z-scored) daily return of the S&P 500 is less than 1, meaning a small price change. Figure 4A.7 provides the conditional distribution of the (z-scored) index for the days when the absolute value of the (z-scored) daily return of the S&P 500 is over 1, meaning that there are sizable shocks in the US stock market. One can see that the values of the distribution are centered around zero with thin tails. That means when there is a small price change in the S&P 500, the volatility index of the US corporate market tends to be around the average level, not an extreme value. But, as one can find in Figure 4A.7, when the absolute value of the z-scored daily return of the S&P 500 is larger than 1, which indicates the US stock market has a sizable positive or negative shock on that day, the conditional distribution of the volatility index for the US corporate bond market displays fat tails. Especially, in Figure 4A.7, when the z-scored S&P 500 daily return is less than -1 (a big negative shock), the mass of distribution is not centered around the average, and the tail of the conditional distribution of the volatility index is fatter compared to the distribution in Figure 4A.6. This implies that when there is a substantial shock to the US stock market, the US corporate bond market tends to show larger volatility on the same day, as it tends to have extreme values, as the fat tail shows. This could provide evidence of the linkage between the US corporate bond market and the US stock market, as Campbell and Taksler (2003) explored similar market linkage with analyzing the effect of equity volatility on corporate bond yields.

4.7 Conclusion

Realized volatility has been an important measure for the volatility of financial assets since its development in Andersen et al. (2001a,b, 2003). I investigate the distributions of the daily realized volatility of US corporate bonds using the RV method, slightly modified for a CPP setting. To describe the price dynamics of US corporate bonds, I employ a CPP setting to accommodate their irregularity and illiquidity, and then I slightly modify the RV to fit the CPP setting for US corporate bonds.

I empirically analyze the distribution of the RV of US corporate bonds using the high frequency transaction data from TRACE of FINRA. Utilizing the advantage of millions of corporate bond

transactions recorded every second, I calculate the daily RV of bonds by using all transactions during a day and construct their unconditional distribution. Linking the transaction data with the bond characteristics data from TRACE via Refinitiv Workspace, I build the conditional distributions of daily RV by bond characteristics. I find that there are significant differences in the shapes of the conditional distributions constructed for the groups with different characteristics. The group of bonds with high yield ratings in credit rating shows a higher mean and median for the mean of the daily RV than the group with investment grades. The group of bonds with a small-size issued amount and the group with a high yield rate display higher mean and median values, less centered distributions, and fatter tails than the group with a large-size issued amount and the group with a low yield rate, respectively.

I also examine linkages between the US corporate bond market and other financial markets. I choose three representative instruments in the US stock market and the US treasury bond market. I calculate the cross sectional mean of daily RV of corporate bonds for the realized volatility index for each day. Then, I find that the group of days with greater price changes in those financial instruments has higher realized volatility for the US corporate bond market than the group of days with relatively small price movements in the financial instruments. I also discover that the conditional distribution of the RV of the corporate bond market for the days with greater shocks in the S&P 500 displays a less centered distribution and fatter tails, especially when the type of the shocks is negative.

BIBLIOGRAPHY

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Ebens, H. (2001a). The distribution of realized stock return volatility. *Journal of financial economics*, 61(1):43–76.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Ebens, H. (2001b). The distribution of realized stock return volatility. *Journal of financial economics*, 61(1):43–76.

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

Barndorff-Nielsen, O. E. and Shephard, N. (2004). Power and bipower variation with stochastic volatility and jumps. *Journal of financial econometrics*, 2(1):1–37.

Campbell, J. Y. and Taksler, G. B. (2003). Equity volatility and corporate bond yields. *The Journal of finance*, 58(6):2321–2350.

Dick-Nielsen, J. (2009). Liquidity biases in trace. *The Journal of Fixed Income*, 19(2):43.

Protter, P. (1990). Stochastic differential equations. *Stochastic Integration and Differential Equations: A New Approach*, pages 187–284.

TABLES AND FIGURES

Table 4A.1 Monte Carlo Simulation Result

MC	Avg Intraday Obs ≈ 26.6 Illiquid Bond Market			Avg Intraday Obs ≈ 60.3 Liquid Bond Market			Avg Intraday Obs ≈ 146.7 Thick Bond Market		
	Case(1)	Case(2)	Case(3)	Case(1)	Case(2)	Case(3)	Case(1)	Case(2)	Case(3)
MAPE	0.0403	0.2865	0.2432	0.0187	0.1935	0.1582	0.0058	0.1255	0.1021

Notes. The simulation is done 5,000 times (days). For the illiquid bond market, the mean value for jumps (intraday transaction observations) is set at 26.6 per day. (The number of jumps in the simulation corresponds to the number of intraday transaction observations in real data). For the liquid bond market, the mean value for jumps is set at 60.3 per day. In the thick bond market, the mean value for jumps is set at 146.7 per day.

Table 4A.2 Example of Cancellation of bond transaction

CUSIP	Company Symbol	Date	Time	Price	Yield	Trade Status	Message No.
17275RAX0	CSCO	20190401	12:47:00	99.833	2.5906	T	49568
17275RAX0	CSCO	20190401	12:47:00	99.833	2.5906	X	49568
17275RAX0	CSCO	20190401	13:33:39	99.789	2.6281	T	57659

Table 4A.3 Example of Correction of bond transaction

CUSIP	Company Symbol	Date	Time	Price	Volume	Trade Status	Message No.	Orig Msg No.
023770AA8	AAL	20190402	12:57:52	98.27	500000	T	57113	
023770AA8	AAL	20190402	12:57:52	98.27	500000	C	57113	
023770AA8	AAL	20190402	12:57:52	98.17	414041.7	R	58249	57113

Table 4A.4 The unconditional daily corporate bond volatility distribution

Bond	Volatility				Log St.dev			
	Mean	Std	Skew	Kurt	Mean	Std	Skew	Kurt
Min	0.0001	0.0002	0.3008	-0.9639	-4.8942	0.2401	-8.2086	-1.3260
0.10	0.0008	0.0007	0.8641	0.8597	-3.8394	0.3523	-3.2892	0.2771
0.25	0.0020	0.0018	1.2330	1.9862	-3.3500	0.4109	-1.7987	1.0867
0.50	0.0042	0.0034	1.8355	4.9924	-2.9313	0.4787	-1.1285	2.9454
0.75	0.0068	0.0054	3.2182	16.2460	-2.6569	0.5547	-0.6820	8.1248
0.90	0.0100	0.0105	5.2471	41.1743	-2.4399	0.6715	-0.2868	21.7149
Max	0.0283	0.0341	14.8560	230.3933	-1.9055	1.5826	0.9421	100.0411
Mean	0.0052	0.0047	2.6434	16.4441	-3.0517	0.5036	-1.4866	7.9331
Std	0.0047	0.0050	2.2749	30.6782	0.5592	0.1582	1.3769	14.0918

Notes. Volatility is computed by the RV method. The log of the standard deviation is the log of the square root of the volatility calculated by the RV method.

Table 4A.5 Daily Corporate Bond Volatility Distribution by Bond Characteristics

Bond	Credit Rating				Issue Size				Yield Rate			
	Invest Grade		High Yield		Large Size		Small Size		Higher than 5%		Lower than 5%	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std
Min	0.0001	0.0002	0.0008	0.0012	0.0001	0.0002	0.0001	0.0004	0.0003	0.0003	0.0001	0.0000
0.10	0.0005	0.0005	0.0023	0.0022	0.0004	0.0005	0.0023	0.0021	0.0029	0.0023	0.0004	0.0005
0.25	0.0015	0.0012	0.0032	0.0029	0.0010	0.0010	0.0036	0.0032	0.0042	0.0033	0.0008	0.0008
0.50	0.0037	0.0030	0.0048	0.0041	0.0024	0.0020	0.0059	0.0043	0.0062	0.0046	0.0019	0.0016
0.75	0.0066	0.0051	0.0076	0.0061	0.0050	0.0038	0.0086	0.0073	0.0095	0.0077	0.0035	0.0024
0.90	0.0100	0.0085	0.0108	0.0123	0.0073	0.0056	0.0128	0.0134	0.0159	0.0143	0.0057	0.0036
Max	0.0253	0.0297	0.0283	0.0341	0.0200	0.0303	0.0283	0.0341	0.0283	0.0341	0.0129	0.0108
Mean	0.0049	0.0041	0.0062	0.0062	0.0035	0.0029	0.0070	0.0064	0.0079	0.0068	0.0025	0.0019
Std	0.0047	0.0043	0.0047	0.0063	0.0033	0.0035	0.0052	0.0057	0.0055	0.0061	0.0023	0.0017

Notes. The criteria for dividing the two groups based on credit rating are outlined in Section 4.6.2. Large size bonds are defined as bonds with an issued amount higher than the median. There are also two groups by yield rate. One is a group of bonds with a yield rate higher than 5%. The other group has a yield rate less than 5%.

Table 4A.6 Daily Bond Market Volatility by Different Market Regime

	S&P 500			CBOE VIX			T-Bill 30Yrs		
	>1%	<-1%	Btw	>7%	<-7%	Btw	>0.025%	<-0.025%	Btw
Mean	0.0063	0.0063	0.0055	0.0061	0.0060	0.0055	0.0059	0.0060	0.0054
Std	0.0036	0.0044	0.0031	0.0040	0.0035	0.0030	0.0031	0.0038	0.0032
Obs (days)	143	125	1242	219	207	1084	378	418	704

Notes. The statistics in the table (mean, standard deviation, number of observations) are for the realized volatility index, which represents the average daily RV of US corporate bonds on each day, for the US corporate bond market under different conditions of other financial instruments. Each column shows the statistics for different conditions of price changes (daily returns) of the S&P 500, VIX, and T-Bill. For example, the first column reports the statistics for the volatility index on the days when the S&P 500 has changed by more than 1%, and the second column shows the statistics for the volatility index on the days when the S&P 500 has changed by less than -1%. The third column shows the statistics for the days when the return of the S&P 500 is between -1% and 1%.

Figure 4A.1 The unconditional distribution of the mean of the daily RV of corporate bonds
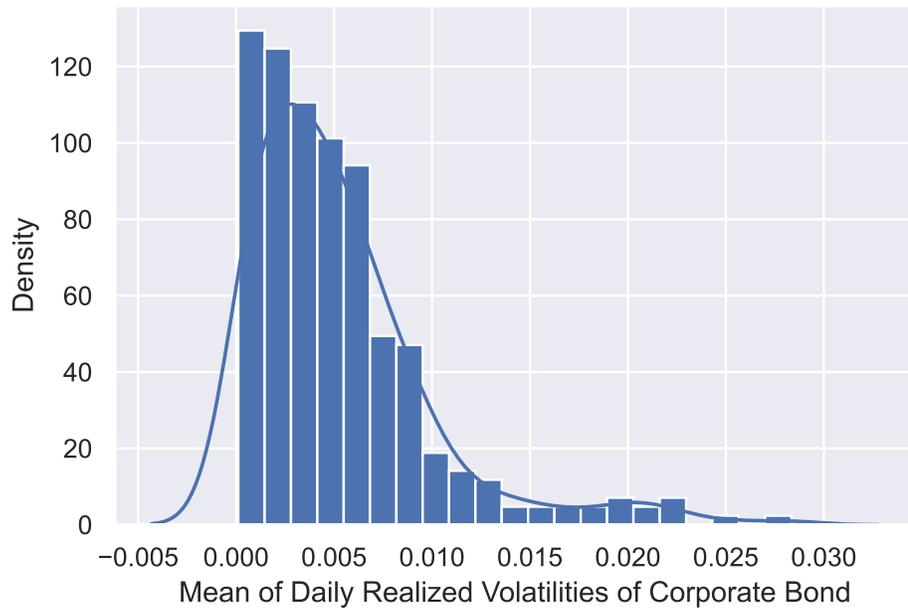
Figure 4A.2 The unconditional distribution of the mean of log of standard deviation
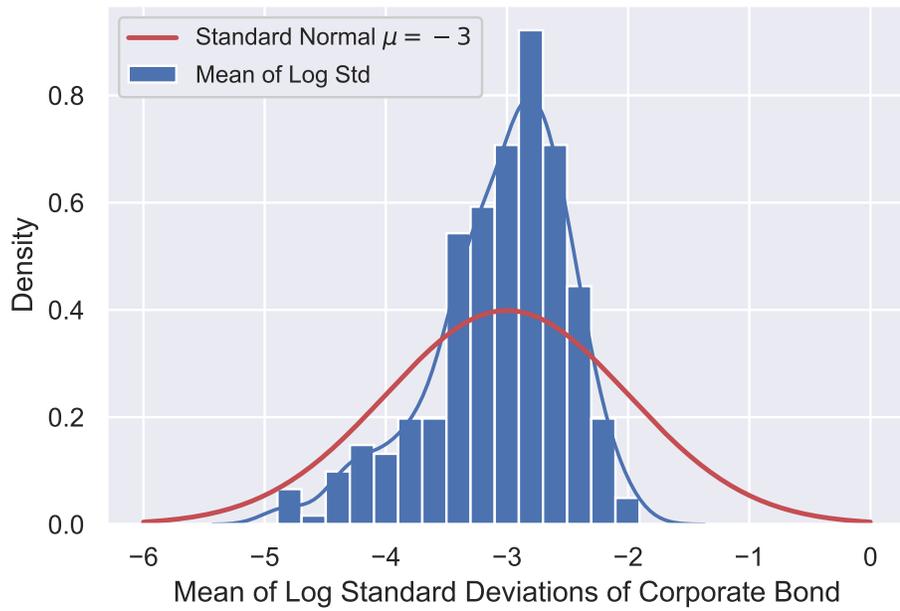
Figure 4A.3 The conditional distributions of the mean of the daily RV by credit rating
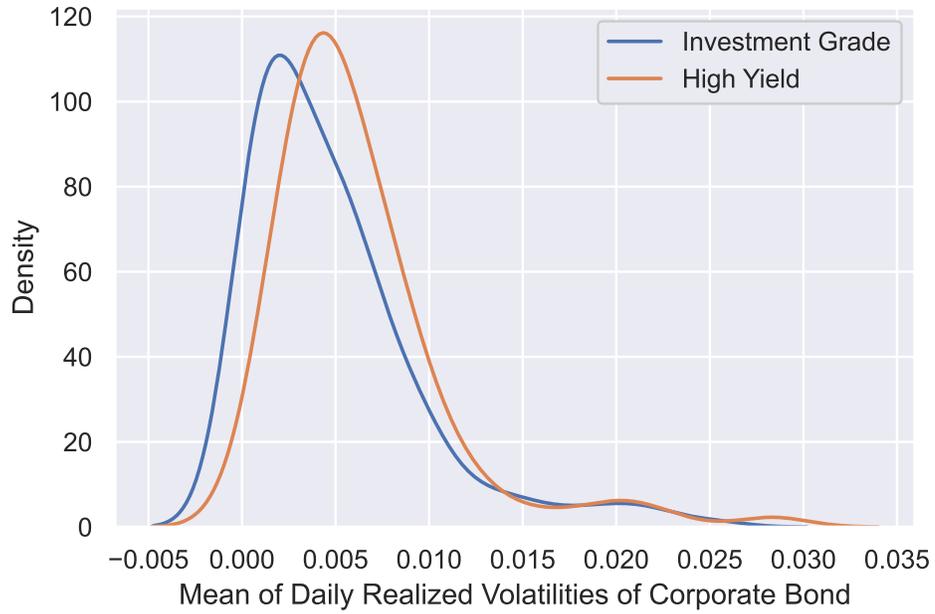
Figure 4A.4 The conditional distributions of the mean of the daily RV by size of issued amount
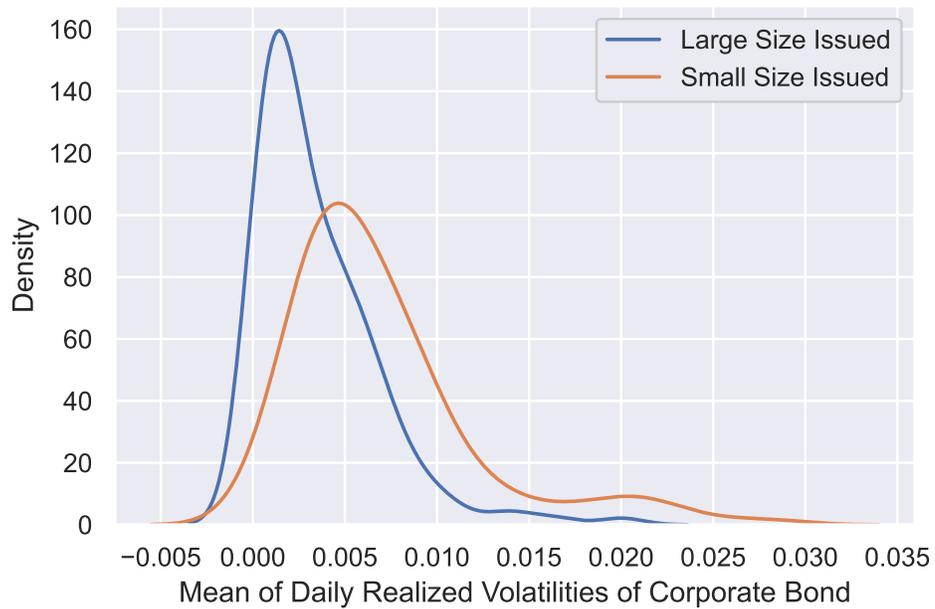
Figure 4A.5 The conditional distributions of the mean of the daily RV by yield rate
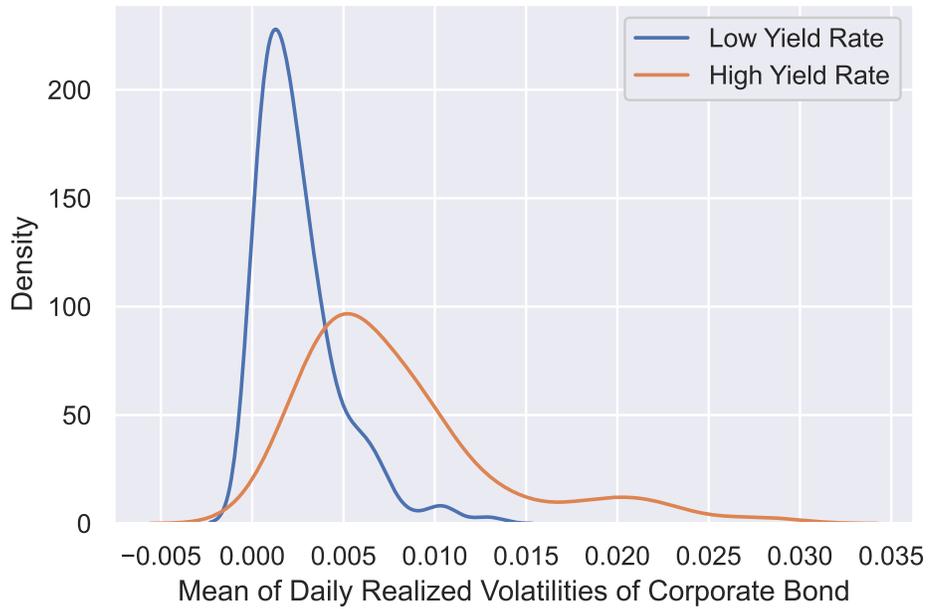
Figure 4A.6 The conditional distributions of the mean of the daily RV of corporate bonds when the return of the S&P 500 is small
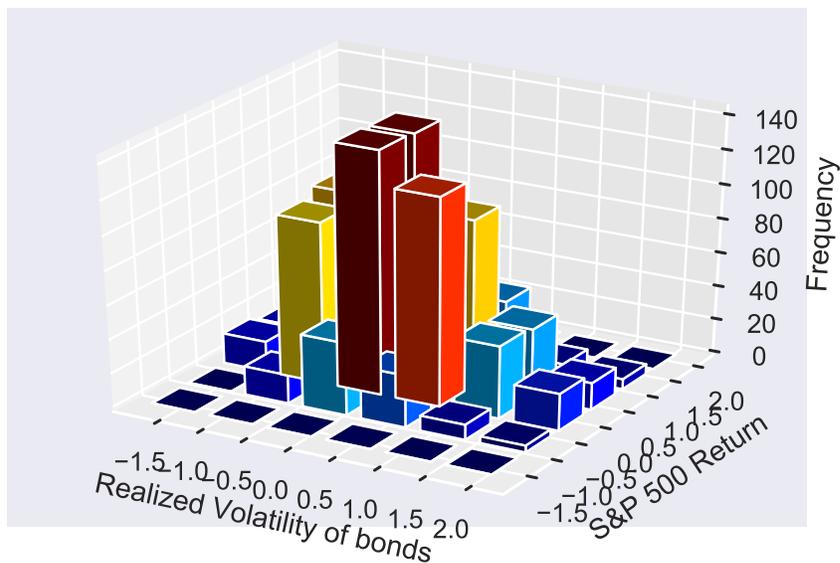
Figure 4A.7 The conditional distributions of the mean of the daily RV of corporate bonds when the return of the S&P 500 is big