THREE ESSAYS ON ROBUST INFERENCE FOR LINEAR PANEL MODELS WITH MANY TIME PERIODS

By

Yu Sun

A DISSERTATION

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

Economics – Doctor of Philosophy

2013

ABSTRACT

THREE ESSAYS ON ROBUST INFERENCE FOR LINEAR PANEL MODELS WITH MANY TIME PERIODS

By

Yu Sun

This dissertation consists of three chapters. The first chapter is a critique on the two-way cluster-robust standard errors. In the presence of both cross-sectional correlation and serial correlation, traditional one-way cluster-robust standard errors are not valid. A new robust variance estimator called two-way cluster-robust standard errors is proposed by Thompson (2011) and Cameron et al. (2011) to conduct accurate inference when double clustering exists. However, this approach does not allow for correlation across different firms in different time periods. If such correlation exists, then the two-way cluster-robust standard errors will fail to work. Monte Carlo simulation results demonstrate that using two-way cluster-robust standard errors may lead to unreliable inference even when there is a simple $AR(1)$ time effect. One solution to address this problem is proposed by Thompson (2011). He has improved the original formula for the two-way cluster-robust standard errors to account for correlation across different firms in different time periods. An alternative solution is the standard errors proposed by Driscoll and Kraay (1998) that are robust to cross-sectional correlation of general and unknown form as well as heteroskedasticity and serial correlation under covariance stationarity and weak dependence. The Driscoll and Kraay, 1998 (DK) standard errors perform well when firm dummies are included. Interestingly, without removing the firm effect, the DK standard errors do not behave well. Simulations results illustrate these interesting findings.

The second chapter provides an analysis of the standard errors proposed by Driscoll and Kraay (1998) in linear Difference-in-Differences (DD) models with fixed effects and individual-specific time trends. The analysis is accomplished within the fixed-b asymptotic framework developed by Kiefer and Vogelsang (2005) for heteroskedasticity and autocorrelation (HAC) robust covariance

matrix estimator based tests. For the fixed-N, large-T case, it is shown that fixed-b asymptotic distributions of test statistics constructed using the DD estimator and the DK standard errors are different from the results found by Kiefer and Vogelsang (2005) and Vogelsang (2012). The newly derived fixed-b asymptotic distributions depend on the date of policy change, λ, individual-specific trend functions as well as the choice of kernel and bandwidth. Whether time period dummies are included does not affect the fixed-b limits. For other regressors that don't have a structural change, the usual fixed-b asymptotic distributions still apply. Monte Carlo simulations illustrate the performance of the fixed-b approximations in practice.

The third chapter studies finite sample properties of the naive moving blocks bootstrap (MBB) tests based on the DK standard errors in linear DD models with individual fixed effects. The naive bootstrap procedure is a bootstrap where the formula used to compute the standard errors on the resampled data is the same as the formula used on the original data. Following the approach in Gonçalves (2011), the so-called "panel MBB" method is used in this chapter. This method applies the standard MBB to the time series of vectors containing all the individual observations at each time period. Monte Carlo simulation results show that the bootstrap is much more accurate than the standard normal approximation, and it closely follows the new fixed-b approximation proposed in the second chapter. This improvement holds for the special case of Bartlett kernel. Results would look similar for other kernels. It even holds when the independent and identically distributed ($i.i.d.$) bootstrap is used, despite potential serial correlation in the data. Simulation results also show that if the block length is appropriately chosen, the bootstrap can outperform the fixed-b approximation when there is strong serial correlation.

To my parents.

ACKNOWLEDGEMENTS

I would like to express the deepest appreciation to my committee chair, Professor Timothy Vogelsang, for his excellent guidance, caring, patience, and support throughout my dissertation. His wisdom, knowledge, and commitment to the highest standards inspired and motivated me. He kept me optimistic and cheered me up during the toughest moments. Without his guidance and persistent help this dissertation would not have been possible.

I want to thank my committee members, Professor Jeffrey Wooldridge and Richard Baillie, for providing me useful comments and great advice that improved my dissertation. Special thank goes to Professor John Jiang, who always shared his valuable experiences with me. I also want to thank Professor Soren Anderson, Steven Haider, Thomas Jeitschko, Peter Schmidt, Byron Brown and many other great faculty at Michigan State University, who helped me navigate life as a Ph.D. student.

I appreciate the assistance given by our friendly staff Belén Feight, Margaret Lynch, Lori Jean Nichols and Jon Glazier. I am grateful to my friend Seunghwa Rho who always supported me and helped me a lot. Many thanks to my friends Sukampon Chongwilaikasaem, Luke Chu, Cheol Keun Cho, Gaoyang Wang, Xiaoni Guo, Yuqing Zhou, Cuicui Lu, Wei Li, Xiaojun Wang and many others for every moment I shared with you at Michigan State University.

Last, but by no means least, I am particularly grateful for the courage, support and endless love I have received from my parents and my grandparents throughout my life. Thank you so much for giving me the opportunity of an education from the best universities and your understanding.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

ROBUST INFERENCE FOR LINEAR PANEL MODELS

1.1 Introduction

Many empirical papers in the accounting and finance literatures use panel data sets with observations on multiple firms over multiple time periods. In such panel data settings, the common assumption of independence in regression errors is likely to be violated. For example, temporary market-wide common shocks will cause correlation across firms in the same time period, and persistent firm characteristics will cause correlation over time. Moreover, persistent common shocks, such as business cycles, will cause correlation across different firms in different time periods. Potential clusterings are big challenges, since if we fail to take into account them, we will underestimate the standard error and hence over-reject the null hypothesis when conducting hypothesis tests. Therefore, how to conduct a robust inference plays a key role in empirical researches. Throughout this chapter, we call one dimension firm and the other time.

Various approaches are available to obtain "robust" standard errors. White (1980) proposed an approach to account for heteroskedasticity in cross-section data. Later White (1984) presented a formula for a multivariate dependent variable. Arellano (1987) proposed the well-known one-way cluster-robust standard errors in linear panel models. Wooldridge (2003) provided an overview of applications of cluster methods. Hansen (2007) investigated asymptotic properties of a robust variance matrix estimator for panel data when T is large. Fama and MacBeth (1973) proposed a method that computes standard errors robust to correlation across firms in the same time period. White standard errors and one-way cluster-robust standard errors are common in econometrics textbooks (e.g., Wooldridge, 2002).

Most papers in the literature only deal with clustering in one dimension and ignore clustering in the other dimension. Methods that control for clustering in one dimension usually assume independence in the other dimension. However, when both cross-sectional and serial correlation exist, the one-way cluster-robust method mis-specifies the error structure and underestimate the true standard error. This will lead to over-rejections in hypothesis testing. One solution is the two-way cluster-robust standard errors proposed by Thompson (2011) and Cameron et al. (2011). This variance estimator is designed to produce robust inference when there is two-way non-nested clustering. Specifically, in finance applications, clustering at the firm level and at the time (e.g. day) level is of interest. This method allows for serial correlation for a given firm and correlation across different firms in the same time period (cross-sectional correlation). However, this approach assumes that there is no correlation across different firms in different time periods. This method generalizes the standard cluster-robust variance estimator for one-way clustering to that for two-way clustering, and relies on similar relatively weak distributional assumptions. It can also be generalized to clustering with more than two dimensions (see Cameron et al., 2011).

Petersen (2009) has compared these robust standard errors and suggested using the two-way cluster-robust standard errors as a robustness check. Gow et al. (2010) find that two-way cluster-robust standard errors are required for valid inference in many accounting applications. However, the two-way clustering method only works for a specific and restricted error structure. In practice, the assumption that there is no correlation across different firms in different time periods is likely to be violated. Suppose now there is a common shock to all the firms in the same industry; it is much more realistic that this shock would affect those firms to some extent in the future rather than completely disappear at the end of the current time period. Hence different firms in different time periods may have some correlation between each other due to the lagged effect. This could happen in a business cycle. If so, then the two-way cluster-robust standard errors will probably fail. There are two solutions available to correct this problem. Thompson (2011) has improved the original formula for the two-way cluster-robust standard errors to account for correlation across different firms in different time periods. We will call it the revised two-way cluster-robust standard errors. Another alternative solution is to use the Driscoll and Kraay, 1998 (DK) standard errors which account for heteroskedasticity, autocorrelation and cross-sectional correlation of general and unknown form. A recent paper by Vogelsang (2012) has shown that fixed-b asymptotic approximations (see Kiefer and Vogelsang, 2005) for the DK standard errors perform substantially better than standard normal asymptotic approximations for either the DK standard errors or the one-way cluster-robust standard errors in the context of linear panel models with individual fixed effects and cross-sectional correlation.

The objective of this chapter is to show that in the presence of both firm effect and time effect, if there is correlation across different firms in different time periods, the two-way cluster-robust method fails. Furthermore, two possible solutions to correct this problem are analyzed using simulations. First, several tables from Petersen (2009) are replicated and similar results are found in simulations. In these tables, the sensitivity of standard error estimates to the presence of either firm effects or time effects is examined. Next, we study the performance of the two-way cluster-robust standard errors in the presence of both firm effects and time effects by comparing them to the White standard errors and the one-way cluster-robust standard errors. In this scenario, the two-way cluster-robust standard errors perform better than the one-way clustering method. Then, we assume that the time effect follows an $AR(1)$ process and analyze the performance of the two-way clustering method. When the absolute value of the autocorrelation parameter, ρ, is close to 1, the two-way clustering method generally fails and leads to over-rejections. Finally, we examine the performance of the revised two-way clustering method and the DK standard errors. The DK standard errors perform well when firm dummies are included. Without removing the firm effect, the DK standard errors do not behave well. Besides, firm dummies should be included if we care about the endogeneity problem.

The rest of this chapter is organized as follows. Section 1.2 describes the model and reviews several estimating methods for standard errors in panel data sets, including White, one-way cluster-robust, FM, original two-way cluster-robust, revised two-way cluster-robust and DK standard errors. Test statistics and their asymptotic distributions are also included in this section. Section 1.3 reports Monte Carlo simulation results. Section 1.3 also has theory for DK tests that explains some strange patterns in simulations. Section 1.4 concludes. Appendix A contains proofs of a theorem that explains the strange pattern of the DK standard errors when firm effects are not removed in the large-N, large-T case. Appendix B contains all simulation result tables.

1.2 The Model and Standard Errors

We follow the definitions for firm effects, time effects and persistent common shocks in Thompson (2011). Firm effect means that the errors have arbitrary serial correlation for a given firm. Time effect means that the errors have arbitrary correlation across different firms in the same time period. Persistent common shock means that the errors have arbitrary correlation across different firms in different time periods. Consider a linear regression model given by

$$y_{it} = x_{it}\beta + \varepsilon_{it}, \tag{1.1}$$

$$i = 1,2,\ldots,N, \quad t = 1,2,\ldots,T,$$

where y_{it}, x_{it} and ε_{it} are scalars. The error ε_{it} and the regressor x_{it} are assumed to have the same structure given by

$$\varepsilon_{it} = \gamma_i + \delta_t + \eta_{it}, \tag{1.2}$$

$$x_{it} = \mu_i + \theta_t + \xi_{it}, \tag{1.3}$$

with

$$\delta_t = \rho\delta_{t-1} + e_t, \tag{1.4}$$

$$\theta_t = \rho\theta_{t-1} + u_t, \tag{1.5}$$

where δ_t and θ_t have the same autocorrelation parameter ρ. γ_i and μ_i are firm effects. δ_t and θ_t are time effects. η_{it} and ξ_{it} are idiosyncratic errors. All error components have zero mean, finite variance, and are independent of each other. It is assumed that γ_i, μ_i, e_t, u_t, η_{it} and ξ_{it} all follow a normal distribution. δ_t and θ_t are serially correlated, and they follow an $AR(1)$ process. They are normal when $\rho = 0$.

The parameter of interest is β, and the estimation method is the ordinary least squares (OLS) estimator

$$\hat{\beta} = \left(\sum_{i=1}^{N}\sum_{t=1}^{T} x_{it}^2\right)^{-1} \sum_{i=1}^{N}\sum_{t=1}^{T} x_{it}y_{it}$$
$$= \beta + \left(\sum_{i=1}^{N}\sum_{t=1}^{T} x_{it}^2\right)^{-1} \sum_{i=1}^{N}\sum_{t=1}^{T} x_{it}\varepsilon_{it}. \tag{1.6}$$

Let $v_{it} = x_{it}\varepsilon_{it}$ and define $\hat{v}_{it} = x_{it}\hat{\varepsilon}_{it}$ where $\hat{\varepsilon}_{it}$ are the OLS residuals given by $\hat{\varepsilon}_{it} = y_{it} - x_{it}\hat{\beta}$. Let $\hat{Q} = \sum_{i=1}^{N}\sum_{t=1}^{T} x_{it}^2$ and $\Omega = \sum_{i,j,t,s} E(v_{it}v_{js})$. We need to estimate the covariance matrix to obtain robust tests. We will focus on the following approaches in this chapter: White standard errors, one-way cluster-robust standard errors, FM standard errors, original and revised two-way cluster-robust standard errors, and DK standard errors. Note that the FM approach also uses a different estimator of β. Details are discussed in subsection 1.2.2.

1.2.1 White and One-Way Cluster-Robust Standard Errors

In order to write down a general notation that nests each one-way approach, we use the group notation in this subsection. With observations grouped into G clusters of N_g observations, for $g \in \{1, \ldots, G\}$, we can rewrite model (1.1) as

$$\mathbf{y}_g = \mathbf{x}_g'\beta + \varepsilon_g,$$

where $\mathbf{y}_g$, $\mathbf{x}_g$ and ε_g are $N_g \times 1$ vectors. The one-way cluster-robust variance estimator is

$$\hat{V}_C = \hat{Q}^{-1}\left(\sum_{g=1}^{G} \hat{\mathbf{v}}_g\hat{\mathbf{v}}_g'\right)\hat{Q}^{-1}, \tag{1.7}$$

where $\hat{\mathbf{v}}_g$ is a $N_g \times 1$ vector containing all $\hat{\mathbf{v}}_{it}$ in cluster g. If each cluster only contains one single observation, then this estimator gives White (1980) standard errors

$$\hat{V}_{White} = \hat{Q}^{-1}\left(\sum_{i=1}^{N}\sum_{t=1}^{T} \hat{v}_{it}^2\right)\hat{Q}^{-1}. \tag{1.8}$$

If we cluster by firm, then $G = N$ and $N_g = T$. If we cluster by time, then $G = T$ and $N_g = N$. This estimator is consistent if

$$G^{-1}\sum_{g=1}^{G}\hat{\mathbf{v}}_g\hat{\mathbf{v}}_g' \xrightarrow{p} E(\mathbf{v}_g\mathbf{v}_g') \text{ as } G \to \infty. \tag{1.9}$$

When either firm effects or time effects exist, White standard errors are not valid. If there are firm effects only, we can cluster by firm. If there are time effects only, we can cluster by time. One-way cluster-robust standard errors allow for correlation of any unknown form within clusters, but the errors are assumed to be uncorrelated across clusters. When both firm effects and time effects are present, the consistency condition (1.9) is violated and thus the one-way clustering method fails to work.

1.2.2 FM Standard Errors

The Fama and MacBeth (1973) approach is originally used in asset pricing models such as the well-known capital asset pricing model (CAPM). Since stocks have weak serial correlation in daily and weekly holding periods, this approach is designed to correct cross-sectional correlation. In the original version of this approach, researchers run T cross-sectional regressions (one for each time period). For each coefficient β_j, the FM estimator is the average of the T estimates

$$\widehat{\beta}_j^{FM} = \frac{1}{T}\sum_{t=1}^{T}\hat{\beta}_{t,j}, \tag{1.10}$$

and the FM variance estimator is given by

$$s^2\left(\hat{\beta}_j^{FM}\right) = \frac{1}{T}\sum_{t=1}^{T}\frac{\left(\hat{\beta}_{t,j} - \hat{\beta}_j^{FM}\right)^2}{T-1}. \tag{1.11}$$

The variance formula assumes no correlation over time. Therefore, when there are only time effects, this approach produces a consistent variance estimator as $T \to \infty$. However, in the presence of firm effects, the assumption does not hold, and hence the FM standard errors tend to be too small.

1.2.3 Original and Revised Two-Way Cluster-Robust Standard Errors

Thompson (2011) and Cameron et al. (2011) have extended one-way cluster-robust standard errors to two-way cluster-robust standard errors that are robust to double clustering by firm and time. The original version just generalizes the one-way clustering method, and assumes no correlation across different firms in different time periods. Thompson (2011) noticed this limitation and proposed a revised version which takes into account correlation across different firms in different time periods. The revised formula is

$$\hat{V}^{r}_{double} = \hat{V}_{firm} + \hat{V}_{time,0} - \hat{V}_{White,0} + \sum_{l=1}^{L}(\hat{V}_{time,l} + \hat{V}'_{time,l}) - \sum_{l=1}^{L}(\hat{V}_{White,l} + \hat{V}'_{White,l}), \tag{1.12}$$

with

$$\hat{V}_{firm} = \hat{Q}^{-1}(\sum_{i=1}^{N}\hat{s}_i^2)\hat{Q}^{-1},$$
$$\hat{V}_{time,l} = \hat{Q}^{-1}(\sum_{t=l+1}^{T}\hat{s}_t\hat{s}_{t-l})\hat{Q}^{-1},$$
$$\hat{V}_{White,l} = \hat{Q}^{-1}(\sum_{i=1}^{N}\sum_{t=l+1}^{T}\hat{v}_{it}\hat{v}_{i,t-l})\hat{Q}^{-1}.$$

$\hat{s}_i = \sum_{t=1}^{T}\hat{v}_{it}$ is the sum of all observations for firm i. $\hat{s}_t = \sum_{i=1}^{N}\hat{v}_{it}$ is the sum of all observations for time t. This estimator is consistent as $\min(N,T) \to \infty$ (see Thompson, 2011). $\hat{V}_{firm}$ is the usual formula for standard errors clustered by firm, $\hat{V}_{time,0}$ is the usual formula for standard errors clustered by time, and $\hat{V}_{White,0}$ is the usual White standard errors. $\hat{V}_{firm}$ accounts for serial correlation for each firm, while $\hat{V}_{time,0}$ accounts for correlation across different firms in the same time period. The terms $\hat{V}_{time,l}$ with $l \geq 1$ account for the correlation across different firms in different time periods. The terms $\hat{V}_{White,l}$ with $l \geq 0$ are subtracted off because of double counting. The original two-way formula only contains the first three terms in (1.12)

$$\hat{V}_{double} = \hat{V}_{firm} + \hat{V}_{time,0} - \hat{V}_{White,0}. \tag{1.13}$$

Suppose there are 3 firms and 3 time periods. Table 1.1 illustrates the sample covariance matrix of the residuals under the assumptions for the original formula. The original version allows for correlation of any unknown form within clusters, clustering either by firm or by time, but it assumes no correlation across different firms in different time periods. The revised version corrects for potential persistent common shocks in the data. In fact, the $\hat{V}_{time,0} + \sum_{l=1}^{L} (\hat{V}_{time,l} + \hat{V}'_{time,l})$ part is exactly the DK standard errors using the truncated kernel with a truncation lag L. We will talk about the DK standard errors in details in the next subsection.

Table 1.1: **Residual cross product matrix:** When standard errors are clustered by both firm and time, correlation of residuals of the same firm in different years and residuals of the same year in different firms may be nonzero. However, correlation of residuals in different firms and different years are assumed to be zero.

	Firm 1			Firm 2			Firm 3		
Firm 1	ε_{11}^2	$\varepsilon_{11}\varepsilon_{12}$	$\varepsilon_{11}\varepsilon_{13}$	$\varepsilon_{11}\varepsilon_{21}$	0	0	$\varepsilon_{11}\varepsilon_{31}$	0	0
	$\varepsilon_{12}\varepsilon_{11}$	ε_{12}^2	$\varepsilon_{12}\varepsilon_{13}$	0	$\varepsilon_{12}\varepsilon_{22}$	0	0	$\varepsilon_{12}\varepsilon_{32}$	0
	$\varepsilon_{13}\varepsilon_{11}$	$\varepsilon_{13}\varepsilon_{12}$	ε_{13}^2	0	0	$\varepsilon_{13}\varepsilon_{23}$	0	0	$\varepsilon_{13}\varepsilon_{33}$
Firm 2	$\varepsilon_{21}\varepsilon_{11}$	0	0	ε_{21}^2	$\varepsilon_{21}\varepsilon_{22}$	$\varepsilon_{21}\varepsilon_{23}$	$\varepsilon_{21}\varepsilon_{31}$	0	0
	0	$\varepsilon_{22}\varepsilon_{12}$	0	$\varepsilon_{22}\varepsilon_{21}$	ε_{22}^2	$\varepsilon_{22}\varepsilon_{23}$	0	$\varepsilon_{22}\varepsilon_{32}$	0
	0	0	$\varepsilon_{23}\varepsilon_{13}$	$\varepsilon_{23}\varepsilon_{21}$	$\varepsilon_{23}\varepsilon_{22}$	ε_{23}^2	0	0	$\varepsilon_{23}\varepsilon_{33}$
Firm 3	$\varepsilon_{31}\varepsilon_{11}$	0	0	$\varepsilon_{31}\varepsilon_{21}$	0	0	ε_{31}^2	$\varepsilon_{31}\varepsilon_{32}$	$\varepsilon_{31}\varepsilon_{33}$
	0	$\varepsilon_{32}\varepsilon_{12}$	0	0	$\varepsilon_{32}\varepsilon_{22}$	0	$\varepsilon_{32}\varepsilon_{31}$	ε_{32}^2	$\varepsilon_{32}\varepsilon_{33}$
	0	0	$\varepsilon_{33}\varepsilon_{13}$	0	0	$\varepsilon_{33}\varepsilon_{23}$	$\varepsilon_{33}\varepsilon_{31}$	$\varepsilon_{33}\varepsilon_{32}$	ε_{33}^2

1.2.4 DK Standard Errors

Driscoll and Kraay (1998) first proposed the heteroskedasticity, autocorrelation and cross-section correlation (HACC) robust variance estimator using the time series of cross-sectional sums of observations. The idea is to first aggregate all the individual observations at each time period and then apply the HAC estimator to the time series of the sums. The first step takes into account potential cross-sectional correlation in the data, and the second step takes into account potential serial correlation in the data. Therefore, the DK standard errors are robust to cross-sectional correlation of unknown form as well as heteroskedasticity and serial correlation, assuming covariance stationarity and weak dependence in the time dimension.

Define $\hat{v}_t = \sum_{i=1}^{N} \hat{v}_{it}$, and let $\hat{\hat{\Gamma}}_j = T^{-1} \sum_{t=j+1}^{T} \hat{v}_t \hat{v}'_{t-j}$. The DK standard errors are given by

$$\hat{V}_{DK} = T\hat{Q}^{-1}\hat{\hat{\Omega}}\hat{Q}^{-1}, \tag{1.14}$$

with

$$\hat{\hat{\Omega}} = \hat{\hat{\Gamma}}_0 + \sum_{j=1}^{T-1} k(\frac{j}{M})(\hat{\hat{\Gamma}}_j + \hat{\hat{\Gamma}}'_j).$$

where $k(x)$ is a kernel function such that $k(x) = k(-x)$, $k(0) = 1$, $|k(x)| \leq 1$, $k(x)$ is continuous at $x = 0$, and $\int_{-\infty}^{\infty} k^2(x) < \infty$. M is the bandwidth parameter, or the truncation lag.

1.2.5 Test Statistics and Asymptotic Distributions

Consider testing the null hypotheses about β of the form $H_0 : \beta = \beta_0$. Define the t-statistic as

$$t = \frac{\hat{\beta} - \beta_0}{\sqrt{\hat{V}}}.$$

If we only assume heteroskedasticity, White standard errors are consistent as $N \to \infty$. If we allow for heteroskedasticity and general forms of serial correlation, firm clustered standard errors are consistent as $N \to \infty$. If we assume independence over time and allow for cross-sectional correlation, FM and time clustered standard errors are consistent as $T \to \infty$. Two-way clustered standard errors are consistent if there are serial correlation for a given firm and cross-sectional correlation at a given time period but no correlation across different firms in different time periods. Consistency of two-way cluster standard errors requires $N, T \to \infty$. So t-statistics based on these standard errors have a limiting standard normal distribution.

For the DK standard errors, the traditional asymptotic approach relies on $\hat{\Omega}$ being a consistent estimator of Ω. Consistency of $\hat{\Omega}$ requires that $M \to \infty$ as $T \to \infty$, but at a slower rate of convergence $\frac{M}{T} \to 0$. Under the traditional approach, the t-statistic has a limiting standard normal distribution. An alternative asymptotic theory has been proposed by Kiefer and Vogelsang (2005).

They model the bandwidth as a fixed proportion of the sample size. That is, $M = bT$ with b a fixed constant in $(0,1]$. Because b is held fixed in this approach, this alternative approach is usually labeled fixed-b asymptotics while the traditional approach is labeled small-b asymptotics. Under the fixed-b approach, $\hat{\Omega}$ converges to a random variable that depends on the kernel function and bandwidth, rather than a constant. As a result, the t-statistic has a nonstandard limiting distribution. This limiting distribution reflects the choice of kernel and bandwidth, but is otherwise pivotal. Fixed-b asymptotics provide more accurate and reliable inference than small-b asymptotics. For each kernel function, fixed-b critical values can be simulated. In particular, in linear panel models with individual fixed effects, Vogelsang (2012) has shown that

$$t \Rightarrow \frac{W_1(1)}{\sqrt{P_1((b)}},$$

where $\Rightarrow$ denotes weak convergence, $W_1(r)$ is the standard Wiener process, and $P_1(b)$ is a random matrix that depends on the kernel function and bandwidth. For example, in the case of Bartlett kernel,

$$P_1(b) = \frac{2}{b}\left(\int_0^1 B_1^2(r)dr - \int_0^{1-b} B_1(r)B_1(r+b)dr\right)$$

where $B_1(r) = W_1(r) - rW_1(1)$.

1.3 Finite Sample Performances

This section compares finite sample performances of the covariance matrix estimators described in section 1.2 under different error structures. First, errors with one-way clusering are considered. We follow Petersen (2009) and analyze the sensitivity of standard errors to the presence of firm effects or time effects. Next, we compare the performance of White, one-way cluster-robust, and original two-way cluster-robust standard errors in the context of double clustering and persistent common shocks. Finally, we examine the performance of revised two-way cluster-robust and DK standard errors in the context of persistent common shocks.

1.3.1 Data Generating Process

The data generating process (DGP) is based on model (1.1). Suppose the structures of ε_{it} and x_{it} satisfy (1.2), (1.3), (1.4) and (1.5). The true slope coefficient β is 1. When there are only firm effects, the correlation structures of ε_{it} and x_{it} take the following form

$$corr\left(x_{it}, x_{js}\right) = \begin{cases} 1, \text{ for } i = j \text{ and } t = s \\ \rho_x = \dfrac{\sigma_\mu^2}{\sigma_x^2}, \text{ for } i = j \text{ and all } t \neq s \\ 0, \text{ for all } i \neq j \end{cases}$$

$$corr\left(\varepsilon_{it}, \varepsilon_{js}\right) = \begin{cases} 1, \text{ for } i = j \text{ and } t = s \\ \rho_\varepsilon = \dfrac{\sigma_\gamma^2}{\sigma_\varepsilon^2}, \text{ for } i = j \text{ and all } t \neq s \\ 0, \text{ for all } i \neq j \end{cases}$$

When there are only time effects, the correlation structures of ε_{it} and x_{it} take the following form

$$corr\left(x_{it}, x_{js}\right) = \begin{cases} 1, \text{ for } i = j \text{ and } t = s \\ \rho_x = \dfrac{\sigma_\theta^2}{\sigma_x^2}, \text{ for } t = s \text{ and all } i \neq j \\ 0, \text{ for all } t \neq s \end{cases}$$

$$corr\left(\varepsilon_{it}, \varepsilon_{js}\right) = \begin{cases} 1, \text{ for } i = j \text{ and } t = s \\ \rho_\varepsilon = \dfrac{\sigma_\delta^2}{\sigma_\varepsilon^2}, \text{ for } t = s \text{ and all } i \neq j \\ 0, \text{ for all } t \neq s \end{cases}$$

So the variance of γ_i (or δ_t), μ_i (or θ_t), η_{it} and ξ_{it} can be written as $\rho_\varepsilon \cdot \sigma_\varepsilon^2$, $\rho_x \cdot \sigma_x^2$, $(1-\rho_\varepsilon) \cdot \sigma_\varepsilon^2$ and $(1-\rho_x) \cdot \sigma_x^2$, respectively. In order to examine the sensitivity of standard errors to the presence of either firm effects or time effects, we set $\sigma_x = 1$, and $\sigma_\varepsilon = 2$. We allow the fraction of the variance of x_{it} and ε_{it} caused by the firm effect, *i.e.* ρ_x and ρ_ε respectively, to vary from 0% to 75%. The simulation results are based on 5,000 random samples with 500 firms and 10 years per firm. The empirical null rejection probabilities of t-statistics built upon White, one-way cluster-robust and FM standard errors are reported at a two-sided significance level 1%.

When there are double clustering and persistent common shocks, we focus on the comparison of the performances of each variance estimator. The DGP follows (1.2) and (1.3), with both firm effects and time effects. Firm effects (γ_i, μ_i) and idiosyncratic errors (η_{it}, ξ_{it}) follow a standard normal distribution. For a special case of double clustering but no persistent common shocks, time effects (δ_t, θ_t) are assumed to follow a standard normal distribution ($\rho = 0$). For a special case of persistent common shocks, time effects (δ_t, θ_t) are assumed to follow an $AR(1)$ process ($\rho > 0$). The (N,T) combinations vary in different simulations, but all simulations are based on 2,000 random samples. In the double clustering case, we allow N and T to vary from 10 to 250 separately. In the persistent common shock case, we allow $N = T = 10, 50, 250$. The autocorrelation parameter, ρ, takes values from -0.95 to 0.95 in Table B.6, B.7 and B.8. $\rho = 0, 0.3, 0.6, 0.9$ in Table B.9 and B.10. For the DK standard errors, we focus on the Bartlett kernel, $k(x) = 1 - |x|$ for $|x| \leq 1$ and $k(x) = 0$ for $|x| \geq 1$. We set the bandwidth $b = 0.1, 0.2, \ldots, 0.9$. The truncation lag in the revised two-way clustering method is set to be the same as the bandwidth in DK. The empirical null rejection probabilities of t-statistics are reported at a two-sided significance level 5%.

1.3.2 Results

Table B.1-B.4 illustrate how sensitive standard errors are to the presence of either firm effects or time effects. The DGP of Table B.1 and B.2 contains firm effects only, and the DGP of Table B.3 and B.4 contains time effects only and $\rho = 0$. Table B.1 and B.3 report empirical null rejection probabilities of t-statistics based on White standard errors and one-way cluster-robust standard errors. Table B.2 and B.4 report empirical null rejection probabilities of t-statistics based on FM standard errors. ρ_X varies across columns while ρ_ε varies across rows. In Table B.1 and B.3, each cell contains the average OLS estimate of β and the standard deviation of $\hat{\beta}$. The third and fifth entry are the average White standard errors and clustered standard errors, respectively. The empirical null rejection probabilities of White and clustered t-statistics at a two-sided significance level 1% are shown in square brackets below the standard error estimates. In Table B.2 and B.4, each cell contains the average FM coefficient estimate and the standard deviation of $\hat{\beta}$. The third entry is average FM standard errors. The empirical null rejection probabilities of FM t-statistics at a two-sided significance level 1% are shown in square brackets below.

For example, consider the case where 50% of the variability in both the error and the regressor is due to the firm effect or the time effect, *i.e.* $\rho_X = \rho_\varepsilon = 0.50$. In Table B.1, the average OLS coefficient estimate is 1.0008 and the standard deviation of the OLS coefficient estimate is 0.0510. The White standard error estimate is 0.0283 and the clustered standard error is 0.0508. 15.98% of the White t-statistics are greater than 2.58 in absolute value, while 1.02% of the clustered t-statistics are greater than 2.58 in absolute value. In Table B.2, the average FM coefficient estimate is 1.0008 and the standard deviation of the FM coefficient estimate is 0.0511. The FM standard error estimate is 0.0239 and 24.98% of the FM t-statistics are greater than 2.58 in absolute value. In Table B.3 , the average OLS coefficient estimate is 0.9966 and the standard deviation of the OLS coefficient estimate is 0.3073. The White standard error estimate is 0.0277 and the clustered standard error estimate is 0.2445. 81.28% of the White t-statistics are greater than 2.58 in absolute value, while 7.40% of the clustered t-statistics are greater than 2.58 in absolute value. In Table B.4, the average FM coefficient estimate is 0.9999 and the standard deviation of the FM coefficient estimate is 0.0282. The FM standard error estimate is 0.0276 and 2.68% of the FM t-statistics are greater than 2.58 in absolute value.

If there are no firm (time) effects in either the error or the regressor, White standard errors work well. As you can see from Table B.1 and B.3, in the first row and first column, the rejection probabilities are around 1%. However, as long as both of the regressor and the error contain firm (time) effects, White standard errors underestimate the variance and lead to over-rejections. As ρ_X and ρ_ε increase, White standard errors remain the same either across columns or across rows, but the true standard errors increase. In contrast, standard errors clustered by firm are very close to the true standard errors. In Table B.1, the rejection probabilities for clustered t-statistics are around 1%, despite the change of ρ_X and ρ_ε. In this setting, one-way cluster-robust standard errors correctly account for the correlation in the data and produce accurate inference. In Table B.3, standard errors clustered by time are much more accurate than White standard errors, but they still underestimate the true standard errors. Moving down the diagonal of Table B.3 from upper left to bottom right, the rejection probabilities for clustered t-statistics at a two-sided significance level 1% go from 4.04% to 9.16%. One possibility is that we have large N and small T ($N = 500$ and $T = 10$) in the DGP. There are only ten clusters if clustered by time, which is not large enough for standard normal approximations to be valid.

The FM approach is designed to account for correlation across different firms in the same time period, so when there are only firm effects, FM standard errors fail to account for serial correlation. From Table B.2, we can see that FM standard errors are biased downward. Moving down the diagonal of Table B.2 from upper left to bottom right, the true standard errors rise while the FM standard errors shrink. In the presence of time effects only, the FM approach works well. FM standard errors are very close to the true standard errors, and the rejection probabilities for FM t-statistics at a two-sided significance level 1% are approximately 3% for all cells in Table B.4.

When there are both firm effects and time effects, one-way cluster-robust standard errors would probably be biased. According to Petersen (2009), a common approach to address double clustering is to include a full set of time dummies and then cluster by firm. If the time effect is constant across firms in the same time period, then time dummies completely eliminate the time effect. What is left in the error term is just the firm effect. However, this approach only works when the correlation is correctly specified. If the time effect is not constant across firms, time dummies will not completely remove the time effect, and thus standard errors clustered by firm would be biased. Another limitation of the inclusion of dummies that empirical researchers care about is that it restricts the types of regressors that can be included. One solution suggested by Petersen (2009) is to cluster by firm and time simultaneously, using the two-way cluster-robust standard errors proposed by Thompson (2011) and Cameron et al. (2011). Table B.5 compares performances of White, one-way cluster-robust and original two-way cluster-robust standard errors.

In Table B.5, the DGP contains firm effects and time effects, but no persistent common shocks ($\rho = 0$). N and T vary from 10 to 250 separately. Column 1 reports the average OLS coefficient estimates, and column 2-5 report the empirical null rejection probabilities for t-statistics based on White, firm clustered, time clustered and original two-way clustered standard errors, respectively, at a two-sided significance level 5%. Rejection probabilities of White and clustered t-statistics are substantially larger than 5%. When N and T are close and both of them are large, the original two-way cluster-robust standard errors work well. Table B.5 shows that when $N = T = 50$, the rejection probability is 7.55%. When $N = 50$ and $T = 100$, the rejection probability is 6.70%. When $N = T = 100$, the rejection probability is 6.65%. When $N = 100$ and $T = 250$, the rejection probability is 4.85%. When $N = T = 250$, the rejection probability is 6.10%. When $N = 250$ and $T = 100$, the rejection probability is 5.60%. The larger the sample size, the greater the improvement.

The limitation of the original two-way clustering method is that although it considers cross-sectional correlation in the same time period, it does not allow for correlation across different firms in different time periods. If persistent common shocks such as business cycles exist, failure to account for them would lead to over-rejections. This approach should take into account cross-section correlation of general form.

Table B.6 to B.8 compare performances of White, one-way cluster-robust and original two-way cluster-robust standard errors when the time effect follows an $AR(1)$ process. We set $N = T = 10, 50, 250$ respectively. Column 1 reports the average OLS coefficient estimates, and column 2-5 report the empirical null rejection probabilities for t-statistics based on White, firm clustered, time clustered and original two-way clustered standard errors, respectively, at a two-sided significance level 5%.

Again, rejection probabilities of White and clustered t-statistics are substantially larger than 5%. When N and T are small, the original two-way clustered standard errors do not work no matter what value ρ takes. Even when $\rho = 0$, this method would produce a rejection probability at 12.85%. This confirms that the two-way clustering approach needs both N and T to be sufficiently large. When $N = T = 50$, different stories happen when ρ is close to zero and when ρ is close to one. When ρ is close to zero, correlation across different firms in different time periods are weak. The original two-way cluster-robust standard errors are still reasonable. For example, when

$\rho = 0.1$, the rejection probability is 6.60%. However, when correlation across different firms in different time periods is strong, the original two-way clustering method over-rejects. When $\rho = 0.7$, the rejection probability is 22.15%. When $\rho = 0.9$, the rejection probability rises to 45.60%. Increase in sample size helps improve the inference if ρ is small ($|\rho| \leq .7$ in the tables). For large ρ, increasing N, T makes it even worse for the two-way approach. As shown in Table B.8, when $\rho = 0.1$, the rejection probability is 5.05%, while in Table B.7, it is 6.60%. When ρ is very close to 1, over-rejection becomes more severe. When $\rho = 0.9$, the rejection probability is 52.65% while in Table B.7 it is 45.60%.

Table B.9 and B.10 compare performances of one-way cluster-robust, original and revised two-way cluster-robust, and DK standard errors when the time effect follows an $AR(1)$ process. Usual fixed-b critical values are used for t-statistics based on the DK standard errors. Table B.9 uses the standard OLS estimator, while Table B.10 uses the fixed-effects OLS estimator. We set $N = T = 50, 250$. There are several interesting findings to note. In both tables, one-way cluster-robust standard errors over-reject a lot. The original double clustering method is okay when T is large and ρ is small. When $N = T = 250$ and $\rho = 0.3$, the rejection probability is 6%. The revised double clustering method has a better performance than the original one only when ρ is large and the truncation lag is not large. However, this revised method still over-rejects. When $N = T = 50$, $\rho = 0.9$, and the truncation lag $L = 5$, the rejection probability of the original version is 52.5% while the rejection probability of the revised version is 29%. When $N = T = 250$, $\rho = 0.9$, and the truncation lag $L = 5$, the rejection probability of the original version is 50.9% while the rejection probability of the revised version is 17.1%. Also, rejection probabilities of the revised method increases as the truncation lag gets bigger. Without including firm dummies, the DK standard errors have a strange pattern. Rejection probabilities of the DK standard errors fall as ρ increases. In Table B.10, rejection probabilities of firm clustered standard errors are substantially larger than 5%. Rejection probabilities of time clustered standard errors and original two-way cluster-robust standard errors are very close, since firm effects are removed by firm dummies. Similar interesting patterns are found for the revised double clustering method. The patterns of the DK standard errors are consistent with those in Vogelsang (2012), and they behave very well. When $N = T = 250$ and $\rho = 0, 0.3$, the rejection probabilities are approximately 5% for all values of the bandwidth b. The DK standard errors still behave well even when $\rho = 0.9$. When $N = T = 250$, $\rho = 0.9$, and $b = .9$, the rejection probability is 8.8%.

The strange pattern of the DK standard errors in Table B.9 is caused by the presence of firm effects. Theoretical evidence is provided in the next subsection. The patterns of the revised double clustering method can be explained in two ways. First, as mentioned in subsection 1.2.3, the part accounts for potential persistent common shocks in the data is exactly the DK standard errors with truncation kernel. The downweighting causes downward bias of the variance estimator, and thus over-rejections. This explains why rejection probabilities of the revised version is bigger than those of the original version. Second, the revised two-way approach relies on the variance estimator being consistent. Using the traditional approach leads to unreliable inference.

1.3.3 Strange Patterns of the DK Standard Errors

This section presents theoretical evidence to explain the strange patterns of the DK standard errors in the large-N, large-T case. All limits are taken as $N, T \to \infty$. Proofs are provided in Appendix A.

Consider model (1.1) with x_{it} and ε_{it} satisfying (1.2), (1.3), (1.4) and (1.5). Consider testing the null hypotheses about β of the form

$$H_0 : \beta = \beta_0.$$

Define the t-statistic as

$$t_{DK} = \frac{\hat{\beta} - \beta_0}{\sqrt{\hat{V}_{DK}}}.$$

The following theorem summarizes the theoretical results for large-N, large-T case when firm dummies are not included in the model.

Theorem 1.1. *Suppose model (1.1) has one regressor x_{it}, and the structures of ε_{it} and x_{it} satisfy (1.2), (1.3), (1.4) and (1.5). Suppose firm dummies are not included in the model. Assume $M = bT$*

where $b \in (0,1]$ is fixed. Assume $N = \phi T$ such that $N \to \infty$ when $T \to \infty$. The Bartlett kernel is considered. As $T \to \infty$,

1. *If the regressor and errors in model (1.1) contain both firm effects and time effects, then*

$$\sqrt{N}\left(\widehat{\beta}-\beta\right) \Rightarrow Q^{-1}\sqrt{1+\phi\sigma^2}Z_1, \tag{1.15}$$

$$t_{DK} \Rightarrow \sqrt{1+\frac{1}{\phi\sigma^2}}\cdot\frac{Z_1}{\sqrt{P(b)}}, \tag{1.16}$$

   where $Z_1 \sim N(0,1)$, and $P(b)$ is a random variable depending on bandwidth. Z_1 is independent of $P(b)$, and σ^2 is the long run variance of $\theta_t\delta_t$.

2. *If the regressor and errors in model (1.1) only contain firm effects, then*

$$\sqrt{N}\left(\widehat{\beta}-\beta\right) \Rightarrow Q^{-1}Z_2, \tag{1.17}$$

$$\left|t_{DK}\right| \to \infty, \tag{1.18}$$

   where $Z_2 \sim N(0,1)$.

3. *If the regressor and errors in model (1.1) only contain time effects, then usual fixed-b limits (see Vogelsang, 2012) are obtained.*

Note that when the model satisfies (1.2), (1.3), (1.4) and (1.5), it is easy to show that $\theta_t\delta_t$ satisfies a Functional Central Limit Theorem (FCLT). However, it is not necessary to assume that the time effects θ_t and δ_t are independent and they both follow $AR(1)$. The assumption can be relaxed to allow for a more general setting. We only need to assume that $\theta_t\delta_t$ satisfies a FCLT. That is, $T^{-\frac{1}{2}}\sum_{t=1}^{[rT]}\theta_t\delta_t \Rightarrow \sigma W(r)$, where $W(r)$ is a standard Wiener process and σ^2 is the long run variance of $\theta_t\delta_t$.

Theorem 1.1 shows that in the presence of firm effects and time effects, if firm dummies are not included, the fixed-b limit of t_{DK} is not asymptotically pivotal as usual. It depends on the ratio, $\phi = \frac{N}{T}$, and the long run variance of $\theta_t\delta_t$, σ^2. The reason is that the firm effect destroys the weak dependence needed for results of Vogelsang (2012) to hold. Result (1.16) indicates that the usual fixed-b critical values have to be scaled by a nuisance parameter which is generally unknown in practice. As a consequence, in practice one would have to either: i) estimate the scaling factor or ii) include firm dummies to get back the asymptotically pivotal limit. Yet another important reason to recommend the inclusion of firm dummies is the problem of endogeneity. Empirical researchers are worried about the regressors that are not time-varying, and want to leave out firm dummies. However, they must be very careful because solving the endogeneity problem should be a priority. Including firm dummies removes the individual heterogeneity that is correlated to the regressors. Furthermore, if the individual heterogeneity is the source that generates cross-sectional correlation, the inclusion of firm dummies would completely eliminate the cross-sectional correlation and thus one-way clustered standard errors would work.

Table B.11 demonstrates the performance of the DK standard errors in the presence of firm effects and $AR(1)$ time effects, using the adjusted fixed-b critical values derived in Theorem 1.1. Patterns look similar to Vogelsang (2012). For a given N, T, ρ combination, rejection probabilities are above 5% with small b and they steadily decline as b increases. For a given value of ρ, as T increases, rejection probabilities approach 5% for all bandwidths. When $T = 250$ and $b = 1$, rejection probabilities are around 7% or 8% when there is strong serial correlation ($\rho = 0.9$). Rejection probabilities rise as ρ increases.

When there are no time effects and only firm effects, the DK standard error estimate tends to decline toward zero, and thus the t-statistic would go to infinity. Table B.12 illustrates the performance of the DK standard errors in this case, using the usual fixed-b critical values. Given N, as T increases, rejection probabilities for the DK standard errors blow up toward 1 for all bandwidths. In contrast, rejection probabilities for firm clustered standard errors are close to 5% when N is large, which is expected because the one-way approach is designed to account for any form of serial correlation assuming independence in the cross section. Also, when both N and T are large, the two-way approach gives similar results as the one-way approach.

1.4 Conclusion

This chapter compares finite sample performances of White, FM, one-way cluster-robust, two-way cluster-robust and DK standard errors using Monte Carlo simulations. If there is only one-way clustering, one-way clustered standard errors could work very well. However, in the presence of two-way clustering, one-way clustered standard errors is not sufficient to take into account all potential correlations in the data. Petersen (2009) suggests applied researchers use original two-way cluster-robust standard errors. When there are no persistent common shocks, this two-way clustering method is valid and it allows for any unknown form of correlation within clusters. The limitation of this method is that it does not take into account correlation across different firms in different time periods. If we assume the time effect to be a simple AR(1) process which generates correlation across different firms in different time periods, the original two-way clustering approach over-rejects when there is strong serial correlation (ρ is large). As a result, we need to find a solution to solve this problem. Thompson (2011) has improved the original formula for the two-way cluster-robust standard errors to account for correlation across different firms in different time periods.

Another alternative solution is to use the DK standard errors which account for heteroskedasticity, autocorrelation and cross-sectional correlation of general and unknown form. The DK standard errors are valid only when firm effects are removed. The presence of firm effects will distort the results and lead to strange outcomes for the DK standard errors. Theoretical evidences indicate that the usual fixed-b critical values have to be scaled by a nuisance parameter which is generally unknown in practice. Therefore, empirical researchers have to choose between estimating the scaling factor and including firm dummies. Another reason to include firm dummies is that they would eliminate the individual heterogeneity that is potentially correlated with the regressors. After firm effects are removed, the DK standard errors produce remarkably better performance than other standard errors.

In sum, using the original two-way cluster-robust standard errors as a robustness check only works in a special case of double clustering. When persistent common shocks are concerned, the DK standard errors should be considered as a robustness check. However, the DK standard errors are valid under the assumptions of covariance stationarity and weak dependence in the time dimension. Also, firm dummies should be included to remove firm effects. Otherwise, one has to estimate the nuisance parameter to adjust the fixed-b critical values.

CHAPTER 2

FIXED-b INFERENCE FOR DIFFERENCE-IN-DIFFERENCES ESTIMATION

2.1 Introduction

This chapter focuses on fixed-b asymptotic distributions of the $Wald$ and t statistics for Difference-in-Differences (DD) estimation in linear panel settings. Recently, DD estimation has become increasingly popular in policy analysis. DD estimation involves identifying a specific intervention or treatment (often a policy change or a passage of a law). Applied researchers then compare the difference in outcomes before and after the intervention for groups affected by the intervention (treatment groups) to the same difference for unaffected groups (control groups). Such panel data sets often contain serial correlation and/or spatial correlation in the cross section. Even though the correlation structure is not of interest, the failure to account for potential serial and spatial correlation may lead to severe distortions in the inference about parameters of interest. After Bertrand et al. (2004) pointed out that standard errors robust to serial correlation should be considered in DD estimation, using clustered standard errors (see Arellano, 1987) has become a standard method to deal with serial correlation in the DD context. Hansen (2007) extended the results for the traditional short panel case, large-N, fixed-T case, to large-N, large-T and fixed-N, large-T cases. The clustered standard errors are valid under the assumption that individuals are uncorrelated with each other. In other words, spatial correlation in the cross section is often ignored. Wooldridge (2003) provided a useful discussion of cluster methods. Sometimes the cross-sectional observations can be divided into groups or clusters where it is assumed that individuals within a cluster are correlated while individuals across clusters are uncorrelated. In this case, standard errors robust to cross-section clustering can be constructed. The number of clusters could be small, though.

In time series econometrics, the nonparametric HAC robust covariance matrix estimator (see Newey and West, 1987) is widely used. To handle the spatial correlation, robust standard errors can be obtained using the approaches of Conley (1999), Kelejian and Prucha (2007), Bester et al. (2008), Bester et al. (2011) or Kim and Sun (2011a) when a distance measure is available. Kim and Sun (2011b) provides results on kernel HAC standard errors in linear panel models with individual and time dummy variables using a distance measure. When a distance measure is either unavailable or unknown for the cross section of the panel, the DK approach can be used to obtain robust standard errors. Driscoll and Kraay (1998) established consistency of these standard errors under mixing conditions. However, the mixing conditions do not hold for the fixed-effects estimator. Fortunately, Gonçalves (2011) has established consistency of the DK standard errors for the fixed-effects estimator in the presence of general forms of cross-sectional correlation. A recent paper by Vogelsang (2012) develops a fixed-b asymptotic theory for test statistics based on the fixed-effects estimator and the DK standard errors following Kiefer and Vogelsang (2005).

This chapter provides an analysis of the DK standard errors in linear DD models with fixed effects and individual-specific time trends. The analysis is accomplished within the fixed-b asymptotic framework proposed by Kiefer and Vogelsang (2005) for HAC estimator based tests. Fixed-b asymptotics are appealing because they reflect the influence of the choice of kernel and bandwidth on the behavior of the standard errors while the traditional asymptotics don't. Large-T framework is required in the fixed-b approach. According to the survey of DD papers in Bertrand et al. (2004), among 92 DD papers they found, 10% have at least 36 time periods and 5% have at least 51 time periods. Therefore, it is feasible to use the DK standard errors for DD estimation to cope with any general forms of spatial correlation in the cross section given covariance stationarity and weak dependence in the time dimension. This chapter only considers fixed-N, large-T case. Simulation results suggest that the asymptotic theory can be extended to large-N, large-T case.

The main objective of this chapter is to derive fixed-b asymptotic distributions of test statistics constructed using the DD estimator and the DK standard errors. It is found that the fixed-b limits are different from those derived by Kiefer and Vogelsang (2005) and Vogelsang (2012). The newly derived fixed-b asymptotic distributions depend on the date of policy change, λ, and individual-specific trend functions in addition to the choice of kernel and bandwidth. For the individual fixed-effects model with no trend, the fixed-b asymptotic distributions are the same as found in a pure time series model with a shift in mean. New critical values are simulated in this study and they have a U-shape with respect to λ. Whether time period dummies are included does not affect the fixed-b asymptotic distributions. For other regressors that don't have a structural break, the fixed-b asymptotic distributions for DK test statistics found in Vogelsang (2012) still apply. The traditional short panel case is not included. With T fixed, there is not sufficient information in the time dimension for the DK approach to work.

The remainder of the chapter is organized as follows. The next section describes the DD models and test statistics. Section 2.3 presents the fixed-b asymptotic results for test statistics constructed using the DD estimator and the DK standard errors, and new critical values for t statistics in two special cases. Finite sample properties are examined in Section 2.4. Section 2.5 concludes. Proofs are given in Appendix C, and tables are given in Appendix D.

Throughout the chapter, x_{it} and β denote the full set of regressors and parameters respectively in each model. "$'$" denotes the transpose, when used in the context of a vector.

2.2 Model Setup and Test Statistics

Consider a DD model with fixed effects and individual-specific deterministic trends given by

$$y_{it} = \mathbf{f}(t)'\mathbf{a_i} + \beta_1 Treat_i + \beta_2 DU_t + \beta_3 Treat_i \cdot DU_t + u_{it}, \tag{2.1}$$
$$i = 1,2,\ldots,N, \quad t = 1,2,\ldots,T,$$

where y_{it} and u_{it} are scalars, $\mathbf{f}(t)$ denotes a $J \times 1$ vector of trend functions, $\mathbf{a_i}$ denotes a $J \times 1$ vector of individual-specific unobservable variables.[1] $Treat_i$ denotes an indicator for individuals in the treatment group which takes one if individual i is in the treatment group. Without loss of generality, we assume that the first kN individuals are in the treatment group. Thus, $Treat_i = \mathbb{1}(i \leq kN)$. DU_t

[1] $\mathbf{a_i}$ could be either random or deterministic. Asymptotic results will not differ because of the de-trending transformation.

denotes an indicator for post-policy-change time periods which takes one after the policy change. That is, $DU_t = \mathbb{1}(t > \lambda T) = \mathbb{1}(r > \lambda)$, where the parameter λ is the relative date of policy change within the time sample. Both k and λ are assumed known. Often time fixed effects are included which gives the model

$$y_{it} = \lambda_{\mathbf{t}} + \mathbf{f}(t)'\mathbf{a_i} + \beta_1 Treat_i + \beta_2 DU_t + \beta_3 Treat_i \cdot DU_t + u_{it}. \tag{2.2}$$

An alternative model includes common time trends instead of time fixed effects. The asymptotic results for the alternative model remain unchanged. A more general model with additional regressors is

$$y_{it} = \mathbf{f}(t)'\mathbf{a_i} + \beta_1 Treat_i + \beta_2 DU_t + \beta_3 Treat_i \cdot DU_t + \mathbf{z_{it}}'\gamma + u_{it}, \tag{2.3}$$

where $\mathbf{z_{it}}$ is a $(K \times 1)$ vector of additional regressors. Including time fixed effects gives the model

$$y_{it} = \lambda_{\mathbf{t}} + \mathbf{f}(t)'\mathbf{a_i} + \beta_1 Treat_i + \beta_2 DU_t + \beta_3 Treat_i \cdot DU_t + \mathbf{z_{it}}'\gamma + u_{it}. \tag{2.4}$$

The focus is on estimation and inference about β_3, which explains the impact of a policy change on y. The ordinary least squares (OLS) estimator of β_3, $\hat{\beta}_3$, is usually referred to as DD estimator. Since we are primarily interested in the DD estimator, we could do a de-trending transformation to get rid of the unobservable variables $\lambda_{\mathbf{t}}$ and $\mathbf{a_i}$, similar to the fixed-effects transformation. Therefore, we will call the de-trended OLS estimator the "fixed-effects OLS estimator" in the remainder. Consider the fixed-effects OLS estimator of β given by

$$\hat{\beta} = \left(\sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{\mathbf{x}}_{it} \tilde{\mathbf{x}}_{it}' \right)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{\mathbf{x}}_{it} \tilde{y}_{it}, \tag{2.5}$$

where in model (2.1)

$$\beta = \begin{bmatrix} \beta_2 \\ \beta_3 \end{bmatrix}, \quad \tilde{\mathbf{x}}_{\mathbf{it}} = \mathbf{x_{it}} - \hat{\mathbf{x}}_{\mathbf{it}} = \begin{bmatrix} \widetilde{DU}_t \\ Treat_i \cdot \widetilde{DU}_t \end{bmatrix}, \quad \tilde{y}_{it} = y_{it} - \hat{y}_{it}, \quad \widetilde{DU}_t = DU_t - \widehat{DU}_t,$$

with $\hat{y}_{it} = \sum_{s=1}^{T} y_{is}\mathbf{f}(s)'\Big(\sum_{s=1}^{T}\mathbf{f}(s)\mathbf{f}(s)'\Big)^{-1}\mathbf{f}(t)$ and $\widehat{DU}_t = \sum_{s=1}^{T} DU_s\mathbf{f}(s)'\Big(\sum_{s=1}^{T}\mathbf{f}(s)\mathbf{f}(s)'\Big)^{-1}\mathbf{f}(t)$. Note that $Treat_i$ drops after the transformation as long as $\mathbf{f}(t)$ has an intercept. In model (2.2) we have

$$
\begin{aligned}
\beta &= \beta_3,\\
\tilde{y}_{it} &= y_{it} - \hat{y}_{it} - \frac{1}{N}\sum_{j=1}^{N}(y_{jt} - \hat{y}_{jt}),\\
\tilde{x}_{it} &= x_{it} - \hat{x}_{it} - \frac{1}{N}\sum_{j=1}^{N}(x_{jt} - \hat{x}_{jt}) = \widetilde{Treat}_i \cdot \widetilde{DU}_t,
\end{aligned}
$$

with

$$\widetilde{Treat}_i = Treat_i - \frac{1}{N}\sum_{j=1}^{N} Treat_j = \mathbb{1}(i \le kN) - k.$$

Let

$$\mathbf{h}_{it} = \begin{bmatrix} \widetilde{DU}_t \\ Treat_i \cdot \widetilde{DU}_t \end{bmatrix}.$$

Here, both $Treat_i$ and DU_t drop after the transformation. In model (2.3) we have the same $\tilde{y}_{it}$ and $\widetilde{DU}_t$ as in model (2.1) but different β and $\tilde{\mathbf{x}}_{\mathbf{it}}$ given by

$$\beta = \begin{bmatrix} \beta_2 \\ \beta_3 \\ \gamma \end{bmatrix}, \quad \tilde{\mathbf{x}}_{\mathbf{it}} = \begin{bmatrix} \mathbf{h}_{it} \\ \tilde{\mathbf{z}}_{it} \end{bmatrix},$$

where $\tilde{\mathbf{z}}_{it} = \mathbf{z}_{it} - \hat{\mathbf{z}}_{it} = \mathbf{z}_{it} - \sum_{s=1}^{T}\mathbf{z}_{is}\mathbf{f}(s)'\Big(\sum_{s=1}^{T}\mathbf{f}(s)\mathbf{f}(s)'\Big)^{-1}\mathbf{f}(t)$. In model (2.4), $\tilde{y}_{it}$, $\tilde{\mathbf{z}}_{it}$, $\widetilde{DU}_t$ and $\widetilde{Treat}_i$ take the same form as in model (2.2). However, β and $\tilde{\mathbf{x}}_{\mathbf{it}}$ now become

$$\beta = \begin{bmatrix} \beta_3 \\ \gamma \end{bmatrix}, \quad \tilde{\mathbf{x}}_{\mathbf{it}} = \begin{bmatrix} \widetilde{Treat}_i \cdot \widetilde{DU}_t \\ \tilde{\mathbf{z}}_{it} \end{bmatrix}.$$

Plugging (2.1), (2.2), (2.3) or (2.4) into (2.5) for $\tilde{y}_{it}$ yields

$$\hat{\beta} - \beta = \Big(\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{\mathbf{x}}_{it}\tilde{\mathbf{x}}_{it}'\Big)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{\mathbf{x}}_{it}u_{it}. \tag{2.6}$$

Let $\tilde{\mathbf{v}}_{it} = \tilde{\mathbf{x}}_{it} u_{it}$ and define $\hat{\mathbf{v}}_{it} = \tilde{\mathbf{x}}_{it} \hat{u}_{it}$ where $\hat{u}_{it}$ are the OLS residuals given by

$$\hat{u}_{it} = \tilde{y}_{it} - \tilde{\mathbf{x}}'_{\mathbf{it}} \hat{\beta}.$$

As shown by Driscoll and Kraay (1998), it is possible to obtain standard errors in a panel model that are robust to spatial correlation of unknown form, as well as heteroskedasticity and serial correlation, under the covariance stationarity and weak dependence conditions. Define

$$\hat{\hat{\mathbf{v}}}_t = \sum_{i=1}^{N} \hat{\mathbf{v}}_{it},$$

and the partial sums of $\hat{\hat{\mathbf{v}}}_t$ as

$$\hat{\hat{S}}_{[rT]} = \sum_{t=1}^{[rT]} \hat{\hat{\mathbf{v}}}_t,$$

where $r \in (0,1]$ and $[rT]$ is the integer part of $[rT]$. Let

$$\hat{\hat{\Gamma}}_j = T^{-1} \sum_{t=j+1}^{T} \hat{\hat{\mathbf{v}}}_t \hat{\hat{\mathbf{v}}}'_{t-j},$$

and then define

$$\hat{\hat{\Omega}} = \hat{\hat{\Gamma}}_0 + \sum_{j=1}^{T-1} k(\frac{j}{M})(\hat{\hat{\Gamma}}_j + \hat{\hat{\Gamma}}'_j),$$

which is the nonparametric kernel HAC estimator using the cross-sectional sum, $\hat{\hat{\mathbf{v}}}_t$, the kernel, $k(x)$, and bandwidth M. An equivalent expression of $\hat{\hat{\Omega}}$ is given by

$$\hat{\hat{\Omega}} = T^{-1} \sum_{t=1}^{T} \sum_{s=1}^{T} K_{ts} \hat{\hat{\mathbf{v}}}_t \hat{\hat{\mathbf{v}}}'_s,$$

where

$$K_{ts} = k(\frac{|t-s|}{M}).$$

When $\hat{\hat{\Omega}}$ is used as the middle term of the sandwich form of the covariance matrix, we obtain the robust covariance matrix estimator proposed by Driscoll and Kraay (1998)

$$\hat{V} = T(\sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{\mathbf{x}}_{it} \tilde{\mathbf{x}}'_{it})^{-1} \hat{\hat{\Omega}} (\sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{\mathbf{x}}_{it} \tilde{\mathbf{x}}'_{it})^{-1}.$$

Consider testing linear hypotheses about β of the form

$$H_0 : R\beta = \mathbf{r},$$

where R is a $q \times K^*$ matrix of known constants with full rank with $q \leq K^*$ and $\mathbf{r}$ is a $q \times 1$ vector of known constants. Define the *Wald* statistics as

$$Wald = (R\hat{\beta} - \mathbf{r})'[R\hat{V}R']^{-1}(R\hat{\beta} - \mathbf{r}).$$

In the case where $q = 1$ we can define the t-statistics

$$t = \frac{R\hat{\beta} - \mathbf{r}}{\sqrt{R\hat{V}R'}}.$$

Note that $q \leq 2$ in model (2.1) and $q = 1$ in model (2.2). In these two cases, the focus is on the asymptotic behavior of the t-statistics under null hypotheses involving restrictions on the DD estimator. For model (2.3) and (2.4), the asymptotic behavior of the *Wald*-statistics under null hypotheses involving linear restrictions on the γ vector is also analyzed.

2.3 Asymptotic Theory and Critical Values

This section analyzes the asymptotic properties of the test statistics under null hypotheses in large-T, fixed-N case. All limits are taken as $T \to \infty$ and N held fixed. Simulated critical values are provided. Throughout, the symbol "$\Rightarrow$" denotes weak convergence. Both "$\xrightarrow{p}$" and "$p\lim$" denote convergence in probability.

The asymptotic distributions of *Wald* and t statistics under null hypotheses are obtained using large-T asymptotics. This approach allows the standard errors to be approximated within the fixed-b asymptotic framework developed by Kiefer and Vogelsang (2005) which captures the choice of kernel and bandwidth in the asymptotic approximation. Moreover, it generates limits that are invariant to general forms of spatial correlation under assumptions of covariance stationarity and weak dependence in the time dimension. The asymptotic distributions of the statistics depend on the form of the kernel used to compute the HAC estimators. Here we focus on Bartlett kernel,

$k(x) = 1 - |x|$ for $|x| \leq 1$ and $k(x) = 0$ for $|x| \geq 1$. Before we proceed, some definitions are required. The random matrices that appear in the asymptotic results are expressed in terms of the following functions and random variables.

Definition 2.1. Let $W(r)$ denote a generic vector of independent standard Wiener processes. Define

$$
\begin{aligned}
H^F(r,\lambda) &= \mathbb{1}(r > \lambda) - \int_\lambda^1 \mathbf{F}(s)'ds\Big(\int_0^1 \mathbf{F}(s)\mathbf{F}(s)'ds\Big)^{-1}\mathbf{F}(r), \\
N^F(W) &= \int_0^1 H^F(r,\lambda)dW(r), \\
Q^F(r,\lambda,W) &= \int_0^r H^F(s,\lambda)dW(s) - \int_0^1 dW(s)F(s)'\Big(\int_0^1 F(s)F(s)'ds\Big)^{-1}\int_0^r F(s)H^F(s,\lambda)ds \\
&\quad - \int_0^r H^F(s,\lambda)^2 ds\Big(\int_0^1 H^F(s,\lambda)^2 ds\Big)^{-1} N^F(W).
\end{aligned}
$$

The following definition defines some random matrices that appear in the asymptotic results.

Definition 2.2. Let $B(r)$ denote a generic vector of Brownian bridges. If $k(x)$ is the Bartlett kernel, let the random matrices, $P^F(b,\lambda,Q^F)$, $P(b,B)$, $P_{21}(b,\lambda,Q^F,B)$ and $P_{21}(b,\lambda,Q^F,B)$ be defined as follows for $b \in (0,1]$

$$
\begin{aligned}
P^F(b,\lambda,Q^F) &= \frac{2}{b}\int_0^1 Q^F(r,\lambda,W)Q^F(r,\lambda,W)'dr \\
&\quad - \frac{1}{b}\int_0^{1-b}[Q^F(r,\lambda,W)Q^F(r+b,\lambda,W)' + Q^F(r+b,\lambda,W)Q^F(r,\lambda,W)']dr, \\
P(b,B) &= \frac{2}{b}\int_0^1 B(r)B(r)'dr - \frac{1}{b}\int_0^{1-b}[B(r)B(r+b)' + B(r+b)B(r)']dr, \\
P_{12}(b,\lambda,Q^F,B) &= \frac{2}{b}\int_0^1 Q^F(r,\lambda,W)B(r)'dr - \frac{1}{b}\int_0^{1-b}[Q^F(r,\lambda,W)B(r+b)' \\
&\quad + Q^F(r+b,\lambda,W)B(r)']dr, \\
P_{21}(b,\lambda,\mathbf{Q}^\mathbf{F}) &= \frac{2}{b}\int_0^1 B(r)Q^F(r,\lambda,W)'dr - \frac{1}{b}\int_0^{1-b}[B(r)Q^F(r+b,\lambda,W)' \\
&\quad + B(r+b)Q^F(r,\lambda,W)']dr.
\end{aligned}
$$

For all models, the following assumption on the trend functions is sufficient to obtain the main results of this chapter.

Assumption 2.1. $\mathbf{f}(t)$ *includes a constant, there exists a* $J \times J$ *diagonal matrix* τ_T *and a vector of functions* $\mathbf{F}$*, such that* $\tau_T \mathbf{f}(t) = \mathbf{F}(\frac{t}{T}) + o_P(1)$*,* $\int_0^1 F_i(r)dr < \infty$*,* $i = 1, \ldots, J$*, and* $det[\int_0^1 \mathbf{F}(r)\mathbf{F}(r)'dr] > 0$.

Assumption 2.1 is fairly standard and is the same as the assumption used by Bunzel and Vogelsang (2005). Note that the standard individual fixed-effects model is a special case with $\mathbf{f}(t) = 1$; the individual specific trend model is a special case with $\mathbf{f}(t) = (1,t)'$.

2.3.1 Models With No Additional Regressors

This subsection investigates the asymptotic properties of the statistics in models (2.1) and (2.2). For a given time period t, stack $u_{1t}, u_{2t}, \ldots, u_{Nt}$ into a $N \times 1$ vector

$$\mathbf{u_t} = \begin{bmatrix} u_{1t} \\ u_{2t} \\ \vdots \\ u_{Nt} \end{bmatrix}$$

The following assumption is sufficient to obtain results for the fixed-effects OLS estimator based on model (2.1) and (2.2).

Assumption 2.2. $T^{-\frac{1}{2}} \sum_{t=1}^{[rT]} \mathbf{u_t} \Rightarrow \Lambda W_N(r)$*, where* $W_N(r)$ *is an* $N \times 1$ *vector of independent standard Wiener processes and* $\Lambda\Lambda'$ *is the* $N \times N$ *long run variance matrix of* $\mathbf{u_t}$.

For a given time period t, stacking the N cross-section errors in the same period into a vector accounts for general forms of spatial correlation. Assumption 2.2 holds under covariance stationarity and weak dependence in the time dimension. It essentially requires that $\mathbf{u_t}$ satisfy a functional central limit theorem (FCLT). Here, $\Lambda\Lambda'$ is not restricted to be diagonal. Therefore, the assumption allows for general forms of spatial correlation. Stationarity is not required in the cross section for large-T, fixed-N case. This is analogous to large-N, fixed-T case where the random sampling in the cross section allows for general forms of serial correlation in model, including nonstationarity.

Before we start to derive the results in model (2.1), it is worth noting that the t-statistics on the DD estimator in the following three models are exactly the same.[2]

1. $y_{it} = a_i + \beta_1 Treat_i + \beta_2 DU_t + \beta_3 Treat_i \cdot DU_t + u_{it}$,

2. $y_{it} = \lambda_t + \beta_1 Treat_i + \beta_2 DU_t + \beta_3 Treat_i \cdot DU_t + u_{it}$,

3. $y_{it} = a_i + \lambda_t + \beta_1 Treat_i + \beta_2 DU_t + \beta_3 Treat_i \cdot DU_t + u_{it}$,

where a_i is a full set of individual dummies, and λ_t is a full set of time period dummies. This exact equivalence result directly implies that whether time period dummies are included does not affect the limit of the t-statistic on the DD estimator in the individual fixed-effects model. Proofs of the exact equivalence result are provided in Appendix C. Furthermore, Monte Carlo simulation results suggest this exact equivalence continue to hold when trend is also included in the model. Proofs are not given for this special case.

Let

$$A = \begin{bmatrix} 1,1,\ldots,1,1,\ldots,1 \\ 1,1,\ldots,1,0,\ldots,0 \end{bmatrix}$$

where A is a $2 \times N$ matrix with all elements in the first row and first kN elements in the second row equal to one. Let $G = AA'$. The following proposition and lemma present the asymptotic distributions of $(\hat{\beta} - \beta)$ and the partial sums in model (2.1).

Proposition 2.1. *Suppose Assumption 2.1 and 2.2 hold. Let $W^*(r)$ denote a 2×1 vector of standard Wiener processes and let Λ^* denote the matrix square root of the matrix $A\Lambda\Lambda' A'$. In model (2.1), for N fixed as $T \to \infty$ the following holds:*

$$\sqrt{T}(\hat{\beta} - \beta) \Rightarrow \left(G \int_0^1 H^F(r,\lambda)^2 dr \right)^{-1} \cdot \Lambda^* \int_0^1 H^F(r,\lambda) dW^*(r).$$

Lemma 2.2. *Suppose Assumption 2.1 and 2.2 hold. Assume $M = bT$ where $b \in (0,1]$ is fixed. Let $W^*(r)$ denote a 2×1 vector of standard Wiener processes and let Λ^* denote the matrix square*

[2] The result also holds when a global intercept is included.

root of the matrix $A\Lambda\Lambda' A'$. In model (2.1), for N fixed as $T \to \infty$ the following holds:

$$T^{-\frac{1}{2}} \hat{\hat{S}}_{[rt]} \Rightarrow \Lambda^* Q^F(r, \lambda, W^*).$$

When $k(x)$ is the Bartlett kernel, from calculations in Hashimzade and Vogelsang (2008a) we have

$$\hat{\hat{\Omega}} = \frac{2}{b} T^{-2} \sum_{t=1}^{T-1} \hat{\hat{S}}_t \hat{\hat{S}}_t' - \frac{1}{b} T^{-2} \sum_{t=1}^{T-M-1} (\hat{\hat{S}}_t \hat{\hat{S}}_{t+M}' + \hat{\hat{S}}_{t+M} \hat{\hat{S}}_t') \tag{2.7}$$

using the fact that $\hat{\hat{S}}_T = 0$. The following proposition presents the fixed-b limit of the HAC estimator.

Proposition 2.3. *Suppose Assumption 2.1 and 2.2 hold. Assume $M = bT$ where $b \in (0,1]$ is fixed. Let $W^*(r)$ denote a 2×1 vector of standard Wiener processes and let Λ^* denote the matrix square root of the matrix $A\Lambda\Lambda' A'$. In model (2.1), for N fixed as $T \to \infty$ the following holds:*

$$\hat{\hat{\Omega}} \Rightarrow \Lambda^* P^F(b, \lambda, Q^F) \Lambda^{*\prime}.$$

Based on Proposition 2.1 and 2.3, the following theorem summarizes the theoretical results for model (2.1).

Theorem 2.1. *Suppose the model does not include time period dummies nor additional regressors. Suppose Assumption 2.1 and 2.2 hold. Assume $M = bT$ where $b \in (0,1]$ is fixed. Let W_q^{**} denote the $q \times 1$ vector of standard Wiener processes. For N fixed as $T \to \infty$,*

$$Wald \Rightarrow N^F(W_q^{**})' P^F(b, \lambda, Q_q^{F**})^{-1} N^F(W_q^{**})$$

$$t \Rightarrow \frac{N^F(W_1^{**})}{\sqrt{P^F(b, \lambda, Q_1^{F**})}}$$

Theorem 2.1 demonstrates that asymptotically pivotal test statistics are obtained within the fixed-b framework in the presence of spatial correlation in the cross section. Therefore, the statistics based on the DK standard errors under fixed-b asymptotics have broader robustness properties with respect to correlation in the model. The limiting distributions differ from those derived by Kiefer and Vogelsang (2005) and Vogelsang (2012) in the following two ways. First, the fixed-b limits here depend on not only the choice of kernel and bandwidth, but also the date of policy change, λ, and individual-specific trend functions. Second, the asymptotic distribution is different from Vogelsang (2012) because DU_t is deterministic and thus there are some extra terms in the asymptotic distribution of partial sums. $N^F(W_q^{**})$ follows a normal distribution, and $P^F(b,\lambda,Q_q^{F**})$ is a random matrix which depends on the date of policy change, trend functions and the choice of kernel and bandwidth. Moreover, $N^F(W_q^{**})$ and $P^F(b,\lambda,Q_q^{F**})$ are independent. The limiting distributions of the test statistics are identical to the results in the pure time series model with a shift in mean and deterministic trends. The limiting distributions are non-standard, but critical values can be obtained using simulation methods.

Corollary 2.2. *Suppose model (2.1) is a standard individual fixed-effects model with no time trends. That is,* $\mathbf{f}(t) = 1$. *Define* $\lambda W(1) - W(\lambda) = (\lambda - 1)\tilde{W}(\frac{\lambda}{\lambda-1})$. *Let* W_q^{**} *denote the* $q \times 1$ *vector of standard Wiener processes. Then*

$$H^F(r,\lambda) = \mathbb{1}(r > \lambda) - (1-\lambda), \quad N^F(W) = \lambda W(1) - W(\lambda) = (\lambda - 1)\tilde{W}(\frac{\lambda}{\lambda - 1}),$$
$$Q^F(r,\lambda,W) = \int_0^r H^F(s,\lambda)dW(s) - W(1)\int_0^r H^F(s,\lambda)ds - \int_0^r H^F(s,\lambda)^2 ds \cdot \Big(\int_0^1 H^F(s,\lambda)^2 ds\Big)^{-1} N^F(W).$$

For N fixed as $T \to \infty$, *the following hold*

$$\sqrt{T}(\hat{\beta} - \beta) \Rightarrow \frac{1}{\lambda(1-\lambda)} G^{-1}\Lambda^*(\lambda - 1)\tilde{W}(\frac{\lambda}{\lambda - 1})$$
$$Wald \Rightarrow N^F(W_q^{**})' P^F(b,\lambda,Q_q^{F**})^{-1} N^F(W_q^{**}), \quad t \Rightarrow \frac{N^F(W_1^{**})}{\sqrt{P^F(b,\lambda,Q_1^{F**})}}$$

Corollary 2.2 provides results for a standard individual fixed-effects DD model. The limits are identical to the results in the pure time series model with a shift in mean.

When time period dummies are also included in the model (2.2), the limiting distributions of the statistics remain the same due to the exact equivalence result. This finding is useful since empirical researchers often put a full set of time period dummies in their model.

2.3.2 Models With Additional Regressors

This subsection analyzes the asymptotic properties of the statistics in models (2.3) and (2.4). Some additional notations in this subsection are needed as follows. Let I_h denote a $h \times h$ identity matrix. Let ι denote an $N \times 1$ vector of ones. Let e_i denote a $N \times 1$ vector with i^{th} element equal to one and zeros otherwise, *i.e*.

$$e_i = (0, 0, \dots, 0, 1, 0, \dots, 0)'.$$

Define a $K \times (K+1)$ matrix B and a $K \times N(K+1)$ matrix A_i as follows

$$B = [\mathbf{0}, I_K], \quad A_i = (e_i \otimes B).$$

Let $\tilde{e}_1$ denote an $(K+1) \times 1$ vector with 1^{st} element equal to one and zeros otherwise, *i.e*.

$$\tilde{e}_1 = (1, 0, \dots, 0)'.$$

Let $\bar{e}_1$ denote an $(NK+1) \times 1$ vector with 1^{st} element equal to one and zeros otherwise, *i.e*.

$$\bar{e}_1 = (1, 0, \dots, 0)'.$$

The following assumption on additional regressors $\mathbf{z_{it}}$ is sufficient to obtain results for the fixed-effects OLS estimator based on models (2.3) and (2.4).

Assumption 2.3. *Suppose there is no structural change for* $\mathbf{z_{it}}$ *within the entire sample periods. Assume that* $p\lim T^{-1} \sum_{t=1}^{T} \mathbf{z_{it}} = \mu_i \equiv E(\mathbf{z_i})$ *and* $p\lim T^{-1} \sum_{t=1}^{[rT]} \tilde{\mathbf{z}}_{it} \tilde{\mathbf{z}}'_{it} = rQ_i$ *for* $r \in (0,1]$ *where* $\bar{Q} = \sum_{i=1}^{N} Q_i$ *and* $\bar{Q}$ *is nonsingular.*

Note that Assumption 2.3 requires that the additional regressors don't have structural change before and after the policy change. In other words, $\mathbf{z_{it}}$ is uncorrelated with $Treat_i$ and DU_t. Under this assumption, $\mathbf{z_{it}}$ is included to reduce the variance of the error. However, empirical researchers are more interested in the case where the additional regressors also have a structural change. In this case, the fixed-b limits for test statistics based on the $\mathbf{z_{it}}$ coefficients may not be the usual fixed-b limits.

To handle the case where additional regressors are also included (model 2.3), Assumption 2.2 needs to be strengthened as follows. Stack the additional regressors $\mathbf{z}_{it}$ and trend functions and consider the reduced form of the $T \times K$ stacked vector $\mathbf{z}_i$–that is, the linear projection of $\mathbf{z}_i$ onto the space spanned by the $T \times J$ stacked vector of trend functions $\mathbf{f}(T)$–with an error term as

$$\mathbf{z}_i = \mathbf{f}(T)\mathbf{b}_i + \mathbf{e}_i,$$

where $\mathbf{e}_i$ is a $T \times 1$ vector and $\mathbf{b}_i$ is a $J \times K$ vector. It is easy to show that $\tilde{\mathbf{z}}_{it}$ are the OLS residuals given by

$$\tilde{\mathbf{z}}_{it} = \mathbf{z}_{it} - \hat{\mathbf{b}}_i' \mathbf{f}(t),$$

where $\hat{\mathbf{b}}_i$ is the OLS estimator of $\mathbf{b}_i$. Define the $(K+1) \times 1$ vector

$$\mathbf{v}_t^{ii} = \begin{bmatrix} u_{it} \\ (\mathbf{z}_{it} - \mathbf{b}_i' \mathbf{f}(t)) u_{it} \end{bmatrix}.$$

Stack the vectors $\mathbf{v}_t^{11}, \ldots, \mathbf{v}_t^{NN}$ to form the $N(K+1) \times 1$ vector of time series

$$\mathbf{v}_t = \begin{bmatrix} \mathbf{v}_t^{11} \\ \mathbf{v}_t^{22} \\ \vdots \\ \mathbf{v}_t^{NN} \end{bmatrix}.$$

Assumption 2.4. *$E(u_{it}|\mathbf{z}_{it}) = 0$ and $T^{-\frac{1}{2}} \sum_{t=1}^{[rT]} \mathbf{v}_t \Rightarrow \dot{\Lambda} W(r)$, where $W(r)$ is an $N(K+1) \times 1$ vector of standard Wiener processes and $\dot{\Lambda}\dot{\Lambda}'$ is the $N(K+1) \times N(K+1)$ long run variance matrix of $\mathbf{v}_t$.*

Assumption 2.3 requires that the sample mean and sample variance-covariance matrix of the additional regressors across time have well-defined limits. The form of Q_i depends on the form of dummies included in the model and the choice of the trend functions. Assumption 2.4 allows weak exogeneity in the cross section and over time and requires a FCLT holds for $\mathbf{v}_t$. Because Q_i is not restricted to be identical for all i and because the form of $\dot{\Lambda}\dot{\Lambda}'$ is not restricted to be block diagonal, the assumptions allow for heterogeneity in the conditional heteroskedasticity and serial correlation as well as general forms of spatial correlation.

The following lemma shows that $\mathbf{h}_{it}$ and $\tilde{\mathbf{z}}_{it}$ are asymptotically uncorrelated.

Lemma 2.4. *Under Assumption 2.1 and 2.3, for N fixed and as $T \to \infty$, the following holds*

$$T^{-1} \sum_{i=1}^{N} \sum_{t=1}^{[rT]} \mathbf{h}_{it} \tilde{\mathbf{z}}_{it}' \xrightarrow{p} 0$$

In particular, when $r = 1$, $T^{-1} \sum\limits_{i=1}^{N} \sum\limits_{t=1}^{T} \mathbf{h}_{it} \tilde{\mathbf{z}}_{it}' \xrightarrow{p} 0$.

Let

$$R = \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix}$$

where R_{11} is a $q_1 \times 2$ matrix, R_{12} is a $q_1 \times K$ matrix, R_{21} is a $q_2 \times 2$ matrix and R_{22} is a $q_2 \times K$ matrix. Usually we pay attention to restrictions either on the DD estimator or on the additional explanatory variables, not on both of them at the same time. In other words, we are interested in the cases when $q_2 = 0$ and $R_{12} = 0$, or when $q_1 = 0$ and $R_{21} = 0$. The next theorem presents the results for model (2.3).

Theorem 2.3. *Suppose the model includes additional regressors but no time period dummies. Suppose Assumption 2.1, 2.3 and 2.4 hold. Assume $M = bT$ where $b \in (0,1]$ is fixed. Let $\bar{W}(r)$ denote a $q_1 \times 1$ vector of standard Wiener processes. Let $W_q(r)$ denote a $q_2 \times 1$ vector of standard Wiener processes. Let $W^*(r)$ denote a 2×1 vector of standard Wiener processes and $\dot{\Lambda}^*$ is the matrix square root of the matrix $(A \otimes \tilde{e}_1')\dot{\Lambda}\dot{\Lambda}'(A \otimes \tilde{e}_1')'$. For N fixed as $T \to \infty$, the following hold:*

$$\sqrt{T}(\hat{\beta} - \beta) \Rightarrow \begin{bmatrix} (G \int_0^1 H^F(r,\lambda)^2 dr)^{-1} (\dot{\Lambda}^* \int_0^1 H^F(r,\lambda) dW^*(r) \\ \bar{Q}^{-1} (\sum_{i=1}^{N} A_i) \dot{\Lambda} W(1) \end{bmatrix}.$$

If $q_2 = 0$ *and* $R_{12} = 0$*, that is, we are testing restrictions on the DD estimator, then* $R = [R_{11}, 0]$.

$$\begin{aligned} Wald &\Rightarrow N^F(\bar{W})'(P^F(b,\lambda,\bar{Q}^F))^{-1}N^F(\bar{W}), \\ t &\Rightarrow \frac{N^F(\bar{W})}{\sqrt{P_1^F(b,\lambda,\bar{Q}_1^F)}}. \end{aligned}$$

If $q_1 = 0$ *and* $R_{21} = 0$*, that is, we are testing restrictions on the additional regressors, then* $R = [0, R_{22}]$.

$$\begin{aligned} Wald &\Rightarrow W_q(1)'P_q(b,B)^{-1}W_q(1), \\ t &\Rightarrow \frac{W_q(1)}{\sqrt{P_q(b,B)}}. \end{aligned}$$

Theorem 2.3 provides some interesting insights into doing inference for DD estimator and the $\mathbf{z}_{it}$ coefficient estimator $\hat{\gamma}$ under fixed-b asymptotics. If we only focus on testing restrictions on DD estimator, the limiting distributions of test statistics are the same as the results in Theorem 2.1. If we only want to test restrictions on $\hat{\gamma}$, the limiting distribution of test statistics are identical to the results in Vogelsang (2012). Note that the limiting distributions of test statistics based on $\hat{\gamma}$ are invariant to trend functions. In either case, the test statistics are asymptotically pivotal. Nevertheless, testing restrictions on both of them at the same time is much more complicated. The test statistics are no longer asymptotically pivotal. General forms of the limits of the test statistics are provided in the proof of Theorem 2.3 in Appendix C.

The most general model including both additional regressors and time period dummies (model 2.4) requires a stronger assumption than Assumption 2.4. To cope with this case, Assumption 2.4 needs to be strengthened in the following way. Define the $K \times 1$ vector $\mathbf{v}_t^{ij} = (\mathbf{z}_{it} - \mathbf{b}_i'\mathbf{f}(t))u_{jt}$. For a given j stack u_{jt} and the vectors $\mathbf{v}_t^{1j}, \mathbf{v}_t^{2j}, \dots, \mathbf{v}_t^{Nj}$ into an $(NK+1) \times 1$ vector

$$\mathbf{v}_t^j = \begin{bmatrix} u_{jt} \\ \mathbf{v}_t^{1j} \\ \mathbf{v}_t^{2j} \\ \vdots \\ \mathbf{v}_t^{Nj} \end{bmatrix},$$

and then stack the vectors $\mathbf{v}_t^1, \mathbf{v}_t^2, \ldots, \mathbf{v}_t^N$ into an $N(NK+1) \times 1$ vector

$$\mathbf{v}_t^{ex} = \begin{bmatrix} \mathbf{v}_t^1 \\ \mathbf{v}_t^2 \\ \vdots \\ \mathbf{v}_t^N \end{bmatrix},$$

where the "ex" superscript denotes an extended vector that includes vectors $\mathbf{v}_t^{ij}$ for $i \neq j$.

Assumption 2.5. *$E(u_{it}|z_{jt}) = 0$ for all $i, j = 1, 2, \ldots, N$ and $T^{-\frac{1}{2}} \sum_{t=1}^{[rT]} v_t^{ex} \Rightarrow \Lambda^{ex} W^{ex}(r)$, where $W^{ex}(r)$ is an $N(NK+1) \times 1$ vector of standard Wiener processes and $\Lambda^{ex}\Lambda^{ex\prime}$ is the $N(NK+1) \times N(NK+1)$ long run variance matrix of v_t.*

Assumption 2.5 requires strict exogeneity in the cross section but allows weak exogeneity over time. It also requires that a FCLT hold for the extended vector $\mathbf{v}_t^{ex}$. Here, $\Lambda^{ex}\Lambda^{ex\prime}$ is not restricted to be block diagonal, which permits general spatial correlation. Assumption 2.4 and 2.5 indicate that the form of exogeneity needed depends on whether or not time period dummies are included in the model. Without time period dummies, only weak exogeneity is required in both the time and cross-section dimensions. When time period dummies are included, strict exogeneity is needed in the cross-section dimension while only weak exogeneity is required in the time dimension.

Like results in model (2.2), including time period dummies does not affect the fixed-b limits. The following theorem summarizes the results for model (2.4). Note that Assumption 2.4 is now replaced with the stronger Assumption 2.5.

Theorem 2.4. *Suppose the model includes both additional regressors and time period dummies. Suppose Assumption 2.1, 2.3 and 2.5 hold. Assume $M = bT$ where $b \in (0,1]$ is fixed. Let $\tilde{A} = [1-k, \ldots, 1-k, -k, \ldots, -k]$ and $\tilde{G} = \tilde{A}\tilde{A}' = \sum_{i=1}^N \widetilde{Treat}_i^2$. Let $W_1^{ex*}(r)$ denote a standard Wiener processes with long run variance $\Lambda_1^{ex*2} = (\tilde{A} \otimes \bar{e}_1')\Lambda^{ex}\Lambda^{ex\prime}(\tilde{A} \otimes \bar{e}_1')'$. For N fixed as $T \to \infty$, the* *following hold:*

$$\sqrt{T}(\hat{\beta}-\beta) \Rightarrow \begin{bmatrix} (\tilde{G}\int_0^1 H^F(r,\lambda)^2 dr)^{-1}\Lambda_1^{ex*}\int_0^1 H^F(r,\lambda) dW_1^{ex*}(r) \\ \bar{Q}^{-1}(\sum_{i=1}^N A_i^{ex})\Lambda^{ex}W^{ex}(1) \end{bmatrix}$$

and the limits of the statistics are the same as given by Theorem 2.3.

Theorem 2.4 demonstrates that results for statistics in Theorem 2.3 continue to hold when time period dummies are included. This is consistent to the findings in model (2.2).

2.3.3 Asymptotic Critical Values

The asymptotic critical values for *Wald* and t statistics based on DD estimator can be obtained through Monte Carlo simulations. To keep the analysis straightforward, we consider the case $q = 1$ and focus on the individual fixed-effects model and the individual-specific trend model. The asymptotic critical values are simulated using $50,000$ replications. The Wiener processes are approximated by normalized sums of $i.i.d.$ $N(0,1)$ errors using 1000 steps. The critical values for t statistics in the standard individual fixed-effects model are presented in Table D.1-D.4. The critical values for t statistics in the individual-specific trend model are presented in Table D.5-D.8. Using the Bartlett kernel, critical values are computed for the percentage points 90%, 95%, 97.5%, and 99%. Right tail critical values are given. The left tail critical values follow from symmetry around zero. The policy change point λ goes from 0.1 to 0.9 with step size 0.1. The bandwidths b starts from 0.02 to 1 with step size 0.02.

The critical values are invariant to the values of k. For a given b, the critical values are symmetric around $\lambda = 0.5$ with respect to λ. The minimum value occurs at $\lambda = 0.5$. As λ approaches zero or one, the critical values increase. This pattern is the same as the pure time series model with a known structural break (see Cho, 2012). For a given λ, with $b = 0.02$, critical values are close to $N(0,1)$ regardless of the choice of trend functions. As b grows, tails get fatter. With $b = 1$ tails are quite fat. For different choices of trend functions, tails get fatter in different rates. For example, when $\lambda = 0.5$, in the standard individual fixed-effects model the critical values at $5\%/2.5\%$ tails with $b = 0.02$ and $b = 1$ are 1.712/2.056 and 4.781/5.958, respectively, while in the individual specific model, the critical values at 5%/2.5% tails with $b = 0.02$ and $b = 1$ are 1.745/2.073 and 5.098/6.395, respectively. Therefore, tails get fatter more quickly in the individual-specific trend model. The critical values predict that if $N(0,1)$ critical values are used for t statistics, then for a given value of T, as bandwidth M increases, b increases and thus t will over-reject.

2.4 Finite Sample Properties

This section analyzes finite sample performances of the DK standard errors using a simulation study. Because using traditional clustered standard errors is the most common method to conduct robust inference for DD estimator, the fixed-b approximations for the DK standard errors given by the theorems are compared with the standard normal approximations for traditional clustered and the DK standard errors. "t_{clus}" denotes t-statistics constructed using traditional clustered standard errors and "t_{DK}" denotes t-statistics constructed using the DK standard errors.

Since applied researchers are interested in the double clustering approach proposed by Cameron et al. (2011) and Thompson (2011), finite sample performances of the two-way clustered standard errors are also included. "t_{double}" denotes t-statistics constructed using the original formula of the double clustering approach, while "t^r_{double}" denotes t-statistics constructed using the revised formula. The revised formula is

$$\hat{V}^r_{double} = \hat{V}_{firm} + \hat{V}_{time,0} - \hat{V}_{White,0} + \sum_{l=1}^{L}(\hat{V}_{time,l} + \hat{V}'_{time,l}) - \sum_{l=1}^{L}(\hat{V}_{White,l} + \hat{V}'_{White,l}), \tag{2.8}$$

with

$$\hat{V}_{firm} = \hat{Q}^{-1}(\sum_{i=1}^{N}\hat{\mathbf{s}}_i\hat{\mathbf{s}}_i')\hat{Q}^{-1},$$
$$\hat{V}_{time,l} = \hat{Q}^{-1}(\sum_{t=l+1}^{T}\hat{\mathbf{s}}_t\hat{\mathbf{s}}_{t-l}')\hat{Q}^{-1},$$
$$\hat{V}_{White,l} = \hat{Q}^{-1}(\sum_{i=1}^{N}\sum_{t=l+1}^{T}\hat{\mathbf{v}}_{it}\hat{\mathbf{v}}_{i,t-l}')\hat{Q}^{-1}.$$

$\hat{\mathbf{s}}_i = \sum_{t=1}^{T} \hat{\mathbf{v}}_{it}$ is the sum of all observations for individual i. $\hat{\mathbf{s}}_t = \sum_{i=1}^{N} \hat{\mathbf{v}}_{it}$ is the sum of all observations for time t. The original formula only contains the first three terms in (2.8)

$$\hat{V}_{double} = \hat{V}_{firm} + \hat{V}_{time,0} - \hat{V}_{White,0}. \tag{2.9}$$

The DGP used for the simulations is very similar to the one used in Vogelsang (2012). The model is

$$y_{it} = c_i + g_i t + \beta_1 Treat_i + \beta_2 DU_t + \beta_3 Treat_i \cdot DU_t + z_{it}\gamma + u_{it}, \tag{2.10}$$

where

$$u_{it} = \rho u_{i,t-1} + \varepsilon_{it}, \quad u_{i0} = 0, \quad \varepsilon_{it} \sim N(0,1), \quad cov(\varepsilon_{it}, \varepsilon_{js}) = 0 \; for \; t \neq s;$$
$$z_{it} = \rho z_{i,t-1} + e_{it}, \quad z_{i0} = 0, \quad e_{it} \sim N(0,1), \quad cov(e_{it}, e_{js}) = 0 \; for \; t \neq s.$$

c_i is the individual fixed effects and $g_i t$ is the individual-specific simple linear trend. In all cases, all coefficients are set to zero. Also set $c_i = 0$, $g_i = 0$, $k = 0.5$ and $\lambda = 0.5$. Note that we can set $c_i = 0$ without loss of generality because the fixed effects OLS estimator is exactly invariant to $c_i = 0$. Only one additional regressor z_{it} is included and it is uncorrelated with u_{it}. z_{it} and u_{it} are modeled as $AR(1)$ processes with the same autoregressive parameter ρ. ε_{it} and e_{it} have spatial correlation in the cross section, though uncorrelated over time. In particular, they are constructed in the following way. For a given time period, t, N $i.i.d.$ $N(0,1)$ random variables are placed on a square grid. At each grid point, ε_{it} is constructed as the weighted sum of the normal random variable at that grid point, the normal random variables that are one step away to the left, right, up or down on the grid with a weight θ and the normal random variables that are two steps away in the same direction with a weight θ^2. Hence, ε_{it} is a spatial $MA(2)$ process with parameter θ and the distance measure is maximum coordinate-wise distance on the grid. e_{it} is constructed in a similar way. In all cases, $\theta = 0.5$.

Results are given for sample sizes $T = 10, 50, 250$ and $N = 10, 50, 250$ for $AR(1)$ errors, and $N = 9, 49, 256$ for spatial $MA(2)$ errors. The number of replications is 2,500 in all cases and the significance level is 5%. Results are reported for the Bartlett kernel. Fixed-effects OLS as discussed in section 2.2 is used to estimate the model. Results for testing the null hypothesis $H_0 : \beta_3 = 0$ against the alternative $H_1 : \beta_3 \neq 0$ are labeled t_{DD}. Results for testing the null hypothesis $H_0 : \gamma = 0$ against the alternative $H_1 : \gamma \neq 0$ are labeled t_z.

Tables D.9–D.11 reports empirical null rejection probabilities for t_{clus} and t_{DK} statistics in the individual fixed-effects model with no additional regressor z_{it}. Tables D.12–D.15 reports empirical null rejection probabilities for t_{clus} and t_{DK} statistics in the individual-specific trend model with no additional regressor z_{it}. Tables D.16–D.17 reports empirical null rejection probabilities for t_{DD} and t_z statistics when one additional regressor z_{it} is included. Table D.18 compares the empirical null rejection probabilities for t_{clus}, t_{double}, t^r_{double} and t_{DK} in the individual fixed-effects model with no additional regressor z_{it}. Tables D.9, D.11, D.12 and D.14 consider $AR(1)$ errors, while the other tables focus on the spatial $MA(2)$ errors. In Tables D.11, D.14 and D.15, a full set of time period dummies is included.

A small selection of bandwidths are considered, $b = 0.02, 0.06, 0.1, 0.4, 0.7, 1$. The autocorrelation parameter $\rho = 0, 0.3, 0.6, 0.9$. For t_{DK} two sets of null rejection probabilities are reported. The first set uses the 5% $N(0,1)$ critical value. The second set uses the new fixed-b critical values (adjusted fixed-b critical values) obtained in subsection 2.3.3. For t_{clus}, t_{double} and t^r_{double}, rejection probabilities are reported using the 5% $N(0,1)$ critical value.

There are several points worth noting. First, looking at Tables D.9 and D.11, the rejection probabilities for each combination of N, T, ρ and b are exactly the same in these two tables. This pattern demonstrates the exact equivalence result shown in subsection 2.3.1. Similar patterns can be found in Table D.12 and D.14 with $AR(1)$ errors, and Table D.13 and D.15 with spatial $MA(2)$ errors. These four tables suggest that the exact equivalence continue to hold in the individual-specific trend model with no additional regressors, despite the correlation structure of the error.

Next, similar patterns for t_{DK} can be found in all tables. Patterns for t_{DK} are quite different when $N(0,1)$ critical value is used compared to when the adjusted fixed-b critical values are used. Using $N(0,1)$ critical value, rejection probabilities tend to be much higher than 5% and this over-rejection problem gets worse as b increases or as ρ increases. Only when b is small, T is large, and ρ is close to zero are rejection probabilities close to 5%. In contrast, when the adjusted fixed-b critical values are used, the over-rejection problem is less severe. For a given N,T,ρ combination, rejection probabilities are above 5% with small b and they steadily decline as b increases. For a given value of ρ, as T increases, rejection probabilities approach 5% for all bandwidths. When $T = 250$ and $b = 1$, rejection probabilities are around 8% or 9% when there is strong serial correlation ($\rho = 0.9$). In the presence of spatial correlation, rejection probabilities for t_{clus} are substantially larger than 5%. This is expected since the traditional clustered standard errors are not robust to the spatial correlation in the cross section. For $AR(1)$ errors in table D.9 and D.12, the traditional clustered standard errors behave well, and can outperform the DK standard errors when there is strong serial correlation and the bandwidth is small.

The patterns in the rejection probabilities of t_{DK} are similar to Vogelsang (2012). As explained in Vogelsang (2012), the bias in $\hat{\hat{\Omega}}$ consists of two parts. One part depends on the strength of the serial correlation and this bias rises as the serial correlation becomes stronger, which explains why the over-rejection problem gets worse as ρ increases. This bias causes over-rejection for either the $N(0,1)$ critical value or the adjusted fixed-b critical values. However, this bias declines as b increases. The other part is captured by the adjusted fixed-b approximations, but not the $N(0,1)$ approximations. Therefore, over-rejection becomes less severe when fixed-b critical values are used. It is shown (see Vogelsang, 2008) that as b increases, bias in $\hat{\hat{\Omega}}$ initially decreases but then increases as b increases further. Because of this, when b is close to one, $\hat{\hat{\Omega}}$ has substantial downward bias and t_{DK} tends to over-reject when the $N(0,1)$ critical value is used. Overall, the $N(0,1)$ approximations do not reflect the influence of the bandwidth, and thus using the $N(0,1)$ critical value may lead to severe distortions in rejections. In contrast, the fixed-b approximations capture most of the bias in $\hat{\hat{\Omega}}$. In addition, the part that they cannot capture decreases as b increases. This demonstrates why the rejection probability of t_{DK} is lowest at $b = 1$ when adjusted fixed-b critical values are used.

Tables D.16 and D.17 report empirical null rejection probabilities in the individual fixed-effects model and individual-specific trend model with one additional regressor z_{it}, respectively. For t_{DD}, the adjusted fixed-b critical values are used. For t_z, the usual fixed-b critical values in Kiefer and Vogelsang (2005) and Vogelsang (2012) are used.

Note that the usual fixed-b critical values are used for t_z because there is no structural break in z_{it}. These critical values are invariant to the choices of trend functions. Patterns of the rejection probabilities are consistent to the findings in Vogelsang (2012). The fixed-b approximation for t_{DD} reflects the change of trend functions when a simple linear trend is included in the model.

Table D.18 reports the null rejection probabilities for the individual fixed-effects model with spatial $MA(2)$ errors. Note that the correlation structure here is different from that used in chapter 1. The results illustrate that the DK standard errors using fixed-b approximations lead to much more accurate inference than the two-way clustered standard errors in the presence of a different form of cross-sectional correlation. The findings are similar to those in chapter 1. The original double clustering method is okay when T is large and ρ is small. The revised double clustering method has a better performance than the original one only when ρ is large and the truncation lag is small. The rejection probabilities of the revised method increases as the truncation lag gets bigger. The DK approach using fixed-b critical values outperform the double clustering approach when the bandwidth is chosen appropriately.

2.5 Conclusion

This chapter derives a fixed-b asymptotic theory for test statistics in DD models with fixed effects and individual specific trends in linear panel settings. The standard errors proposed by Driscoll and Kraay (1998) that are robust to heteroskedasticity, autocorrelation and spatial correlation of general form are analyzed. This chapter establishes the conditions under which the DK standard errors lead to valid tests in linear DD models with fixed effects and individual-specific time trends for fixed-N, large-T case. It is shown that the fixed-b asymptotics for tests on the DD estimator are different from the limits in Vogelsang (2012), but they are identical to the limits in the pure time series model with a shift in mean for the individual fixed-effects model. The tests on additional regressors without a structural break have the same fixed-b asymptotic distributions as in Vogelsang (2012).

The exact equivalence result is found for the cases when only individual dummies are included, when only time period dummies are included and when both sets of dummies are included. As a result, whether time period dummies are included in the model does not affect the asymptotic distribution. It is also shown that the fixed-b asymptotics for tests on DD estimator depend on the individual-specific deterministic trends included and the date of policy change λ. New critical values are simulated for individual fixed-effects model and individual specific trend model. For each value of bandwidth, the adjusted critical values shows a U-shaped pattern in λ. Tails get fatter in different rates for different trend functions. Simulation results illustrate that the use of fixed-b critical values will lead to much more reliable inference in practice in the presence of spatial correlation.

In a more interesting case where the additional regressors also have a structural change, the fixed-b limits of test statistics on the z_{it} parameter would change. The conjecture of the fixed-b asymptotic distributions in this case would be similar to the findings in the pure time series model with a structural break (see Cho, 2012).

CHAPTER 3

FINITE SAMPLE PERFORMANCES OF THE MOVING BLOCKS BOOTSTRAP FOR LINEAR DIFFERENCE-IN-DIFFERENCES MODELS WITH INDIVIDUAL FIXED EFFECTS

3.1 Introduction

This chapter studies finite sample performances of the bootstrap procedure for linear Difference-in-Differences (DD) models with individual fixed effects. The bootstrap method consists of randomly resampling the original data many times and then using the quantities computed from the simulated pseudo-data to make inference from the original observed data. This chapter discusses bootstrap methods in the context of hypothesis testing. Bootstrap methods are widely used in empirical studies, especially when distributions of test statistics are nonstandard and critical values are complicated to compute, or difficult to derive theoretically. Moreover, it is not even necessary for us to know the asymptotic distribution when applying the bootstrap method.

What determines the reliability of the bootstrap is how well the bootstrap data generating process (DGP) mimics the features of the true DGP. The bootstrap has originally been proposed by Efron (1979) for independent and identically distributed (*i.i.d.*) data. Later, the wild bootstrap has been proposed by Wu (1986) to take into account heteroskedasticity. It becomes more complicated to implement bootstrap methods for dependent data. Several bootstrap procedures have been proposed for time series data, including the moving blocks bootstrap (MBB) proposed by Kunsch (1989) and Liu and Singh (1992). More recently, the bootstrap is applied to panel data models. Following the approach in Gonçalves (2011), the so-called "panel MBB" method is used in this chapter. This method applies the standard MBB to the time series of vectors containing all the individual observations at each time period. Since this method only resamples the vectors at each time period, it preserves the potential cross-sectional correlation structure in the data. Therefore, the panel MBB allows for inference that is robust to heteroskedasticity, serial correlation and cross-sectional correlation of unknown form. Also, we use the naive bootstrap where the formula used to compute the standard errors on the resampled data is the same as the formula used on the original data.

The DD coefficient is of interest and the estimation method is the fixed-effects ordinary least squares (OLS) estimator. The main focus is on the tests based on the DD estimator and the DK standard errors. In particular, we consider panels with many time periods where the Driscoll and Kraay, 1998 (DK) standard errors are valid. The DD estimator becomes more and more popular in recent empirical researches because it allows us to evaluate the causal effects of a policy change. Researchers are concerned with the reliability of the inference based on the DD estimator. There has been an extensive research to seek robust inference for DD models. As pointed out in Bertrand, Duflo, and Mullainathan, 2004 (BDM), ignoring the presence of serial correlation leads to very unreliable inference. Wooldridge (2003) and other econometricians had already been strongly suggesting the use of clustered standard errors. Motivated by the results in BDM, using clustered standard errors has become a common method in empirical works. Alternatively, Bertrand et al. (2004) also suggested using the blocks bootstrap method where each cluster is a block. Take a state-level data for example, this method first stacks residuals for each state into vectors and then randomly draws with replacement for each state a new residual vector from this distribution, leaving residuals within each state unchanged. The bootstrap method is straightforward and easy to implement. However, both of these two methods lead to biased inference when the number of clusters is small. Based on the work of BDM, Cameron, Gelbach, and Miller, 2008 (CGM) proposed a wild bootstrap-based procedure. Following CGM, applied researchers use the wild cluster bootstrap method to obtain improved inference. Usually it is assumed that data are independent in the cross section dimension, or are independent across clusters, but are correlated in the time dimension. This chapter explores improved inference that is robust to cross-sectional correlation of more general form.

In linear panel models with individual fixed effects, a recent paper by Gonçalves (2011) has provided both theoretical and simulation evidences indicating that the panel MBB, including the

$i.i.d.$ bootstrap, outperforms the standard normal approximation and closely mimics the fixed-b approximation proposed in Vogelsang (2012) when a standard nonparametric heteroskedasticity and autocorrelation consistent (HAC) variance estimator is used to compute test statistics. Gonçalves and Vogelsang (2011) have also found similar results in pure time series models. Following the approach of Kiefer and Vogelsang (2005) and Vogelsang (2012), in chapter 2 we have derived the asymptotic distributions of test statistics based on the DD estimator and the DK standard errors, assuming that the bandwidth is a fixed proportion of the sample size in time dimension. This new fixed-b limiting distribution is different from the one proposed in Vogelsang (2012). Therefore, the first-order asymptotic validity of the panel MBB needs to be examined in linear DD models.

The main goal of this chapter is to analyze finite sample properties of the panel MBB in linear DD models with individual fixed effects using Monte Carlo simulations. Simulation results show that the panel MBB performs very well, even when there is strong serial correlation. The bootstrap is much more accurate than the standard normal approximation, and it closely follows the new fixed-b approximation proposed in chapter 2. This improvement holds for the special case of Bartlett kernel. Results would look similar for other kernels. The improvement even holds when the $i.i.d.$ bootstrap is used, despite potential serial correlation in the data. Simulations results also show that if the block length is appropriately chosen, the panel MBB could outperform the fixed-b approximation when there is strong serial correlation. Theoretical evidences are not provided in this chapter, but can directly follow Gonçalves (2011).

The remainder of this chapter is organized as follows. In the next section we describe the model and test statistics. We also review the fixed-b asymptotic approximation. Section 3.3 describes the bootstrap method. Section 3.4 reports simulation results which compare the standard normal approximation, the fixed-b approximation and the bootstrap. Section 3.5 concludes. Appendix E contains all figures.

3.2 The Difference-in-Differences Model

3.2.1 The Model and DD Estimator

Consider a DD model with individual fixed effects given by

$$y_{it} = c_i + \beta_1 Treat_i + \beta_2 DU_t + \beta_3 Treat_i \cdot DU_t + \mathbf{z}_{it}\gamma + u_{it}, \tag{3.1}$$
$$i = 1, 2, \ldots, N, \quad t = 1, 2, \ldots, T,$$

where y_{it} and u_{it} are scalars, c_i denotes the unobserved individual heterogeneity. $Treat_i$ denotes an indicator for individuals in the treatment group which takes one if individual i is in the treatment group. Without loss of generality, we assume that the first kN individuals are in the treatment group. Thus, $Treat_i = \mathbb{1}(i \leq kN)$. DU_t denotes an indicator for post-policy-change time periods which takes one after the policy change. That is, $DU_t = \mathbb{1}(t > \lambda T) = \mathbb{1}(r > \lambda)$, where the parameter λ is the relative date of the policy change within the time sample. Both k and λ are assumed known. $\mathbf{z}_{it}$ is a $K \times 1$ vector of additional regressors.

The parameter of interest is β_3, which evaluates the impact of a policy change on y. The estimation method is the fixed-effects ordinary least squares (OLS) estimator, or the DD estimator

$$\hat{\beta} = \left(\sum_{i=1}^{N} \sum_{t=1}^{T} (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)' \right)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)(y_{it} - \bar{y}_i), \tag{3.2}$$

where

$$\beta = \begin{bmatrix} \beta_2 \\ \beta_3 \\ \gamma \end{bmatrix}, \quad \mathbf{x}_{it} = \begin{bmatrix} DU_t \\ Treat_i \cdot DU_t \\ \mathbf{z}_{it} \end{bmatrix}, \quad \bar{y}_i = T^{-1} \sum_{t=1}^{T} y_{it}, \quad \bar{\mathbf{x}}_i = T^{-1} \sum_{t=1}^{T} \mathbf{x}_{it}.$$

3.2.2 The DK Standard Errors

Driscoll and Kraay (1998) first proposed the HAC type robust variance estimator using the time series of sums of all the individual observations at each time period. The idea is to first aggregate all the individual observations at each time period and then apply the HAC estimator to the time series of the sums. The first step takes into account potential cross-sectional correlation in the data, and the second step takes into account potential serial correlation in the data. Therefore, the DK standard errors are robust to cross-sectional correlation of unknown form as well as heteroskedasticity and serial correlation, assuming covariance stationarity and weak dependence in the time dimension.

Let $\tilde{\mathbf{v}}_{it} = \tilde{\mathbf{x}}_{it}u_{it}$ and define $\hat{\mathbf{v}}_{it} = \tilde{\mathbf{x}}_{it}\hat{u}_{it}$ where $\tilde{\mathbf{x}}_{it} = \mathbf{x}_{it} - \bar{\mathbf{x}}_i$, $\tilde{y}_{it} = y_{it} - \bar{y}_i$, $\hat{u}_{it}$ are the OLS residuals given by $\hat{u}_{it} = \tilde{y}_{it} - \tilde{\mathbf{x}}'_{it}\hat{\beta}$. Define $\hat{\hat{\mathbf{v}}}_t = \sum_{i=1}^{N} \hat{\mathbf{v}}_{it}$,, and let $\hat{\hat{\Gamma}}_j = T^{-1} \sum\limits_{t=j+1}^{T} \hat{\hat{\mathbf{v}}}_t \hat{\hat{\mathbf{v}}}'_{t-j}$.

Let $\Omega = \lim_{T\to\infty} Var(T^{-\frac{1}{2}} \sum\limits_{t=1}^{T} \sum\limits_{i=1}^{N} \tilde{\mathbf{v}}_{it})$. Following the approach of Driscoll and Kraay (1998), the estimation of Ω is implemented with the nonparametric kernel HAC estimator given by

$$\hat{\hat{\Omega}} = \hat{\hat{\Gamma}}_0 + \sum_{j=1}^{T-1} k(\frac{j}{M})(\hat{\hat{\Gamma}}_j + \hat{\hat{\Gamma}}'_j),$$

where $k(x)$ is a kernel function such that $k(x) = k(-x)$, $k(0) = 1$, $|k(x)| \leq 1$, $k(x)$ is continuous at $x = 0$, and $\int_{-\infty}^{\infty} k^2(x) < \infty$. M is the bandwidth parameter. When $\hat{\hat{\Omega}}$ is used as the middle term of the sandwich form of the covariance matrix, we obtain the robust covariance matrix estimator proposed by Driscoll and Kraay (1998)

$$\hat{V} = T(\sum_{i=1}^{N}\sum_{t=1}^{T} \tilde{\mathbf{x}}_{it}\tilde{\mathbf{x}}'_{it})^{-1}\hat{\hat{\Omega}}(\sum_{i=1}^{N}\sum_{t=1}^{T} \tilde{\mathbf{x}}_{it}\tilde{\mathbf{x}}'_{it})^{-1}.$$

3.2.3 Test Statistics and Asymptotic Distributions

Consider testing linear hypotheses about β of the form

$$H_0 : R\beta = \mathbf{r},$$

where R is a $q \times (K+2)$ matrix of known constants with full rank with $q \leq (K+2)$ and $\mathbf{r}$ is a $q \times 1$ vector of known constants. In the case where $q = 1$ we can define the t-statistic

$$t = \frac{R\hat{\beta} - \mathbf{r}}{\sqrt{R\hat{V}R'}}.$$

The main focus is on the asymptotic behavior of t-statistics based on the DD estimator. For comparison purposes, t-statistics based on $\hat{\gamma}$ are also considered in models with additional regressors.

The traditional asymptotic approach relies on $\hat{\hat{\Omega}}$ being a consistent estimator of Ω. Consistency of $\hat{\hat{\Omega}}$ requires that $M \to \infty$ as $T \to \infty$, but at a slower rate of convergence $\frac{M}{T} \to 0$. Under the traditional approach, the t-statistic has a limiting standard normal distribution.

An alternative asymptotic theory has been proposed by Kiefer and Vogelsang (2005). They model the bandwidth as a fixed proportion of the sample size. That is, $M = bT$ with b a fixed constant in $(0,1]$. Because b is held fixed in this approach, this new alternative approach is usually labeled fixed-b asymptotics while the traditional approach is labeled small-b asymptotics. Under the fixed-b approach, $\hat{\hat{\Omega}}$ converges to a random matrix rather than a constant. In Vogelsang (2012), the random matrix depends on the kernel function and the bandwidth. In chapter 2, the random matrix also depends on the date of the policy change, λ, in DD models. As a result, the t-statistic has a nonstandard limiting distribution. This limiting distribution reflects the date of the policy change and the choice of kernel and bandwidth, but is otherwise pivotal. Fixed-b asymptotics provide more accurate and reliable inference than small-b asymptotics. For a given date of the policy change, kernel function and bandwidth, fixed-b critical values can be simulated.

In linear DD models with individual fixed effects as in chapter 2, we have shown that

$$t \Rightarrow \frac{N^F(W_1^{**})}{\sqrt{P^F(b, \lambda, Q_1^{F**})}},$$

where $\Rightarrow$ denotes weak convergence, W_1^{**} is the standard Wiener process, and $P^F(b, \lambda, Q_1^{F**})$ is the random matrix that depends on the date of the policy change λ, kernel function and bandwidth.

In the special case of Bartlett kernel, $k(x) = 1 - |x|$ for $|x| \leq 1$ and $k(x) = 0$ for $|x| \geq 1$, we have

$$
\begin{aligned}
H^F(r,\lambda) &= \mathbb{1}(r > \lambda) - (1-\lambda), \quad N^F(W) = \lambda W(1) - W(\lambda) = (\lambda - 1)\tilde{W}(\frac{\lambda}{\lambda - 1}), \\
Q_1^{F**} &= Q^F(r,\lambda,W_1^{**}) = \int_0^r H^F(s,\lambda) dW_1^{**}(s) - W_1^{**}(1) \int_0^r H^F(s,\lambda) ds \\
&\quad - \int_0^r H^F(s,\lambda)^2 ds \Big(\int_0^1 H^F(s,\lambda)^2 ds \Big)^{-1} N^F(W_1^{**}), \\
P^F(b,\lambda,Q^F) &= \frac{2}{b} \int_0^1 Q^F(r,\lambda,W) Q^F(r,\lambda,W)' dr \\
&\quad - \frac{1}{b} \int_0^{1-b} [Q^F(r,\lambda,W) Q^F(r+b,\lambda,W)' + Q^F(r+b,\lambda,W) Q^F(r,\lambda,W)'] dr.
\end{aligned}
$$

3.3 Bootstrap Methods

Another alternative to asymptotic approximations is the bootstrap. In order to obtain heteroskedasticity, autocorrelation and cross-sectional correlation robust inference, we follow the panel MBB approach proposed by Gonçalves (2011). Motivated by the idea of Driscoll and Kraay (1998), Gonçalves (2011) proposed the panel MBB which is an extension of the standard MBB to linear panel models. The panel MBB first stacks all the individual observations at each time period into vectors and then applies the standard MBB to the time series of these vectors. Gonçalves (2011) has proved that this method is robust to heteroskedasticity, serial correlation and cross-sectional correlation of unknown form when the fixed-effects OLS estimator is used, under the assumption that N is an arbitrary nondecreasing function of T and $T \to \infty$. Weak dependence in the time dimension is required for the MBB to be valid, but we allow the dependence in the cross section dimension to be either weak or strong.

Define the bootstrap fixed-effects OLS estimator $\hat{\beta}^*$ as

$$
\hat{\beta}^* = \Big(\sum_{i=1}^N \sum_{t=1}^T (\mathbf{x}_{it}^* - \bar{\mathbf{x}}_i^*)(\mathbf{x}_{it}^* - \bar{\mathbf{x}}_i^*)' \Big)^{-1} \sum_{i=1}^N \sum_{t=1}^T (\mathbf{x}_{it}^* - \bar{\mathbf{x}}_i^*)(y_{it}^* - \bar{y}_i^*), \tag{3.3}
$$

where

$$
\bar{y}_i^* = T^{-1} \sum_{t=1}^T y_{it}^*, \quad \bar{\mathbf{x}}_i^* = T^{-1} \sum_{t=1}^T \mathbf{x}_{it}^*.
$$

Note that (3.3) is calculated using the bootstrap data $(y_{it}^*, \mathbf{x}_{it}^*)$. The method to construct the pseudo-data using the panel MBB is described below.

The first step is to run the pooled OLS regression to obtain the fixed-effects OLS estimator $\hat{\beta}$ and the residuals $\hat{u}_{it}$. Define the $(K+1)\times 1$ vector $\omega_{it} = (\mathbf{z}_{it}', \hat{u}_{it})'$ which collects the additional regressors and the OLS residual for each observation in model (3.1). Let $\omega_t = (\omega_{1t}', \omega_{2t}', \ldots, \omega_{Nt}')'$ denote the $N(K+1)\times 1$ vector containing the N cross-sectional observations at a given time period t. Let $l \in \mathbb{N}$ $(1 \le l < T)$ be the block length, and let $B_{t,l} = \{\omega_t, \omega_{t+1}, \ldots, \omega_{t+l-1}\}$ be the block of l consecutive observations starting at ω_t. For simplicity, assume $T = hl$. Note that $l = 1$ is just the standard $i.i.d.$ bootstrap case. The MBB randomly draws $h = \frac{T}{l}$ blocks with replacement from the set of overlapping blocks $\{B_{1,l}, B_{2,l}, \ldots, B_{T-l+1,l}\}$. Thus the pseudo-data ω_t^* take the form

$$
\begin{aligned}
&\omega_1^* = \omega_{I_1+1}, \omega_2^* = \omega_{I_1+2}, \ldots, \omega_l^* = \omega_{I_1+l},\\
&\omega_{l+1}^* = \omega_{I_2+1}, \ldots, \omega_{2l}^* = \omega_{I_2+l},\\
&\quad\vdots\\
&\omega_{(h-1)l+1}^* = \omega_{I_h+1}, \ldots, \omega_{hl}^* = \omega_{I_h+l},
\end{aligned}
$$

where the indices $I_1, I_2, \ldots, I_h$ are $i.i.d.$ random variables distributed uniformly on $\{0, 1, \ldots, T-l\}$. Let $\mathbf{x}_{it}^* = (DU_t, Treat_i \cdot DU_t, \mathbf{z}_{it}^{*\prime})'$. Pseudo-values y_{it}^* are given by

$$
y_{it}^* = \mathbf{x}_{it}^{*\prime}\hat{\beta} + \hat{u}_{it}^*. \tag{3.4}
$$

It is worth noting that the bootstrap data generating process (DGP) is a bit different from that in Gonçalves (2011). Gonçalves (2011) uses the pairs bootstrap where the bootstrap data $(y_{it}^*, \mathbf{x}_{it}^*)$ are directly drawn from the original data $(y_{it}, \mathbf{x}_{it})$ without a first-step regression to obtain the OLS residuals. The pairs bootstrap does not work in DD models because it may mix the pre and post policy change values and thus lead to a biased estimator $\hat{\beta}^*$.

One might want to do the pairs bootstrap within the pre/post policy change subgroup. However, if testing the additional regressors is of interest, this method gives biased estimators for the additional regressors. Therefore, a combination of the residual bootstrap and the pairs bootstrap is used in this chapter. Since DU_t and $Treat_i \cdot DU_t$ are indicators, they are not resampled in the bootstrap procedure. Only the pairs of additional regressors and the residuals are resampled. New pseudo-values of the dependent variable are computed using (3.4).

For example, consider a simple time series model with one random regressor z:

$$y_t = \mu + \beta z_t + u_t.$$

We have

$$y_t = \hat{\mu} + \hat{\beta} z_t + \hat{u}_t, \tag{3.5}$$

where $\hat{\mu}$ and $\hat{\beta}$ are the OLS estimators, and $\hat{u}_t$ is the OLS residual. Equation (3.5) holds for all (y_t, z_t). For each bootstrap sample (y_t^*, z_t^*),

$$y_t^* = \hat{\mu} + \hat{\beta} z_t^* + \hat{u}_t^* \tag{3.6}$$

is always true. Equation (3.5) is the "population model" for the bootstrap sample, and $\hat{\mu}$ and $\hat{\beta}$ are the "population coefficients". As usual in the bootstrap literature, let E^* denote the expected value induced by the bootstrap resampling, conditional on a realization of the original time series. We have

$$E^*(\hat{u}_t^*) = \frac{1}{T} \sum_{t=1}^{T} \hat{u}_t = 0,$$

because $\hat{u}_t^*$ is uniformly distributed on $\{\hat{u}_1, \ldots, \hat{u}_T\}$ conditional on the original sample. The second equation holds because of the normal equation of the OLS estimator. Similarly, we have

$$E^*(z_t^* \hat{u}_t^*) = \frac{1}{T} \sum_{t=1}^{T} z_t \hat{u}_t = 0.$$

These two conditions guarantee that the OLS estimators $\hat{\mu}^*$ and $\hat{\beta}^*$ can consistently estimate $\hat{\mu}$ and $\hat{\beta}$, respectively. This explains why the bootstrap would work intuitively. If we resample (y_t, z_t) within the pre/post policy change subgroup, the expected value $E^*(z_t^* \hat{u}_t^*)$ becomes

$$E^*(z_t^* \hat{u}_t^*) = \frac{1}{\lambda T} \sum_{t=1}^{\lambda T} z_t \hat{u}_t + \frac{1}{(1-\lambda)T} \sum_{t=1}^{(1-\lambda)T} z_t \hat{u}_t \neq 0.$$

This method causes z_t^* to be correlated with $\hat{u}_t^*$ and thus leads to a biased OLS estimator.

Next, consider model (3.1). Without loss of generality, we can set $c_i = 0$ and $\beta_1 = 0$. We have

$$y_{it} = \hat{\beta}_2 DU_t + \hat{\beta}_3 Treat_i \cdot DU_t + \mathbf{z}'_{it}\hat{\gamma} + \hat{u}_{it}.$$

If we directly draw $(y_{it}, \mathbf{z}'_{it})$ from the original data, it is possible that the pre/post policy change values are mixed in the bootstrap sample. For example, suppose a original post-policy-change pair $(y_{is}, \mathbf{z}'_{is})$ appears as a pre-policy-change pair in the bootstrap data. Then in the original data we have $y_{is} = \hat{\beta}_2 + \hat{\beta}_3 Treat_i + \mathbf{z}'_{is}\hat{\gamma} + \hat{u}_{is}$, while in the bootstrap data $y_{is} = \mathbf{z}'_{is}\hat{\gamma} + \hat{u}^*_{is}$. $\hat{u}^*_{is}$ is no longer the original OLS residual $\hat{u}_{is}$ associated with $(y_{is}, \mathbf{z}'_{is})$. This will cause $\mathbf{z}^*_{it}$ to be correlated with $\hat{u}^*_{it}$ and thus leads to a biased OLS estimator. Therefore, we have to resample $(\mathbf{z}'_{it}, \hat{u}_{it})$ and re-construct y_{it} using (3.4). In (3.4), we have

$$E^*(\mathbf{z}^*_{it}\hat{u}^*_{it}) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{z}_{it}\hat{u}_{it} = 0.$$

The OLS estimator of (3.4) can consistently estimate $\hat{\beta}$.

Given a bootstrap sample $(y^*_{it}, \mathbf{x}^*_{it})$, let

$$\begin{aligned}
&\tilde{\mathbf{x}}^*_{it} = \mathbf{x}^*_{it} - \bar{\mathbf{x}}^*_i, \quad \hat{\mathbf{v}}^*_{it} = \tilde{\mathbf{x}}^*_{it}\hat{u}^*_{it}, \quad \hat{\hat{\mathbf{v}}}^*_t = \sum_{i=1}^{N}\hat{\mathbf{v}}^*_{it}, \\
&\hat{\hat{\Gamma}}^*_j = T^{-1}\sum_{t=j+1}^{T}\hat{\hat{\mathbf{v}}}^*_t\hat{\hat{\mathbf{v}}}^{*\prime}_{t-j}, \\
&\hat{\hat{\Omega}}^* = \hat{\hat{\Gamma}}^*_0 + \sum_{j=1}^{T-1}k(\frac{j}{M})(\hat{\hat{\Gamma}}^*_j + \hat{\hat{\Gamma}}^{*\prime}_j), \\
&\hat{V}^* = T(\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{\mathbf{x}}^*_{it}\tilde{\mathbf{x}}^{*\prime}_{it})^{-1}\hat{\hat{\Omega}}^*(\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{\mathbf{x}}^*_{it}\tilde{\mathbf{x}}^{*\prime}_{it})^{-1}.
\end{aligned}$$

The naive bootstrap t-statistic t^* can be defined as

$$t^* = \frac{R\hat{\beta}^* - \mathbf{r}^*}{\sqrt{R\hat{V}^*R'}},$$

where $\mathbf{r}^* = R\hat{\beta}$.

To obtain the bootstrap critical value t^*_c for a test with a significance level α, we generate B bootstrap samples indexed by j and compute t^*_j. We sort t^*_j from the smallest to the largest and then calculate $t^*_c = t^*_{[\alpha(B+1)]}$, where $[\alpha(B+1)]$ is the integer part of $\alpha(B+1)$.

3.4 Finite Sample Performances

This section compares finite sample performances of the standard normal asymptotic approximation, the fixed-b asymptotic approximation and the naive panel MBB using Monte Carlo simulations. We first present results for the simplest DD model without additional regressors, and then add one additional regressor into the model and report the results. The interesting patterns found in Gonçalves (2011) and Gonçalves and Vogelsang (2011) hold in the simplest DD model. They continue to hold after one additional regressor is added to the model.

The DGP used for simulations is very similar to the one used in Vogelsang (2012). The model is

$$y_{it} = c_i + \beta_1 Treat_i + \beta_2 DU_t + \beta_3 Treat_i \cdot DU_t + z_{it}\gamma + u_{it}, \tag{3.7}$$

where

$$u_{it} = \rho u_{i,t-1} + \varepsilon_{it}, \quad u_{i0} = 0, \quad \varepsilon_{it} \sim N(0,1), \quad cov(\varepsilon_{it}, \varepsilon_{js}) = 0 \text{ for } t \neq s;$$
$$z_{it} = \rho z_{i,t-1} + e_{it}, \quad z_{i0} = 0, \quad e_{it} \sim N(0,1), \quad cov(e_{it}, e_{js}) = 0 \text{ for } t \neq s.$$

c_i is the unobserved individual fixed effects. Only one additional regressor z_{it} is included and it is uncorrelated with u_{it}. z_{it} and u_{it} are modeled as $AR(1)$ processes with the same autoregressive parameter. ε_{it} and e_{it} have spatial correlation in the cross section dimension, though uncorrelated over time. In particular, they are constructed in the following way. For a given time period t, N $i.i.d.$ standard normal random variables are placed on a square grid. At each grid point, ε_{it} is constructed as the weighted sum of the normal random variable at that grid point, the normal random variables that are one step away to the left, right, up or down on the grid with a weight θ and the normal random variables that are two steps away in the same direction with a weight θ^2. Hence, ε_{it} is a spatial $MA(2)$ process with parameter θ and the distance measure is maximum coordinate-wise distance on the grid. e_{it} is constructed in a similar way.

We consider testing the null hypothesis that $H_0 : \beta_3 = 0$ against the alternative $H_1 : \beta_3 \neq 0$

with a significance level of 5% using the t-statistic

$$t_{DD} = \frac{\hat{\beta_3}}{se(\hat{\beta_3})},$$

where $se(\hat{\beta_3})$ is the DK standard error estimate. In the cases where the additional regressor z_{it} is included, we also consider testing the null hypothesis that $H_0 : \gamma = 0$ against the alternative $H_1 : \gamma \neq 0$ with a significance level of 5% using the t-statistic

$$t_z = \frac{\hat{\gamma}}{se(\hat{\gamma})},$$

where $se(\hat{\gamma})$ is the DK standard error estimate.

In all cases, $\beta_1, \beta_2, \beta_3$ and γ are set to zero. Also set $c_i = 0$, $\theta = 0.5$, $k = 0.5$ and $\lambda = 0.5$ unless otherwise specified. Note that we can set $c_i = 0$ without loss of generality because the fixed-effects OLS estimator is exactly invariant to $c_i = 0$. Results are reported for sample sizes $T = 50, 250$ and $N = 50, 250$ when there is no cross-sectional correlation, $T = 50, 250$ and $N = 49, 256$ when there is spatial correlation. In the simulations, $1,000$ random samples are generated for each pair of (N, T). We consider three values for the AR parameter, ρ: 0.0, 0.3 and 0.9, and four values for the bandwidth: $b = 0.02$, 0.1, 0.5 and 0.7. We only consider the Bartlett kernel. We reject the null hypothesis whenever $t_{DD} > t_{c1}$ or $t_z > t_{c2}$, where t_{c1} and t_{c2} are critical values. In particular, $t_{c1} = t_{c2} = 1.96$ is used for the standard normal asymptotic approximation. For the fixed-b asymptotic approximation, t_{c1} is the 97.5% percentile of the fixed-b asymptotic distribution derived in chapter 2, while t_{c2} is the 97.5% percentile of the fixed-b asymptotic distribution derived by Kiefer and Vogelsang (2005). For the naive panel MBB, both t_{c1} and t_{c2} are the 97.5% bootstrap percentile of the corresponding bootstrap t-statistics. For each sample, the bootstrap tests are based on 499 replications. In most cases, we consider the block length $l = 1$, *i.e.* the *i.i.d.* bootstrap. Results for the block length $l = 25$ when $T = 250$ are reported in the case of spatial correlation.

All results are shown in figures. (See Appendix E.) Figures E.1 and E.2 illustrate the empirical null rejection probabilities as a function of λ, given that there is no cross-sectional correlation and $N = 100$, $T = 250$, $\rho = 0.3$ and $b = 0.02$ and 0.5, respectively. We consider five values for λ: 0.1, 0.3, 0.5, 0.7 and 0.9. The standard *i.i.d.* bootstrap is used. In both figures, the standard normal asymptotic approximation leads to over-rejection. The empirical null rejection probabilities using the standard normal asymptotic approximation show a U-shape with the bottom at $\lambda = 0.5$. The over-rejection problem gets worse when λ approaches either 0 or 1. In contrast, the naive panel MBB is more accurate than the standard normal approximation. The improvement is remarkable. The larger the bandwidth b, the bigger the improvement. In fact, the bootstrap closely follows the fixed-b asymptotic approximation, and thus reflects the date of the policy change λ. The bootstrap rejection probabilities do not vary much for different values of λ.

Figures E.3–E.20 each contains two columns. Each column contains three graphs corresponding to the three values of ρ. Every sub-figure illustrates the empirical null rejection probabilities as a function of the bandwidth b given $\lambda = 0.5$. Figures E.3–E.12 present results for the simplest DD model without the additional regressor. Figures E.3, E.5, E.7 and E.9 present results for models without cross-sectional correlation, while Figures E.4, E.6, E.8, and Figure E.10-E.12 present results for models with spatial $MA(2)$ correlation.

Figures E.3 and E.5 focus on cases when $N = 50$ and $N = 250$, respectively. Figures E.4 and E.6 focus on cases when $N = 49$ and $N = 256$, respectively. In each figure, the first column presents results for $T = 50$ while the second column presents results for $T = 250$. Several interesting patterns can be found here. For the standard normal approximation, rejection probabilities tend to be much larger than 5%. The over-rejection problem gets worse when b increases. In contrast, the $i.i.d.$ bootstrap is always much more accurate than the standard normal approximation. The larger the bandwidth b, the bigger the improvement. The improvement becomes larger as the sample size T increases. This improvement holds for $N = 50$ and $N = 250$. The improvement holds regardless of potential cross-sectional correlation in the data. The $i.i.d.$ bootstrap tends to closely mimic the fixed-b approximation for all DGPs, all (N, T) combinations, and all bandwidths, despite potential serial correlation in the data. Looking at Figures E.4 and E.6, where spatial $MA(2)$ correlation exists, when $\rho = 0$, $i.e.$ there is no serial correlation but cross-sectional correlation only, the bootstrap rejection probabilities are very close to 5%. Even when there is strong serial correlation, $i.e.$ $\rho = 0.9$, if the bandwidth is large enough, the bootstrap rejection probabilities could still be around 10% or less.

Figures E.7-E.10 illustrate how different values of N would affect the improvement of the $i.i.d.$ bootstrap over the standard normal approximation. Figures E.7 and E.8 focus on cases when $T = 50$, and Figures E.9 and E.10 focus on cases when $T = 250$. In Figures E.7 and E.9, the first column presents results for $N = 50$ while the second column presents results for $N = 250$. In Figures E.8 and E.10, the first column presents results for $N = 49$ while the second column presents results for $N = 256$. Across all DGPs, all (N,T) combinations and all values of ρ, no significant improvement of the $i.i.d.$ bootstrap over the standard normal approximation is observed as N increases.

Figures E.11 and E.12 compare the performance of the bootstrap with different block lengths. In each figure, the first column presents results for the block length $l = 25$ while the second column presents results for $l = 1$, the $i.i.d.$ bootstrap. Figure E.11 focuses on the case when $N = 49$ and $T = 250$. Figure E.12 focuses on the case when $N = 256$ and $T = 250$. It is worth noting that when there is strong serial correlation ($e.g.$, $\rho = 0.9$), increasing the block length to 25 helps further improve the inference, and the bootstrap is likely to outperform the fixed-b approximation across all the bandwidths. But when there is no serial correlation in the data ($\rho = 0$), yet we set the block length to be 25, the bootstrap can over-reject a little bit. When $N = 49$ and $l = 25$, the improvement over the fixed-b approximation is very small. However, when N increases from 49 to 256, significant improvement can be found in Figure E.12. The results suggest that if the block length is appropriately chosen, the panel MBB can outperform the fixed-b approximation when there is strong serial correlation.

Figures E.13–E.20 present results for the DD model with one additional regressor z. Since we are interested in the performance of the bootstrap when the cross-sectional correlation exists, all DGPs include the spatial $MA(2)$ correlation in the cross section. Figures E.13–E.16 illustrate the empirical null rejection probabilities for tests based on β_3 and γ. The first column shows results for β_3, and the second column shows results for γ. (N,T) combinations $(49,50)$, $(49,250)$, $(256,50)$, and $(256,250)$ are considered in Figures E.13–E.16, respectively. In other words, (large-T, small-

N), (small-T, large-N) and (large-T, large-N) cases are included. Figures E.17–E.20 compare the performance of the bootstrap with different block lengths. Figures E.17 and E.19 focuses on $(N,T) = (49,250)$. Figures E.18 and E.20 focuses on $(N,T) = (256,250)$. The patterns for the DD estimator found in the simplest DD model continue to hold after the additional regressor z is added. Similar patterns also hold for inference on the z coefficient, which is consistent with findings in Gonçalves (2011).

3.5 Conclusion

In this chapter we use Monte Carlo simulations to investigate finite sample performances of the naive panel MBB applied to heteroskedasticity, autocorrelation and cross-sectional correlation robust tests based on the DD estimator and the DK standard errors. Simulation results show that the naive panel MBB outperforms the standard normal approximation in the special case of Bartlett kernel. This improvement even holds for the $i.i.d.$ bootstrap, despite potential serial correlation in the data. The results suggest that the finite sample performance of the naive panel bootstrap closely follow the performance of the fixed-b approximation to the first order. In addition, the results also suggest that the bootstrap can be more accurate than the fixed-b approximation when appropriate block length is chosen. Results would look similar for other kernels.

Gonçalves and Vogelsang (2011) have shown that the naive MBB, including the $i.i.d.$ bootstrap, has the same limiting distribution as the fixed-b asymptotic distribution. For the special case of a location model, Gonçalves and Vogelsang (2011) have proved that the $i.i.d.$ bootstrap can produce more accurate inference than the standard normal approximation depending on the choice of the bandwidth and the number of finite moments in the data. Given the patterns in the simulations, we can conjecture that the asymptotic equivalence of the panel MBB and the fixed-b distribution holds in our settings. The improvement of the $i.i.d.$ bootstrap over the standard normal approximation could also be extended to panel models and inference on the DD parameter. Theoretical explanations can be included in future research.

APPENDICES

Appendix A

PROOFS IN CHAPTER 1

Proofs of Theorem 1.1 is provided.

Proofs of Theorem 1.1. First, we need to show that sample variance of x_{it} has a well-defined limit.

$$\begin{aligned}
\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{[rT]}x_{it}^2 &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{[rT]}(\mu_i+\theta_t+\xi_{it})^2 \\
&= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{[rT]}(\mu_i^2+\theta_t^2+\xi_{it}^2+2\mu_i\theta_t+2\mu_i\xi_{it}+2\theta_t\xi_{it}) \\
&= \frac{[rT]}{T}\cdot\frac{1}{N}\sum_{i=1}^{N}\mu_i^2+\frac{1}{T}\sum_{t=1}^{[rT]}\theta_t^2+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{[rT]}\xi_{it}^2+2\big(\frac{1}{N}\sum_{i=1}^{N}\mu_i\big)\big(\frac{1}{T}\sum_{t=1}^{[rT]}\theta_t\big) \\
&\quad+\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{[rT]}\mu_i\xi_{it}+\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{[rT]}\theta_t\xi_{it} \\
&\xrightarrow{p} r\big(E(\mu_i^2)+E(\theta_t^2)+E(\xi_{it}^2)+2E(\mu_i)E(\theta_t)+2E(\mu_i\xi_{it})+2E(\theta_t\xi_{it})\big) \\
&= rQ
\end{aligned}$$

where $Q=E(\mu_i^2)+E(\theta_t^2)+E(\xi_{it}^2)$.

Next, we prove (1.15) and (1.16). We have to show that $\theta_t\delta_t$ is a zero mean covariance stationary process and thus it can be represented in the form of a $MA(\infty)$ process according to Wold's theorem. Therefore, $\theta_t\delta_t$ satisfies a FCLT, and $T^{-\frac{1}{2}}\sum_{t=1}^{[rT]}\theta_t\delta_t \Rightarrow \sigma W(r)$, where $W(r)$ is a standard Wiener process and σ^2 is the long run variance of $\theta_t\delta_t$. It is straightforward to get

$$\begin{aligned}
&E(\theta_t\delta_t)=E(\theta_t)E(\delta_t)=0, \\
&\gamma_j=cov(\theta_t\delta_t,\theta_{t-j}\delta_{t-j})=E(\theta_t\delta_t\theta_{t-j}\delta_{t-j})=E(\theta_t\theta_{t-j})E(\delta_t\delta_{t-j})=\frac{\rho^{2j}}{(1-\rho^2)^2}.
\end{aligned}$$

Some algebra yields

$$
\begin{aligned}
& N^{-\frac{1}{2}} T^{-1} \sum_{i=1}^{N} \sum_{t=1}^{[rT]} v_{it} \\
&= N^{-\frac{1}{2}} T^{-1} \sum_{i=1}^{N} \sum_{t=1}^{[rT]} \left(\mu_i + \theta_t + \xi_{it}\right)\left(\gamma_i + \delta_t + \eta_{it}\right) \\
&= N^{-\frac{1}{2}} T^{-1} \sum_{i=1}^{N} \sum_{t=1}^{[rT]} \left(\mu_i\gamma_i + \mu_i\delta_t + \mu_i\eta_{it} + \gamma_i\theta_t + \theta_t\delta_t + \theta_t\eta_{it} + \gamma_i\xi_{it} + \xi_{it}\delta_t + \eta_{it}\xi_{it}\right) \\
&= N^{-\frac{1}{2}} T^{-1} \big([rT] \sum_{i=1}^{N} \mu_i\gamma_i + \sum_{i=1}^{N} \mu_i \sum_{t=1}^{[rT]} \delta_t + \sum_{i=1}^{N} \sum_{t=1}^{[rT]} \mu_i\eta_{it} + \sum_{i=1}^{N} \gamma_i \sum_{t=1}^{[rT]} \theta_t + N \sum_{t=1}^{[rT]} \theta_t\delta_t \\
&\quad + \sum_{i=1}^{N} \sum_{t=1}^{[rT]} \theta_t\eta_{it} + \sum_{i=1}^{N} \sum_{t=1}^{[rT]} \gamma_i\xi_{it} + \sum_{i=1}^{N} \sum_{t=1}^{[rT]} \xi_{it}\delta_t + \sum_{i=1}^{N} \sum_{t=1}^{[rT]} \eta_{it}\xi_{it} \big) \\
&= \frac{[rT]}{T} N^{-\frac{1}{2}} \sum_{i=1}^{N} \mu_i\gamma_i + T^{-\frac{1}{2}} \big(N^{-\frac{1}{2}} \sum_{i=1}^{N} \mu_i\big)\big(T^{-\frac{1}{2}} \sum_{t=1}^{[rT]} \delta_t\big) + T^{-\frac{1}{2}} \big((NT)^{-\frac{1}{2}} \sum_{i=1}^{N} \sum_{t=1}^{[rT]} \mu_i\eta_{it}\big) \\
&\quad + T^{-\frac{1}{2}} \big(N^{-\frac{1}{2}} \sum_{i=1}^{N} \gamma_i\big)\big(T^{-\frac{1}{2}} \sum_{t=1}^{[rT]} \theta_t\big) + \phi^{\frac{1}{2}} \big(T^{-\frac{1}{2}} \sum_{t=1}^{[rT]} \theta_t\delta_t\big) + T^{-\frac{1}{2}} \big((NT)^{-\frac{1}{2}} \sum_{i=1}^{N} \sum_{t=1}^{[rT]} \theta_t\eta_{it}\big) \\
&\quad + T^{-\frac{1}{2}} \big((NT)^{-\frac{1}{2}} \sum_{i=1}^{N} \sum_{t=1}^{[rT]} \gamma_i\xi_{it}\big) + T^{-\frac{1}{2}} \big((NT)^{-\frac{1}{2}} \sum_{i=1}^{N} \sum_{t=1}^{[rT]} \delta_t\xi_{it}\big) \\
&\quad + T^{-\frac{1}{2}} \big((NT)^{-\frac{1}{2}} \sum_{i=1}^{N} \sum_{t=1}^{[rT]} \eta_{it}\xi_{it}\big) \\
&= \frac{[rT]}{T} N^{-\frac{1}{2}} \sum_{i=1}^{N} \mu_i\gamma_i + \phi^{\frac{1}{2}} \big(T^{-\frac{1}{2}} \sum_{t=1}^{[rT]} \theta_t\delta_t\big) + o_p(1) \\
&\Rightarrow rZ_1^* + \phi^{\frac{1}{2}} \sigma W^*(r),
\end{aligned}
$$

where $Z_1^* \sim N(0,1)$, and $W^*(r)$ is a standard Wiener process. Z_1^* is independent with $W^*(r)$ because $\mu_i,\ \gamma_i$ are independent with $\theta_t,\ \delta_t$. Therefore,

$$
\begin{aligned}
\sqrt{N}(\hat{\beta} - \beta) &= \Big(\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} x_{it}^2\Big)^{-1} \cdot \frac{1}{\sqrt{NT^2}} \sum_{i=1}^{N} \sum_{t=1}^{T} v_{it} \\
&\Rightarrow Q^{-1}\big(Z_1^* + \phi^{\frac{1}{2}} \sigma W^*(1)\big) = Q^{-1} \sqrt{1 + \phi\sigma^2} Z_1,
\end{aligned}
$$

where $Z_1 \sim N(0,1)$. Define the partial sums of $\hat{\hat{v}}_t$ as

$$\hat{\hat{S}}_{[rT]} = \sum_{t=1}^{[rT]} \hat{\hat{v}}_t,$$

where $r \in (0,1]$ and $[rT]$ is the integer part of rT. The limiting distribution of $\hat{\hat{S}}_{[rT]}$ is

$$\begin{aligned}
\frac{1}{\sqrt{NT^2}}\hat{\hat{S}}_{[rT]} &= \frac{1}{\sqrt{NT^2}}\sum_{t=1}^{[rT]} \hat{\hat{v}}_t = \frac{1}{\sqrt{NT^2}}\sum_{i=1}^{N}\sum_{t=1}^{[rT]} x_{it}\hat{\varepsilon}_{it} \\
&= \frac{1}{\sqrt{NT^2}}\sum_{i=1}^{N}\sum_{t=1}^{[rT]} x_{it}\left(y_{it} - x_{it}\hat{\beta}\right) \\
&= \frac{1}{\sqrt{NT^2}}\sum_{i=1}^{N}\sum_{t=1}^{[rT]} x_{it}\left(\varepsilon_{it} - x_{it}\left(\hat{\beta} - \beta\right)\right) \\
&= \frac{1}{\sqrt{NT^2}}\sum_{i=1}^{N}\sum_{t=1}^{[rT]} v_{it} - \left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{[rT]} x_{it}^2\right)\cdot\sqrt{N}\left(\hat{\beta} - \beta\right) \\
&\Rightarrow rZ_1^* + \phi^{\frac{1}{2}}\sigma W^*(r) - (rQ)\cdot Q^{-1}\left(Z_1^* + \phi^{\frac{1}{2}}\sigma W^*(1)\right) \\
&= rZ_1^* + \phi^{\frac{1}{2}}\sigma W^*(r) - rZ_1^* - r\phi^{\frac{1}{2}}\sigma W^*(1) \\
&= \phi^{\frac{1}{2}}\sigma(W^*(r) - rW^*(1)) \equiv \phi^{\frac{1}{2}}\sigma B(r)
\end{aligned}$$

where $B(r)$ is a Brownian bridge.

Following the approach of Kiefer and Vogelsang (2005), rewrite the $\hat{\hat{\Omega}}$ in terms of the partial sums of $\hat{\hat{v}}_t$. Consider the Bartlett kernel

$$K(x) = \begin{cases} 1 - |x| & |x| \leq 1 \\ 0 & |x| > 1, \end{cases}$$

Algebra from Hashimzade and Vogelsang (2008b) gives

$$\begin{aligned}
T\hat{\hat{\Omega}} &= \sum_{t=1}^{T}\sum_{s=1}^{T} K_{ts}\hat{\hat{v}}_t\hat{\hat{v}}_s' \\
&= \frac{2}{M}\sum_{t=1}^{T-1}\hat{\hat{S}}_t\hat{\hat{S}}_t' - \frac{1}{M}\sum_{t=1}^{T-M-1}\left(\hat{\hat{S}}_t\hat{\hat{S}}_{t+M}' + \hat{\hat{S}}_{t+M}\hat{\hat{S}}_t'\right) - \frac{1}{M}\sum_{t=T-M}^{T-1}\left(\hat{\hat{S}}_t\hat{\hat{S}}_T' + \hat{\hat{S}}_T\hat{\hat{S}}_t'\right) + \hat{\hat{S}}_T\hat{\hat{S}}_T' \\
&= \frac{2}{M}\sum_{t=1}^{T-1}\hat{\hat{S}}_t\hat{\hat{S}}_t' - \frac{1}{M}\sum_{t=1}^{T-M-1}\left(\hat{\hat{S}}_t\hat{\hat{S}}_{t+M}' + \hat{\hat{S}}_{t+M}\hat{\hat{S}}_t'\right)
\end{aligned}$$

using the fact that $\hat{\hat{S}}_T = 0$ by the OLS normal equations. Note that in this setting, $\hat{\hat{S}}_t$ is a scalar and $M = bT$. Continuing the algebra,

$$T\hat{\hat{\Omega}} = \frac{2}{bT}\sum_{t=1}^{T-1}\hat{\hat{S}}_t^2 - \frac{2}{bT}\sum_{t=1}^{T-bT-1}\hat{\hat{S}}_t\hat{\hat{S}}_{t+M}$$

Then

$$\begin{aligned}
\frac{1}{NT^2}\cdot T\hat{\hat{\Omega}} &= \frac{2}{bT}\sum_{t=1}^{T-1}\frac{1}{\sqrt{NT^2}}\hat{\hat{S}}_t\frac{1}{\sqrt{NT^2}}\hat{\hat{S}}_t - \frac{2}{bT}\sum_{t=1}^{T-bT-1}\frac{1}{\sqrt{NT^2}}\hat{\hat{S}}_t\frac{1}{\sqrt{NT^2}}\hat{\hat{S}}_{t+M} \\
&\Rightarrow \frac{2}{b}\left(\int_0^1 \phi\sigma^2 B(r)^2 dr - \int_0^{1-b}\phi\sigma^2 B(r)B(r+b)dr\right) \\
&= \phi\sigma^2 P(b)
\end{aligned}$$

where $P(b) = \frac{2}{b}\left(\int_0^1 B(r)^2 dr - \int_0^{1-b} B(r)B(r+b)dr\right)$. It directly follows that

$$\begin{aligned}
N\cdot\hat{V}_{DK} &= NT\left(\sum_{i=1}^N\sum_{t=1}^T x_{it}^2\right)^{-1}\hat{\hat{\Omega}}\left(\sum_{i=1}^N\sum_{t=1}^T x_{it}^2\right)^{-1} \\
&= \left(\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T x_{it}^2\right)^{-1}\frac{1}{NT^2}\cdot T\hat{\hat{\Omega}}\left(\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T x_{it}^2\right)^{-1} \\
&\Rightarrow Q^{-1}\cdot\phi\sigma^2 P(b)\cdot Q^{-1} = Q^{-2}\phi\sigma^2 P(b)
\end{aligned}$$

Therefore,

$$t_{DK} = \frac{\hat{\beta}-\beta_0}{\sqrt{\hat{V}_{DK}}} = \frac{\sqrt{N}(\hat{\beta}-\beta)}{\sqrt{N\cdot\hat{V}_{DK}}} \Rightarrow \frac{Q^{-1}\sqrt{1+\phi\sigma^2}Z_1}{\sqrt{Q^{-2}\phi\sigma^2 P(b)}} = \sqrt{1+\frac{1}{\phi\sigma^2}}\cdot\frac{Z_1}{\sqrt{P(b)}}.$$

Next, we prove (1.17) and (1.18) following the same steps as above.

$$\begin{aligned}
\frac{1}{\sqrt{NT^2}}\sum_{i=1}^N\sum_{t=1}^{[rT]} v_{it} &= \frac{1}{\sqrt{NT^2}}\sum_{i=1}^N\sum_{t=1}^{[rT]}(\mu_i+\xi_{it})(\gamma_i+\eta_{it}) \\
&= \frac{1}{\sqrt{NT^2}}\sum_{i=1}^N\sum_{t=1}^{[rT]}(\mu_i\gamma_i+\mu_i\eta_{it}+\gamma_i\xi_{it}+\eta_{it}\xi_{it}) \\
&= \frac{[rT]}{T}N^{-\frac{1}{2}}\sum_{i=1}^N\mu_i\gamma_i + op(1) \\
&\Rightarrow rZ_2,
\end{aligned}$$

where $Z_2 \sim N(0,1)$. Therefore,

$$\sqrt{N}\left(\widehat{\beta}-\beta\right)=\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}x_{it}^{2}\right)^{-1}\cdot\frac{1}{\sqrt{NT^{2}}}\sum_{i=1}^{N}\sum_{t=1}^{T}v_{it}\Rightarrow Q^{-1}Z_{2}$$

The limiting distribution of $\hat{\hat{S}}_{[rT]}$ is

$$\begin{aligned}\frac{1}{\sqrt{NT^{2}}}\hat{\hat{S}}_{[rT]}&=\frac{1}{\sqrt{NT^{2}}}\sum_{i=1}^{N}\sum_{t=1}^{[rT]}v_{it}-\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{[rT]}x_{it}^{2}\right)\cdot\sqrt{N}\left(\hat{\beta}-\beta\right)\\&\Rightarrow rZ_{2}-(rQ)\cdot Q^{-1}Z_{2}=0\end{aligned}$$

Therefore,

$$\begin{aligned}\frac{1}{NT^{2}}\cdot T\hat{\hat{\Omega}}&=\frac{2}{bT}\sum_{t=1}^{T-1}\frac{1}{\sqrt{NT^{2}}}\hat{\hat{S}}_{t}\frac{1}{\sqrt{NT^{2}}}\hat{\hat{S}}_{t}-\frac{2}{bT}\sum_{t=1}^{T-bT-1}\frac{1}{\sqrt{NT^{2}}}\hat{\hat{S}}_{t}\frac{1}{\sqrt{NT^{2}}}\hat{\hat{S}}_{t+M}\\&\Rightarrow\frac{2}{b}\left(\int_{0}^{1}0\cdot 0dr-\int_{0}^{1-b}0\cdot 0dr\right)=0\end{aligned}$$

It directly follows that

$$\begin{aligned}N\cdot\hat{V}_{DK}&=\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}x_{it}^{2}\right)^{-1}\frac{1}{NT^{2}}\cdot T\hat{\hat{\Omega}}\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}x_{it}^{2}\right)^{-1}\\&\Rightarrow Q^{-1}\cdot 0\cdot Q^{-1}=0\end{aligned}$$

Therefore,

$$\left|t_{DK}\right|=\frac{\left|\hat{\beta}-\beta_{0}\right|}{\sqrt{\hat{V}_{DK}}}=\frac{\left|\sqrt{N}\left(\hat{\beta}-\beta\right)\right|}{\sqrt{N\cdot\hat{V}_{DK}}}\to\infty.$$

$\square$

Appendix B

TABLES IN CHAPTER 1

Table B.1: Estimating coefficient, standard errors and null rejection probabilities with firm effects: OLS and one-way clustered standard errors.

		Source of regressor volatility			
Avg(β_{OLS}) Std(β_{OLS}) Avg(SE_{White}) % Sig(t_{White}) Avg(SE_C^f) % Sig(t_C^f)		0%	25%	50%	75%
Source of error volatility	0%	1.0003	1.0004	1.0004	1.0004
		0.0285	0.0283	0.0283	0.0283
		0.0283	0.0283	0.0283	0.0283
		[0.0108]	[0.0098]	[0.0086]	[0.0078]
		0.0282	0.0282	0.0282	0.0282
		[0.0108]	[0.0098]	[0.0086]	[0.0090]
	25%	1.0001	1.0005	1.0007	1.0008
		0.0284	0.0353	0.0411	0.0463
		0.0283	0.0283	0.0283	0.0283
		[0.0094]	[0.0402]	[0.0756]	[0.1180]
		0.0282	0.0352	0.0411	0.0462
		[0.0090]	[0.0108]	[0.0092]	[0.0104]

Continued on next page.

Table B.1 (cont'd)

		Source of regressor volatility			
Avg(β_{OLS}) Std(β_{OLS}) Avg(SE_{White}) % Sig(t_{White}) Avg(SE_C^f) % Sig(t_C^f)		0%	25%	50%	75%
	50%	1	1.0006	1.0008	1.0009
		0.0283	0.0412	0.051	0.0592
		0.0283	0.0283	0.0283	0.0283
		[0.0110]	[0.0762]	[0.1598]	[0.2262]
		0.0282	0.0411	0.0508	0.0589
		[0.0112]	[0.0100]	[0.0102]	[0.0098]
	75%	0.9999	1.0006	1.0008	1.0010
		0.0283	0.0464	0.0593	0.0699
		0.0283	0.0282	0.0282	0.0282
		[0.0120]	[0.1156]	[0.2218]	[0.3068]
		0.0282	0.0462	0.0589	0.0694
		[0.0112]	[0.0090]	[0.0088]	[0.0102]

Table B.2: Estimating coefficient, standard errors and null rejection probabilities with firm effects: FM standard errors.

		Source of regressor volatility			
Avg(β_{FM}) Std(β_{FM}) Avg(SE_{FM}) % Sig(t_{FM})		0%	25%	50%	75%
Source of error volatility	0%	1.0003	1.0004	1.0004	1.0004
		0.0286	0.0284	0.0283	0.0283
		0.0276	0.0275	0.0275	0.0275
		[0.0322]	[0.0304]	[0.0282]	[0.0284]
	25%	1.0001	1.0006	1.0007	1.0008
		0.0285	0.0355	0.0412	0.0463
		0.0276	0.0267	0.0258	0.0248
		[0.0304]	[0.0766]	[0.1302]	[0.1902]
	50%	1	1.0006	1.0008	1.001
		0.0285	0.0414	0.0511	0.0593
		0.0276	0.0258	0.0239	0.0218
		[0.0316]	[0.1336]	[0.2498]	[0.3662]
	75%	0.9999	1.0006	1.0008	1.001
		0.0284	0.0466	0.0594	0.07
		0.0276	0.0249	0.0218	0.0183
		[0.0290]	[0.1928]	[0.3660]	[0.5134]

Table B.3: Estimating coefficient, standard errors and null rejection probabilities with time effects: OLS and clustered standard errors.

		Source of regressor volatility			
Avg(β_{OLS}) Std(β_{OLS}) Avg(SE_{White}) % Sig(t_{White}) Avg(SE_C^t) % Sig(t_C^t)		0%	25%	50%	75%
Source of error volatility	0%	1.0005	1.0005	1.0005	1.0005
		0.0285	0.0289	0.0298	0.0312
		0.0283	0.0287	0.0294	0.0305
		0.01	0.01	0.0098	0.0102
		0.026	0.026	0.0259	0.0257
		0.0404	0.0406	0.0476	0.0642
	25%	1.0003	0.999	0.9978	0.9961
		0.028	0.1518	0.2181	0.2831
		0.0279	0.0281	0.0286	0.0295
		0.0116	0.6208	0.7292	0.7904
		0.0254	0.124	0.1739	0.2202
		0.0396	0.0524	0.0734	0.0908
	50%	1.0002	0.9984	0.9966	0.9942
		0.0276	0.213	0.3073	0.3994
		0.0275	0.0274	0.0277	0.0283
		0.0096	0.7344	0.8128	0.8540

Continued on next page.

Table B.3 (cont'd)

		Source of regressor volatility			
Avg(β_{OLS})					
Std(β_{OLS})					
Avg(SE_{White})					
% Sig(t_{White})					
Avg(SE_C^t)					
% Sig(t_C^t)		0%	25%	50%	75%
		0.0245	0.1732	0.2445	0.3103
		0.0412	0.0526	0.074	0.0910
	75%	1	0.9978	0.9957	0.9927
		0.0272	0.2602	0.376	0.4889
		0.0269	0.0266	0.0267	0.0269
		0.0092	0.7856	0.853	0.8806
		0.0235	0.2113	0.2989	0.3796
		0.0364	0.052	0.0738	0.0916

Table B.4: Estimating coefficient, standard errors and null rejection probabilities with time effects: FM standard errors.

		Source of regressor volatility			
Avg(β_{FM}) Std(β_{FM}) Avg(SE_{FM}) % Sig(t_{FM})		0%	25%	50%	75%
Source of error volatility	0%	1.0006	0.9999	0.9995	0.9986
		0.0285	0.0323	0.0405	0.0561
		0.0275	0.0316	0.0389	0.0551
		[0.0308]	[0.0300]	[0.0348]	[0.0306]
	25%	1.0003	0.9994	0.9999	0.999
		0.0247	0.0285	0.0348	0.0492
		0.0237	0.0275	0.0337	0.0476
		[0.0344]	[0.0300]	[0.0272]	[0.0318]
	50%	1	0.9996	0.9999	0.9999
		0.0199	0.0232	0.0282	0.0391
		0.0195	0.0225	0.0276	0.0394
		[0.0258]	[0.0296]	[0.0268]	[0.0236]
	75%	0.9997	1.0001	1.0005	0.9998
		0.0143	0.0166	0.0202	0.0281
		0.0138	0.0159	0.0195	0.0277
		[0.0322]	[0.0292]	[0.0308]	[0.0280]

Table B.5: Comparing performances of White, one-way cluster-robust and two-way cluster-robust standard errors in the presence of both firm effects and time effects when N, T varies seperately. For time effects with $\rho = 0$.

N	T	β_{OLS}	SE_{White}	SE_C^f	SE_C^t	SE_{double}
10	10	0.9999	0.2645	0.23	0.241	0.181
10	25	0.9996	0.3735	0.209	0.271	0.137
10	50	0.9977	0.463	0.1875	0.346	0.1395
10	100	1.0004	0.566	0.166	0.4345	0.13
10	250	1.0014	0.694	0.1395	0.5915	0.1175
25	10	0.997	0.383	0.262	0.211	0.145
25	25	0.999	0.423	0.1945	0.192	0.0845
25	50	1.0013	0.52	0.1405	0.241	0.0775
25	100	1.0014	0.603	0.1295	0.35	0.0815
25	250	1.0005	0.7225	0.104	0.5205	0.08
50	10	0.9964	0.4565	0.3325	0.18	0.1295
50	25	1.0019	0.5295	0.2495	0.154	0.084
50	50	1.0001	0.554	0.1845	0.194	0.0755
50	100	1.0004	0.635	0.1385	0.2645	0.067
50	250	0.9998	0.7255	0.1065	0.4075	0.0715
100	10	1.0031	0.563	0.4395	0.166	0.133
100	25	1.002	0.604	0.336	0.131	0.0745
100	50	1.0012	0.6425	0.258	0.1485	0.078
100	100	1.0006	0.67	0.1865	0.1825	0.0665
100	250	0.9999	0.7485	0.108	0.291	0.0485
250	10	0.9962	0.7065	0.611	0.146	0.1315

Continued on next page.

Table B.5 (cont'd)

N	T	β_{OLS}	SE_{White}	SE_C^f	SE_C^t	SE_{double}
250	25	1.0016	0.7165	0.497	0.104	0.0825
250	50	1.0004	0.7315	0.3945	0.1015	0.0755
250	100	1.0011	0.7575	0.2935	0.1145	0.056
250	250	1.0003	0.7925	0.1735	0.176	0.061

Table B.6: Comparing performances of White, one-way cluster-robust and two-way cluster-robust standard errors in the presence of firm effects and $AR(1)$ time effects when $N = T = 10$.

ρ	β_{OLS}	SE_{White}	SE_C^f	SE_C^t	SE_{double}
-0.95	0.9984	0.6215	0.644	0.5035	0.499
-0.9	1.0053	0.5855	0.599	0.451	0.442
-0.7	1.0059	0.3945	0.393	0.2895	0.265
-0.5	1.0065	0.283	0.27	0.2145	0.181
-0.3	0.9928	0.2275	0.2155	0.1815	0.1365
-0.1	0.996	0.203	0.1745	0.1715	0.1205
0	0.9995	0.2135	0.1805	0.1855	0.138
0.1	1.0066	0.219	0.1785	0.1875	0.1365
0.3	1.0029	0.2195	0.186	0.1995	0.142
0.5	0.9973	0.2395	0.2035	0.2075	0.163
0.7	1.0025	0.28	0.257	0.238	0.1985
0.9	1.0035	0.3465	0.3125	0.273	0.2365
0.95	0.992	0.424	0.403	0.348	0.316

Table B.7: Comparing performances of White, one-way cluster-robust and two-way cluster-robust standard errors in the presence of firm effects and $AR(1)$ time effects when $N = T = 50$.

ρ	β_{OLS}	SE_{White}	SE_C^f	SE_C^t	SE_{double}
-0.95	0.992	0.927	0.9225	0.6525	0.6485
-0.9	0.9953	0.896	0.846	0.531	0.518
-0.7	1.0007	0.7655	0.5415	0.3105	0.2465
-0.5	1.0037	0.645	0.3295	0.2135	0.113
-0.3	1.0029	0.563	0.203	0.1725	0.0695
-0.1	0.9974	0.566	0.198	0.183	0.074
0	1.0015	0.565	0.1655	0.166	0.055
0.1	0.9979	0.5765	0.184	0.191	0.066
0.3	1.0019	0.5715	0.2025	0.176	0.074
0.5	0.9995	0.6255	0.2785	0.197	0.1125
0.7	0.9989	0.72	0.4825	0.2915	0.2215
0.9	1.0005	0.8505	0.766	0.4835	0.456
0.95	0.9966	0.887	0.8345	0.5525	0.536

Table B.8: Comparing performances of White, one-way cluster-robust and two-way cluster-robust standard errors in the presence of firm effects and $AR(1)$ time effects when $N = T = 250$.

ρ	β_{OLS}	SE_{White}	SE_C^f	SE_C^t	SE_{double}
-0.95	0.9979	0.9665	0.954	0.6635	0.662
-0.9	0.9971	0.943	0.8865	0.5275	0.52
-0.7	0.9987	0.888	0.5755	0.276	0.219
-0.5	1.0003	0.853	0.3245	0.198	0.107
-0.3	0.9996	0.8235	0.21	0.1745	0.0675
-0.1	0.999	0.7865	0.1815	0.1755	0.053
0	1.0002	0.788	0.1705	0.1635	0.049
0.1	1.0008	0.81	0.179	0.1695	0.0505
0.3	0.9991	0.8225	0.2195	0.1765	0.056
0.5	1.0005	0.811	0.3065	0.184	0.096
0.7	0.9998	0.892	0.557	0.2805	0.2205
0.9	1.0004	0.9495	0.881	0.536	0.5265
0.95	1.0063	0.976	0.9525	0.666	0.6635

Table B.9: Comparing performances of one-way cluster-robust, two-way cluster-robust and DK standard errors in the presence of firm effects and $AR(1)$ time effects when $N = T = 50$ and $N = T = 250$. No firm dummies.

		SE_C^f	SE_C^t	SE_{double}	SE_{double}^r								SE_{DK} Using Usual Fixed-b Critical Values							
					values of b								values of b							
N,T	ρ				.1	.2	.3	.4	.5	.6	.7	.8	.1	.2	.3	.4	.5	.6	.7	.8
50	.0	.174	.186	.071	.123	.188	.253	.325	.399	.485	.536	.622	.160	.148	.137	.135	.132	.130	.133	.131
	.3	.224	.179	.084	.126	.206	.283	.348	.419	.472	.545	.647	.138	.131	.126	.125	.121	.124	.120	.122
	.6	.374	.245	.151	.150	.229	.297	.355	.422	.470	.551	.653	.125	.108	.100	.092	.091	.094	.095	.096
	.9	.772	.558	.525	.290	.363	.443	.491	.544	.613	.681	.781	.310	.221	.190	.183	.180	.180	.179	.181
250	.0	.171	.172	.059	.103	.166	.236	.314	.397	.465	.530	.606	.155	.147	.143	.138	.138	.137	.136	.136
	.3	.198	.164	.060	.097	.184	.247	.311	.381	.441	.516	.589	.121	.116	.105	.104	.106	.110	.106	.106
	.6	.402	.229	.159	.140	.216	.277	.319	.373	.423	.476	.546	.087	.084	.080	.082	.075	.075	.073	.076
	.9	.848	.520	.509	.171	.246	.316	.351	.400	.443	.509	.599	.105	.086	.076	.073	.073	.072	.072	.070

Table B.10: Comparing performances of one-way cluster-robust, two-way cluster-robust and DK standard errors in the presence of firm effects and $AR(1)$ time effects when $N = T = 50$ and $N = T = 250$. Firm dummies.

		SE_C^f	SE_C^t	SE_{double}	SE_{double}^r								SE_{DK} Using Usual Fixed-b Critical Values							
					values of b								values of b							
N,T	ρ				.1	.2	.3	.4	.5	.6	.7	.8	.1	.2	.3	.4	.5	.6	.7	.8
50	.0	.631	.075	.082	.182	.259	.302	.316	.361	.402	.445	.537	.068	.066	.062	.062	.060	.062	.060	.064
	.3	.674	.091	.102	.184	.250	.310	.328	.376	.420	.469	.552	.067	.063	.055	.056	.059	.059	.058	.057
	.6	.786	.191	.196	.203	.287	.334	.378	.417	.458	.520	.603	.103	.090	.087	.082	.082	.085	.083	.083
	.9	.933	.516	.525	.328	.439	.496	.526	.565	.587	.648	.727	.283	.233	.219	.202	.201	.198	.192	.190
250	.0	.830	.049	.050	.152	.217	.255	.292	.337	.368	.412	.500	.048	.048	.046	.047	.047	.048	.048	.048
	.3	.840	.072	.073	.150	.219	.253	.305	.334	.368	.408	.478	.048	.051	.050	.050	.048	.050	.049	.050
	.6	.906	.190	.192	.163	.242	.293	.329	.362	.410	.454	.530	.069	.064	.061	.064	.062	.060	.061	.058
	.9	.980	.534	.535	.202	.289	.341	.394	.420	.479	.542	.625	.120	.102	.092	.094	.090	.093	.092	.089

Table B.11: Comparing performances of one-way cluster-robust, two-way cluster-robust and DK standard errors in the presence of firm effects and $AR(1)$ time effects. No firm dummies.

					SE_{DK} Using Adjusted Fixed-b Critical Values									
					values of b									
N,T	ρ	SE_C^f	SE_C^t	SE_{double}	.1	.2	.3	.4	.5	.6	.7	.8	.9	1.0
50,50	.0	.174	.186	.071	.051	.049	.052	.052	.051	.049	.051	.055	.052	.053
	.3	.224	.179	.084	.073	.064	.063	.059	.062	.062	.065	.068	.064	.066
	.6	.374	.245	.151	.100	.085	.079	.072	.068	.074	.071	.073	.074	.075
	.9	.772	.558	.525	.310	.220	.188	.183	.180	.180	.179	.180	.181	.186
50,100	.0	.128	.264	.066	.052	.053	.053	.051	.050	.054	.051	.055	.053	.055
	.3	.150	.251	.067	.049	.049	.050	.047	.047	.047	.049	.048	.047	.048
	.6	.273	.258	.121	.070	.064	.063	.062	.056	.056	.058	.056	.055	.058
	.9	.756	.547	.505	.187	.139	.121	.118	.121	.121	.118	.121	.123	.125
50,250	.0	.093	.403	.065	.043	.048	.046	.044	.044	.047	.049	.047	.044	.046
	.3	.090	.380	.055	.041	.043	.040	.043	.041	.042	.042	.043	.043	.043
	.6	.188	.350	.121	.068	.059	.061	.060	.061	.064	.064	.064	.066	.067
	.9	.713	.535	.472	.102	.089	.080	.081	.077	.080	.081	.084	.082	.083
100,50	.0	.254	.137	.077	.073	.064	.066	.065	.067	.065	.064	.061	.063	.066
	.3	.288	.141	.083	.072	.061	.060	.059	.058	.060	.059	.059	.060	.060
	.6	.498	.224	.176	.094	.082	.080	.074	.076	.080	.077	.080	.080	.082
	.9	.831	.565	.545	.299	.218	.186	.173	.165	.170	.176	.179	.178	.181
100,100	.0	.179	.180	.063	.063	.066	.059	.060	.059	.059	.060	.058	.060	.061
	.3	.197	.168	.073	.056	.056	.054	.054	.051	.051	.050	.055	.054	.056
	.6	.401	.246	.156	.081	.072	.073	.074	.070	.071	.070	.072	.070	.072
	.9	.828	.555	.532	.187	.137	.124	.118	.117	.117	.115	.116	.115	.117

Continued on next page.

Table B.11 (cont'd)

					SE_{DK} Using Adjusted Fixed-b Critical Values									
					values of b									
N,T	ρ	SE_C^f	SE_C^t	SE_{double}	.1	.2	.3	.4	.5	.6	.7	.8	.9	1.0
100,250	.0	.103	.281	.051	.042	.042	.047	.044	.047	.049	.046	.047	.045	.047
	.3	.137	.296	.061	.053	.052	.048	.047	.044	.047	.044	.047	.045	.048
	.6	.266	.273	.117	.050	.052	.049	.051	.051	.053	.053	.053	.052	.055
	.9	.787	.538	.505	.097	.077	.072	.070	.071	.070	.067	.066	.068	.069
250,50	.0	.395	.092	.068	.061	.053	.054	.051	.051	.052	.052	.051	.049	.051
	.3	.446	.102	.071	.058	.055	.057	.058	.055	.056	.055	.059	.058	.058
	.6	.645	.207	.188	.098	.085	.076	.070	.070	.070	.070	.071	.072	.073
	.9	.891	.546	.539	.291	.204	.177	.168	.164	.163	.165	.166	.170	.171
250,100	.0	.299	.106	.055	.053	.056	.055	.059	.058	.060	.060	.059	.059	.063
	.3	.344	.129	.078	.066	.066	.064	.061	.059	.059	.060	.057	.058	.059
	.6	.569	.205	.170	.072	.071	.071	.070	.067	.071	.071	.072	.074	.075
	.9	.878	.545	.535	.185	.145	.128	.124	.113	.118	.121	.123	.124	.125
250,250	.0	.171	.172	.059	.060	.053	.051	.050	.051	.053	.051	.054	.053	.055
	.3	.198	.164	.060	.049	.046	.043	.045	.045	.049	.048	.046	.047	.048
	.6	.401	.229	.159	.067	.066	.064	.062	.059	.059	.058	.056	.057	.059
	.9	.848	.520	. 509	.103	.086	.075	.072	.073	.072	.071	.068	.072	.073

Table B.12: Comparing performances of one-way cluster-robust, two-way cluster-robust and DK standard errors in the presence of a firm effect. No firm dummies.

					SE_{DK} Using Usual Fixed-b Critical Values									
					values of b									
N	T	SE_C^f	SE_C^t	SE_{double}	.1	.2	.3	.4	.5	.6	.7	.8	.9	1.0
10	10	.118	.365	.158	.330	.301	.291	.275	.266	.270	.274	.271	.270	.275
	25	.112	.525	.135	.489	.467	.447	.429	.417	.419	.415	.416	.416	.418
	50	.122	.623	.134	.598	.572	.558	.553	.546	.539	.540	.537	.541	.542
	100	.117	.733	.140	.716	.698	.673	.667	.658	.654	.652	.651	.652	.653
	250	.114	.826	.133	.814	.801	.787	.780	.772	.772	.771	.772	.772	.774
25	10	.075	.376	.103	.344	.319	.296	.284	.278	.279	.280	.279	.281	.284
	25	.078	.513	.089	.491	.460	.452	.446	.440	.435	.436	.434	.433	.436
	50	.073	.623	.082	.607	.589	.571	.555	.546	.544	.541	.544	.542	.544
	100	.076	.717	.086	.705	.679	.659	.648	.635	.633	.632	.628	.626	.630
	250	.084	.845	.090	.831	.822	.815	.811	.803	.801	.799	.799	.797	.799
50	10	.059	.370	.077	.336	.313	.296	.276	.268	.263	.268	.265	.264	.270
	25	.068	.550	.076	.521	.495	.473	.458	.446	.437	.438	.439	.437	.442
	50	.057	.626	.061	.599	.573	.559	.550	.537	.535	.534	.534	.532	.536
	100	.069	.739	.073	.726	.708	.696	.685	.678	.679	.675	.673	.674	.678
	250	.059	.825	.061	.816	.800	.796	.791	.784	.778	.775	.777	.775	.778
100	10	.058	.362	.076	.331	.313	.307	.292	.284	.283	.282	.275	.275	.278
	25	.063	.526	.069	.492	.466	.448	.429	.420	.414	.410	.412	.413	.417
	50	.070	.628	.073	.612	.596	.575	.561	.548	.545	.543	.541	.540	.542
	100	.057	.750	.060	.737	.718	.698	.692	.682	.676	.674	.675	.673	.678
	250	.059	.824	.060	.813	.806	.798	.791	.784	.780	.774	.775	.776	.778

Continued on next page.

Table B.12 (cont'd)

					SE_{DK} Using Usual Fixed-b Critical Values									
					values of b									
N	T	SE_C^f	SE_C^t	SE_{double}	.1	.2	.3	.4	.5	.6	.7	.8	.9	1.0
250	10	.056	.346	.070	.311	.294	.271	.268	.257	.264	.253	.252	.251	.255
	25	.045	.517	.051	.489	.466	.446	.439	.431	.426	.424	.428	.429	.431
	50	.046	.642	.048	.617	.595	.583	.571	.565	.559	.555	.554	.558	.561
	100	.053	.749	.054	.723	.709	.695	.693	.681	.676	.672	.672	.672	.673
	250	.053	.847	.054	.822	.806	.795	.785	.782	.776	.775	.774	.777	.779

Appendix C

PROOFS IN CHAPTER 2

Proofs of the exact equivalence result, Proposition 2.1 and 2.3, Lemma 2.2 and 2.4, Theorem 2.1–2.3 are provided in this Appendix.

Proof of the exact equivalence result. It is straightforward to obtain

$$\sum_{t=1}^{T} \widetilde{DU}_t^2 = \lambda(1-\lambda)T,$$

$$\widetilde{DU}_t \widetilde{DU}_s = DU_t DU_s - (1-\lambda)DU_t - (1-\lambda)DU_s + (1-\lambda)^2,$$

$$\sum_{i=1}^{N} \widetilde{Treat}_i^2 = k(1-k)N.$$

Define

$$\eta = \lambda \sum_{i=1}^{kN} \sum_{t=1}^{T} u_{it} - \sum_{i=1}^{kN} \sum_{t=1}^{\lambda T} u_{it} - k\lambda \sum_{i=1}^{N} \sum_{t=1}^{T} u_{it} + k \sum_{i=1}^{N} \sum_{t=1}^{\lambda T} u_{it}$$

$$\xi = k^2 S_t^N S_s^N - k S_t^{kN} S_s^N - k S_t^N S_s^{kN} + S_t^{kN} S_s^{kN} = (S_t^{kN} - k S_t^N)(S_s^{kN} - k S_s^N)$$

$$S_t^N = \sum_{i=1}^{N} u_{it}, \quad S_t^k N = \sum_{i=1}^{kN} u_{it}$$

Recall $t = \frac{\hat{\beta}_3 - \beta_3}{s.e.(\hat{\beta}_3)}$.

Consider the individual dummies case. We have

$$\hat{\beta} - \beta = \Big(\sum_{i=1}^{N} \sum_{t=1}^{T} \begin{bmatrix} \widetilde{DU}_t \\ Treat_i \cdot \widetilde{DU}_t \end{bmatrix} [\widetilde{DU}_t, Treat_i \cdot \widetilde{DU}_t] \Big)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} \begin{bmatrix} \widetilde{DU}_t \\ Treat_i \cdot \widetilde{DU}_t \end{bmatrix} u_{it} \quad \text{(C.1)}$$

Simple algebra yields

$$
\begin{aligned}
&\Big(\sum_{i=1}^{N}\sum_{t=1}^{T}\begin{bmatrix}\widetilde{DU}_t \\ Treat_i\cdot\widetilde{DU}_t\end{bmatrix}[\widetilde{DU}_t, Treat_i\cdot\widetilde{DU}_t]\Big)^{-1} \\
=&\Big(\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{DU}_t^2\begin{bmatrix}1 & Treat_i \\ Treat_i & Treat_i\end{bmatrix}\Big)^{-1} = \Big(\lambda(1-\lambda)NT\begin{bmatrix}1 & k \\ k & k\end{bmatrix}\Big)^{-1} \\
=&\frac{1}{\lambda k(1-\lambda)(1-k)NT}\begin{bmatrix}k & -k \\ -k & 1\end{bmatrix}
\end{aligned}
\tag{C.2}
$$

$$
\begin{aligned}
\sum_{i=1}^{N}\sum_{t=1}^{T}\begin{bmatrix}\widetilde{DU}_t \\ Treat_i\cdot\widetilde{DU}_t\end{bmatrix}u_{it} &= \sum_{i=1}^{N}\begin{bmatrix}1 \\ Treat_i\end{bmatrix}\sum_{t=1}^{T}\widetilde{DU}_t u_{it} = \sum_{i=1}^{N}\begin{bmatrix}1 \\ Treat_i\end{bmatrix}\Big(\lambda\sum_{t=1}^{T}u_{it} - \sum_{t=1}^{\lambda T}u_{it}\Big) \\
&= \begin{bmatrix}\lambda\sum_{i=1}^{N}\sum_{t=1}^{T}u_{it} - \sum_{i=1}^{N}\sum_{t=1}^{\lambda T}u_{it} \\ \lambda\sum_{i=1}^{kN}\sum_{t=1}^{T}u_{it} - \sum_{i=1}^{kN}\sum_{t=1}^{\lambda T}u_{it}\end{bmatrix}
\end{aligned}
\tag{C.3}
$$

Plugging (C.2) and (C.3) into (C.1), it directly follows

$$
\hat{\beta}-\beta = \frac{1}{\lambda k(1-\lambda)(1-k)NT}\begin{bmatrix}k(\lambda\sum_{i>kN}\sum_{t=1}^{T}u_{it} - \sum_{i>kN}\sum_{t=1}^{\lambda T}u_{it}) \\ \eta\end{bmatrix}
$$

In particular, we have

$$
\hat{\beta}_3-\beta_3 = \frac{\eta}{\lambda k(1-\lambda)(1-k)NT}. \tag{C.4}
$$

Next, consider the standard error matrix. We know

$$
\hat{\hat{v}}_t = \sum_{i=1}^{N}\begin{bmatrix}\widetilde{DU}_t \\ Treat_i\cdot\widetilde{DU}_t\end{bmatrix}u_{it} = \widetilde{DU}_t\begin{bmatrix}S_t^N \\ S_t^{kN}\end{bmatrix}
$$

Therefore,

$$\hat{\hat{\Omega}} = T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T} K_{ts}\hat{\hat{v}}_t\hat{\hat{v}}_s' = T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T} K_{ts}\widetilde{\widetilde{DU}}_t\widetilde{\widetilde{DU}}_s \begin{bmatrix} S_t^N \\ S_t^{kN} \end{bmatrix} [S_s^N, S_s^{kN}]$$
$$= T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T} K_{ts}\widetilde{\widetilde{DU}}_t\widetilde{\widetilde{DU}}_s \begin{bmatrix} S_t^N S_s^N & S_t^N S_s^{kN} \\ S_t^{kN} S_s^N & S_t^{kN} S_s^{kN} \end{bmatrix}$$

Using this formula, it follows

$$\Big(\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T}\begin{bmatrix} \widetilde{\widetilde{DU}}_t \\ Treat_i \cdot \widetilde{\widetilde{DU}}_t \end{bmatrix}[\widetilde{\widetilde{DU}}_t, Treat_i \cdot \widetilde{\widetilde{DU}}_t]\Big)^{-1}\hat{\hat{\Omega}}\Big(\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T}\begin{bmatrix} \widetilde{\widetilde{DU}}_t \\ Treat_i \cdot \widetilde{\widetilde{DU}}_t \end{bmatrix}$$
$$\cdot [\widetilde{\widetilde{DU}}_t, Treat_i \cdot \widetilde{\widetilde{DU}}_t]\Big)^{-1}$$
$$= \Big(\frac{1}{\lambda k(1-\lambda)(1-k)N}\Big)^2 \begin{bmatrix} k & -k \\ -k & 1 \end{bmatrix} \hat{\hat{\Omega}} \begin{bmatrix} k & -k \\ -k & 1 \end{bmatrix}$$
$$= \Big(\frac{1}{\lambda k(1-\lambda)(1-k)N}\Big)^2 \begin{bmatrix} * & * \\ * & T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T} K_{ts}\widetilde{\widetilde{DU}}_t\widetilde{\widetilde{DU}}_s\xi \end{bmatrix}$$

Specifically, we have

$$s.e.(\hat{\beta}_3) = \frac{1}{T(\lambda k(1-\lambda)(1-k)N)^2}\sum_{t=1}^{T}\sum_{s=1}^{T} K_{ts}\widetilde{\widetilde{DU}}_t\widetilde{\widetilde{DU}}_s\xi. \tag{C.5}$$

Now consider the individual and time dummies case. Similarly we can derive

$$\hat{\beta}_3 - \beta_3 = \Big(\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{\widetilde{Treat}}_i^2 \cdot \widetilde{\widetilde{DU}}_t^2\Big)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{\widetilde{Treat}}_i\widetilde{\widetilde{DU}}_t u_{it}$$
$$= \frac{1}{\lambda k(1-\lambda)(1-k)NT}\sum_{i=1}^{N}\widetilde{\widetilde{Treat}}_i\Big(\lambda\sum_{t=1}^{T}u_{it} - \sum_{t=1}^{\lambda T}u_{it}\Big) \tag{C.6}$$
$$= \frac{\eta}{\lambda k(1-\lambda)(1-k)NT}$$

For the standard error matrix, it is easy to show

$$\hat{\hat{v}}_t = \sum_{i=1}^{N}\widetilde{\widetilde{Treat}}_i\widetilde{\widetilde{DU}}_t u_{it} = \widetilde{\widetilde{DU}}_t(S_t^{kN} - kS_t^N),$$

and

$$\hat{\hat{\Omega}} = T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T} K_{ts}\hat{\hat{v}}_t\hat{\hat{v}}_s' = T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T} K_{ts}\widetilde{DU}_t\widetilde{DU}_s(S_t^{kN}-kS_t^N)(S_s^{kN}-kS_s^N)$$
$$= T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T} K_{ts}\widetilde{DU}_t\widetilde{DU}_s\xi.$$

Thus, it follows

$$s.e.(\hat{\beta}_3) = \Big(\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{Treat}_i^2\cdot\widetilde{DU}_t^2\Big)^{-2}\hat{\hat{\Omega}} = \frac{1}{T(\lambda k(1-\lambda)(1-k)N)^2}\sum_{t=1}^{T}\sum_{s=1}^{T} K_{ts}\widetilde{DU}_t\widetilde{DU}_s\xi. \tag{C.7}$$

From above, we know the top and the bottom of t statistics are exactly equivalent in these two cases. As a result, t statistics are exact equivalent in these cases. By symmetry, it is easy to show that this exact equivalence result holds in the case when only time period dummies are included. □

Proof of Proposition 2.1. $\sqrt{T}(\hat{\beta}-\beta) = (T^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{x}_{it}\tilde{x}_{it}')^{-1}(T^{-\frac{1}{2}}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{x}_{it}u_{it})$. Using Assumption 2.1 and 2.2, it can be shown that

$$T^{-\frac{1}{2}}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{x}_{it}u_{it} = T^{-\frac{1}{2}}\sum_{t=1}^{T}\left(\tilde{x}_{1t},\ldots,\tilde{x}_{Nt}\right)u_t = T^{-\frac{1}{2}}\sum_{t=1}^{T} A\cdot\widetilde{DU}_t\cdot u_t$$
$$= A\cdot T^{-\frac{1}{2}}\sum_{t=1}^{T}\Big[DU_t - T^{-1}\sum_{s=1}^{T} DU_s\mathbf{f}(s)'\tau_T\Big(T^{-1}\sum_{s=1}^{T}\tau_T\mathbf{f}(s)\mathbf{f}(s)'\tau_T\Big)^{-1}$$
$$\cdot\,\tau_T\mathbf{f}(t)\Big]u_t$$
$$\Rightarrow A\Lambda\int_0^1\Big[\mathbb{1}(r>\lambda) - \int_\lambda^1\mathbf{F}(s)'ds\Big(\int_0^1\mathbf{F}(s)\mathbf{F}(s)'ds\Big)^{-1}\mathbf{F}(r)\Big]dW(r)$$
$$= \Lambda^*\int_0^1 H^F(r,\lambda)dW^*(r)$$

$$T^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{x}_{it}\tilde{x}_{it}' = T^{-1}\sum_{t=1}^{T} A\cdot\widetilde{DU}_t\widetilde{DU}_t\cdot A' = G\cdot T^{-1}\sum_{t=1}^{T}\widetilde{DU}_t^2$$
$$= G\cdot T^{-1}\sum_{t=1}^{T}\Big[DU_t - T^{-1}\sum_{s=1}^{T} DU_s\mathbf{f}(s)'\tau_T\Big(T^{-1}\sum_{s=1}^{T}\tau_T\mathbf{f}(s)\mathbf{f}(s)'\tau_T\Big)^{-1}$$
$$\cdot\,\tau_T\mathbf{f}(t)\Big]^2$$
$$\Rightarrow G\cdot\int_0^1\Big[\mathbb{1}(r>\lambda) - \int_\lambda^1\mathbf{F}(s)'ds\Big(\int_0^1\mathbf{F}(s)\mathbf{F}(s)'ds\Big)^{-1}\mathbf{F}(r)\Big]^2 dr$$
$$= G\int_0^1 H^F(r,\lambda)^2 dr$$

Therefore,

$$\sqrt{T}(\hat{\beta}-\beta) \Rightarrow (G\int_0^1 H^F(r,\lambda)^2 dr)^{-1} \cdot \Lambda^* \int_0^1 H^F(r,\lambda) dW^*(r) \qquad \square$$

Proof of Lemma 2.2. Using Assumption 2.1, 2.2 and Proposition 2.1, we obtain

$$\begin{aligned}
T^{-\frac{1}{2}}\hat{\hat{S}}_{[rt]} &= T^{-\frac{1}{2}}\sum_{t=1}^{[rT]}\hat{v}_t = T^{-\frac{1}{2}}\sum_{t=1}^{[rT]}\sum_{i=1}^{N}\tilde{x}_{it}\hat{u}_{it} = T^{-\frac{1}{2}}\sum_{t=1}^{[rT]}\sum_{i=1}^{N}\tilde{x}_{it}[\tilde{u}_{it}-\tilde{x}'_{it}(\hat{\beta}-\beta)] \\
&= T^{-\frac{1}{2}}\sum_{t=1}^{[rT]}\sum_{i=1}^{N}\tilde{x}_{it}u_{it} - T^{-\frac{1}{2}}\sum_{t=1}^{[rT]}\sum_{i=1}^{N}\tilde{x}_{it}\ddot{u}_{it} - \Big(T^{-1}\sum_{t=1}^{[rT]}\sum_{i=1}^{N}\tilde{x}_{it}\tilde{x}'_{it}\Big)\sqrt{T}(\hat{\beta}-\beta) \\
&= A\cdot T^{-\frac{1}{2}}\sum_{t=1}^{[rT]}\widetilde{DU}_t u_t - A\cdot T^{-\frac{1}{2}}\sum_{s=1}^{T}u_s\mathbf{f}(s)'\tau_T\cdot\Big(\frac{1}{T}\sum_{s=1}^{T}\tau_T\mathbf{f}(s)\mathbf{f}(s)'\tau_T\Big)^{-1} \\
&\quad \cdot\frac{1}{T}\sum_{t=1}^{[rT]}\tau_T\mathbf{f}(t)\widetilde{DU}_t - \Big(G\cdot T^{-1}\sum_{t=1}^{[rT]}\widetilde{DU}_t^2\Big)\sqrt{T}(\hat{\beta}-\beta) \\
&\Rightarrow \Lambda^*\Big[\int_0^r H^F(s,\lambda)dW^*(s) - \int_0^1 dW(s)F(s)'\Big(\int_0^1 F(s)F(s)'ds\Big)^{-1} \\
&\quad \cdot\int_0^r F(s)H^F(s,\lambda)ds - \int_0^r H^F(s,\lambda)^2 ds\Big(\int_0^1 H^F(s,\lambda)^2 ds\Big)^{-1}N^F(W^*)\Big] \\
&= \Lambda^* Q^F(r,\lambda,W^*)
\end{aligned}$$

because

$$\begin{aligned}
T^{-\frac{1}{2}}\sum_{t=1}^{[rT]}\sum_{i=1}^{N}\tilde{x}_{it}\hat{u}_{it} &= T^{-\frac{1}{2}}\sum_{t=1}^{[rT]}\sum_{i=1}^{N}\tilde{x}_{it}\cdot(\sum_{s=1}^{T}u_{is}\mathbf{f}(s)')(\sum_{s=1}^{T}\mathbf{f}(s)\mathbf{f}(s)')^{-1}\mathbf{f}(t) \\
&= T^{-\frac{1}{2}}\sum_{t=1}^{[rT]}(\sum_{s=1}^{T}\sum_{i=1}^{N}\tilde{x}_{it}u_{is}\mathbf{f}(s)')(\sum_{s=1}^{T}\mathbf{f}(s)\mathbf{f}(s)')^{-1}\mathbf{f}(t) \\
&= T^{-\frac{1}{2}}\sum_{t=1}^{[rT]}(\sum_{s=1}^{T}A\widetilde{DU}_t\mathbf{u}_s\mathbf{f}(s)')(\sum_{s=1}^{T}\mathbf{f}(s)\mathbf{f}(s)')^{-1}\mathbf{f}(t) \\
&= A\cdot T^{-\frac{1}{2}}\sum_{s=1}^{T}\mathbf{u}_s\mathbf{f}(s)'\tau_T\cdot(\frac{1}{T}\sum_{s=1}^{T}\tau_T\mathbf{f}(s)\mathbf{f}(s)'\tau_T)^{-1}\frac{1}{T}\sum_{t=1}^{[rT]}\tau_T\mathbf{f}(t)\widetilde{DU}_t \qquad \square
\end{aligned}$$

Proof of Proposition 2.3. It directly follows from (2.7), Lemma 2.2 and the continuous mapping theorem that

$$\hat{\hat{\Omega}} = \frac{2}{b}T^{-1}\sum_{t=1}^{T-1} T^{-\frac{1}{2}}\hat{\hat{S}}_t \cdot T^{-\frac{1}{2}}\hat{\hat{S}}_t' - \frac{1}{b}T^{-1}\sum_{t=1}^{T-M-1}(T^{-\frac{1}{2}}\hat{\hat{S}}_t \cdot T^{-\frac{1}{2}}\hat{\hat{S}}_{t+M}' + T^{-\frac{1}{2}}\hat{\hat{S}}_{t+M} \cdot T^{-\frac{1}{2}}\hat{\hat{S}}_t')$$
$$\Rightarrow \frac{2}{b}\int_0^1 \Lambda^* Q^F(r,\lambda,W^*)Q^F(r,\lambda,W^*)'\Lambda^{*'}dr$$
$$-\frac{1}{b}\int_0^{1-b}\Lambda^*\Big[Q^F(r,\lambda,W^*)Q^F(r+b,\lambda,W^*)' + Q^F(r+b,\lambda,W^*)Q^F(r,\lambda,W^*)'\Big]\Lambda^{*'}dr$$
$$= \Lambda^* P^F(b,\lambda,Q^F)\Lambda^{*'}$$

□

Proof of Theorem 2.1. Using Proposition 2.3, it directly follows that

$$R\Big(T^{-1}\sum_{i=1}^N\sum_{t=1}^T \tilde{x}_{it}\tilde{x}_{it}'\Big)^{-1}\hat{\hat{\Omega}}\Big(T^{-1}\sum_{i=1}^N\sum_{t=1}^T \tilde{x}_{it}\tilde{x}_{it}'\Big)^{-1}R'$$
$$\Rightarrow R\Big(G\int_0^1 H^F(r,\lambda)^2 dr\Big)^{-1}\Lambda^* P^F(b,\lambda,Q^F)\Lambda^{*'}\Big(G\int_0^1 H^F(r,\lambda)^2 dr\Big)^{-1}R'$$
$$= P^F\Big(b,\lambda,Q^F\Big(r,\lambda,R(G\int_0^1 H^F(r,\lambda)^2 dr)^{-1}\Lambda^* W^*\Big)\Big)$$
$$= \Lambda_q^{**}P^F(b,\lambda,Q^F(r,\lambda,W_q^{**}))\Lambda_q^{**'} = \Lambda_q^{**}P^F(b,\lambda,Q_q^{F**})\Lambda_q^{**'} \tag{C.8}$$

Using Proposition 2.1, we have

$$R\sqrt{T}(\hat{\beta}-\beta) \Rightarrow R\Big(G\int_0^1 H^F(r,\lambda)^2 dr\Big)^{-1}\cdot\Lambda^*\int_0^1 H^F(r,\lambda)dW^*(r) = \Lambda_q^{**}\int_0^1 H^F(r,\lambda)dW_q^{**}(r) \tag{C.9}$$

With (C.8) and (C.9), it follows that

$$Wald = (R\hat{\beta}-r)'[R\hat{V}R']^{-1}(R\hat{\beta}-r)$$
$$= (R\sqrt{T}(\hat{\beta}-\beta))'\Big[R\Big(T^{-1}\sum_{i=1}^N\sum_{t=1}^T \tilde{x}_{it}\tilde{x}_{it}'\Big)^{-1}\hat{\hat{\Omega}}\Big(T^{-1}\sum_{i=1}^N\sum_{t=1}^T \tilde{x}_{it}\tilde{x}_{it}'\Big)^{-1}R'\Big]^{-1}$$
$$\cdot R\sqrt{T}(\hat{\beta}-\beta)$$
$$\Rightarrow (\Lambda_q^{**}\int_0^1 H^F(r,\lambda)d\mathbf{W}_q^{**}(r))'[\Lambda_q^{**}\mathbf{P^F}(b,\lambda,\mathbf{Q^F}_q^{**})\Lambda_q^{**'}]^{-1}\Lambda_q^{**}\int_0^1 H^F(r,\lambda)d\mathbf{W}_q^{**}(r)$$
$$= N^F(W_q^{**})'\mathbf{P^F}(b,\lambda,\mathbf{Q^F}_q^{**})^{-1}N^F(W_q^{**})$$

When $q=1$, it directly follows that $t \Rightarrow \dfrac{N^F(W_1^{**})}{\sqrt{P^F(b,\lambda,Q_1^{F**})}}$. □

Proof of Lemma 2.4.

$$T^{-1}\sum_{t=1}^{T}\widehat{DU}_t\tilde{\mathbf{z}}_{it}' = T^{-1}\sum_{s=\lambda T+1}^{T}\mathbf{f}(s)'\Big(\sum_{s=1}^{T}\mathbf{f}(s)\mathbf{f}(s)'\Big)^{-1}\sum_{t=1}^{T}\mathbf{f}(t)\tilde{\mathbf{z}}_{it}' = 0 \tag{C.10}$$

using the fact that $\sum_{t=1}^{T}\mathbf{f}(t)\tilde{\mathbf{z}}_{it}' = 0$. Hence, $T^{-1}\sum_{t=1}^{[rT]}\widehat{DU}_t\tilde{\mathbf{z}}_{it}' = o_p(1)$. If $r>\lambda$, then

$$\begin{aligned}
T^{-1}\sum_{t=1}^{[rT]} DU_t\tilde{\mathbf{z}}_{it}' &= T^{-1}\sum_{t=\lambda+1}^{[rT]}\tilde{\mathbf{z}}_{it}' \\
&= T^{-1}\sum_{t=\lambda+1}^{[rT]}\mathbf{z}_{it}' - T^{-1}\sum_{t=\lambda+1}^{[rT]}\mathbf{f}(t)'\tau_T\Big(T^{-1}\sum_{s=1}^{T}\tau_T\mathbf{f}(s)\mathbf{f}(s)'\tau_T\Big)^{-1} \\
&\quad \cdot T^{-1}\sum_{s=1}^{T}\tau_T\mathbf{f}(s)\mathbf{z}_{is}' \\
&\xrightarrow{p}(r-\lambda)\Big(\mu_i' - \int_0^1\mathbf{F}(r)'dr\Big(\int_0^1\mathbf{F}(r)\mathbf{F}(r)'dr\Big)^{-1}(\mu_i',0,\ldots,0)\Big) \\
&= (r-\lambda)(\mu_i'-\mu_i') = 0
\end{aligned} \tag{C.11}$$

If $r\le\lambda$, then

$$T^{-1}\sum_{t=1}^{[rT]} DU_t\tilde{\mathbf{z}}_{it}' = 0 \tag{C.12}$$

From (C.10), (C.11) and (C.12), it directly follows that

$$T^{-1}\sum_{t=1}^{[rT]}\widetilde{DU}_t\tilde{\mathbf{z}}_{it}' = T^{-1}\sum_{t=1}^{[rT]}(DU_t-\widehat{DU}_t)\tilde{\mathbf{z}}_{it}' \xrightarrow{p} 0 \tag{C.13}$$

and thus

$$T^{-1}\sum_{i=1}^{N}\sum_{t=1}^{[rT]}h_{it}\tilde{\mathbf{z}}_{it}' = \sum_{i=1}^{N}\begin{bmatrix}1\\ Treat_i\end{bmatrix}T^{-1}\sum_{t=1}^{[rT]}\widetilde{DU}_t\tilde{\mathbf{z}}_{it}' \xrightarrow{p} 0.$$

□

Proof of Theorem 2.3. The $K\times 1$ vector $\tilde{z}_{it}u_{it}$ can be written in terms of the $N(K+1)\times 1$ vector v_t as follows

$$\begin{aligned}
\tilde{z}_{it}u_{it} &= (z_{it}-\hat{b}_i'\mathbf{f}(t))u_{it} = ((z_{it}-b_i'\mathbf{f}(t)) - (\hat{b}_i'\mathbf{f}(t)-b_i'\mathbf{f}(t)))u_{it} = Bv_t^{ii} - (\hat{b}_i-b_i)'\mathbf{f}(t)u_{it} \\
&= A_iv_t - (\tau_T^{-1}(\hat{b}_i-b_i))'\tau_T\mathbf{f}(t)u_{it}
\end{aligned}$$

Using this formula it is easy to show that

$$
\begin{aligned}
T^{-\frac{1}{2}}\sum_{t=1}^{[rT]}\tilde{z}_{it}u_{it} &= T^{-\frac{1}{2}}\sum_{t=1}^{[rT]}(A_i v_t - (\tau_T^{-1}(\hat{b}_i - b_i))'\tau_T \mathbf{f}(t)u_{it}) \\
&= A_i T^{-\frac{1}{2}}\sum_{t=1}^{[rT]} v_t - T^{-\frac{1}{2}}(\sqrt{T}\tau_T^{-1}(\hat{b}_i - b_i))' \cdot T^{-\frac{1}{2}}\sum_{t=1}^{[rT]}\tau_T \mathbf{f}(t)u_{it} \\
&= A_i T^{-\frac{1}{2}}\sum_{t=1}^{[rT]} v_t + T^{-\frac{1}{2}} O_p(1)\cdot O_p(1) = A_i T^{-\frac{1}{2}}\sum_{t=1}^{[rT]} v_t + o_p(1) \\
&\Rightarrow A_i \dot{\Lambda} W(r) \qquad \text{(C.14)}
\end{aligned}
$$

using Assumption 2.1 and 2.4. With Assumption 2.1, 2.3, 2.4, Lemma 2.4 and (C.14), simple algebra gives

$$
\begin{aligned}
\sqrt{T}(\hat{\beta} - \beta) &= \Big(\sum_{i=1}^{N} T^{-1}\sum_{t=1}^{T}\tilde{x}_{it}\tilde{x}_{it}'\Big)^{-1}\Big(\sum_{i=1}^{N} T^{-\frac{1}{2}}\sum_{t=1}^{T}\tilde{x}_{it}u_{it}\Big) \\
&= \begin{bmatrix} \sum_{i=1}^{N} T^{-1}\sum_{t=1}^{T} h_{it}h_{it}' & \sum_{i=1}^{N} T^{-1}\sum_{t=1}^{T} h_{it}\tilde{z}_{it}' \\ \sum_{i=1}^{N} T^{-1}\sum_{t=1}^{T} \tilde{z}_{it}h_{it}' & \sum_{i=1}^{N} T^{-1}\sum_{t=1}^{T} \tilde{z}_{it}\tilde{z}_{it}' \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^{N} T^{-\frac{1}{2}}\sum_{t=1}^{T} h_{it}u_{it} \\ \sum_{i=1}^{N} T^{-\frac{1}{2}}\sum_{t=1}^{T} \tilde{z}_{it}u_{it} \end{bmatrix} \\
&\Rightarrow \begin{bmatrix} (G\int_0^1 H^F(r,\lambda)^2 dr)^{-1} & \mathbf{0} \\ \mathbf{0} & \bar{Q}^{-1} \end{bmatrix} \\
&\quad \cdot \begin{bmatrix} (A\otimes\tilde{e}_1')\dot{\Lambda}\int_0^1[\mathbb{1}(r>\lambda) - \int_\lambda^1 \mathbf{F}(s)'ds(\int_0^1 \mathbf{F}(s)\mathbf{F}(s)'ds)^{-1}\mathbf{F}(r)]dW(r) \\ (\sum_{i=1}^{N} A_i)\dot{\Lambda}W(1) \end{bmatrix} \\
&= \begin{bmatrix} (G\int_0^1 H^F(r,\lambda)^2 dr)^{-1}(\dot{\Lambda}^*\int_0^1 H^F(r,\lambda)dW^*(r) \\ (\bar{Q}^{-1}(\sum_{i=1}^{N} A_i)\dot{\Lambda}W(1) \end{bmatrix}
\end{aligned}
$$

Let

$$
\dot{\Lambda}^{**} = \begin{bmatrix} \dot{\Lambda}^* & \mathbf{0} \\ \mathbf{0} & (\sum_{i=1}^{N} A_i)\dot{\Lambda} \end{bmatrix}
$$

which is a $(K+2)\times(K+2)$ block diagonal matrix. Using the fact that $\sum_{t=1}^{T}\tilde{z}_{it}\mathbf{f}(t)' = 0$, it follows that

$$\sum_{t=1}^{[rT]}\sum_{i=1}^{N}\tilde{z}_{it}\hat{u}_{it} = \sum_{i=1}^{N}\sum_{t=1}^{[rT]}\tilde{z}_{it}\sum_{s=1}^{T}u_{is}\mathbf{f}(s)'\Big(\sum_{s=1}^{T}\mathbf{f}(s)\mathbf{f}(s)'\Big)^{-1}\mathbf{f}(t)$$

$$= \sum_{i=1}^{N}\Big(\sum_{t=1}^{[rT]}\tilde{z}_{it}\mathbf{f}(t)'\Big)\Big(\sum_{s=1}^{T}\mathbf{f}(s)\mathbf{f}(s)'\Big)^{-1}\sum_{s=1}^{T}u_{is}\mathbf{f}(s) \tag{C.15}$$

$$= \sum_{i=1}^{N}op(1)\cdot\Big(\sum_{s=1}^{T}\mathbf{f}(s)\mathbf{f}(s)'\Big)^{-1}\sum_{s=1}^{T}u_{is}\mathbf{f}(s) \xrightarrow{p} 0$$

The limits of the partial sums $\hat{\hat{S}}_{[rT]}$ are easy to obtain

$$T^{-\frac{1}{2}}\hat{\hat{S}}_{[rT]} = T^{-\frac{1}{2}}\sum_{t=1}^{[rT]}\sum_{i=1}^{N}\tilde{x}_{it}\tilde{u}_{it} - (T^{-1}\sum_{t=1}^{[rT]}\sum_{i=1}^{N}\tilde{x}_{it}\tilde{x}_{it}')\sqrt{T}(\hat{\beta}-\beta)$$

$$= \begin{bmatrix} T^{-\frac{1}{2}}\sum_{t=1}^{[rT]}\sum_{i=1}^{N}h_{it}(u_{it}-\hat{u}_{it}) \\ T^{-\frac{1}{2}}\sum_{t=1}^{[rT]}\sum_{i=1}^{N}\tilde{z}_{it}(u_{it}-\hat{u}_{it}) \end{bmatrix} - \begin{bmatrix} T^{-1}\sum_{t=1}^{[rT]}\sum_{i=1}^{N}h_{it}h_{it}' & T^{-1}\sum_{t=1}^{[rT]}\sum_{i=1}^{N}h_{it}\tilde{z}_{it}' \\ T^{-1}\sum_{t=1}^{[rT]}\sum_{i=1}^{N}\tilde{z}_{it}h_{it}' & T^{-1}\sum_{t=1}^{[rT]}\sum_{i=1}^{N}\tilde{z}_{it}\tilde{z}_{it}' \end{bmatrix}$$

$$\cdot\sqrt{T}(\hat{\beta}-\beta)$$

$$\Rightarrow \begin{bmatrix} \dot{\Lambda}^*[\int_0^r H^F(s,\lambda)dW^*(s) - \int_0^1 dW(s)F(s)'\Big(\int_0^1 F(s)F(s)'ds\Big)^{-1} \\ \cdot\int_0^r F(s)H^F(s,\lambda)ds] \\ (\sum_{i=1}^{N}A_i)\dot{\Lambda}W(r) \end{bmatrix}$$

$$- \begin{bmatrix} G\int_0^r H^F(s,\lambda)^2 ds & \mathbf{0} \\ \mathbf{0} & r\bar{Q} \end{bmatrix} \cdot \begin{bmatrix} (G\int_0^1 H^F(r,\lambda)^2 dr)^{-1}(\dot{\Lambda}^*\int_0^1 H^F(r,\lambda)dW^*(r) \\ \bar{Q}^{-1}(\sum_{i=1}^{N}A_i)\dot{\Lambda}W(1) \end{bmatrix}$$

$$= \begin{bmatrix} \dot{\Lambda}^* Q^F(r,\lambda,W^*) \\ (\sum_{i=1}^{N}A_i)\dot{\Lambda}B(r) \end{bmatrix} = \dot{\Lambda}^{**}\begin{bmatrix} Q^F(r,\lambda,W^*) \\ B(r) \end{bmatrix}$$

The limit of $\hat{\hat{\Omega}}$ can be written as

$$
\begin{aligned}
\hat{\hat{\Omega}} &= \frac{2}{b}T^{-1}\sum_{t=1}^{T-1} T^{-\frac{1}{2}}\hat{\hat{S}}_t \cdot T^{-\frac{1}{2}}\hat{\hat{S}}_t' - \frac{1}{b}T^{-1}\sum_{t=1}^{T-M-1}(T^{-\frac{1}{2}}\hat{\hat{S}}_t \cdot T^{-\frac{1}{2}}\hat{\hat{S}}_{t+M}' + T^{-\frac{1}{2}}\hat{\hat{S}}_{t+M} \cdot T^{-\frac{1}{2}}\hat{\hat{S}}_t') \\
&\Rightarrow \frac{2}{b}\int_0^1 \dot{\Lambda}^{**}\begin{bmatrix} Q^F(r,\lambda,W^*) \\ B(r) \end{bmatrix}\begin{bmatrix} Q^F(r,\lambda,W^*) \\ B(r) \end{bmatrix}' \dot{\Lambda}^{**\prime} dr - \frac{1}{b}\int_0^{1-b}\dot{\Lambda}^{**}\Bigg(\begin{bmatrix} Q^F(r,\lambda,W^*) \\ B(r) \end{bmatrix} \\
&\quad \cdot \begin{bmatrix} Q^F(r+b,\lambda,W^*) \\ B(r+b) \end{bmatrix}' - \begin{bmatrix} Q^F(r+b,\lambda,W^*) \\ B(r+b) \end{bmatrix}\begin{bmatrix} Q^F(r,\lambda,W^*) \\ B(r) \end{bmatrix}'\Bigg)\dot{\Lambda}^{**\prime} dr \\
&= \dot{\Lambda}^{**}\begin{bmatrix} P^F(b,\lambda,Q^F) & P_{12}(b,\lambda,Q^F,B) \\ P_{21}(b,\lambda,Q^F,B) & P(b,B) \end{bmatrix}\dot{\Lambda}^{**\prime}
\end{aligned}
$$

$$
\begin{aligned}
&\mathbf{R}(T^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{x}_{it}\tilde{x}_{it}')^{-1}\hat{\hat{\Omega}}(T^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{x}_{it}\tilde{x}_{it}')^{-1}\mathbf{R}' \\
&\Rightarrow \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix}\begin{bmatrix} (G\int_0^1 H^F(r,\lambda)^2 dr)^{-1} & \mathbf{0} \\ \mathbf{0} & \bar{Q}^{-1} \end{bmatrix} \cdot \dot{\Lambda}^{**}\begin{bmatrix} P^F(b,\lambda,Q^F) & P_{12}(b,\lambda,Q^F,B) \\ P_{21}(b,\lambda,Q^F,B) & P(b,B) \end{bmatrix} \\
&\quad \cdot \dot{\Lambda}^{**\prime}\begin{bmatrix} (G\int_0^1 H^F(r,\lambda)^2 dr)^{-1} & \mathbf{0} \\ \mathbf{0} & \bar{Q}^{-1} \end{bmatrix}\begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix}' \\
&= \begin{bmatrix} R_{11}(G\int_0^1 H^F(r,\lambda)^2 dr)^{-1}\dot{\Lambda}^* & R_{12}\bar{Q}^{-1}(\Sigma_{i=1}^N A_i)\dot{\Lambda} \\ R_{21}(G\int_0^1 H^F(r,\lambda)^2 dr)^{-1}\dot{\Lambda}^* & R_{22}\bar{Q}^{-1}(\Sigma_{i=1}^N A_i)\dot{\Lambda} \end{bmatrix} \\
&\quad \cdot \begin{bmatrix} P^F(b,\lambda,Q^F) & P_{12}(b,\lambda,Q^F,B) \\ P_{21}(b,\lambda,Q^F,B) & P(b,B) \end{bmatrix} \\
&\quad \cdot \begin{bmatrix} R_{11}(G\int_0^1 H^F(r,\lambda)^2 dr)^{-1}\dot{\Lambda}^* & R_{12}\bar{Q}^{-1}(\Sigma_{i=1}^N A_i)\dot{\Lambda} \\ R_{21}(G\int_0^1 H^F(r,\lambda)^2 dr)^{-1}\dot{\Lambda}^* & R_{22}\bar{Q}^{-1}(\Sigma_{i=1}^N A_i)\dot{\Lambda} \end{bmatrix}'
\end{aligned}
\tag{C.16}
$$

$$\mathbf{R}\sqrt{T}(\hat{\beta}-\beta) \Rightarrow \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix} \begin{bmatrix} (G\int_0^1 H^F(r,\lambda)^2 dr)^{-1}(\dot{\Lambda}^* \int_0^1 H^F(r,\lambda)dW^*(r) \\ \bar{Q}^{-1}(\Sigma_{i=1}^N A_i)\dot{\Lambda}W(1) \end{bmatrix} \tag{C.17}$$

If $q_2 = 0$ and $R_{12} = 0$, that is, we are testing restrictions on the DD estimator, then $R = [R_{11}, 0]$ and the limits of (C.16) and (C.17) are simplified as follows

$$R\Big(T^{-1}\sum_{i=1}^N\sum_{t=1}^T \tilde{x}_{it}\tilde{x}_{it}'\Big)^{-1}\hat{\hat{\Omega}}\Big(T^{-1}\sum_{i=1}^N\sum_{t=1}^T \tilde{x}_{it}\tilde{x}_{it}'\Big)^{-1}R'$$
$$\Rightarrow R_{11}(G\int_0^1 H^F(r,\lambda)^2 dr)^{-1}\dot{\Lambda}^* P^F(b,\lambda,Q^F)\dot{\Lambda}^{*\prime}(G\int_0^1 H^F(r,\lambda)^2 dr)^{-1}R_{11}'$$
$$= \bar{\Lambda}_1 P^F(b,\lambda,\bar{Q}^F)\bar{\Lambda}_1'$$

and

$$R\sqrt{T}(\hat{\beta}-\beta) \Rightarrow R_{11}(G\int_0^1 H^F(r,\lambda)^2 dr)^{-1}\dot{\Lambda}^*\int_0^1 H^F(r,\lambda)dW^*(r) = \bar{\Lambda}_1\int_0^1 H^F(r,\lambda)d\bar{W}(r)$$

where $\bar{\mathbf{W}}(r)$ is a $q_1 \times 1$ vector of standard Wiener processes and $\bar{\Lambda}_1$ is the matrix square root of the matrix

$$R_{11}(G\int_0^1 H^F(r,\lambda)^2 dr)^{-1}\dot{\Lambda}^*\dot{\Lambda}^{*\prime}(G\int_0^1 H^F(r,\lambda)^2 dr)^{-1}R_{11}'.$$

It directly follows that

$$Wald \Rightarrow (\bar{\Lambda}_1\int_0^1 H^F(r,\lambda)d\bar{W}(r))'(\bar{\Lambda}_1 P^F(b,\lambda,\bar{Q}^F)\bar{\Lambda}_1)^{-1}\bar{\Lambda}_1\int_0^1 H^F(r,\lambda)d\bar{W}(r)$$
$$= (\int_0^1 H^F(r,\lambda)d\bar{W}(r))'(P^F(b,\lambda,\bar{Q}^F))^{-1}\int_0^1 H^F(r,\lambda)d\bar{W}(r)$$

If $q_1 = 0$ and $R_{21} = 0$, that is, we are testing restrictions on the additional regressors, then $R = [0, R_{22}]$ and the limits of (C.16) and (C.17) are simplified as follows

$$R(T^{-1}\sum_{i=1}^N\sum_{t=1}^T \tilde{x}_{it}\tilde{x}_{it}')^{-1}\hat{\hat{\Omega}}(T^{-1}\sum_{i=1}^N\sum_{t=1}^T \tilde{x}_{it}\tilde{x}_{it}')^{-1}R'$$
$$\Rightarrow R_{22}(\sum_{i=1}^N Q_i)^{-1}(\sum_{i=1}^N A_i)\dot{\Lambda}P(b,B)\dot{\Lambda}'(\sum_{i=1}^N A_i)'(\sum_{i=1}^N Q_i)^{-1}R_{22}' = \bar{\Lambda}_2 P(b,B)\bar{\Lambda}_2'$$
$$R\sqrt{T}(\hat{\beta}-\beta) \Rightarrow R_{22}(\sum_{i=1}^N Q_i)^{-1}(\sum_{i=1}^N A_i)\dot{\Lambda}W(1) = \bar{\Lambda}_2 W_q(1)$$

where $W_q(1)$ is a $q_2 \times 1$ vector of standard Wiener processes and $\bar{\Lambda}_2$ is the matrix square root of the matrix

$$R_{22}(\sum_{i=1}^{N} Q_i)^{-1}(\sum_{i=1}^{N} A_i)\dot{\Lambda}\dot{\Lambda}'(\sum_{i=1}^{N} A_i)'(\sum_{i=1}^{N} Q_i)^{-1}R_{22}'$$

It directly follows that

$$Wald \Rightarrow (\bar{\Lambda}_2 W_q(1))'(\bar{\Lambda}_2 P(b,B)\bar{\Lambda}_2)^{-1}\bar{\Lambda}_2 W_q(1) = W_q(1)'P_q(b,B)^{-1}W_q(1) \qquad \square$$

Proof of Theorem 2.4. The key step is to show that the limits of $\sqrt{T}(\hat{\beta}-\beta)$ and $T^{-\frac{1}{2}}\hat{\hat{S}}_{[rT]}$ take the same form as in Theorem 2.3. Once these results are obtained, the rest of the proof closely follows the proof in Theorem 2.3 and details are omitted. With both trend functions and time period dummies in the model it follows that

$$\begin{aligned}
\tilde{z}_{it}u_{it} &= (z_{it} - \hat{b}_i'\mathbf{f}(t))u_{it} - N^{-1}\sum_{j=1}^{N}(z_{jt} - \hat{b}_j'\mathbf{f}(t))u_{it} \\
&= ((z_{it} - b_i'\mathbf{f}(t)) - (\hat{b}_i'\mathbf{f}(t) - b_i'\mathbf{f}(t)))u_{it} - N^{-1}\sum_{j=1}^{N}((z_{jt} - b_j'\mathbf{f}(t)) - (\hat{b}_j'\mathbf{f}(t) - b_j'\mathbf{f}(t)))u_{it} \\
&= (z_{it} - b_i'\mathbf{f}(t))u_{it} - N^{-1}\sum_{j=1}^{N}(z_{jt} - b_j'\mathbf{f}(t))u_{it} - (\hat{b}_i'\mathbf{f}(t) - b_i'\mathbf{f}(t))u_{it} \\
&\quad + N^{-1}\sum_{j=1}^{N}(\hat{b}_j'\mathbf{f}(t) - b_j'\mathbf{f}(t))u_{it} \\
&= v_t^{ii} - N^{-1}\sum_{j=1}^{N} v_t^{ji} - (\hat{b}_i - b_i)'\mathbf{f}(t)u_{it} + N^{-1}\sum_{j=1}^{N}(\hat{b}_j - b_j)'\mathbf{f}(t)u_{it} \\
&= ([\mathbf{0}, e_i' \otimes I_K] - \frac{1}{N}[\mathbf{0}, \iota' \otimes I_K])(e_i' \otimes I_{NK+1})v_t^{ex} - (\hat{b}_i - b_i)'\mathbf{f}(t)u_{it} \\
&\quad + N^{-1}\sum_{j=1}^{N}(\hat{b}_j - b_j)'\mathbf{f}(t)u_{it} \\
&= A_i^{ex}v_t^{ex} - (\tau_T^{-1}(\hat{b}_i - b_i))'\tau_T\mathbf{f}(t)u_{it} + N^{-1}\sum_{j=1}^{N}(\tau_T^{-1}(\hat{b}_j - b_j))'\tau_T\mathbf{f}(t)u_{it}
\end{aligned}$$

Using this formula it directly follows that

$$T^{-\frac{1}{2}}\sum_{t=1}^{[rT]}\tilde{z}_{it}u_{it}=A_i^{ex}T^{-\frac{1}{2}}\sum_{t=1}^{[rT]}v_t^{ex}-T^{-\frac{1}{2}}(\sqrt{T}\tau_T^{-1}(\hat{b}_i-b_i))'\cdot T^{-\frac{1}{2}}\sum_{t=1}^{[rT]}\tau_T\mathbf{f}(t)u_{it}$$

$$+N^{-1}\sum_{j=1}^{N}T^{-\frac{1}{2}}(\sqrt{T}\tau_T^{-1}(\hat{b}_j-b_j))'\cdot T^{-\frac{1}{2}}\sum_{t=1}^{[rT]}\tau_T\mathbf{f}(t)u_{it} \tag{C.18}$$

$$=A_i^{ex}T^{-\frac{1}{2}}\sum_{t=1}^{[rT]}v_t^{ex}+op(1)\Rightarrow A_i^{ex}\Lambda^{ex}W^{ex}(r)$$

using Assumption 2.1 and 2.5. Using (C.13), we have

$$T^{-1}\sum_{t=1}^{[rT]}\sum_{i=1}^{N}\widetilde{Treat}_i\widetilde{DU}_t\cdot\tilde{\mathbf{z}}_{it}'=T^{-1}\sum_{t=1}^{[rT]}\sum_{i=1}^{N}\widetilde{Treat}_i\widetilde{DU}_t[\mathbf{z}_{it}-\hat{\mathbf{z}}_{it}-\frac{1}{N}\sum_{j=1}^{N}(\mathbf{z}_{jt}-\hat{\mathbf{z}}_{jt})]$$

$$=T^{-1}\sum_{t=1}^{[rT]}\sum_{i=1}^{N}\widetilde{Treat}_i\widetilde{DU}_t(\mathbf{z}_{it}-\hat{\mathbf{z}}_{it})-T^{-1}\sum_{t=1}^{[rT]}\frac{1}{N}\sum_{j=1}^{N}(\mathbf{z}_{jt}-\hat{\mathbf{z}}_{jt})$$

$$\cdot\sum_{i=1}^{N}\widetilde{Treat}_i\widetilde{DU}_t$$

$$=T^{-1}\sum_{t=1}^{[rT]}\sum_{i=1}^{N}\widetilde{Treat}_i\widetilde{DU}_t(\mathbf{z}_{it}-\hat{\mathbf{z}}_{it})-T^{-1}\sum_{t=1}^{[rT]}\frac{1}{N}\sum_{j=1}^{N}(\mathbf{z}_{jt}-\hat{\mathbf{z}}_{jt})$$

$$\cdot 0$$

$$=\sum_{i=1}^{N}\widetilde{Treat}_iT^{-1}\sum_{t=1}^{[rT]}\widetilde{DU}_t(\mathbf{z}_{it}-\hat{\mathbf{z}}_{it})\xrightarrow{p}0$$

Using Assumption 2.1, 2.3 and (C.18) it immediately follows that

$$\sqrt{T}(\hat{\beta}-\beta)\Rightarrow\begin{bmatrix}(\tilde{G}\int_0^1H^F(r,\lambda)^2dr)^{-1} & \mathbf{0}\\ \mathbf{0} & \bar{Q}^{-1}\end{bmatrix}$$

$$\cdot\begin{bmatrix}(\tilde{A}\otimes\bar{e}_1')\Lambda^{ex}\int_0^1[\mathbb{1}(r>\lambda)-\mathbf{F}(r)'(\int_0^1\mathbf{F}(s)\mathbf{F}(s)'ds)^{-1}\int_\lambda^1\mathbf{F}(s)ds]dW^{ex}(r)\\ (\sum_{i=1}^{N}A_i^{ex})\Lambda^{ex}W^{ex}(1)\end{bmatrix}$$

$$=\begin{bmatrix}(\tilde{G}\int_0^1H^F(r,\lambda)^2dr)^{-1}\Lambda^{ex*}\int_0^1H^F(r,\lambda)dW^{ex*}(r)\\ \bar{Q}^{-1}(\sum_{i=1}^{N}A_i^{ex})\Lambda^{ex}W^{ex}(1)\end{bmatrix}$$

The result for $T^{-\frac{1}{2}}\hat{\tilde{S}}_{[rT]}$ is given next. From (C.15) we know $\sum_{t=1}^{[rT]}\sum_{i=1}^{N}(z_{it}-\hat{z}_{it})\hat{u}_{it}=o_p(1)$. Similarly, it can be shown that

$$
\begin{aligned}
\sum_{t=1}^{[rT]}\sum_{j=1}^{N}(z_{jt}-\hat{z}_{jt})\hat{u}_{it} &= \sum_{j=1}^{N}\Big(\sum_{t=1}^{[rT]}(z_{jt}-\hat{z}_{jt})\mathbf{f}(t)'\Big)\Big(\sum_{s=1}^{T}\mathbf{f}(s)\mathbf{f}(s)'\Big)^{-1}\sum_{s=1}^{T}u_{is}\mathbf{f}(s)\\
&= \sum_{j=1}^{N}o_p(1)\cdot\Big(\sum_{s=1}^{T}\mathbf{f}(s)\mathbf{f}(s)'\Big)^{-1}\sum_{s=1}^{T}u_{is}\mathbf{f}(s)\xrightarrow{p}0
\end{aligned}
$$

Direct calculation gives

$$
\begin{aligned}
&T^{-\frac{1}{2}}\sum_{t=1}^{[rT]}\sum_{i=1}^{N}\tilde{z}_{it}\tilde{u}_{it}\\
=&T^{-\frac{1}{2}}\sum_{t=1}^{[rT]}\sum_{i=1}^{N}\tilde{z}_{it}(u_{it}-\hat{u}_{it}-\frac{1}{N}\sum_{j=1}^{N}(u_{jt}-\hat{u}_{jt}))=T^{-\frac{1}{2}}\sum_{t=1}^{[rT]}\sum_{i=1}^{N}\tilde{z}_{it}(u_{it}-\hat{u}_{it})\\
=&T^{-\frac{1}{2}}\sum_{t=1}^{[rT]}\sum_{i=1}^{N}\tilde{z}_{it}u_{it}-T^{-\frac{1}{2}}\sum_{t=1}^{[rT]}\sum_{i=1}^{N}(z_{it}-\hat{z}_{it}-\frac{1}{N}\sum_{j=1}^{N}(z_{jt}-\hat{z}_{jt}))\hat{u}_{it}\\
=&T^{-\frac{1}{2}}\sum_{t=1}^{[rT]}\sum_{i=1}^{N}\tilde{z}_{it}u_{it}-T^{-\frac{1}{2}}\sum_{t=1}^{[rT]}\sum_{i=1}^{N}(z_{it}-\hat{z}_{it})\hat{u}_{it}+T^{-\frac{1}{2}}\sum_{t=1}^{[rT]}\sum_{i=1}^{N}\frac{1}{N}\sum_{j=1}^{N}(z_{jt}-\hat{z}_{jt})\hat{u}_{it}\\
=&T^{-\frac{1}{2}}\sum_{t=1}^{[rT]}\sum_{i=1}^{N}\tilde{z}_{it}u_{it}+o_p(1)
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
T^{-\frac{1}{2}}\hat{\hat{S}}_{[rT]} &= T^{-\frac{1}{2}}\sum_{t=1}^{[rT]}\sum_{i=1}^{N}\tilde{x}_{it}\tilde{u}_{it} - (T^{-1}\sum_{t=1}^{[rT]}\sum_{i=1}^{N}\tilde{x}_{it}\tilde{x}_{it}')\sqrt{T}(\hat{\beta}-\beta) \\
&= \begin{bmatrix} T^{-\frac{1}{2}}\sum_{t=1}^{[rT]}\sum_{i=1}^{N}\widetilde{Treat}_i\widetilde{DU}_t\tilde{u}_{it} \\ T^{-\frac{1}{2}}\sum_{t=1}^{[rT]}\sum_{i=1}^{N}\tilde{z}_{it}u_{it} + o_p(1) \end{bmatrix} \\
&\quad - \begin{bmatrix} T^{-1}\sum_{t=1}^{[rT]}\sum_{i=1}^{N}(\widetilde{Treat}_i\widetilde{DU}_t)^2 & T^{-1}\sum_{t=1}^{[rT]}\sum_{i=1}^{N}\widetilde{Treat}_i\widetilde{DU}_t\tilde{z}_{it}' \\ T^{-1}\sum_{t=1}^{[rT]}\sum_{i=1}^{N}\tilde{z}_{it}\widetilde{Treat}_i\widetilde{DU}_t & T^{-1}\sum_{t=1}^{[rT]}\sum_{i=1}^{N}\tilde{z}_{it}\tilde{z}_{it}' \end{bmatrix}\sqrt{T}(\hat{\beta}-\beta) \\
&\Rightarrow \begin{bmatrix} \Lambda^{ex*}\Big[\int_0^r H^F(s,\lambda)dW^{ex*}(s) - \int_0^1 dW(s)F(s)'\Big(\int_0^1 F(s)F(s)'ds\Big)^{-1} \\ \cdot\int_0^r F(s)H^F(s,\lambda)ds\Big] \\ (\sum_{i=1}^{N}A_i^{ex})\Lambda^{ex}W^{ex}(r) \end{bmatrix} \\
&\quad - \begin{bmatrix} \tilde{G}\int_0^r H^F(s,\lambda)^2 ds & \mathbf{0} \\ \mathbf{0} & r\bar{Q} \end{bmatrix} \\
&\quad \cdot \begin{bmatrix} (\tilde{G}\int_0^1 H^F(r,\lambda)^2 dr)^{-1}(\Lambda^{ex*}\int_0^1 H^F(s,\lambda)dW^{ex*}(s) \\ \bar{Q}^{-1}(\sum_{i=1}^{N}A_i^{ex})\Lambda^{ex}W^{ex}(1) \end{bmatrix} \\
&= \begin{bmatrix} \Lambda^{ex*}Q^F(r,\lambda,W^{ex*}) \\ (\sum_{i=1}^{N}A_i^{ex})\Lambda^{ex}B^{ex}(r) \end{bmatrix} = \Lambda^{ex**}\begin{bmatrix} \mathbf{Q^F}(r,\lambda,W^{ex*}) \\ B^{ex}(r) \end{bmatrix} \qquad \square
\end{aligned}
$$

Appendix D

TABLES IN CHAPTER 2

Table D.1: 90% Asymptotic Critical Values for t_{DD} (Bartlett Kernel) Without Trend.

b =	0.02	0.04	0.06	0.08	0.1	0.12	0.14	0.16	0.18	0.2
$\lambda = 0.1$	1.506	1.728	1.953	2.148	2.325	2.485	2.624	2.744	2.864	2.975
0.2	1.380	1.476	1.571	1.663	1.752	1.843	1.940	2.024	2.107	2.185
0.3	1.335	1.390	1.449	1.506	1.569	1.629	1.689	1.751	1.808	1.873
0.4	1.322	1.360	1.409	1.454	1.499	1.545	1.594	1.645	1.699	1.747
0.5	1.325	1.370	1.415	1.458	1.506	1.547	1.599	1.647	1.697	1.750
0.6	1.326	1.374	1.411	1.457	1.501	1.556	1.606	1.658	1.712	1.768
0.7	1.342	1.402	1.463	1.526	1.586	1.649	1.714	1.774	1.838	1.899
0.8	1.377	1.469	1.570	1.663	1.753	1.845	1.932	2.022	2.107	2.186
0.9	1.505	1.732	1.953	2.143	2.318	2.472	2.611	2.745	2.862	2.970
b =	0.22	0.24	0.26	0.28	0.3	0.32	0.34	0.36	0.38	0.4
$\lambda = 0.1$	3.076	3.174	3.269	3.357	3.442	3.529	3.605	3.686	3.768	3.847
0.2	2.267	2.345	2.416	2.484	2.554	2.621	2.684	2.751	2.814	2.881
0.3	1.938	2.001	2.064	2.131	2.197	2.253	2.313	2.369	2.420	2.476
0.4	1.805	1.862	1.922	1.978	2.036	2.094	2.147	2.200	2.257	2.313
0.5	1.801	1.857	1.916	1.971	2.026	2.086	2.141	2.194	2.247	2.301
0.6	1.822	1.879	1.934	1.990	2.045	2.105	2.158	2.214	2.272	2.329
0.7	1.962	2.025	2.089	2.155	2.218	2.281	2.338	2.394	2.450	2.505
0.8	2.261	2.337	2.403	2.473	2.540	2.607	2.670	2.737	2.800	2.862
0.9	3.067	3.175	3.274	3.371	3.449	3.534	3.619	3.703	3.788	3.867

Continued on next page.

Table D.1 (cont'd)

b =	0.42	0.44	0.46	0.48	0.5	0.52	0.54	0.56	0.58	0.6
λ = 0.1	3.926	3.997	4.087	4.163	4.228	4.303	4.381	4.448	4.517	4.585
0.2	2.946	3.009	3.071	3.122	3.174	3.228	3.287	3.339	3.385	3.443
0.3	2.528	2.578	2.633	2.682	2.739	2.797	2.846	2.898	2.947	2.992
0.4	2.370	2.424	2.482	2.536	2.585	2.635	2.686	2.734	2.781	2.830
0.5	2.361	2.416	2.472	2.528	2.577	2.628	2.674	2.719	2.765	2.812
0.6	2.382	2.440	2.496	2.541	2.589	2.643	2.686	2.733	2.773	2.824
0.7	2.562	2.619	2.670	2.727	2.781	2.837	2.888	2.940	2.986	3.028
0.8	2.916	2.979	3.034	3.096	3.156	3.214	3.271	3.326	3.384	3.432
0.9	3.943	4.034	4.106	4.177	4.250	4.320	4.383	4.450	4.526	4.591
b =	0.62	0.64	0.66	0.68	0.7	0.72	0.74	0.76	0.78	0.8
λ = 0.1	4.656	4.725	4.796	4.859	4.923	4.996	5.065	5.130	5.191	5.255
0.2	3.490	3.546	3.602	3.656	3.711	3.756	3.817	3.862	3.911	3.962
0.3	3.038	3.088	3.130	3.174	3.220	3.262	3.308	3.347	3.389	3.426
0.4	2.876	2.914	2.954	2.994	3.033	3.074	3.114	3.152	3.189	3.223
0.5	2.852	2.896	2.942	2.983	3.025	3.062	3.104	3.145	3.182	3.226
0.6	2.869	2.912	2.961	3.000	3.041	3.078	3.116	3.155	3.199	3.237
0.7	3.073	3.120	3.164	3.209	3.252	3.292	3.334	3.381	3.426	3.468
0.8	3.486	3.537	3.588	3.630	3.688	3.742	3.792	3.845	3.899	3.947
0.9	4.661	4.732	4.800	4.873	4.944	5.012	5.070	5.136	5.199	5.262
b =	0.82	0.84	0.86	0.88	0.9	0.92	0.94	0.96	0.98	1.0
λ = 0.1	5.306	5.377	5.436	5.501	5.551	5.607	5.670	5.727	5.789	5.850
0.2	4.004	4.045	4.090	4.139	4.186	4.226	4.277	4.320	4.366	4.411
0.3	3.468	3.500	3.541	3.584	3.625	3.661	3.702	3.742	3.778	3.817
0.4	3.262	3.288	3.323	3.361	3.396	3.434	3.474	3.511	3.547	3.582

Continued on next page.

Table D.1 (cont'd)

0.5	3.267	3.298	3.337	3.376	3.411	3.448	3.484	3.521	3.560	3.597
0.6	3.277	3.318	3.354	3.384	3.420	3.457	3.493	3.531	3.567	3.605
0.7	3.507	3.544	3.584	3.630	3.674	3.712	3.748	3.788	3.827	3.865
0.8	3.989	4.039	4.082	4.124	4.169	4.205	4.259	4.305	4.349	4.393
0.9	5.319	5.388	5.449	5.522	5.590	5.646	5.687	5.754	5.812	5.872

Table D.2: 95% Asymptotic Critical Values for t_{DD} (Bartlett Kernel) Without Trend.

b =	0.02	0.04	0.06	0.08	0.1	0.12	0.14	0.16	0.18	0.2
$\lambda = 0.1$	1.980	2.313	2.618	2.873	3.104	3.302	3.476	3.642	3.803	3.952
0.2	1.775	1.915	2.052	2.190	2.314	2.448	2.571	2.686	2.794	2.902
0.3	1.720	1.801	1.883	1.975	2.057	2.145	2.236	2.324	2.407	2.498
0.4	1.710	1.773	1.833	1.900	1.971	2.036	2.111	2.188	2.268	2.345
0.5	1.712	1.766	1.831	1.902	1.965	2.032	2.102	2.172	2.253	2.324
0.6	1.704	1.767	1.839	1.904	1.969	2.044	2.120	2.193	2.265	2.346
0.7	1.719	1.810	1.893	1.993	2.070	2.160	2.254	2.345	2.436	2.525
0.8	1.788	1.922	2.066	2.195	2.325	2.456	2.569	2.686	2.797	2.907
0.9	1.983	2.325	2.621	2.890	3.126	3.341	3.522	3.678	3.830	3.986
b =	0.22	0.24	0.26	0.28	0.3	0.32	0.34	0.36	0.38	0.4
$\lambda = 0.1$	4.093	4.216	4.330	4.460	4.574	4.681	4.791	4.891	5.018	5.124
0.2	3.005	3.102	3.194	3.289	3.379	3.471	3.564	3.644	3.733	3.814
0.3	2.579	2.663	2.750	2.843	2.931	3.009	3.086	3.156	3.246	3.321
0.4	2.423	2.501	2.581	2.662	2.745	2.816	2.893	2.971	3.047	3.125
0.5	2.399	2.470	2.551	2.632	2.710	2.781	2.863	2.936	3.010	3.086
0.6	2.431	2.501	2.574	2.656	2.732	2.808	2.887	2.970	3.053	3.125
0.7	2.617	2.703	2.790	2.874	2.967	3.044	3.124	3.208	3.276	3.350
0.8	3.011	3.114	3.219	3.301	3.391	3.482	3.572	3.656	3.739	3.813
0.9	4.125	4.254	4.376	4.503	4.622	4.737	4.857	4.962	5.076	5.183
b =	0.42	0.44	0.46	0.48	0.5	0.52	0.54	0.56	0.58	0.6
$\lambda = 0.1$	5.230	5.338	5.455	5.564	5.668	5.762	5.855	5.946	6.050	6.138
0.2	3.899	3.974	4.053	4.134	4.199	4.278	4.355	4.434	4.516	4.581
0.3	3.386	3.457	3.530	3.617	3.694	3.766	3.838	3.902	3.965	4.026

Continued on next page.

Table D.2 (cont'd)

0.4	3.198	3.266	3.329	3.417	3.479	3.547	3.606	3.680	3.737	3.795
0.5	3.161	3.231	3.306	3.380	3.448	3.521	3.588	3.663	3.714	3.780
0.6	3.206	3.273	3.352	3.426	3.496	3.567	3.640	3.696	3.754	3.814
0.7	3.429	3.497	3.570	3.644	3.709	3.783	3.851	3.910	3.974	4.045
0.8	3.893	3.968	4.055	4.143	4.224	4.295	4.362	4.444	4.519	4.596
0.9	5.294	5.399	5.505	5.616	5.714	5.814	5.923	6.007	6.101	6.186

b =	0.62	0.64	0.66	0.68	0.7	0.72	0.74	0.76	0.78	0.8
λ = 0.1	6.219	6.327	6.419	6.507	6.593	6.669	6.757	6.849	6.944	7.032
0.2	4.647	4.723	4.787	4.878	4.950	5.017	5.095	5.150	5.214	5.267
0.3	4.084	4.144	4.207	4.264	4.322	4.368	4.417	4.477	4.533	4.577
0.4	3.859	3.917	3.980	4.034	4.094	4.134	4.184	4.226	4.279	4.337
0.5	3.835	3.893	3.945	3.989	4.045	4.096	4.138	4.186	4.235	4.290
0.6	3.870	3.926	3.987	4.052	4.090	4.143	4.190	4.235	4.282	4.332
0.7	4.121	4.187	4.239	4.289	4.353	4.415	4.462	4.511	4.566	4.622
0.8	4.681	4.755	4.814	4.895	4.970	5.041	5.115	5.179	5.244	5.310
0.9	6.281	6.370	6.467	6.551	6.639	6.743	6.840	6.925	7.012	7.106

b =	0.82	0.84	0.86	0.88	0.9	0.92	0.94	0.96	0.98	1.0
λ = 0.1	7.114	7.205	7.279	7.375	7.454	7.535	7.612	7.701	7.780	7.862
0.2	5.328	5.382	5.440	5.503	5.556	5.608	5.667	5.723	5.783	5.840
0.3	4.627	4.684	4.726	4.782	4.842	4.886	4.938	4.985	5.036	5.087
0.4	4.383	4.427	4.481	4.530	4.582	4.630	4.682	4.729	4.774	4.823
0.5	4.337	4.387	4.444	4.489	4.531	4.577	4.635	4.687	4.735	4.781
0.6	4.387	4.442	4.493	4.541	4.587	4.633	4.680	4.721	4.772	4.821
0.7	4.683	4.746	4.786	4.828	4.887	4.941	4.995	5.044	5.099	5.149
0.8	5.389	5.433	5.486	5.535	5.585	5.639	5.707	5.768	5.829	5.886

Continued on next page.

Table D.2 (cont'd)

0.9	7.189	7.274	7.349	7.432	7.502	7.590	7.660	7.732	7.810	7.890

Table D.3: 97.5% Asymptotic Critical Values for t_{DD} (Bartlett Kernel) Without Trend.

b =	0.02	0.04	0.06	0.08	0.1	0.12	0.14	0.16	0.18	0.2
λ = 0.1	2.440	2.861	3.245	3.560	3.835	4.075	4.289	4.490	4.674	4.852
0.2	2.132	2.328	2.508	2.684	2.850	2.987	3.137	3.281	3.421	3.555
0.3	2.054	2.165	2.289	2.409	2.532	2.642	2.764	2.873	2.973	3.088
0.4	2.037	2.128	2.220	2.320	2.407	2.499	2.601	2.708	2.799	2.901
0.5	2.056	2.130	2.214	2.286	2.375	2.469	2.566	2.669	2.766	2.865
0.6	2.040	2.120	2.206	2.296	2.401	2.490	2.577	2.675	2.792	2.902
0.7	2.064	2.186	2.300	2.427	2.530	2.641	2.765	2.878	2.982	3.098
0.8	2.140	2.325	2.506	2.687	2.868	3.016	3.163	3.311	3.451	3.594
0.9	2.413	2.862	3.235	3.547	3.835	4.086	4.320	4.530	4.720	4.900
b =	0.22	0.24	0.26	0.28	0.3	0.32	0.34	0.36	0.38	0.4
λ = 0.1	5.032	5.195	5.348	5.500	5.651	5.791	5.947	6.083	6.221	6.370
0.2	3.695	3.813	3.935	4.045	4.164	4.274	4.403	4.515	4.619	4.714
0.3	3.199	3.312	3.424	3.536	3.648	3.751	3.855	3.961	4.064	4.153
0.4	3.003	3.095	3.198	3.306	3.397	3.501	3.608	3.713	3.800	3.893
0.5	2.958	3.060	3.163	3.268	3.360	3.456	3.560	3.658	3.744	3.829
0.6	2.994	3.087	3.190	3.293	3.397	3.499	3.590	3.692	3.783	3.883
0.7	3.210	3.317	3.426	3.530	3.641	3.749	3.849	3.946	4.047	4.146
0.8	3.726	3.867	3.985	4.103	4.205	4.341	4.459	4.587	4.685	4.778
0.9	5.090	5.276	5.432	5.584	5.735	5.884	6.035	6.204	6.334	6.454
b =	0.42	0.44	0.46	0.48	0.5	0.52	0.54	0.56	0.58	0.6
λ = 0.1	6.509	6.636	6.773	6.896	7.035	7.150	7.256	7.362	7.483	7.610
0.2	4.811	4.909	5.026	5.109	5.200	5.285	5.364	5.474	5.567	5.658
0.3	4.251	4.333	4.414	4.502	4.589	4.671	4.769	4.842	4.941	5.021

Continued on next page.

Table D.3 (cont'd)

0.4	3.988	4.076	4.176	4.259	4.334	4.423	4.502	4.595	4.672	4.756
0.5	3.916	4.006	4.099	4.211	4.302	4.390	4.466	4.546	4.631	4.690
0.6	3.980	4.068	4.149	4.237	4.315	4.395	4.479	4.548	4.638	4.716
0.7	4.238	4.341	4.442	4.542	4.646	4.728	4.812	4.890	4.966	5.047
0.8	4.889	4.997	5.120	5.215	5.300	5.394	5.496	5.590	5.670	5.762
0.9	6.601	6.749	6.880	6.997	7.106	7.237	7.368	7.509	7.645	7.773

b =	0.62	0.64	0.66	0.68	0.7	0.72	0.74	0.76	0.78	0.8
λ = 0.1	7.725	7.852	7.981	8.079	8.213	8.338	8.429	8.539	8.648	8.741
0.2	5.749	5.833	5.923	5.992	6.080	6.167	6.246	6.324	6.412	6.480
0.3	5.098	5.191	5.264	5.343	5.420	5.481	5.546	5.595	5.675	5.732
0.4	4.822	4.885	4.942	5.007	5.081	5.147	5.216	5.280	5.348	5.410
0.5	4.770	4.846	4.913	4.972	5.029	5.084	5.159	5.213	5.277	5.350
0.6	4.772	4.854	4.926	5.000	5.069	5.119	5.181	5.226	5.294	5.353
0.7	5.115	5.205	5.277	5.356	5.424	5.475	5.551	5.624	5.684	5.751
0.8	5.859	5.944	6.011	6.111	6.194	6.293	6.368	6.441	6.526	6.608
0.9	7.882	8.009	8.122	8.249	8.368	8.485	8.576	8.692	8.771	8.896

b =	0.82	0.84	0.86	0.88	0.9	0.92	0.94	0.96	0.98	1.0
λ = 0.1	8.869	8.976	9.066	9.151	9.272	9.361	9.440	9.549	9.633	9.729
0.2	6.554	6.635	6.696	6.763	6.841	6.920	6.993	7.068	7.151	7.228
0.3	5.805	5.882	5.944	6.003	6.074	6.123	6.189	6.255	6.316	6.380
0.4	5.483	5.529	5.577	5.640	5.706	5.766	5.829	5.889	5.953	6.014
0.5	5.398	5.461	5.526	5.588	5.647	5.718	5.775	5.837	5.899	5.958
0.6	5.409	5.478	5.541	5.603	5.663	5.728	5.787	5.842	5.898	5.959
0.7	5.826	5.888	5.958	6.032	6.083	6.142	6.213	6.275	6.341	6.405
0.8	6.688	6.759	6.827	6.896	6.968	7.048	7.097	7.171	7.246	7.319

Continued on next page.

Table D.3 (cont'd)

0.9	8.994	9.109	9.215	9.317	9.415	9.506	9.602	9.696	9.790	9.881

Table D.4: 99% Asymptotic Critical Values for t_{DD} (Bartlett Kernel) Without Trend.

b =	0.02	0.04	0.06	0.08	0.1	0.12	0.14	0.16	0.18	0.2
λ = 0.1	2.952	3.546	3.998	4.375	4.716	5.015	5.287	5.556	5.809	6.073
0.2	2.586	2.838	3.092	3.305	3.522	3.725	3.906	4.103	4.259	4.424
0.3	2.479	2.649	2.799	2.959	3.117	3.297	3.443	3.604	3.754	3.884
0.4	2.451	2.563	2.690	2.808	2.950	3.066	3.213	3.357	3.492	3.629
0.5	2.446	2.537	2.646	2.747	2.871	3.003	3.144	3.266	3.391	3.505
0.6	2.438	2.546	2.674	2.794	2.924	3.059	3.212	3.324	3.462	3.597
0.7	2.474	2.630	2.793	2.959	3.107	3.243	3.389	3.530	3.689	3.844
0.8	2.588	2.860	3.090	3.312	3.527	3.741	3.936	4.136	4.286	4.472
0.9	2.962	3.569	4.053	4.436	4.782	5.101	5.387	5.629	5.884	6.113
b =	0.22	0.24	0.26	0.28	0.3	0.32	0.34	0.36	0.38	0.4
λ = 0.1	6.308	6.506	6.702	6.923	7.099	7.313	7.505	7.681	7.832	7.992
0.2	4.592	4.731	4.907	5.061	5.224	5.382	5.512	5.600	5.740	5.894
0.3	4.031	4.167	4.295	4.444	4.594	4.733	4.842	4.973	5.080	5.202
0.4	3.744	3.876	4.013	4.120	4.244	4.383	4.546	4.675	4.791	4.915
0.5	3.627	3.756	3.895	4.038	4.168	4.304	4.412	4.541	4.666	4.770
0.6	3.719	3.829	3.962	4.071	4.195	4.339	4.465	4.590	4.709	4.831
0.7	3.989	4.141	4.260	4.409	4.536	4.680	4.800	4.927	5.029	5.149
0.8	4.613	4.781	4.938	5.065	5.228	5.377	5.528	5.683	5.823	5.965
0.9	6.344	6.550	6.713	6.927	7.155	7.305	7.481	7.671	7.857	8.026
b =	0.42	0.44	0.46	0.48	0.5	0.52	0.54	0.56	0.58	0.6
λ = 0.1	8.148	8.336	8.522	8.659	8.783	8.923	9.078	9.188	9.352	9.508
0.2	6.020	6.144	6.273	6.383	6.522	6.639	6.747	6.857	6.977	7.087
0.3	5.324	5.444	5.559	5.664	5.766	5.863	5.931	6.080	6.169	6.275

Continued on next page.

Table D.4 (cont'd)

0.4	5.034	5.138	5.238	5.347	5.471	5.567	5.665	5.773	5.880	6.029
0.5	4.877	4.971	5.081	5.192	5.302	5.431	5.527	5.631	5.711	5.797
0.6	4.951	5.050	5.191	5.297	5.387	5.472	5.587	5.686	5.777	5.849
0.7	5.261	5.358	5.498	5.626	5.725	5.865	5.977	6.106	6.198	6.265
0.8	6.127	6.249	6.381	6.538	6.667	6.775	6.871	7.011	7.109	7.199
0.9	8.197	8.393	8.575	8.718	8.857	9.021	9.170	9.316	9.456	9.597
b =	0.62	0.64	0.66	0.68	0.7	0.72	0.74	0.76	0.78	0.8
λ = 0.1	9.700	9.871	9.996	10.115	10.288	10.405	10.606	10.750	10.861	11.013
0.2	7.204	7.293	7.406	7.515	7.598	7.717	7.818	7.932	8.024	8.128
0.3	6.391	6.467	6.546	6.659	6.780	6.870	6.975	7.039	7.100	7.172
0.4	6.081	6.146	6.222	6.348	6.416	6.499	6.558	6.646	6.723	6.806
0.5	5.873	6.002	6.102	6.182	6.283	6.368	6.428	6.488	6.551	6.621
0.6	5.940	6.019	6.105	6.194	6.280	6.359	6.460	6.538	6.623	6.685
0.7	6.357	6.463	6.548	6.637	6.768	6.848	6.933	7.018	7.089	7.171
0.8	7.308	7.413	7.516	7.628	7.739	7.872	7.985	8.124	8.191	8.264
0.9	9.750	9.872	10.025	10.184	10.305	10.430	10.620	10.755	10.895	11.003
b =	0.82	0.84	0.86	0.88	0.9	0.92	0.94	0.96	0.98	1.0
λ = 0.1	11.118	11.259	11.408	11.569	11.711	11.775	11.905	12.014	12.098	12.221
0.2	8.198	8.305	8.404	8.451	8.537	8.636	8.719	8.808	8.887	8.982
0.3	7.260	7.364	7.461	7.541	7.639	7.723	7.785	7.863	7.944	8.014
0.4	6.889	6.968	7.036	7.099	7.190	7.260	7.324	7.399	7.487	7.560
0.5	6.725	6.813	6.861	6.936	7.010	7.085	7.168	7.237	7.307	7.380
0.6	6.758	6.826	6.879	6.970	7.038	7.120	7.197	7.277	7.353	7.427
0.7	7.249	7.346	7.413	7.507	7.591	7.655	7.730	7.824	7.900	7.976
0.8	8.395	8.458	8.524	8.633	8.733	8.814	8.913	8.998	9.089	9.184

Continued on next page.

Table D.4 (cont'd)

0.9	11.170	11.242	11.364	11.519	11.671	11.737	11.831	11.946	12.111	12.236

Table D.5: 90% Asymptotic Critical Values for t_{DD} (Bartlett Kernel) With A Simple Trend.

b =	0.02	0.04	0.06	0.08	0.1	0.12	0.14	0.16	0.18	0.2
$\lambda = 0.1$	1.438	1.591	1.736	1.881	2.015	2.136	2.247	2.349	2.446	2.547
0.2	1.362	1.443	1.522	1.608	1.693	1.783	1.870	1.959	2.052	2.142
0.3	1.346	1.422	1.495	1.577	1.660	1.748	1.824	1.904	1.991	2.076
0.4	1.364	1.440	1.520	1.599	1.681	1.757	1.837	1.916	1.988	2.060
0.5	1.358	1.432	1.507	1.586	1.660	1.740	1.819	1.894	1.968	2.043
0.6	1.340	1.413	1.492	1.568	1.644	1.727	1.803	1.890	1.965	2.048
0.7	1.360	1.431	1.513	1.592	1.669	1.762	1.844	1.930	2.008	2.094
0.8	1.366	1.443	1.526	1.616	1.699	1.789	1.881	1.975	2.067	2.158
0.9	1.439	1.600	1.752	1.887	2.018	2.142	2.250	2.356	2.455	2.550
b =	0.22	0.24	0.26	0.28	0.3	0.32	0.34	0.36	0.38	0.4
$\lambda = 0.1$	2.638	2.724	2.802	2.886	2.977	3.055	3.139	3.203	3.269	3.340
0.2	2.223	2.302	2.376	2.443	2.515	2.581	2.647	2.715	2.769	2.822
0.3	2.156	2.232	2.308	2.377	2.448	2.512	2.575	2.633	2.700	2.753
0.4	2.139	2.213	2.280	2.344	2.395	2.450	2.506	2.558	2.607	2.647
0.5	2.112	2.168	2.223	2.280	2.327	2.383	2.426	2.465	2.505	2.552
0.6	2.115	2.185	2.250	2.308	2.366	2.424	2.475	2.522	2.566	2.607
0.7	2.171	2.250	2.329	2.398	2.462	2.529	2.592	2.651	2.706	2.761
0.8	2.240	2.321	2.401	2.478	2.545	2.616	2.674	2.736	2.793	2.849
0.9	2.651	2.742	2.838	2.920	3.002	3.078	3.152	3.226	3.306	3.375
b =	0.42	0.44	0.46	0.48	0.5	0.52	0.54	0.56	0.58	0.6
$\lambda = 0.1$	3.407	3.469	3.533	3.600	3.660	3.716	3.778	3.832	3.887	3.943
0.2	2.879	2.935	2.990	3.034	3.085	3.137	3.180	3.228	3.272	3.316
0.3	2.799	2.852	2.897	2.943	2.988	3.030	3.082	3.117	3.170	3.221

Continued on next page.

Table D.5 (cont'd)

0.4	2.686	2.731	2.779	2.826	2.873	2.912	2.964	3.009	3.055	3.096
0.5	2.596	2.632	2.664	2.704	2.739	2.786	2.832	2.879	2.922	2.973
0.6	2.652	2.697	2.746	2.795	2.844	2.886	2.935	2.977	3.019	3.061
0.7	2.810	2.855	2.904	2.950	2.996	3.040	3.083	3.132	3.182	3.226
0.8	2.900	2.948	3.003	3.054	3.103	3.151	3.197	3.249	3.293	3.338
0.9	3.439	3.506	3.575	3.639	3.698	3.761	3.819	3.871	3.921	3.982
b =	0.62	0.64	0.66	0.68	0.7	0.72	0.74	0.76	0.78	0.8
λ = 0.1	3.996	4.043	4.095	4.147	4.193	4.251	4.304	4.362	4.410	4.466
0.2	3.360	3.408	3.457	3.501	3.556	3.600	3.657	3.700	3.747	3.800
0.3	3.268	3.313	3.368	3.407	3.452	3.502	3.548	3.602	3.644	3.691
0.4	3.140	3.191	3.233	3.282	3.329	3.372	3.419	3.465	3.511	3.555
0.5	3.017	3.063	3.103	3.153	3.193	3.238	3.281	3.320	3.362	3.400
0.6	3.103	3.144	3.193	3.246	3.285	3.333	3.378	3.424	3.466	3.509
0.7	3.281	3.327	3.374	3.426	3.481	3.532	3.587	3.629	3.678	3.724
0.8	3.381	3.428	3.476	3.524	3.573	3.625	3.674	3.722	3.771	3.816
0.9	4.033	4.084	4.141	4.191	4.243	4.292	4.349	4.398	4.453	4.507
b =	0.82	0.84	0.86	0.88	0.9	0.92	0.94	0.96	0.98	1.0
λ = 0.1	4.524	4.576	4.629	4.685	4.732	4.779	4.842	4.894	4.942	4.992
0.2	3.847	3.894	3.934	3.980	4.025	4.068	4.112	4.158	4.201	4.242
0.3	3.738	3.785	3.834	3.879	3.928	3.972	4.020	4.063	4.104	4.145
0.4	3.596	3.638	3.677	3.719	3.761	3.801	3.841	3.882	3.924	3.965
0.5	3.442	3.483	3.522	3.562	3.599	3.635	3.674	3.717	3.755	3.792
0.6	3.550	3.593	3.636	3.677	3.717	3.757	3.795	3.832	3.874	3.913
0.7	3.771	3.810	3.854	3.905	3.949	3.993	4.036	4.079	4.120	4.163
0.8	3.856	3.908	3.953	3.992	4.043	4.083	4.129	4.176	4.216	4.262

Continued on next page.

Table D.5 (cont'd)

0.9	4.565	4.612	4.663	4.720	4.781	4.824	4.879	4.928	4.976	5.028

Table D.6: 95% Asymptotic Critical Values for t_{DD} (Bartlett Kernel) With A Simple Trend.

b =	0.02	0.04	0.06	0.08	0.1	0.12	0.14	0.16	0.18	0.2
$\lambda = 0.1$	1.877	2.095	2.306	2.497	2.679	2.840	2.986	3.143	3.279	3.408
0.2	1.755	1.864	1.976	2.102	2.225	2.349	2.467	2.601	2.734	2.859
0.3	1.739	1.843	1.953	2.066	2.181	2.294	2.414	2.526	2.646	2.757
0.4	1.755	1.873	1.980	2.101	2.213	2.325	2.435	2.549	2.661	2.749
0.5	1.745	1.856	1.961	2.081	2.199	2.305	2.409	2.524	2.619	2.713
0.6	1.735	1.844	1.965	2.073	2.190	2.306	2.416	2.533	2.636	2.734
0.7	1.749	1.857	1.971	2.092	2.203	2.323	2.433	2.556	2.675	2.792
0.8	1.748	1.866	1.989	2.115	2.247	2.371	2.497	2.629	2.760	2.873
0.9	1.874	2.091	2.301	2.489	2.665	2.832	2.982	3.142	3.278	3.414
b =	0.22	0.24	0.26	0.28	0.3	0.32	0.34	0.36	0.38	0.4
$\lambda = 0.1$	3.539	3.671	3.788	3.909	4.013	4.127	4.227	4.321	4.427	4.505
0.2	2.985	3.088	3.196	3.287	3.379	3.466	3.563	3.644	3.724	3.798
0.3	2.861	2.970	3.067	3.166	3.261	3.353	3.436	3.521	3.603	3.685
0.4	2.857	2.945	3.030	3.117	3.195	3.271	3.345	3.413	3.476	3.550
0.5	2.799	2.879	2.957	3.030	3.115	3.173	3.238	3.310	3.371	3.432
0.6	2.832	2.925	3.022	3.108	3.184	3.256	3.328	3.398	3.465	3.540
0.7	2.898	3.003	3.112	3.211	3.298	3.383	3.458	3.542	3.622	3.696
0.8	2.979	3.089	3.186	3.276	3.363	3.444	3.529	3.609	3.690	3.763
0.9	3.555	3.687	3.799	3.921	4.022	4.124	4.228	4.324	4.417	4.511
b =	0.42	0.44	0.46	0.48	0.5	0.52	0.54	0.56	0.58	0.6
$\lambda = 0.1$	4.601	4.678	4.757	4.842	4.931	5.012	5.081	5.153	5.228	5.301
0.2	3.870	3.929	3.992	4.055	4.119	4.186	4.256	4.328	4.397	4.465
0.3	3.759	3.821	3.877	3.948	4.019	4.083	4.152	4.220	4.282	4.343

Continued on next page.

Table D.6 (cont'd)

0.4	3.613	3.676	3.739	3.800	3.867	3.940	3.996	4.059	4.127	4.196
0.5	3.489	3.547	3.598	3.648	3.706	3.766	3.824	3.893	3.948	4.010
0.6	3.602	3.667	3.721	3.780	3.852	3.911	3.980	4.043	4.107	4.164
0.7	3.767	3.841	3.911	3.977	4.043	4.108	4.168	4.231	4.307	4.371
0.8	3.833	3.909	3.972	4.042	4.118	4.189	4.246	4.304	4.369	4.429
0.9	4.596	4.691	4.776	4.864	4.938	5.001	5.083	5.160	5.232	5.305
b =	0.62	0.64	0.66	0.68	0.7	0.72	0.74	0.76	0.78	0.8
λ = 0.1	5.374	5.447	5.525	5.601	5.668	5.738	5.800	5.874	5.952	6.027
0.2	4.534	4.591	4.662	4.722	4.782	4.851	4.914	4.984	5.044	5.114
0.3	4.401	4.470	4.541	4.604	4.679	4.742	4.815	4.886	4.945	5.011
0.4	4.257	4.324	4.377	4.441	4.501	4.560	4.615	4.675	4.733	4.788
0.5	4.072	4.136	4.193	4.249	4.304	4.361	4.410	4.465	4.523	4.577
0.6	4.233	4.301	4.356	4.413	4.471	4.533	4.592	4.648	4.707	4.764
0.7	4.439	4.511	4.582	4.654	4.715	4.785	4.859	4.927	4.985	5.045
0.8	4.505	4.571	4.641	4.701	4.767	4.819	4.890	4.949	5.020	5.087
0.9	5.392	5.468	5.543	5.617	5.686	5.753	5.820	5.892	5.978	6.049
b =	0.82	0.84	0.86	0.88	0.9	0.92	0.94	0.96	0.98	1.0
λ = 0.1	6.114	6.167	6.238	6.314	6.379	6.456	6.513	6.583	6.652	6.723
0.2	5.166	5.231	5.289	5.341	5.405	5.475	5.539	5.594	5.653	5.711
0.3	5.071	5.129	5.186	5.242	5.302	5.368	5.418	5.475	5.535	5.590
0.4	4.841	4.899	4.961	5.019	5.078	5.131	5.188	5.243	5.293	5.349
0.5	4.631	4.683	4.735	4.789	4.838	4.892	4.944	4.995	5.048	5.098
0.6	4.822	4.872	4.926	4.981	5.037	5.089	5.146	5.198	5.256	5.310
0.7	5.112	5.179	5.237	5.298	5.359	5.415	5.480	5.537	5.591	5.649
0.8	5.150	5.208	5.281	5.337	5.396	5.458	5.512	5.567	5.624	5.681

Continued on next page.

Table D.6 (cont'd)

0.9	6.128	6.176	6.261	6.317	6.390	6.468	6.531	6.589	6.651	6.722

Table D.7: 97.5% Asymptotic Critical Values for t_{DD} (Bartlett Kernel) With A Simple Trend.

b =	0.02	0.04	0.06	0.08	0.1	0.12	0.14	0.16	0.18	0.2
λ = 0.1	2.289	2.577	2.845	3.074	3.307	3.527	3.714	3.906	4.067	4.240
0.2	2.121	2.269	2.420	2.587	2.756	2.912	3.060	3.227	3.386	3.538
0.3	2.095	2.243	2.378	2.533	2.677	2.827	2.972	3.127	3.279	3.419
0.4	2.117	2.252	2.393	2.551	2.696	2.852	3.009	3.162	3.301	3.421
0.5	2.073	2.227	2.387	2.535	2.676	2.831	2.978	3.124	3.233	3.351
0.6	2.095	2.239	2.388	2.547	2.690	2.837	2.983	3.127	3.259	3.395
0.7	2.104	2.251	2.394	2.541	2.683	2.834	2.971	3.125	3.270	3.414
0.8	2.111	2.267	2.423	2.591	2.752	2.922	3.081	3.229	3.365	3.516
0.9	2.244	2.538	2.798	3.053	3.298	3.500	3.689	3.863	4.031	4.205
b =	0.22	0.24	0.26	0.28	0.3	0.32	0.34	0.36	0.38	0.4
λ = 0.1	4.403	4.559	4.714	4.853	5.003	5.132	5.262	5.396	5.534	5.659
0.2	3.666	3.806	3.949	4.067	4.180	4.270	4.378	4.486	4.580	4.666
0.3	3.546	3.688	3.810	3.934	4.059	4.162	4.266	4.360	4.456	4.547
0.4	3.533	3.635	3.758	3.859	3.956	4.062	4.159	4.248	4.325	4.403
0.5	3.455	3.565	3.676	3.771	3.849	3.938	4.014	4.098	4.185	4.257
0.6	3.536	3.644	3.748	3.873	3.975	4.080	4.173	4.263	4.343	4.435
0.7	3.546	3.676	3.804	3.925	4.069	4.166	4.283	4.381	4.480	4.579
0.8	3.679	3.815	3.941	4.068	4.201	4.306	4.398	4.520	4.619	4.720
0.9	4.374	4.526	4.667	4.807	4.946	5.088	5.219	5.367	5.471	5.591
b =	0.42	0.44	0.46	0.48	0.5	0.52	0.54	0.56	0.58	0.6
λ = 0.1	5.778	5.872	5.976	6.092	6.187	6.271	6.374	6.467	6.576	6.666
0.2	4.754	4.852	4.942	5.044	5.133	5.219	5.281	5.361	5.454	5.536
0.3	4.648	4.747	4.838	4.911	4.979	5.071	5.158	5.244	5.331	5.401

Continued on next page.

Table D.7 (cont'd)

0.4	4.491	4.584	4.658	4.753	4.844	4.920	5.003	5.082	5.173	5.250
0.5	4.339	4.403	4.474	4.542	4.608	4.688	4.770	4.854	4.929	5.004
0.6	4.506	4.592	4.659	4.750	4.828	4.901	4.978	5.058	5.156	5.241
0.7	4.676	4.762	4.859	4.952	5.032	5.106	5.185	5.251	5.328	5.413
0.8	4.810	4.909	5.002	5.094	5.186	5.277	5.338	5.422	5.502	5.578
0.9	5.695	5.813	5.910	6.011	6.113	6.213	6.297	6.382	6.475	6.570

b =	0.62	0.64	0.66	0.68	0.7	0.72	0.74	0.76	0.78	0.8
λ = 0.1	6.748	6.845	6.947	7.041	7.136	7.229	7.336	7.433	7.519	7.618
0.2	5.606	5.679	5.759	5.833	5.932	6.020	6.117	6.204	6.297	6.365
0.3	5.492	5.567	5.655	5.749	5.833	5.917	5.998	6.075	6.143	6.208
0.4	5.338	5.409	5.483	5.555	5.650	5.718	5.806	5.886	5.949	6.018
0.5	5.079	5.153	5.234	5.320	5.393	5.468	5.535	5.611	5.684	5.750
0.6	5.334	5.407	5.466	5.545	5.629	5.689	5.765	5.829	5.900	5.965
0.7	5.509	5.601	5.686	5.769	5.866	5.958	6.037	6.117	6.186	6.270
0.8	5.669	5.756	5.838	5.926	6.018	6.078	6.146	6.220	6.314	6.399
0.9	6.664	6.759	6.850	6.962	7.052	7.128	7.237	7.332	7.421	7.520

b =	0.82	0.84	0.86	0.88	0.9	0.92	0.94	0.96	0.98	1.0
λ = 0.1	7.698	7.786	7.894	7.967	8.062	8.165	8.261	8.327	8.414	8.500
0.2	6.438	6.527	6.602	6.661	6.719	6.788	6.861	6.933	7.005	7.068
0.3	6.288	6.361	6.434	6.508	6.578	6.653	6.720	6.795	6.872	6.941
0.4	6.088	6.154	6.234	6.309	6.376	6.448	6.502	6.577	6.642	6.708
0.5	5.824	5.890	5.946	6.015	6.072	6.141	6.204	6.263	6.327	6.395
0.6	6.039	6.112	6.182	6.255	6.317	6.388	6.453	6.522	6.588	6.657
0.7	6.332	6.417	6.488	6.553	6.634	6.701	6.774	6.848	6.919	6.989
0.8	6.485	6.570	6.618	6.702	6.784	6.854	6.926	6.997	7.060	7.134

Continued on next page.

Table D.7 (cont'd)

0.9	7.586	7.675	7.783	7.888	7.971	8.052	8.116	8.211	8.296	8.382

Table D.8: 99% Asymptotic Critical Values for t_{DD} (Bartlett Kernel) With A Simple Trend.

b =	0.02	0.04	0.06	0.08	0.1	0.12	0.14	0.16	0.18	0.2
$\lambda = 0.1$	2.762	3.157	3.518	3.795	4.071	4.369	4.605	4.852	5.075	5.293
0.2	2.528	2.742	2.932	3.158	3.344	3.555	3.763	3.946	4.157	4.393
0.3	2.529	2.734	2.929	3.125	3.316	3.529	3.731	3.923	4.125	4.282
0.4	2.511	2.704	2.912	3.108	3.286	3.502	3.697	3.895	4.052	4.243
0.5	2.482	2.688	2.882	3.080	3.275	3.454	3.641	3.818	3.964	4.116
0.6	2.532	2.728	2.924	3.133	3.289	3.505	3.705	3.900	4.098	4.230
0.7	2.490	2.664	2.851	3.052	3.261	3.469	3.637	3.835	4.029	4.229
0.8	2.533	2.753	2.969	3.171	3.389	3.587	3.804	4.016	4.211	4.419
0.9	2.746	3.122	3.459	3.774	4.051	4.340	4.591	4.831	5.059	5.254
b =	0.22	0.24	0.26	0.28	0.3	0.32	0.34	0.36	0.38	0.4
$\lambda = 0.1$	5.533	5.737	5.915	6.041	6.262	6.423	6.589	6.740	6.871	7.052
0.2	4.567	4.726	4.858	5.015	5.132	5.290	5.442	5.567	5.686	5.818
0.3	4.462	4.619	4.773	4.924	5.068	5.230	5.381	5.504	5.586	5.697
0.4	4.399	4.523	4.643	4.790	4.905	5.011	5.122	5.252	5.370	5.498
0.5	4.262	4.398	4.517	4.650	4.770	4.878	5.007	5.102	5.222	5.322
0.6	4.384	4.539	4.682	4.839	4.940	5.062	5.180	5.293	5.396	5.509
0.7	4.408	4.568	4.703	4.868	5.011	5.154	5.305	5.429	5.554	5.662
0.8	4.592	4.764	4.926	5.090	5.236	5.345	5.477	5.603	5.745	5.871
0.9	5.459	5.657	5.835	5.977	6.161	6.320	6.493	6.669	6.818	6.975
b =	0.42	0.44	0.46	0.48	0.5	0.52	0.54	0.56	0.58	0.6
$\lambda = 0.1$	7.181	7.339	7.503	7.654	7.775	7.914	8.041	8.159	8.290	8.429
0.2	5.908	6.025	6.129	6.239	6.371	6.459	6.584	6.709	6.814	6.928
0.3	5.839	5.944	6.060	6.185	6.288	6.386	6.525	6.655	6.774	6.868

Continued on next page.

Table D.8 (cont'd)

0.4	5.573	5.669	5.788	5.875	5.982	6.086	6.214	6.333	6.440	6.532
0.5	5.427	5.522	5.634	5.733	5.817	5.937	6.026	6.147	6.264	6.362
0.6	5.616	5.737	5.844	5.992	6.098	6.207	6.290	6.399	6.499	6.619
0.7	5.777	5.899	6.020	6.138	6.254	6.360	6.454	6.566	6.668	6.796
0.8	6.016	6.111	6.208	6.311	6.442	6.556	6.690	6.818	6.911	7.036
0.9	7.096	7.230	7.370	7.509	7.647	7.748	7.872	7.989	8.112	8.230

b =	0.62	0.64	0.66	0.68	0.7	0.72	0.74	0.76	0.78	0.8
$\lambda = 0.1$	8.581	8.752	8.829	8.933	9.063	9.148	9.280	9.402	9.548	9.642
0.2	7.011	7.092	7.204	7.313	7.419	7.550	7.618	7.713	7.777	7.911
0.3	6.950	7.087	7.186	7.332	7.422	7.559	7.653	7.751	7.827	7.902
0.4	6.623	6.716	6.826	6.927	7.016	7.103	7.213	7.281	7.375	7.465
0.5	6.467	6.586	6.647	6.750	6.826	6.907	7.017	7.103	7.189	7.283
0.6	6.733	6.813	6.929	6.991	7.079	7.196	7.282	7.390	7.485	7.551
0.7	6.908	7.017	7.123	7.237	7.339	7.461	7.552	7.643	7.745	7.849
0.8	7.146	7.228	7.361	7.451	7.561	7.639	7.757	7.860	7.973	8.080
0.9	8.347	8.470	8.591	8.704	8.812	8.930	9.020	9.114	9.243	9.348

b =	0.82	0.84	0.86	0.88	0.9	0.92	0.94	0.96	0.98	1.0
$\lambda = 0.1$	9.731	9.861	9.969	10.072	10.168	10.301	10.360	10.475	10.600	10.714
0.2	8.036	8.144	8.239	8.328	8.431	8.519	8.606	8.691	8.770	8.859
0.3	8.010	8.113	8.197	8.291	8.391	8.495	8.601	8.697	8.787	8.877
0.4	7.543	7.631	7.705	7.807	7.900	7.989	8.075	8.160	8.253	8.338
0.5	7.372	7.458	7.551	7.636	7.711	7.789	7.888	7.956	8.059	8.144
0.6	7.646	7.747	7.841	7.938	8.028	8.111	8.193	8.274	8.355	8.440
0.7	7.938	8.057	8.138	8.217	8.302	8.389	8.485	8.573	8.659	8.748
0.8	8.176	8.261	8.344	8.441	8.571	8.646	8.766	8.826	8.901	8.984

Continued on next page.

Table D.8 (cont'd)

0.9	9.467	9.583	9.741	9.874	9.963	10.051	10.104	10.208	10.311	10.424

Table D.9: Null Rejection Probabilities, 5% level, t_{DD} (Bartlett Kernel). No trend or additional regressors. $\lambda = .5$, $k = .5$. $AR(1)$ error. Two-Tailed Test of $H_0 : \beta_3 = 0$.

			$N(0,1)$ CV						Adjusted Fixed-b CV					
			t_{DK}, values of b						t_{DK}, values of b					
N,T	ρ	t_{clus}	.02	.06	.1	.4	.7	1.0	.02	.06	.1	.4	.7	1.0
10,10	.0	.102	.127	.127	.127	.276	.397	.465	.111	.088	.070	.065	.067	.068
	.3	.102	.221	.221	.221	.367	.470	.539	.202	.173	.153	.120	.114	.115
	.6	.102	.347	.347	.347	.470	.565	.617	.328	.298	.260	.189	.184	.178
	.9	.102	.503	.503	.503	.572	.659	.709	.485	.451	.423	.278	.275	.281
10,50	.0	.105	.060	.076	.098	.272	.401	.469	.049	.040	.044	.041	.044	.041
	.3	.104	.167	.127	.133	.300	.422	.488	.147	.084	.067	.056	.054	.056
	.6	.102	.344	.225	.207	.342	.460	.533	.326	.172	.127	.088	.090	.082
	.9	.101	.654	.503	.446	.508	.604	.651	.640	.447	.347	.228	.218	.217
10,250	.0	.093	.049	.068	.096	.254	.371	.443	.039	.040	.044	.046	.044	.039
	.3	.091	.070	.078	.104	.262	.381	.448	.054	.048	.048	.050	.047	.046
	.6	.089	.123	.098	.116	.269	.386	.454	.104	.060	.066	.060	.054	.051
	.9	.087	.378	.216	.194	.332	.442	.515	.354	.170	.131	.098	.092	.091
50,10	.0	.056	.113	.113	.113	.273	.381	.447	.096	.080	.068	.061	.060	.060
	.3	.057	.213	.213	.213	.354	.472	.537	.195	.165	.142	.113	.107	.106
	.6	.062	.363	.363	.363	.479	.571	.626	.342	.304	.267	.185	.186	.181
	.9	.056	.508	.508	.508	.586	.658	.704	.489	.453	.420	.277	.281	.282
50,50	.0	.060	.068	.085	.112	.269	.395	.466	.053	.051	.054	.052	.054	.050
	.3	.059	.176	.136	.146	.294	.416	.488	.156	.094	.076	.069	.066	.067
	.6	.058	.353	.227	.211	.348	.466	.535	.330	.181	.137	.098	.092	.093
	.9	.057	.640	.498	.443	.506	.593	.647	.626	.452	.356	.225	.216	.214

Continued on next page.

Table D.9 (cont'd)

			$N(0,1)$ CV						Adjusted Fixed-b CV					
			t_{DK}, values of b						t_{DK}, values of b					
N,T	ρ	t_{clus}	.02	.06	.1	.4	.7	1.0	.02	.06	.1	.4	.7	1.0
50,250	.0	.056	.054	.076	.096	.247	.363	.435	.044	.045	.044	.048	.047	.042
	.3	.056	.079	.085	.103	.253	.366	.440	.062	.053	.050	.049	.049	.045
	.6	.057	.125	.102	.117	.260	.372	.451	.108	.066	.058	.057	.058	.054
	.9	.056	.370	.224	.200	.320	.435	.510	.345	.172	.125	.092	.091	.088
250,10	.0	.053	.112	.112	.112	.278	.394	.459	.102	.082	.064	.061	.060	.058
	.3	.055	.216	.216	.216	.356	.464	.530	.198	.168	.144	.114	.111	.108
	.6	.056	.352	.352	.352	.449	.543	.602	.330	.295	.266	.195	.194	.192
	.9	.050	.508	.508	.508	.568	.656	.708	.486	.457	.417	.271	.266	.265
250,50	.0	.057	.065	.083	.101	.251	.375	.445	.050	.049	.050	.046	.047	.046
	.3	.058	.164	.126	.135	.278	.390	.473	.147	.085	.072	.064	.064	.062
	.6	.054	.337	.212	.195	.326	.442	.517	.316	.168	.127	.092	.090	.092
	.9	.051	.654	.494	.438	.508	.599	.650	.638	.440	.345	.224	.212	.211
250,250	.0	.048	.053	.074	.093	.257	.379	.455	.042	.045	.049	.044	.046	.048
	.3	.046	.071	.081	.097	.264	.386	.459	.060	.050	.051	.048	.049	.050
	.6	.048	.119	.099	.110	.274	.388	.470	.103	.063	.064	.052	.054	.054
	.9	.047	.381	.229	.204	.335	.448	.523	.362	.171	.126	.091	.093	.091

Table D.10: Null Rejection Probabilities, 5% level, t_{DD} (Bartlett Kernel). No trend or additional regressors. $\lambda = .5$, $k = .5$. MA(2) spatial correlation in cross-section. $\theta = 0.5$. Two-Tailed Test of $H_0 : \beta_3 = 0$.

			$N(0,1)$ CV						Adjusted Fixed-b CV					
			t_{DK}, values of b						t_{DK}, values of b					
N,T	ρ	t_{clus}	.02	.06	.1	.4	.7	1.0	.02	.06	.1	.4	.7	1.0
9,10	.0	.339	.110	.110	.110	.280	.396	.473	.101	.082	.069	.072	.070	.068
	.3	.339	.221	.221	.221	.368	.471	.537	.200	.175	.152	.123	.122	.123
	.6	.346	.361	.361	.361	.456	.556	.613	.341	.310	.276	.202	.199	.197
	.9	.334	.484	.484	.484	.556	.654	.702	.470	.430	.402	.271	.267	.263
9,50	.0	.340	.064	.076	.094	.255	.366	.441	.051	.047	.045	.050	.045	.042
	.3	.337	.165	.118	.127	.274	.392	.461	.148	.080	.070	.066	.064	.063
	.6	.337	.333	.215	.194	.327	.433	.503	.310	.170	.127	.093	.089	.091
	.9	.342	.644	.484	.428	.500	.588	.644	.628	.430	.339	.220	.219	.210
9,250	.0	.368	.059	.072	.094	.262	.386	.460	.050	.048	.048	.052	.046	.048
	.3	.368	.078	.081	.099	.270	.390	.467	.064	.052	.050	.053	.048	.050
	.6	.369	.126	.099	.112	.280	.401	.472	.111	.063	.060	.056	.053	.054
	.9	.366	.390	.232	.206	.343	.456	.527	.370	.178	.129	.092	.087	.088
49,10	.0	.577	.108	.108	.108	.274	.400	.482	.094	.072	.056	.055	.057	.053
	.3	.568	.219	.219	.219	.370	.489	.553	.200	.165	.134	.104	.101	.099
	.6	.566	.342	.342	.342	.472	.574	.635	.318	.288	.258	.186	.181	.180
	.9	.562	.508	.508	.508	.574	.659	.704	.490	.458	.426	.278	.277	.276
49,50	.0	.565	.057	.073	.095	.260	.380	.455	.049	.045	.048	.050	.051	.050
	.3	.557	.162	.114	.130	.292	.409	.478	.146	.079	.066	.065	.064	.064
	.6	.553	.341	.218	.199	.342	.449	.519	.319	.171	.122	.094	.092	.092

Continued on next page.

Table D.10 (cont'd)

			$N(0,1)$ CV						Adjusted Fixed-b CV					
			t_{DK}, values of b						t_{DK}, values of b					
N,T	ρ	t_{clus}	.02	.06	.1	.4	.7	1.0	.02	.06	.1	.4	.7	1.0
	.9	.566	.654	.501	.441	.516	.603	.656	.642	.445	.349	.219	.209	.204
49,250	.0	.578	.056	.072	.092	.268	.379	.452	.047	.045	.045	.050	.050	.050
	.3	.572	.077	.081	.098	.273	.384	.458	.064	.049	.050	.054	.053	.050
	.6	.572	.129	.100	.118	.280	.394	.465	.113	.064	.060	.060	.058	.057
	.9	.580	.387	.226	.202	.333	.455	.524	.361	.178	.125	.096	.092	.093
256,10	.0	.612	.125	.125	.125	.272	.385	.460	.110	.090	.074	.070	.068	.069
	.3	.619	.222	.222	.222	.350	.465	.528	.200	.171	.149	.114	.114	.113
	.6	.621	.350	.350	.350	.454	.557	.622	.328	.296	.262	.190	.185	.182
	.9	.636	.508	.508	.508	.569	.662	.712	.488	.451	.422	.268	.271	.266
256,50	.0	.639	.060	.072	.092	.269	.387	.469	.047	.044	.043	.040	.042	.042
	.3	.635	.174	.123	.129	.302	.415	.489	.151	.081	.068	.058	.060	.052
	.6	.631	.370	.232	.208	.348	.464	.532	.344	.178	.131	.096	.092	.092
	.9	.635	.662	.498	.441	.503	.600	.656	.640	.446	.358	.228	.217	.219
256,250	.0	.623	.050	.074	.093	.252	.384	.460	.038	.039	.045	.051	.049	.050
	.3	.625	.071	.082	.100	.259	.387	.464	.058	.048	.051	.053	.053	.055
	.6	.626	.125	.097	.110	.270	.395	.468	.104	.062	.059	.059	.058	.058
	.9	.624	.373	.216	.193	.333	.437	.510	.350	.165	.126	.099	.093	.094

Table D.11: Null Rejection Probabilities, 5% level, t_{DD} (Bartlett Kernel). No trend or additional regressors. Time dummies. $\lambda = .5$, $k = .5$. $AR(1)$ error. Two-Tailed Test of $H_0 : \beta_3 = 0$.

			$N(0,1)$ CV						Adjusted Fixed-b CV					
			t_{DK}, values of b						t_{DK}, values of b					
N,T	ρ	t_{clus}	.02	.06	.1	.4	.7	1.0	.02	.06	.1	.4	.7	1.0
10,10	.0	.102	.127	.127	.127	.276	.397	.465	.111	.088	.070	.065	.067	.068
	.3	.102	.221	.221	.221	.367	.470	.539	.202	.173	.153	.120	.114	.115
	.6	.102	.347	.347	.347	.470	.565	.617	.328	.298	.260	.189	.184	.178
	.9	.102	.503	.503	.503	.572	.659	.709	.485	.451	.423	.278	.275	.281
10,50	.0	.105	.060	.076	.098	.272	.401	.469	.049	.040	.044	.041	.044	.041
	.3	.104	.167	.127	.133	.300	.422	.488	.147	.084	.067	.056	.054	.056
	.6	.102	.344	.225	.207	.342	.460	.533	.326	.172	.127	.088	.090	.082
	.9	.101	.654	.503	.446	.508	.604	.651	.640	.447	.347	.228	.218	.217
10,250	.0	.093	.049	.068	.096	.254	.371	.443	.039	.040	.044	.046	.044	.039
	.3	.091	.070	.078	.104	.262	.381	.448	.054	.048	.048	.050	.047	.046
	.6	.089	.123	.098	.116	.269	.386	.454	.104	.060	.066	.060	.054	.051
	.9	.087	.378	.216	.194	.332	.442	.515	.354	.170	.131	.098	.092	.091
50,10	.0	.056	.113	.113	.113	.273	.381	.447	.096	.080	.068	.061	.060	.060
	.3	.057	.213	.213	.213	.354	.472	.537	.195	.165	.142	.113	.107	.106
	.6	.062	.363	.363	.363	.479	.571	.626	.342	.304	.267	.185	.186	.181
	.9	.056	.508	.508	.508	.586	.658	.704	.489	.453	.420	.277	.281	.282
50,50	.0	.060	.068	.085	.112	.269	.395	.466	.053	.051	.054	.052	.054	.050
	.3	.059	.176	.136	.146	.294	.416	.488	.156	.094	.076	.069	.066	.067
	.6	.058	.353	.227	.211	.348	.466	.535	.330	.181	.137	.098	.092	.093
	.9	.057	.640	.498	.443	.506	.593	.647	.626	.452	.356	.225	.216	.214

Continued on next page.

Table D.11 (cont'd)

			$N(0,1)$ CV						Adjusted Fixed-b CV					
			t_{DK}, values of b						t_{DK}, values of b					
N,T	ρ	t_{clus}	.02	.06	.1	.4	.7	1.0	.02	.06	.1	.4	.7	1.0
50,250	.0	.056	.054	.076	.096	.247	.363	.435	.044	.045	.044	.048	.047	.042
	.3	.056	.079	.085	.103	.253	.366	.440	.062	.053	.050	.049	.049	.045
	.6	.057	.125	.102	.117	.260	.372	.451	.108	.066	.058	.057	.058	.054
	.9	.056	.370	.224	.200	.320	.435	.510	.345	.172	.125	.092	.091	.088
250,10	.0	.053	.112	.112	.112	.278	.394	.459	.102	.082	.064	.061	.060	.058
	.3	.055	.216	.216	.216	.356	.464	.530	.198	.168	.144	.114	.111	.108
	.6	.056	.352	.352	.352	.449	.543	.602	.330	.295	.266	.195	.194	.192
	.9	.050	.508	.508	.508	.568	.656	.708	.486	.457	.417	.271	.266	.265
250,50	.0	.057	.065	.083	.101	.251	.375	.445	.050	.049	.050	.046	.047	.046
	.3	.058	.164	.126	.135	.278	.390	.473	.147	.085	.072	.064	.064	.062
	.6	.054	.337	.212	.195	.326	.442	.517	.316	.168	.127	.092	.090	.092
	.9	.051	.654	.494	.438	.508	.599	.650	.638	.440	.345	.224	.212	.211
250,250	.0	.048	.053	.074	.093	.257	.379	.455	.042	.045	.049	.044	.046	.048
	.3	.046	.071	.081	.097	.264	.386	.459	.060	.050	.051	.048	.049	.050
	.6	.048	.119	.099	.110	.274	.388	.470	.103	.063	.064	.052	.054	.054
	.9	.047	.381	.229	.204	.335	.448	.523	.362	.171	.126	.091	.093	.091

Table D.12: Null Rejection Probabilities, 5% level, t_{DD} (Bartlett Kernel). Trend. No additional regressors. $\lambda = .5$, $k = .5$. $AR(1)$ errors. Two-Tailed Test of $H_0 : \beta_3 = 0$.

			$N(0,1)$ CV						Adjusted Fixed-b CV					
			t_{DK}, values of b						t_{DK}, values of b					
N,T	ρ	t_{clus}	.02	.06	.1	.4	.7	1.0	.02	.06	.1	.4	.7	1.0
10,10	.0	.092	.186	.186	.186	.331	.414	.479	.168	.122	.091	.081	.078	.079
	.3	.098	.215	.215	.215	.354	.432	.494	.192	.140	.104	.085	.080	.080
	.6	.089	.212	.212	.212	.351	.426	.489	.192	.143	.116	.087	.084	.085
	.9	.090	.210	.210	.210	.328	.386	.447	.188	.144	.105	.070	.069	.070
10,50	.0	.100	.070	.102	.140	.312	.401	.477	.056	.057	.056	.049	.051	.053
	.3	.100	.163	.140	.161	.333	.423	.486	.139	.086	.076	.067	.064	.062
	.6	.102	.315	.211	.212	.365	.444	.512	.282	.134	.104	.080	.084	.083
	.9	.110	.495	.336	.300	.401	.470	.538	.471	.243	.152	.098	.095	.094
10,250	.0	.102	.068	.096	.133	.307	.392	.460	.056	.048	.050	.050	.052	.052
	.3	.102	.087	.106	.138	.308	.398	.462	.073	.056	.052	.050	.052	.052
	.6	.107	.135	.124	.155	.322	.408	.476	.116	.066	.060	.055	.058	.057
	.9	.095	.349	.220	.215	.361	.448	.522	.326	.135	.100	.083	.081	.082
50,10	.0	.053	.180	.180	.180	.328	.406	.472	.160	.119	.094	.074	.064	.065
	.3	.054	.207	.207	.207	.341	.426	.487	.190	.147	.116	.089	.086	.087
	.6	.057	.220	.220	.220	.340	.417	.476	.201	.155	.120	.088	.086	.086
	.9	.060	.219	.219	.219	.338	.400	.453	.196	.149	.112	.077	.072	.072
50,50	.0	.063	.077	.108	.142	.303	.406	.475	.066	.057	.060	.059	.058	.059
	.3	.063	.165	.146	.170	.328	.422	.488	.142	.085	.075	.067	.069	.071
	.6	.066	.314	.226	.222	.364	.450	.515	.286	.137	.099	.080	.080	.081
	.9	.058	.497	.333	.288	.399	.475	.539	.472	.238	.146	.098	.096	.097

Continued on next page.

Table D.12 (cont'd)

			$N(0,1)$ CV						Adjusted Fixed-b CV					
			t_{DK}, values of b						t_{DK}, values of b					
N,T	ρ	t_{clus}	.02	.06	.1	.4	.7	1.0	.02	.06	.1	.4	.7	1.0
50,250	.0	.058	.069	.106	.138	.316	.402	.476	.055	.055	.054	.048	.051	.048
	.3	.058	.094	.119	.144	.322	.410	.475	.076	.060	.059	.053	.054	.052
	.6	.054	.143	.131	.156	.330	.419	.484	.121	.072	.067	.058	.059	.058
	.9	.056	.346	.223	.212	.356	.441	.511	.324	.138	.098	.078	.073	.075
250,10	.0	.054	.200	.200	.200	.356	.434	.488	.177	.131	.098	.085	.083	.082
	.3	.057	.226	.226	.226	.363	.442	.512	.205	.158	.123	.095	.091	.091
	.6	.055	.228	.228	.228	.360	.439	.502	.209	.159	.125	.090	.087	.086
	.9	.050	.214	.214	.214	.335	.408	.470	.189	.144	.112	.078	.073	.070
250,50	.0	.052	.077	.112	.145	.318	.406	.473	.062	.060	.052	.055	.056	.053
	.3	.055	.168	.152	.176	.340	.426	.490	.150	.088	.076	.062	.064	.063
	.6	.051	.312	.212	.214	.365	.450	.526	.284	.137	.105	.081	.079	.080
	.9	.044	.494	.329	.291	.390	.472	.538	.468	.228	.146	.097	.095	.095
250,250	.0	.048	.068	.105	.141	.312	.414	.486	.053	.055	.055	.057	.052	.052
	.3	.051	.090	.117	.151	.314	.415	.499	.076	.059	.056	.056	.054	.054
	.6	.051	.146	.136	.169	.320	.422	.499	.123	.068	.064	.060	.063	.062
	.9	.050	.343	.212	.212	.362	.443	.514	.318	.135	.099	.081	.081	.084

Table D.13: Null Rejection Probabilities, 5% level, t_{DD} (Bartlett Kernel). Trend. No additional regressors. $\lambda = .5$, $k = .5$. MA(2) spatial correlation in cross-section. $\theta = 0.5$. Two-Tailed Test of $H_0 : \beta_3 = 0$.

			$N(0,1)$ CV						Adjusted Fixed-b CV					
			t_{DK}, values of b						t_{DK}, values of b					
N,T	ρ	t_{clus}	.02	.06	.1	.4	.7	1.0	.02	.06	.1	.4	.7	1.0
9,10	.0	.350	.185	.185	.185	.341	.435	.504	.168	.122	.092	.078	.075	.075
	.3	.368	.216	.216	.216	.366	.456	.524	.194	.146	.113	.088	.084	.084
	.6	.364	.238	.238	.238	.377	.464	.526	.216	.155	.124	.087	.084	.084
	.9	.345	.230	.230	.230	.343	.421	.484	.208	.152	.115	.077	.073	.074
9,50	.0	.366	.072	.104	.147	.322	.424	.495	.057	.055	.054	.051	.049	.051
	.3	.360	.174	.152	.180	.343	.428	.498	.150	.088	.076	.062	.060	.058
	.6	.354	.314	.228	.232	.362	.442	.514	.290	.143	.106	.085	.080	.077
	.9	.349	.473	.325	.284	.386	.461	.528	.450	.219	.138	.082	.080	.081
9,250	.0	.354	.082	.114	.145	.314	.410	.491	.065	.059	.057	.060	.056	.059
	.3	.354	.105	.124	.152	.320	.413	.489	.089	.070	.069	.062	.062	.063
	.6	.350	.155	.144	.171	.328	.415	.487	.139	.087	.080	.068	.062	.062
	.9	.361	.362	.240	.240	.373	.453	.518	.338	.160	.121	.089	.083	.083
49,10	.0	.567	.179	.179	.179	.345	.433	.504	.160	.118	.089	.076	.073	.071
	.3	.558	.215	.215	.215	.366	.450	.516	.196	.147	.106	.086	.082	.082
	.6	.560	.229	.229	.229	.370	.438	.513	.206	.153	.120	.088	.088	.086
	.9	.588	.222	.222	.222	.351	.424	.488	.202	.147	.116	.080	.071	.071
49,50	.0	.573	.068	.101	.135	.307	.402	.480	.052	.044	.047	.060	.059	.058
	.3	.568	.162	.140	.162	.330	.428	.487	.138	.076	.064	.067	.068	.070
	.6	.548	.303	.212	.211	.360	.444	.516	.277	.125	.093	.082	.078	.076

Continued on next page.

Table D.13 (cont'd)

			$N(0,1)$ CV						Adjusted Fixed-b CV					
			t_{DK}, values of b						t_{DK}, values of b					
N,T	ρ	t_{clus}	.02	.06	.1	.4	.7	1.0	.02	.06	.1	.4	.7	1.0
	.9	.558	.462	.304	.265	.376	.447	.520	.438	.218	.136	.086	.081	.083
49,250	.0	.589	.071	.112	.144	.325	.413	.486	.057	.057	.056	.058	.057	.055
	.3	.585	.098	.123	.156	.329	.426	.489	.078	.064	.058	.062	.060	.059
	.6	.582	.149	.142	.169	.334	.430	.501	.128	.080	.071	.066	.065	.064
	.9	.572	.364	.233	.231	.367	.454	.533	.334	.153	.113	.089	.087	.087
256,10	.0	.613	.199	.199	.199	.337	.425	.491	.180	.138	.101	.084	.082	.082
	.3	.629	.222	.222	.222	.356	.438	.497	.204	.157	.123	.092	.086	.088
	.6	.632	.218	.218	.218	.353	.431	.495	.196	.142	.106	.076	.076	.076
	.9	.631	.206	.206	.206	.316	.390	.462	.187	.141	.103	.063	.058	.060
256,50	.0	.630	.068	.102	.139	.314	.403	.478	.049	.048	.048	.053	.053	.054
	.3	.626	.168	.145	.169	.337	.427	.497	.146	.076	.067	.063	.060	.063
	.6	.633	.324	.216	.220	.372	.456	.528	.290	.140	.103	.080	.082	.083
	.9	.613	.477	.314	.276	.391	.472	.530	.453	.227	.138	.082	.084	.085
256,250	.0	.629	.068	.100	.135	.325	.414	.491	.053	.048	.048	.051	.055	.054
	.3	.623	.088	.106	.136	.327	.417	.493	.073	.057	.054	.056	.054	.052
	.6	.630	.130	.122	.151	.332	.433	.503	.111	.066	.060	.060	.058	.058
	.9	.619	.365	.224	.217	.374	.460	.525	.333	.135	.096	.077	.074	.077

Table D.14: Null Rejection Probabilities, 5% level, t_{DD} (Bartlett Kernel). Trend. Time Dummies. No additional regressors. $\lambda = .5$, $k = .5$. $AR(1)$ errors. Two-Tailed Test of $H_0 : \beta_3 = 0$.

			$N(0,1)$ CV						Adjusted Fixed-b CV					
			t_{DK}, values of b						t_{DK}, values of b					
N,T	ρ	t_{clus}	.02	.06	.1	.4	.7	1.0	.02	.06	.1	.4	.7	1.0
10,10	.0	.092	.186	.186	.186	.331	.414	.479	.168	.122	.091	.081	.078	.079
	.3	.098	.215	.215	.215	.354	.432	.494	.192	.140	.104	.085	.080	.080
	.6	.089	.212	.212	.212	.351	.426	.489	.192	.143	.116	.087	.084	.085
	.9	.090	.210	.210	.210	.328	.386	.447	.188	.144	.105	.070	.069	.070
10,50	.0	.100	.070	.102	.140	.312	.401	.477	.056	.057	.056	.049	.051	.053
	.3	.010	.163	.140	.161	.333	.423	.486	.139	.086	.076	.067	.064	.062
	.6	.102	.315	.211	.212	.365	.444	.512	.282	.134	.104	.080	.084	.083
	.9	.110	.495	.336	.300	.401	.470	.538	.471	.243	.152	.098	.095	.094
10,250	.0	.102	.068	.096	.133	.307	.392	.460	.056	.048	.050	.050	.052	.052
	.3	.102	.087	.106	.138	.308	.398	.462	.073	.056	.052	.050	.052	.052
	.6	.107	.135	.124	.155	.322	.408	.476	.116	.066	.060	.055	.058	.057
	.9	.095	.349	.220	.215	.361	.448	.522	.326	.135	.100	.083	.081	.082
50,10	.0	.053	.180	.180	.180	.328	.406	.472	.160	.119	.094	.074	.064	.065
	.3	.054	.207	.207	.207	.341	.426	.487	.190	.147	.116	.089	.086	.087
	.6	.057	.220	.220	.220	.340	.417	.476	.201	.155	.120	.088	.086	.086
	.9	.060	.219	.219	.219	.338	.400	.453	.196	.149	.112	.077	.072	.072
50,50	.0	.063	.077	.108	.142	.303	.406	.475	.066	.057	.060	.059	.058	.059
	.3	.063	.165	.146	.170	.328	.422	.488	.142	.085	.075	.067	.069	.071
	.6	.066	.314	.226	.222	.364	.450	.515	.286	.137	.099	.080	.080	.081
	.9	.058	.497	.333	.288	.399	.475	.539	.472	.238	.146	.098	.096	.097

Continued on next page.

Table D.14 (cont'd)

			$N(0,1)$ CV						Adjusted Fixed-b CV					
			t_{DK}, values of b						t_{DK}, values of b					
N,T	ρ	t_{clus}	.02	.06	.1	.4	.7	1.0	.02	.06	.1	.4	.7	1.0
50,250	.0	.058	.069	.106	.138	.316	.402	.476	.055	.055	.054	.048	.051	.048
	.3	.058	.094	.119	.144	.322	.410	.475	.076	.060	.059	.053	.054	.052
	.6	.054	.143	.131	.156	.330	.419	.484	.121	.072	.067	.058	.059	.058
	.9	.056	.346	.223	.212	.356	.441	.511	.324	.138	.098	.078	.073	.075
250,10	.0	.054	.200	.200	.200	.356	.434	.488	.177	.131	.098	.085	.083	.082
	.3	.057	.226	.226	.226	.363	.442	.512	.205	.158	.123	.095	.091	.091
	.6	.055	.228	.228	.228	.360	.439	.502	.209	.159	.125	.090	.087	.086
	.9	.050	.214	.214	.214	.335	.408	.470	.189	.144	.112	.078	.073	.070
250,50	.0	.052	.077	.112	.145	.318	.406	.473	.062	.060	.052	.055	.056	.053
	.3	.055	.168	.152	.176	.340	.426	.490	.150	.088	.076	.062	.064	.063
	.6	.051	.312	.212	.214	.365	.450	.526	.284	.137	.105	.081	.079	.080
	.9	.044	.494	.329	.291	.390	.472	.538	.468	.228	.146	.097	.095	.095
250,250	.0	.048	.068	.105	.141	.312	.414	.486	.053	.055	.055	.057	.052	.052
	.3	.051	.090	.117	.151	.314	.415	.499	.076	.059	.056	.056	.054	.054
	.6	.051	.146	.136	.169	.320	.422	.499	.123	.068	.064	.060	.063	.062
	.9	.050	.343	.212	.212	.362	.443	.514	.318	.135	.099	.081	.081	.084

Table D.15: Null Rejection Probabilities, 5% level, t_{DD} (Bartlett Kernel). Trend. Time Dummies. No additional regressors. $\lambda = .5$, $k = .5$. MA(2) spatial correlation in cross-section. $\theta = 0.5$. Two-Tailed Test of $H_0 : \beta_3 = 0$.

			$N(0,1)$ CV						Adjusted Fixed-b CV					
			t_{DK}, values of b						t_{DK}, values of b					
N,T	ρ	t_{clus}	.02	.06	.1	.4	.7	1.0	.02	.06	.1	.4	.7	1.0
9,10	.0	.350	.185	.185	.185	.341	.435	.504	.168	.122	.092	.078	.075	.075
	.3	.368	.216	.216	.216	.366	.456	.524	.194	.146	.113	.088	.084	.084
	.6	.364	.238	.238	.238	.377	.464	.526	.216	.155	.124	.087	.084	.084
	.9	.345	.230	.230	.230	.343	.421	.484	.208	.152	.115	.077	.073	.074
9,50	.0	.366	.072	.104	.147	.322	.424	.495	.057	.055	.054	.051	.049	.051
	.3	.360	.174	.152	.180	.343	.428	.498	.150	.088	.076	.062	.060	.058
	.6	.354	.314	.228	.232	.362	.442	.514	.290	.143	.106	.085	.080	.077
	.9	.349	.473	.325	.284	.386	.461	.528	.450	.219	.138	.082	.080	.081
9,250	.0	.354	.082	.114	.145	.314	.410	.491	.065	.059	.057	.060	.056	.059
	.3	.354	.105	.124	.152	.320	.413	.489	.089	.070	.069	.062	.062	.063
	.6	.350	.155	.144	.171	.328	.415	.487	.139	.087	.080	.068	.062	.062
	.9	.361	.362	.240	.240	.373	.453	.518	.338	.160	.121	.089	.083	.083
49,10	.0	.567	.179	.179	.179	.345	.433	.504	.160	.118	.089	.076	.073	.071
	.3	.558	.215	.215	.215	.366	.450	.516	.196	.147	.106	.086	.082	.082
	.6	.560	.229	.229	.229	.370	.438	.513	.206	.153	.120	.088	.088	.086
	.9	.588	.222	.222	.222	.351	.424	.488	.202	.147	.116	.080	.071	.071
49,50	.0	.573	.068	.101	.135	.307	.402	.480	.052	.044	.047	.060	.059	.058
	.3	.568	.162	.140	.162	.330	.428	.487	.138	.076	.064	.067	.068	.070
	.6	.548	.303	.212	.211	.360	.444	.516	.277	.125	.093	.082	.078	.076

Continued on next page.

			$N(0,1)$ CV						Adjusted Fixed-b CV					
			t_{DK}, values of b						t_{DK}, values of b					
N,T	ρ	t_{clus}	.02	.06	.1	.4	.7	1.0	.02	.06	.1	.4	.7	1.0
	.9	.558	.462	.304	.265	.376	.447	.520	.438	.218	.136	.086	.081	.083
49,250	.0	.589	.071	.112	.144	.325	.413	.486	.057	.057	.056	.058	.057	.055
	.3	.585	.098	.123	.156	.329	.426	.489	.078	.064	.058	.062	.060	.059
	.6	.582	.149	.142	.169	.334	.430	.501	.128	.080	.071	.066	.065	.064
	.9	.572	.364	.233	.231	.367	.454	.533	.334	.153	.113	.089	.087	.087
256,10	.0	.613	.199	.199	.199	.337	.425	.491	.180	.138	.101	.084	.082	.082
	.3	.629	.222	.222	.222	.356	.438	.497	.204	.157	.123	.092	.086	.088
	.6	.632	.218	.218	.218	.353	.431	.495	.196	.142	.106	.076	.076	.076
	.9	.631	.206	.206	.206	.316	.390	.462	.187	.141	.103	.063	.058	.060
256,50	.0	.630	.068	.102	.139	.314	.403	.478	.049	.048	.048	.053	.053	.054
	.3	.626	.168	.145	.169	.337	.427	.497	.146	.076	.067	.063	.060	.063
	.6	.633	.324	.216	.220	.372	.456	.528	.290	.140	.103	.080	.082	.083
	.9	.613	.477	.314	.276	.391	.472	.530	.453	.227	.138	.082	.084	.085
256,250	.0	.629	.068	.100	.135	.325	.414	.491	.053	.048	.048	.051	.055	.054
	.3	.623	.088	.106	.136	.327	.417	.493	.073	.057	.054	.056	.054	.052
	.6	.630	.130	.122	.151	.332	.433	.503	.111	.066	.060	.060	.058	.058
	.9	.619	.365	.224	.217	.374	.460	.525	.333	.135	.096	.077	.074	.077

Table D.15 (cont'd)

Table D.16: Null Rejection Probabilities, 5% level, t_{DD} (Bartlett Kernel). One additional regressor. No trend. $\lambda = .5$, $k = .5$. MA(2) spatial correlation in cross-section. $\theta = 0.5$. Two-Tailed Test of $H_0 : \beta_3 = 0$ and $H_0 : \gamma = 0$.

		Adjusted Fixed-b CV						Usual Fixed-b CV					
		t_{DD}, values of b						t_z, values of b					
N,T	ρ	.02	.06	.1	.4	.7	1.0	.02	.06	.1	.4	.7	1.0
9,10	.0	.103	.085	.070	.066	.060	.060	.186	.167	.150	.121	.117	.120
	.3	.193	.164	.143	.110	.110	.105	.202	.184	.164	.130	.125	.130
	.6	.320	.281	.254	.174	.164	.154	.240	.214	.190	.149	.140	.143
	.9	.449	.419	.387	.244	.241	.236	.301	.281	.261	.192	.174	.174
9,50	.0	.048	.049	.052	.049	.044	.046	.062	.059	.059	.060	.060	.058
	.3	.143	.086	.073	.064	.055	.059	.081	.070	.068	.064	.062	.063
	.6	.312	.160	.119	.086	.078	.082	.189	.118	.110	.088	.084	.086
	.9	.605	.416	.328	.205	.194	.188	.442	.295	.246	.178	.162	.166
9,250	.0	.047	.044	.048	.047	.044	.042	.056	.054	.054	.050	.050	.052
	.3	.064	.049	.052	.049	.048	.046	.061	.056	.056	.052	.050	.050
	.6	.109	.060	.058	.055	.052	.050	.087	.072	.072	.064	.064	.067
	.9	.359	.172	.123	.089	.086	.086	.243	.145	.133	.101	.093	.096
49,10	.0	.097	.074	.059	.054	.057	.054	.141	.121	.108	.092	.085	.088
	.3	.204	.174	.144	.111	.107	.103	.160	.142	.125	.112	.105	.108
	.6	.320	.290	.262	.186	.174	.176	.226	.203	.184	.148	.144	.146
	.9	.474	.442	.410	.274	.271	.272	.319	.296	.274	.203	.191	.191
49,50	.0	.054	.050	.046	.049	.048	.045	.056	.056	.055	.053	.051	.052
	.3	.148	.085	.074	.063	.066	.064	.080	.069	.061	.056	.055	.058
	.6	.328	.171	.128	.100	.092	.094	.178	.114	.098	.076	.079	.082

Continued on next page.

Table D.16 (cont'd)

		Adjusted Fixed-*b* CV						Usual Fixed-*b* CV					
		t_{DD}, values of b						t_z, values of b					
N,T	ρ	.02	.06	.1	.4	.7	1.0	.02	.06	.1	.4	.7	1.0
	.9	.635	.441	.344	.228	.217	.213	.463	.290	.231	.164	.158	.160
49,250	.0	.045	.045	.044	.044	.046	.050	.048	.052	.051	.051	.053	.051
	.3	.062	.050	.049	.050	.049	.050	.058	.058	.056	.051	.055	.054
	.6	.114	.061	.059	.057	.056	.056	.075	.063	.061	.055	.060	.061
	.9	.361	.176	.129	.093	.095	.092	.243	.129	.115	.099	.095	.099
256,10	.0	.121	.098	.079	.070	.069	.066	.111	.098	.082	.066	.070	.069
	.3	.203	.178	.156	.117	.119	.116	.138	.117	.099	.082	.079	.078
	.6	.334	.301	.273	.198	.192	.196	.222	.198	.174	.128	.115	.115
	.9	.483	.448	.415	.272	.271	.266	.335	.314	.290	.210	.197	.194
256,50	.0	.050	.040	.040	.041	.044	.043	.055	.053	.052	.046	.047	.050
	.3	.152	.085	.070	.057	.064	.057	.076	.062	.059	.057	.056	.056
	.6	.347	.179	.129	.097	.095	.092	.187	.105	.090	.071	.076	.074
	.9	.649	.458	.366	.227	.216	.213	.487	.298	.238	.172	.158	.158
256,250	.0	.040	.040	.044	.048	.044	.050	.053	.050	.047	.045	.045	.045
	.3	.059	.046	.051	.052	.050	.050	.060	.057	.052	.043	.049	.050
	.6	.105	.066	.056	.061	.056	.056	.084	.067	.060	.059	.060	.060
	.9	.346	.165	.129	.100	.094	.092	.229	.114	.094	.085	.081	.086

Table D.17: Null Rejection Probabilities, 5% level, t_{DD} (Bartlett Kernel). Trend and one additional regressor. $\lambda = .5$, $k = .5$. MA(2) spatial correlation in cross-section. $\theta = 0.5$. Two-Tailed Test of $H_0 : \beta_3 = 0$ and $H_0 : \gamma = 0$.

		Adjusted Fixed-b CV						Usual Fixed-b CV					
		t_{DD}, values of b						t_z, values of b					
N,T	ρ	.02	.06	.1	.4	.7	1.0	.02	.06	.1	.4	.7	1.0
9,10	.0	.160	.112	.085	.082	.074	.078	.208	.190	.169	.120	.121	.120
	.3	.189	.140	.105	.082	.081	.083	.225	.200	.180	.137	.136	.139
	.6	.203	.148	.114	.081	.080	.080	.272	.248	.228	.188	.181	.192
	.9	.194	.142	.108	.076	.073	.073	.347	.323	.299	.250	.244	.249
9,50	.0	.060	.057	.058	.054	.060	.057	.063	.064	.061	.062	.055	.058
	.3	.152	.084	.074	.064	.064	.063	.082	.078	.072	.064	.066	.069
	.6	.289	.146	.111	.083	.081	.080	.194	.129	.114	.089	.088	.090
	.9	.444	.213	.140	.086	.089	.087	.442	.313	.274	.219	.216	.224
9,250	.0	.054	.049	.049	.048	.044	.046	.054	.053	.055	.051	.052	.055
	.3	.073	.057	.058	.054	.050	.052	.060	.060	.058	.056	.056	.056
	.6	.122	.073	.068	.061	.058	.056	.087	.075	.070	.062	.060	.064
	.9	.313	.144	.104	.081	.077	.078	.241	.152	.140	.097	.095	.098
49,10	.0	.181	.134	.100	.087	.083	.081	.160	.150	.133	.103	.098	.103
	.3	.201	.154	.112	.094	.089	.086	.186	.168	.152	.118	.117	.119
	.6	.201	.151	.114	.087	.083	.082	.226	.204	.186	.156	.154	.154
	.9	.190	.140	.111	.074	.068	.070	.302	.278	.254	.230	.215	.220
49,50	.0	.066	.057	.055	.058	.060	.058	.058	.056	.055	.051	.052	.052
	.3	.152	.088	.072	.070	.070	.070	.078	.066	.064	.058	.061	.064
	.6	.284	.136	.101	.086	.086	.084	.176	.114	.104	.079	.080	.087

Continued on next page.

Table D.17 (cont'd)

		Adjusted Fixed-*b* CV						Usual Fixed-*b* CV					
		t_{DD}, values of b						t_z, values of b					
N,T	ρ	.02	.06	.1	.4	.7	1.0	.02	.06	.1	.4	.7	1.0
	.9	.452	.216	.139	.089	.086	.088	.446	.302	.259	.191	.183	.187
49,250	.0	.056	.056	.052	.058	.058	.058	.047	.049	.050	.049	.050	.052
	.3	.074	.060	.052	.060	.061	.059	.055	.054	.057	.056	.053	.054
	.6	.118	.072	.064	.067	.067	.066	.080	.067	.066	.059	.059	.063
	.9	.338	.161	.118	.095	.092	.092	.251	.141	.122	.094	.094	.098
256,10	.0	.186	.134	.104	.083	.084	.084	.127	.111	.096	.079	.078	.080
	.3	.208	.164	.134	.098	.094	.094	.146	.126	.111	.097	.095	.097
	.6	.217	.163	.128	.093	.090	.090	.206	.188	.169	.141	.136	.138
	.9	.200	.155	.114	.077	.069	.069	.281	.259	.233	.197	.179	.187
256,50	.0	.051	.050	.048	.056	.054	.055	.060	.056	.054	.055	.051	.054
	.3	.147	.077	.070	.063	.066	.066	.085	.068	.063	.057	.061	.062
	.6	.294	.139	.103	.086	.086	.086	.187	.114	.105	.085	.080	.084
	.9	.448	.231	.140	.088	.086	.084	.442	.285	.243	.189	.170	.174
256,250	.0	.051	.047	.049	.047	.051	.050	.054	.048	.046	.044	.044	.044
	.3	.067	.056	.054	.053	.051	.050	.060	.055	.051	.047	.047	.048
	.6	.113	.066	.063	.060	.057	.060	.082	.066	.066	.059	.058	.059
	.9	.334	.139	.098	.077	.076	.076	.231	.122	.105	.082	.083	.086

Table D.18: Null Rejection Probabilities, 5% level, t_{DD} (Bartlett Kernel). No trend and additional regressors. $\lambda = .5$, $k = .5$. MA(2) spatial correlation in cross-section. $\theta = 0.5$. Two-Tailed Test of $H_0 : \beta_3 = 0$.

				$N(0,1)$ CV				$N(0,1)$ CV				Adjusted Fixed-b CV			
				t^r_{double}, values of b				t_{DK}, values of b				t_{DK}, values of b			
N,T	ρ	t_{clus}	t_{double}	.02	.1	.4	.7	.02	.1	.4	.7	.02	.1	.4	.7
49,50	.0	.565	.063	.093	.223	.405	.466	.057	.095	.260	.380	.049	.048	.050	.051
	.3	.557	.155	.108	.226	.454	.498	.162	.130	.292	.409	.146	.066	.065	.064
	.6	.553	.288	.186	.227	.496	.538	.341	.199	.342	.449	.319	.122	.094	.092
	.9	.566	.479	.381	.327	.632	.666	.654	.441	.516	.603	.642	.349	.219	.209
256,250	.0	.623	.045	.066	.192	.403	.468	.050	.093	.252	.384	.038	.045	.051	.049
	.3	.625	.140	.066	.193	.414	.469	.071	.100	.259	.387	.058	.051	.053	.053
	.6	.626	.289	.073	.194	.433	.481	.125	.110	.270	.395	.104	.059	.059	.058
	.9	.624	.514	.201	.202	.494	.534	.373	.193	.333	.437	.350	.126	.099	.093

Appendix E

FIGURES IN CHAPTER 3

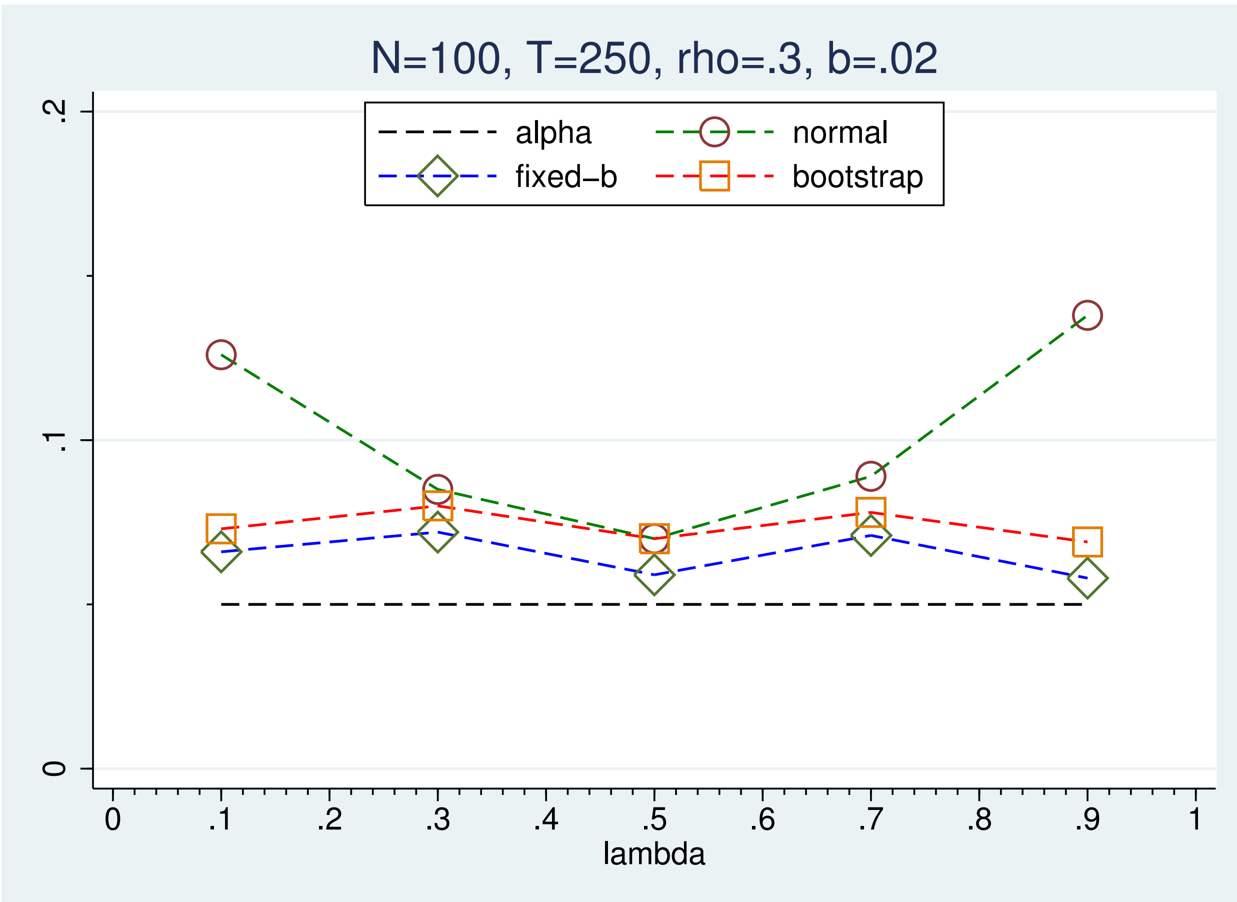


Figure E.1: Empirical null rejection probabilities, no spatial correlation, Bartlett kernel, $N = 100$, $T = 250$, $\rho = 0.3$, $b = 0.02$. For interpretation of the references to color in this and all other figures, the reader is refered to the electronic version of this dissertation.

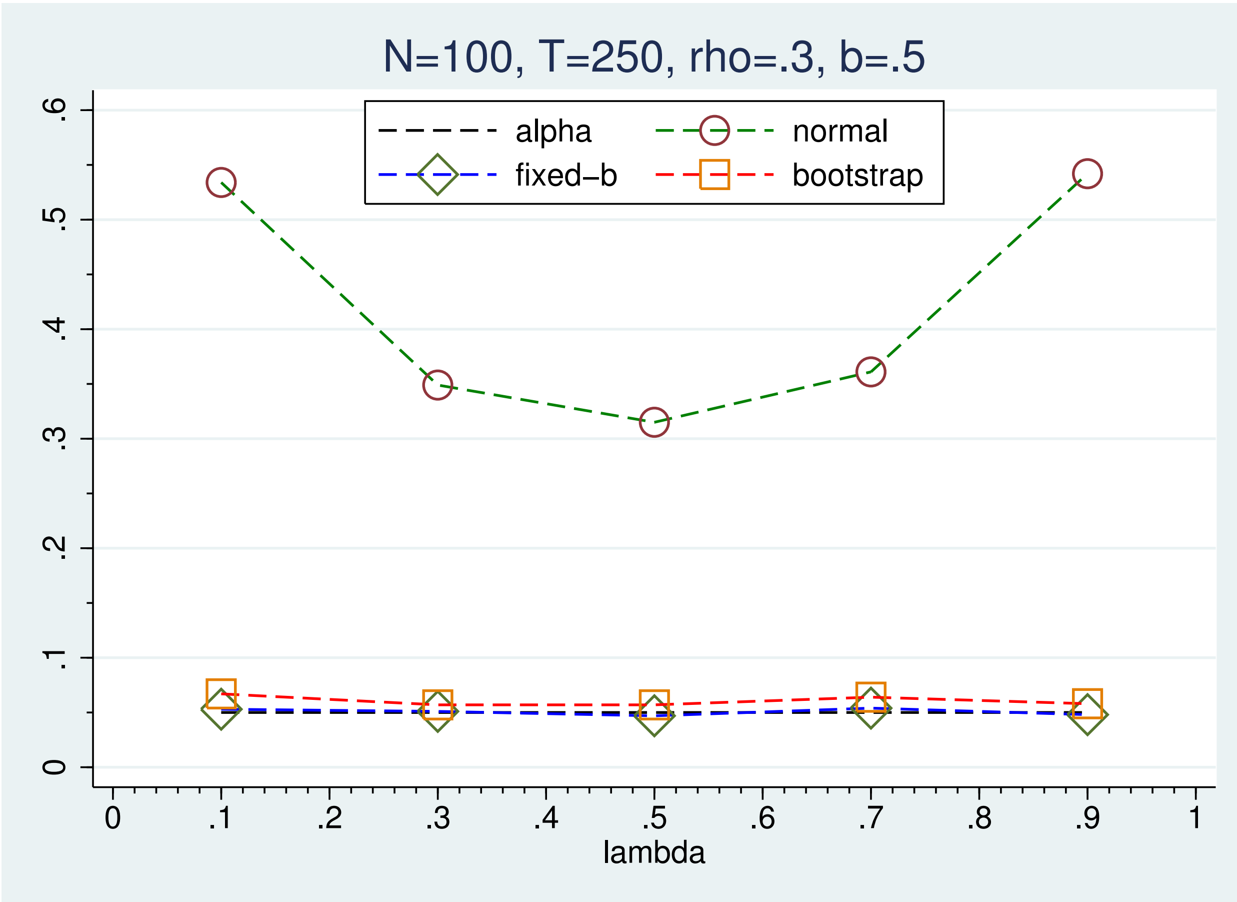
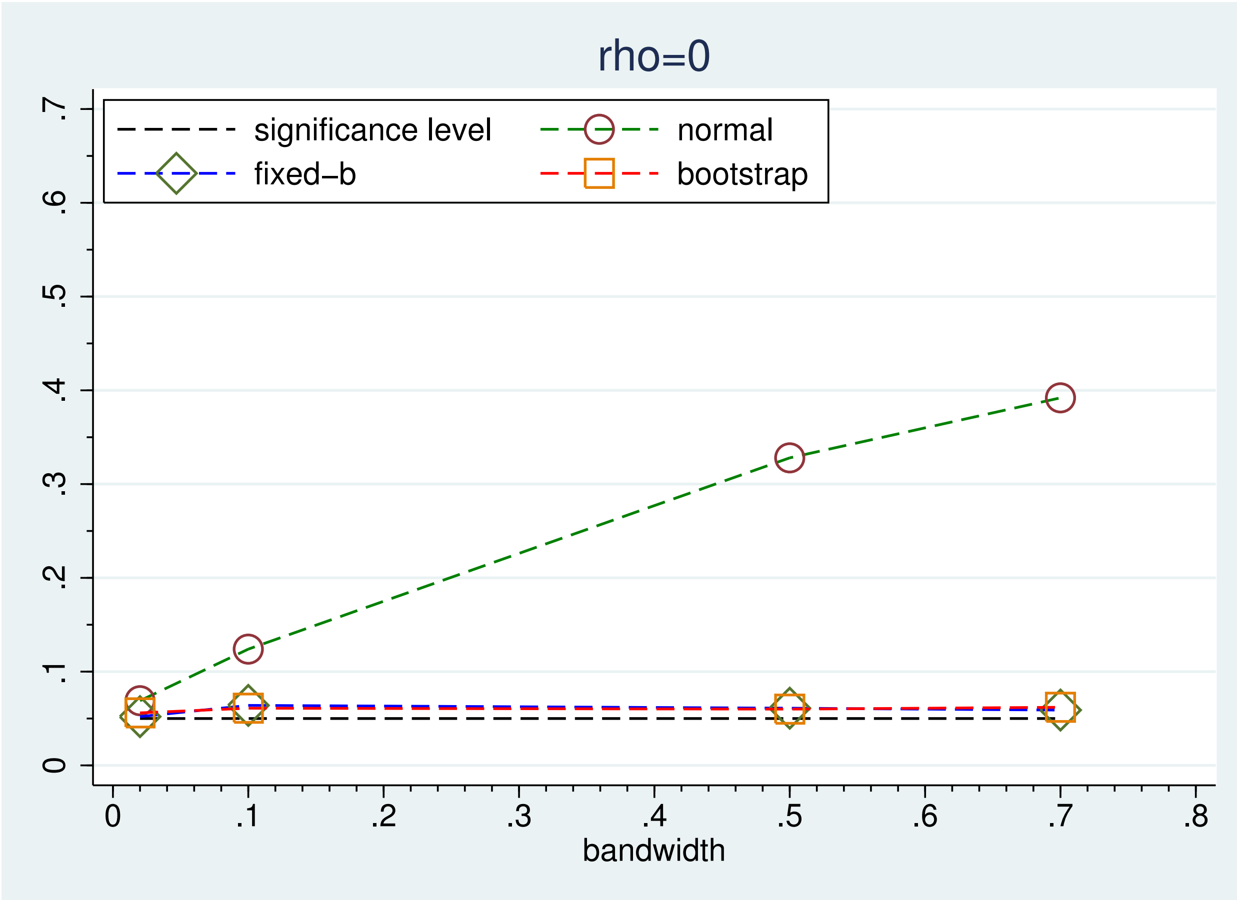


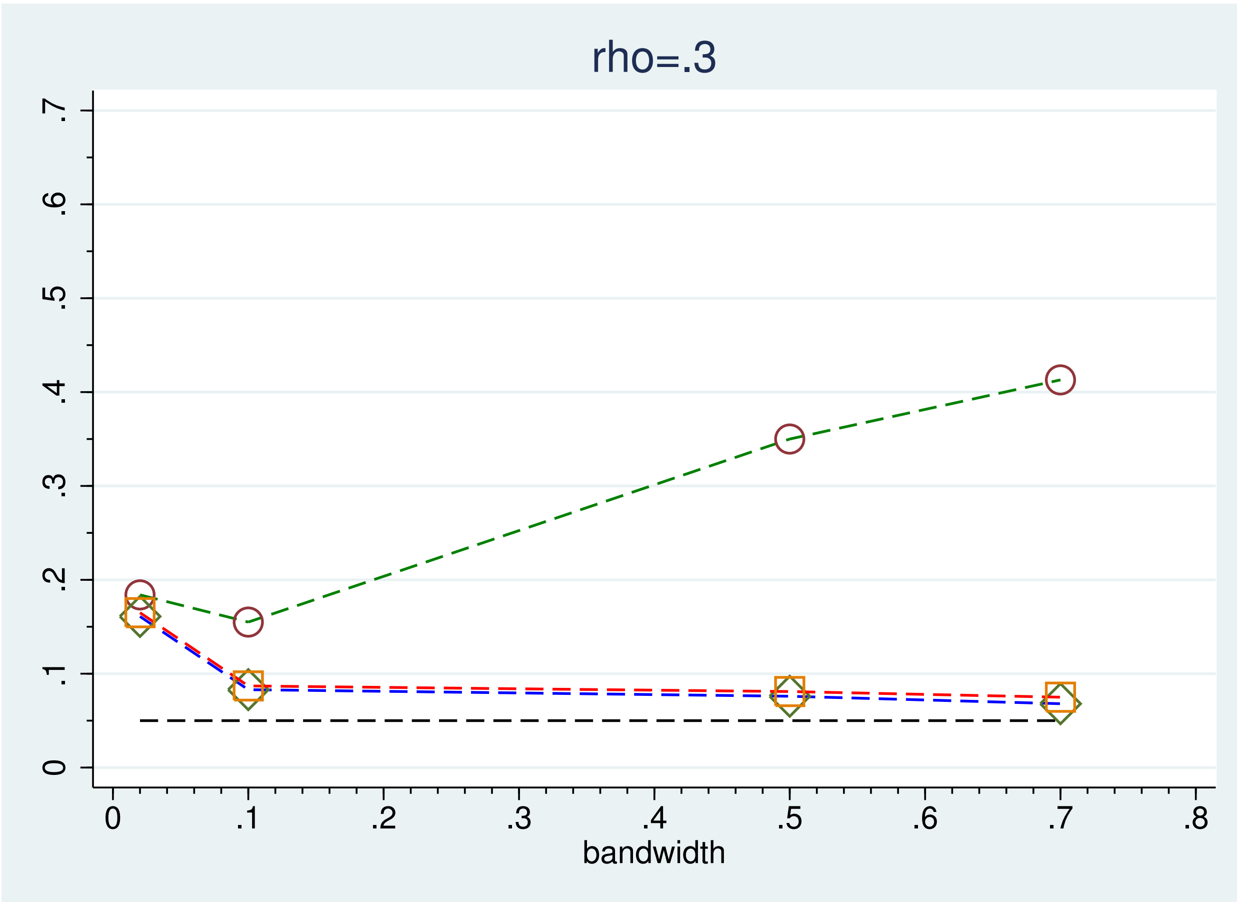
Figure E.2: Empirical null rejection probabilities, no spatial correlation, Bartlett kernel, $N = 100$,$T = 250$, $\rho = 0.3$, $b = 0.5$.



(a) N=50, T=50

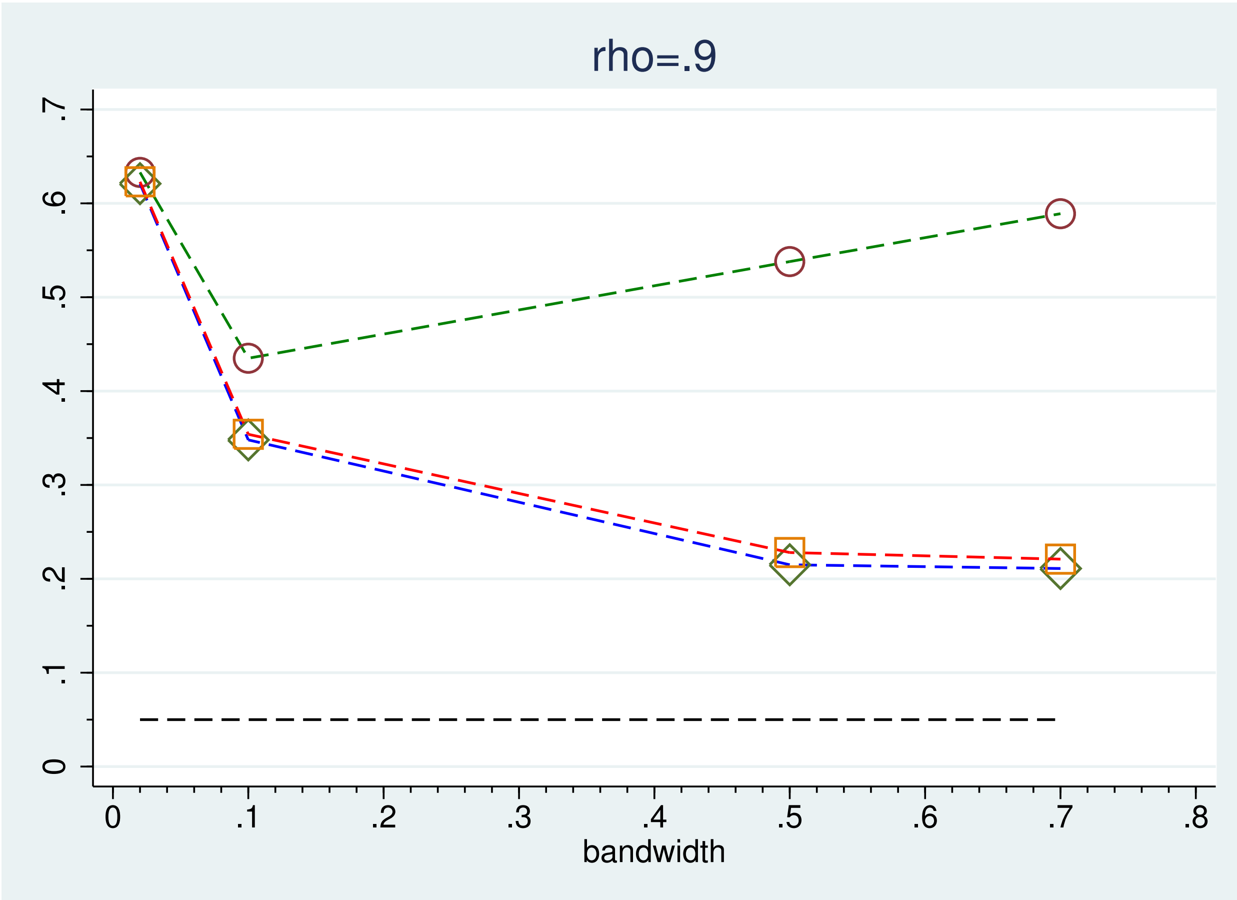
Figure E.3: Empirical null rejection probabilities, no spatial correlation, Bartlett kernel, $N = 50$, $\lambda = 0.5$.

Figure E.3: (cont'd)



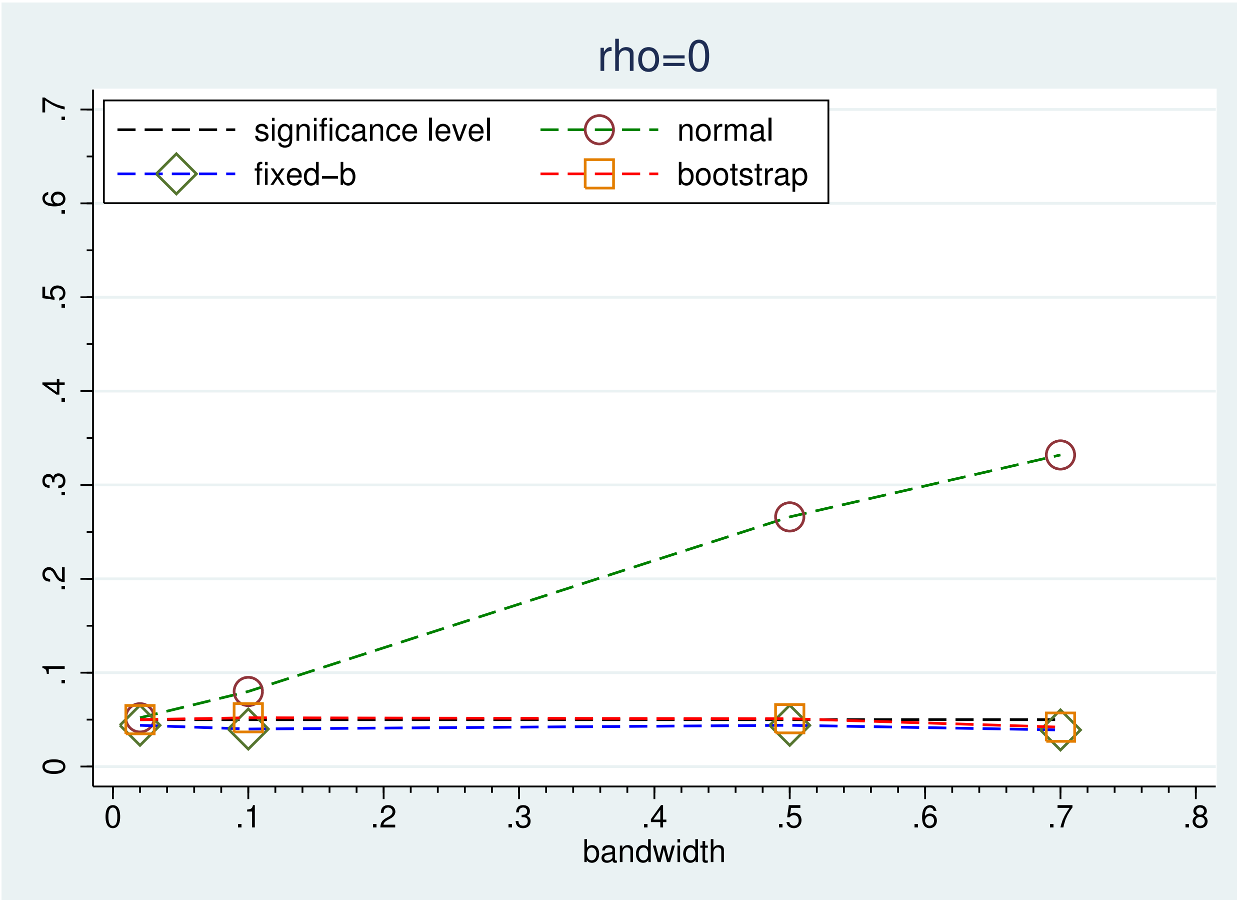
(b) N=50, T=50

Figure E.3: (cont'd)



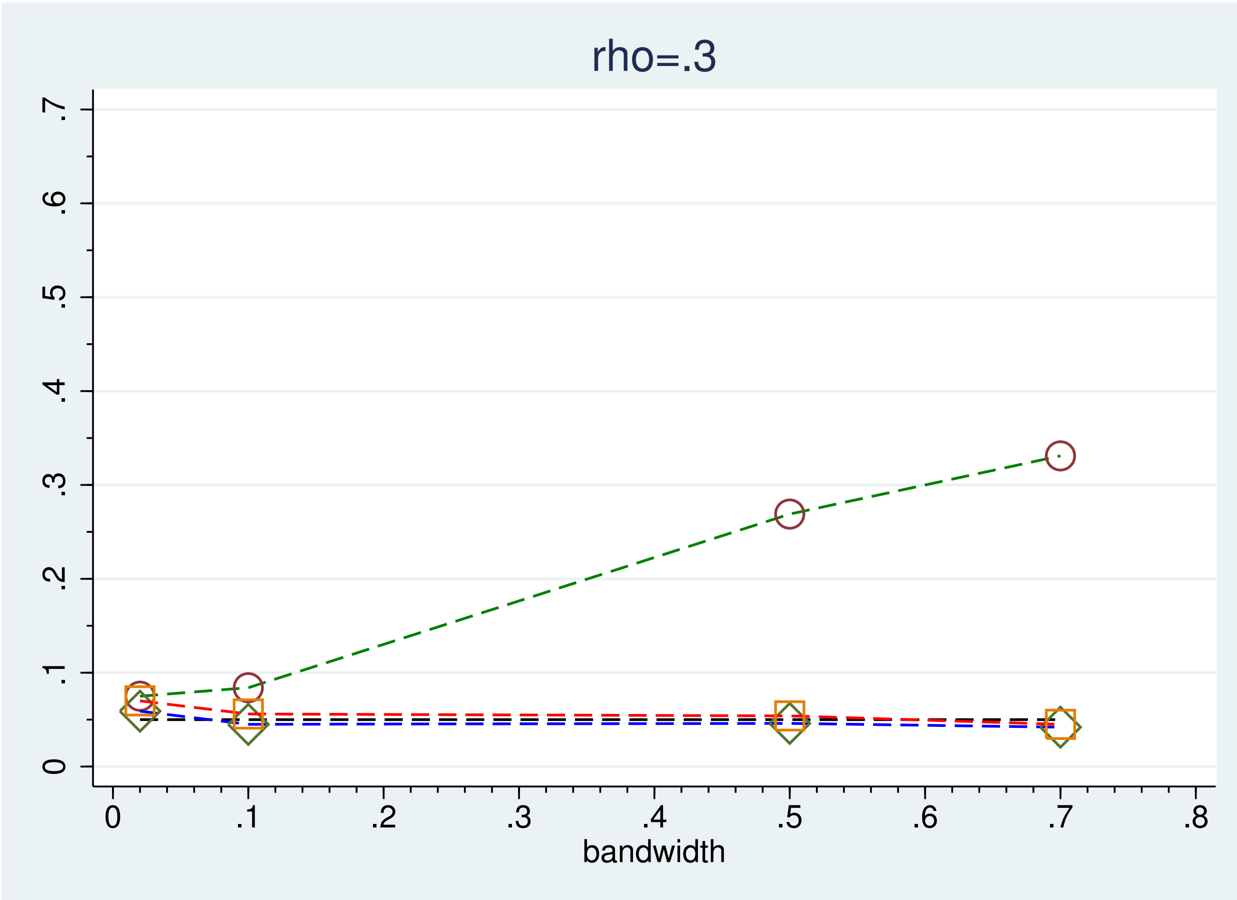
(c) N=50, T=50

Figure E.3: (cont'd)



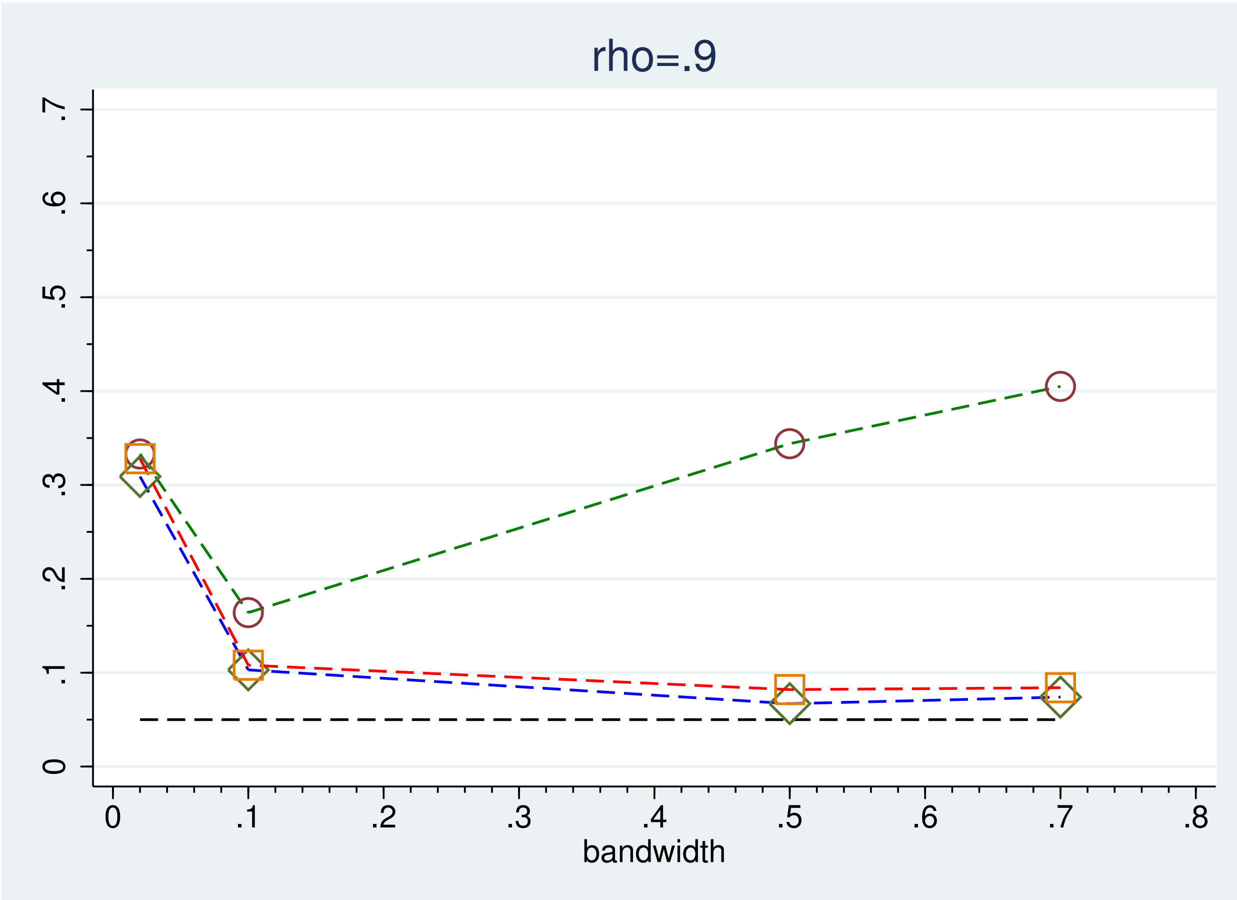
(d) N=50, T=250

Figure E.3: (cont'd)

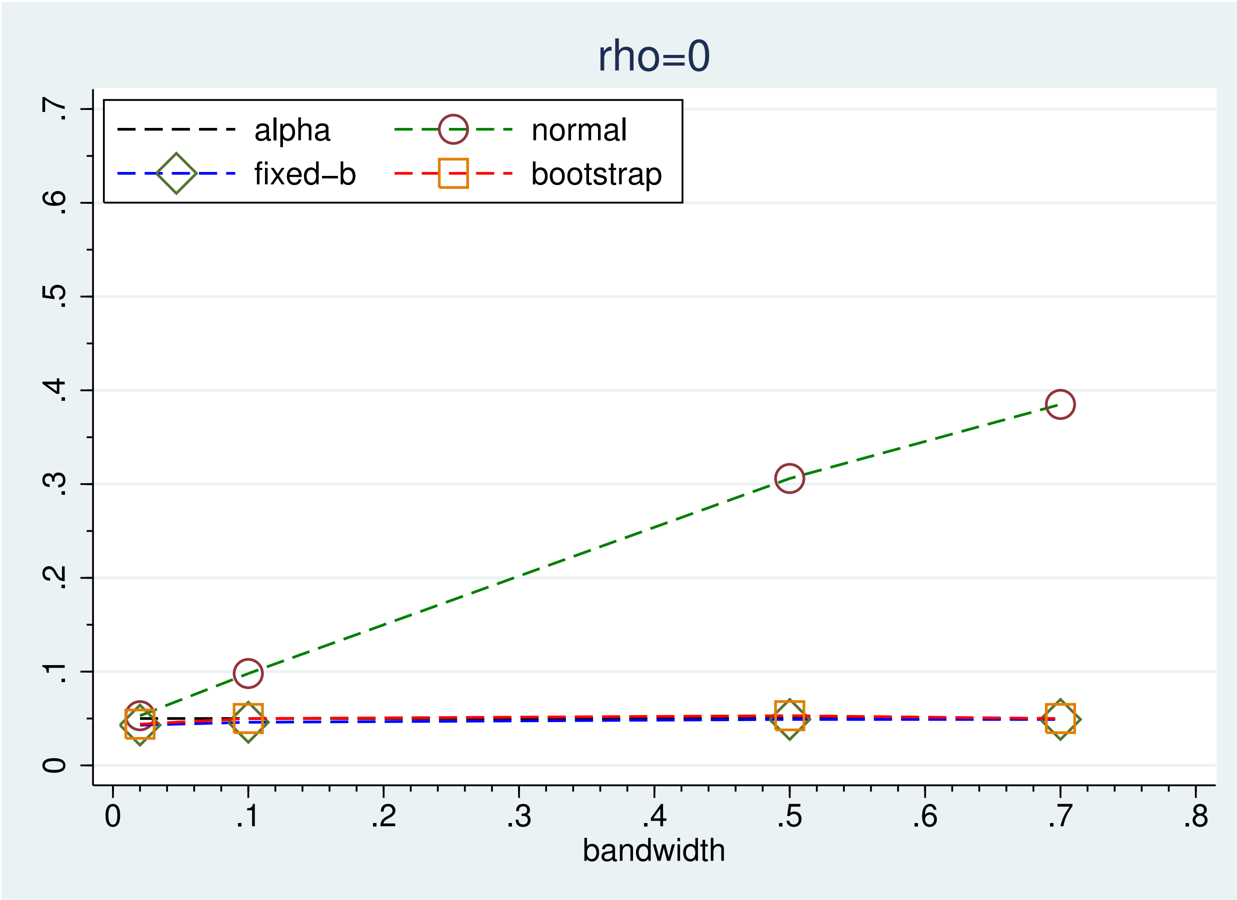


(e) N=50, T=250

Figure E.3: (cont'd)



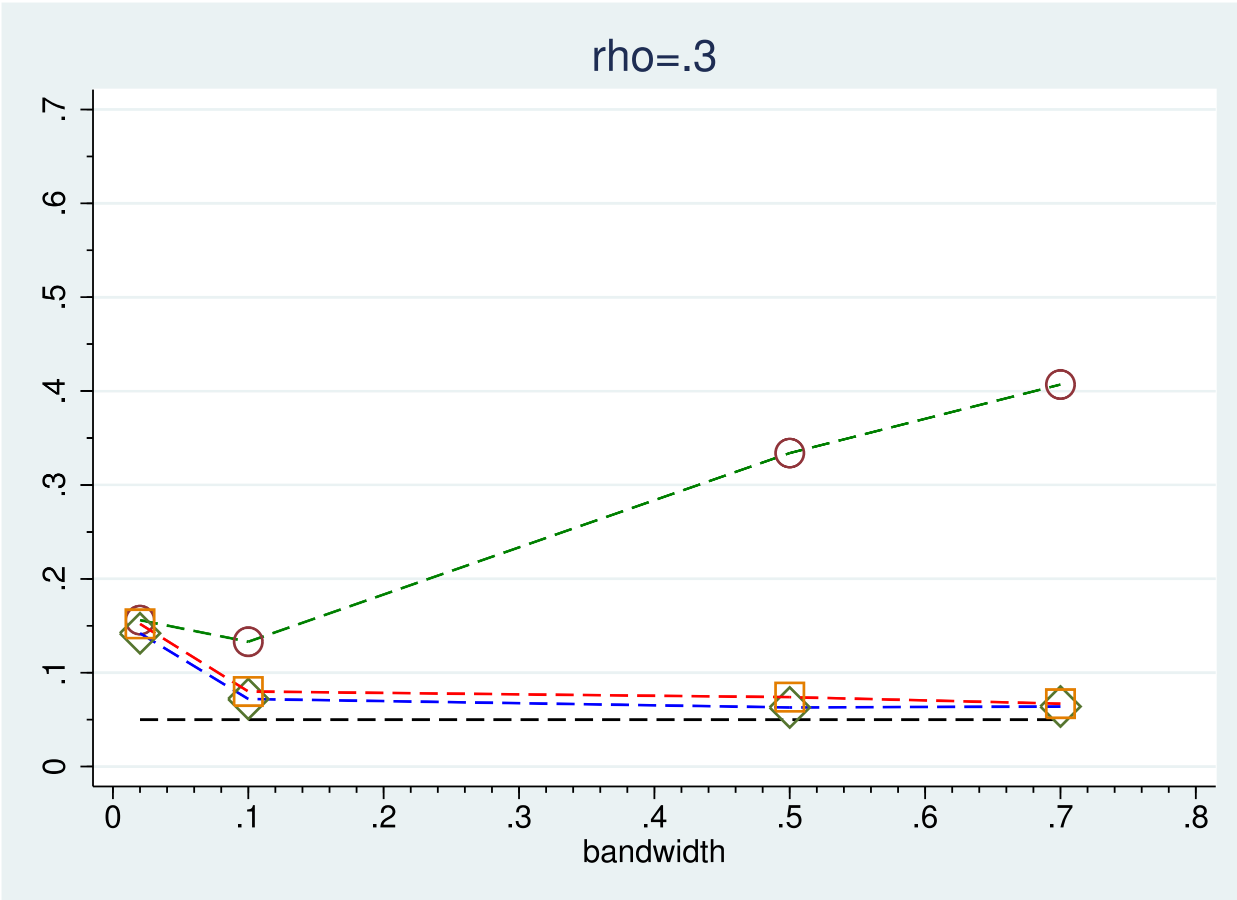
(f) N=50, T=250



(a) N=49, T=50

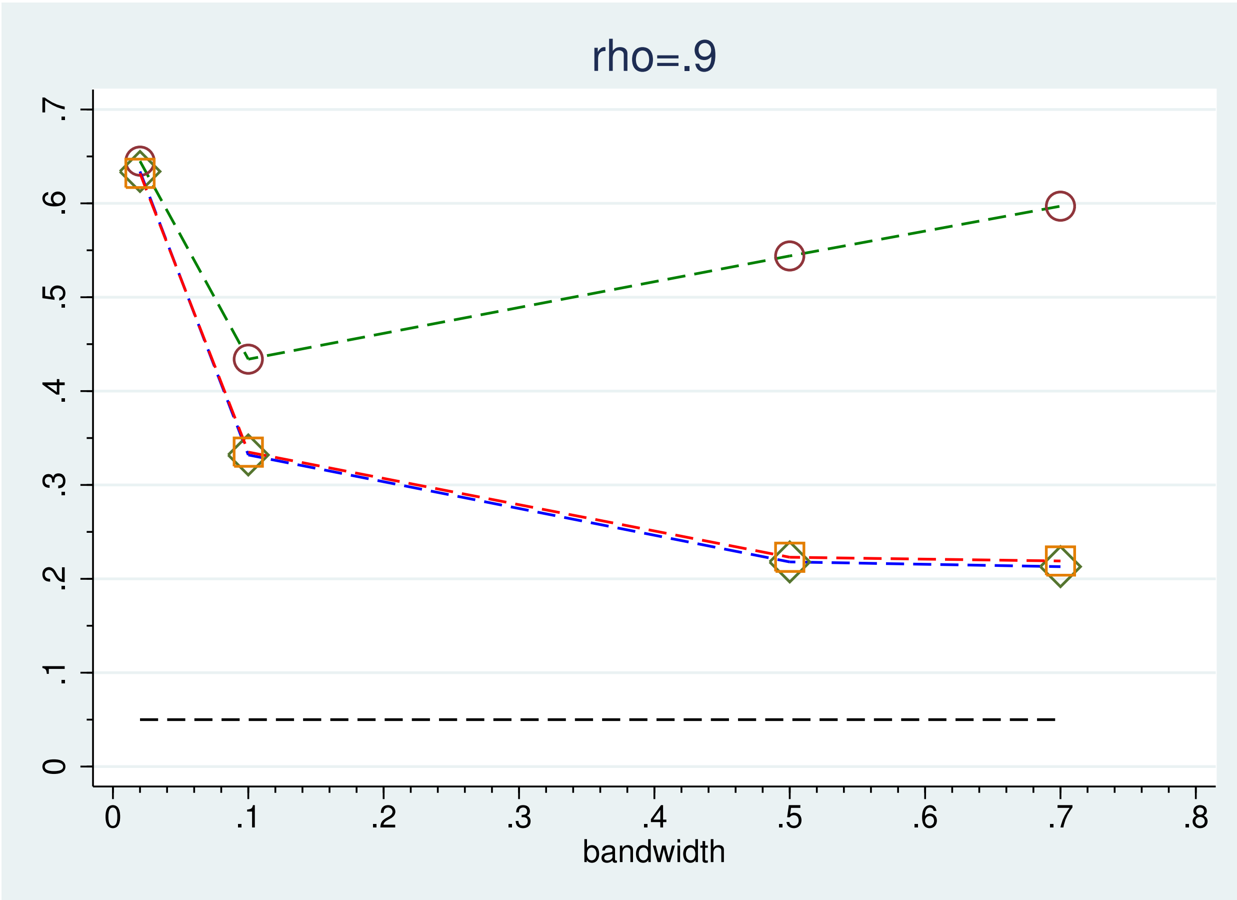
Figure E.4: Empirical null rejection probabilities, spatial $MA(2)$, Bartlett kernel, $N = 49$, $\lambda = 0.5$.

Figure E.4: (cont'd)



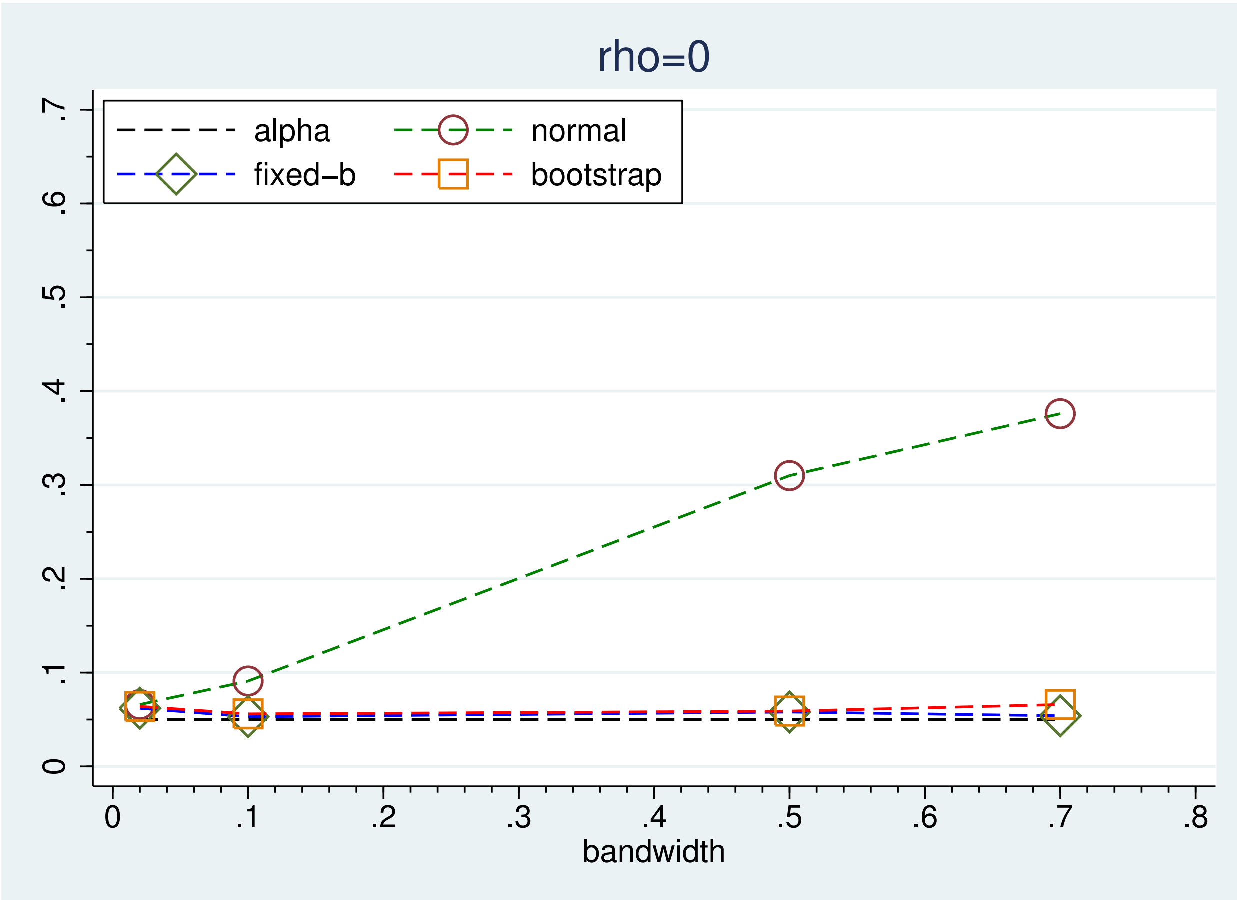
(b) N=49, T=50

Figure E.4: (cont'd)



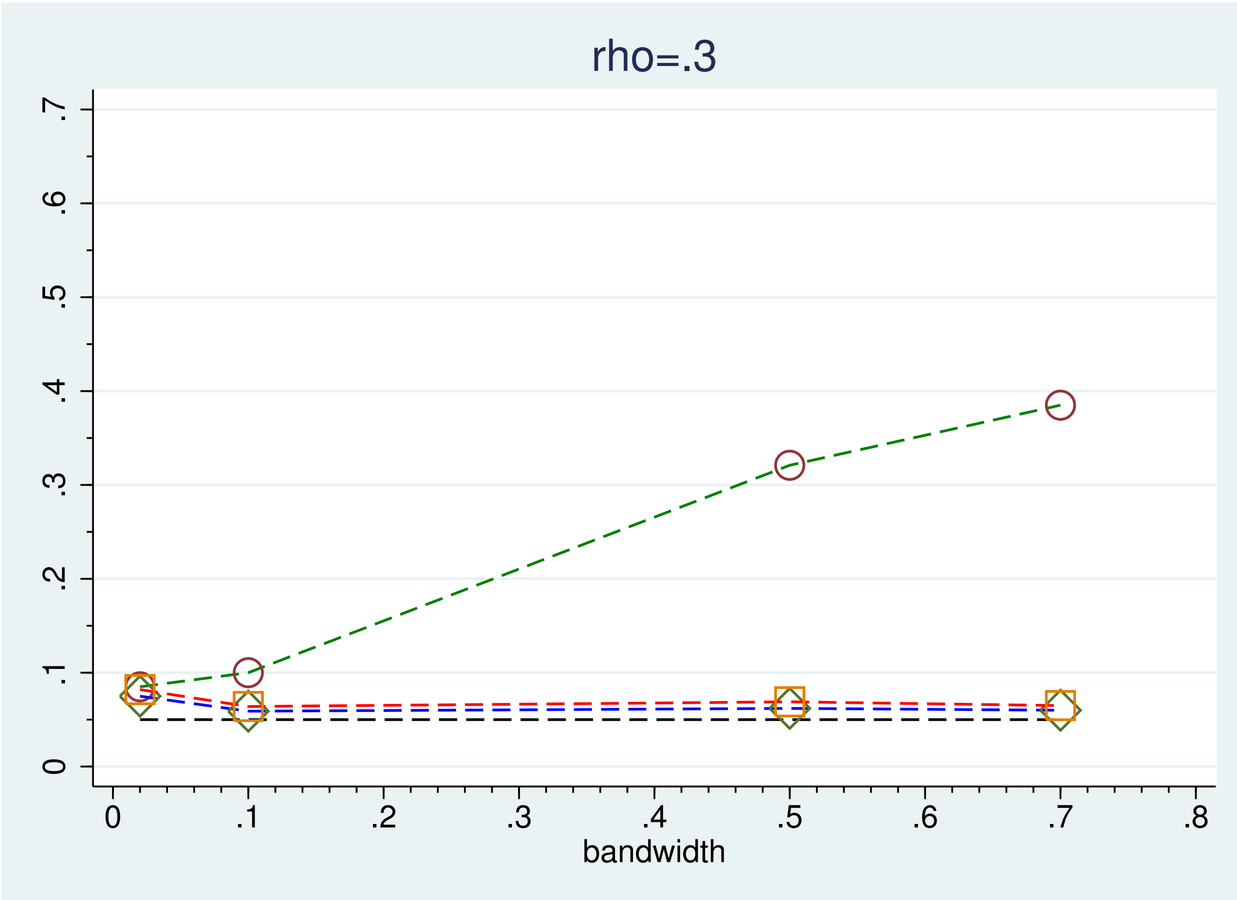
(c) N=49, T=50

Figure E.4: (cont'd)



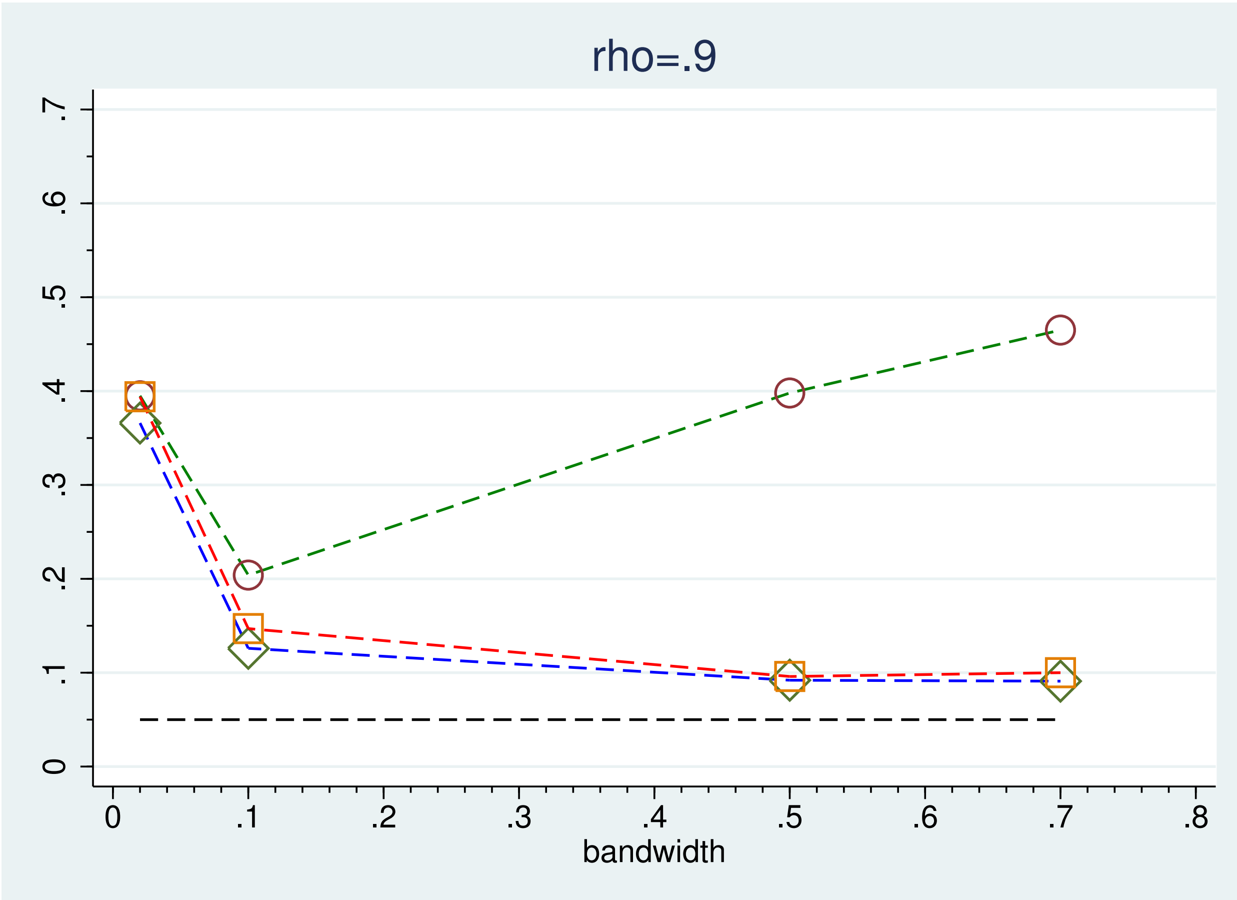
(d) N=49, T=250

Figure E.4: (cont'd)

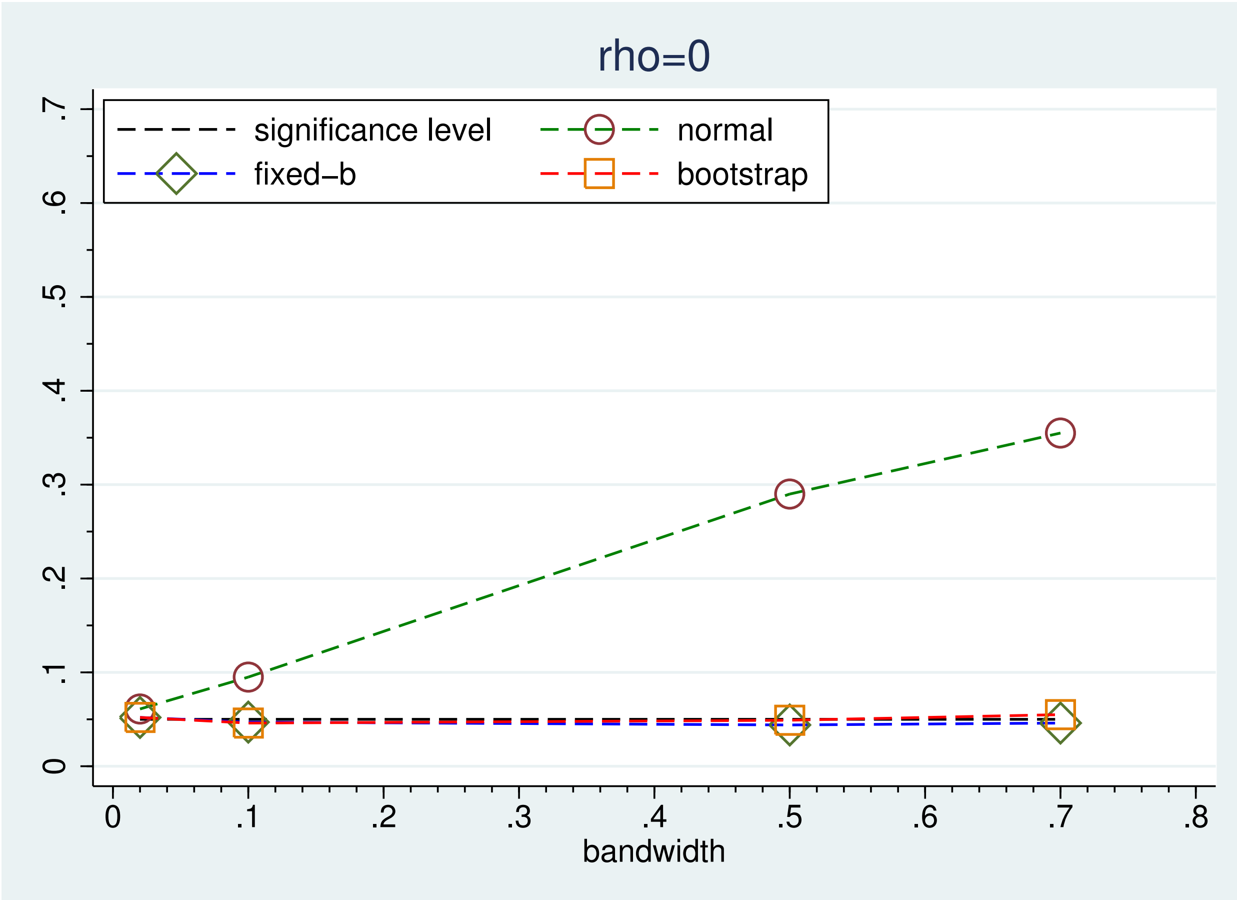


(e) N=49, T=250

Figure E.4: (cont'd)



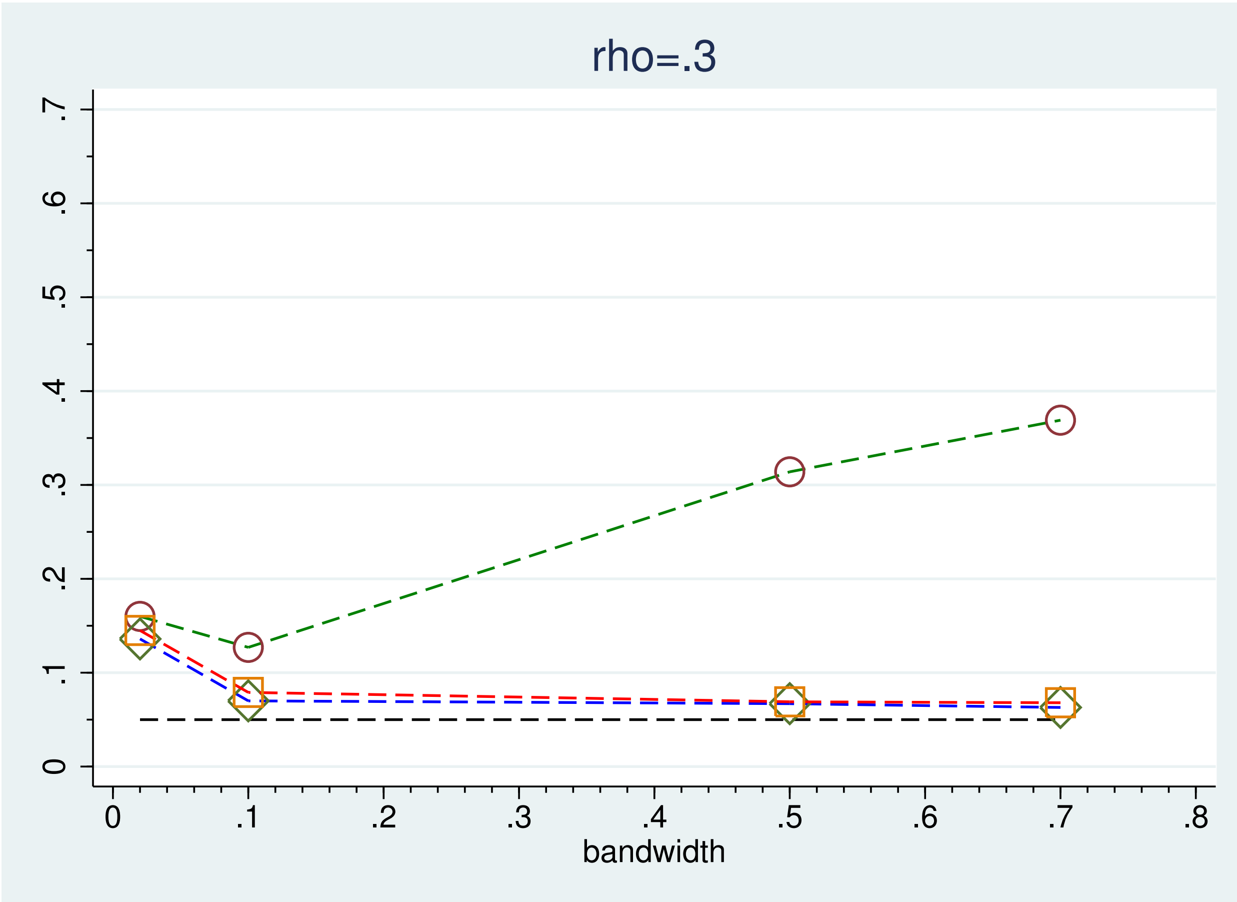
(f) N=49, T=250



(a) N=250, T=50

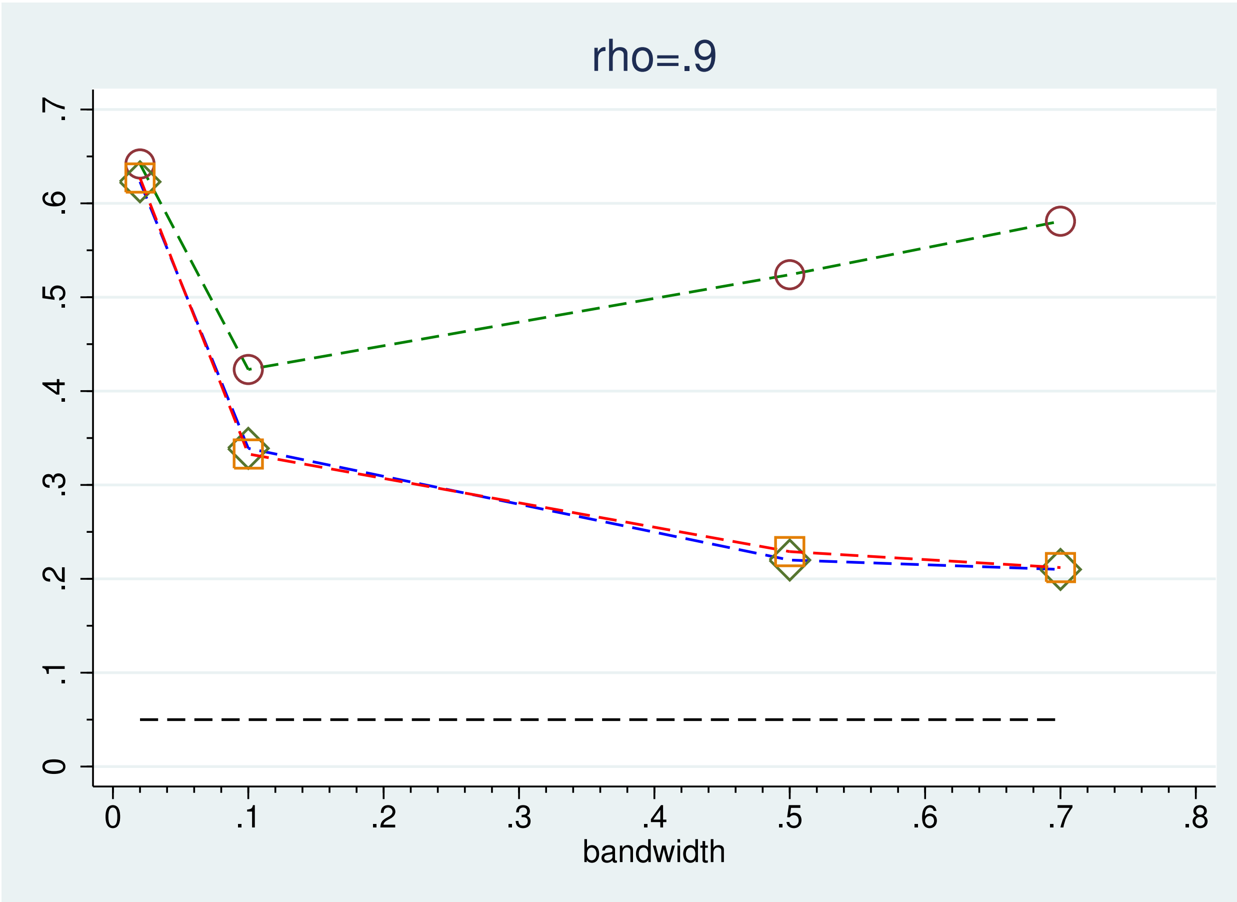
Figure E.5: Empirical null rejection probabilities, no spatial correlation, Bartlett kernel, $N = 250$, $\lambda = 0.5$.

Figure E.5: (cont'd)



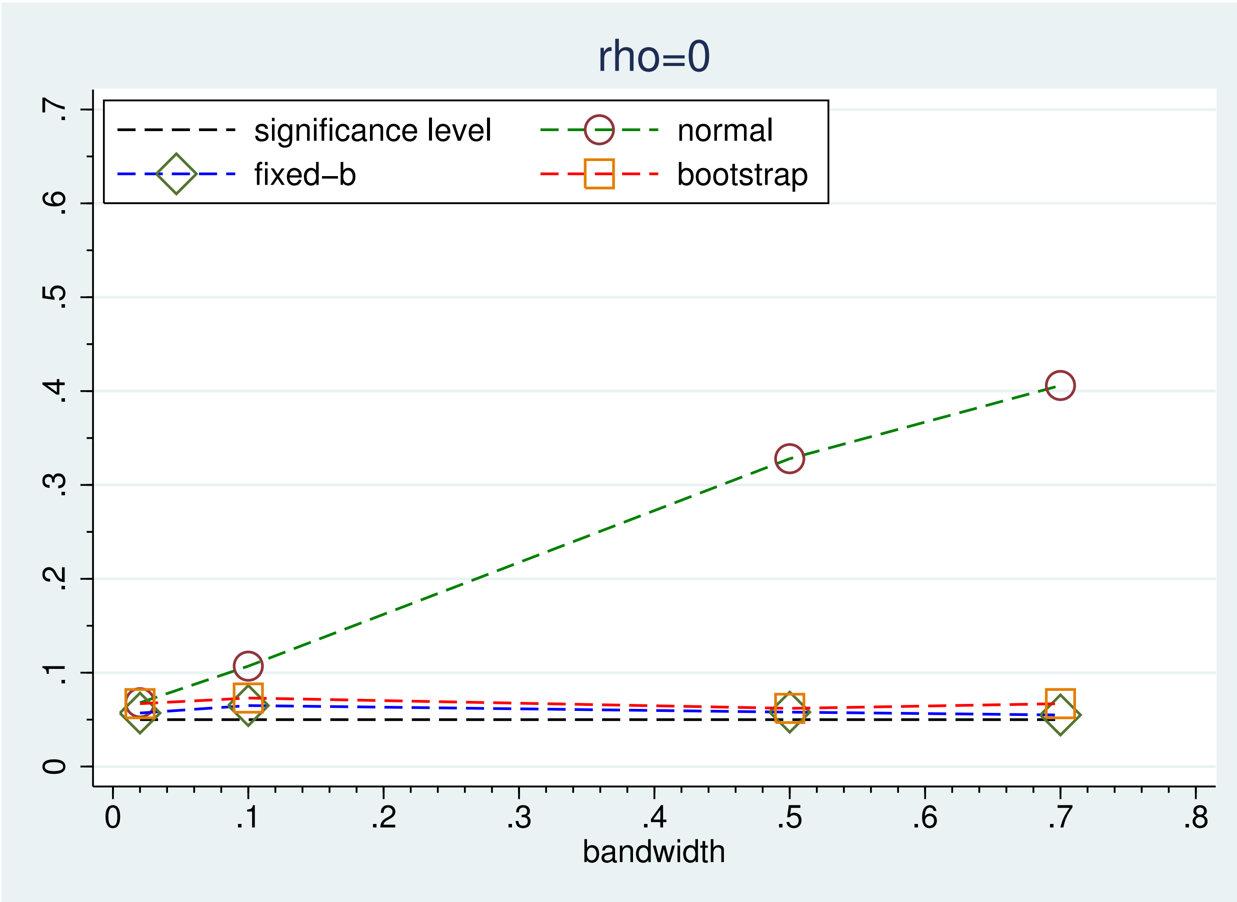
(b) N=250, T=50

Figure E.5: (cont'd)



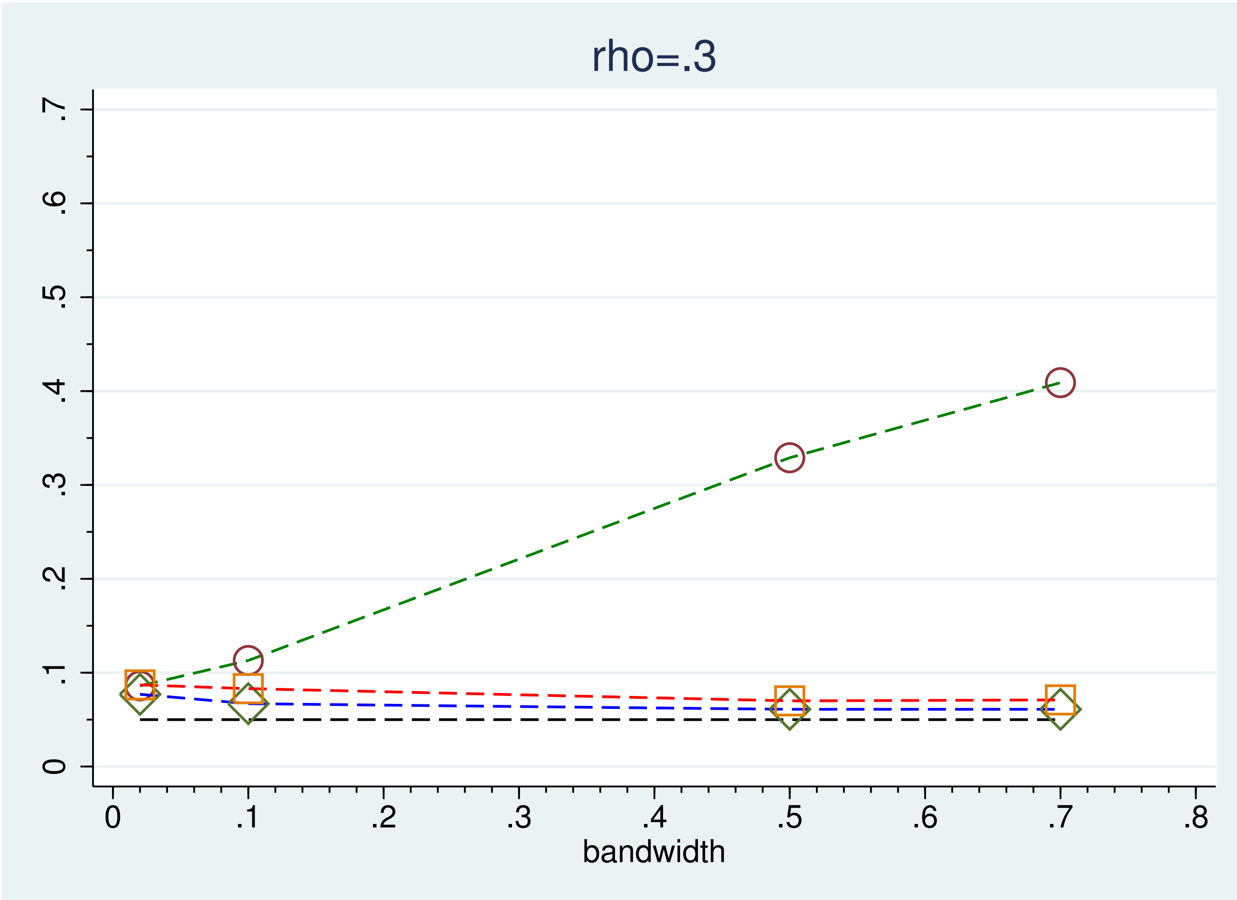
(c) N=250, T=50

Figure E.5: (cont'd)



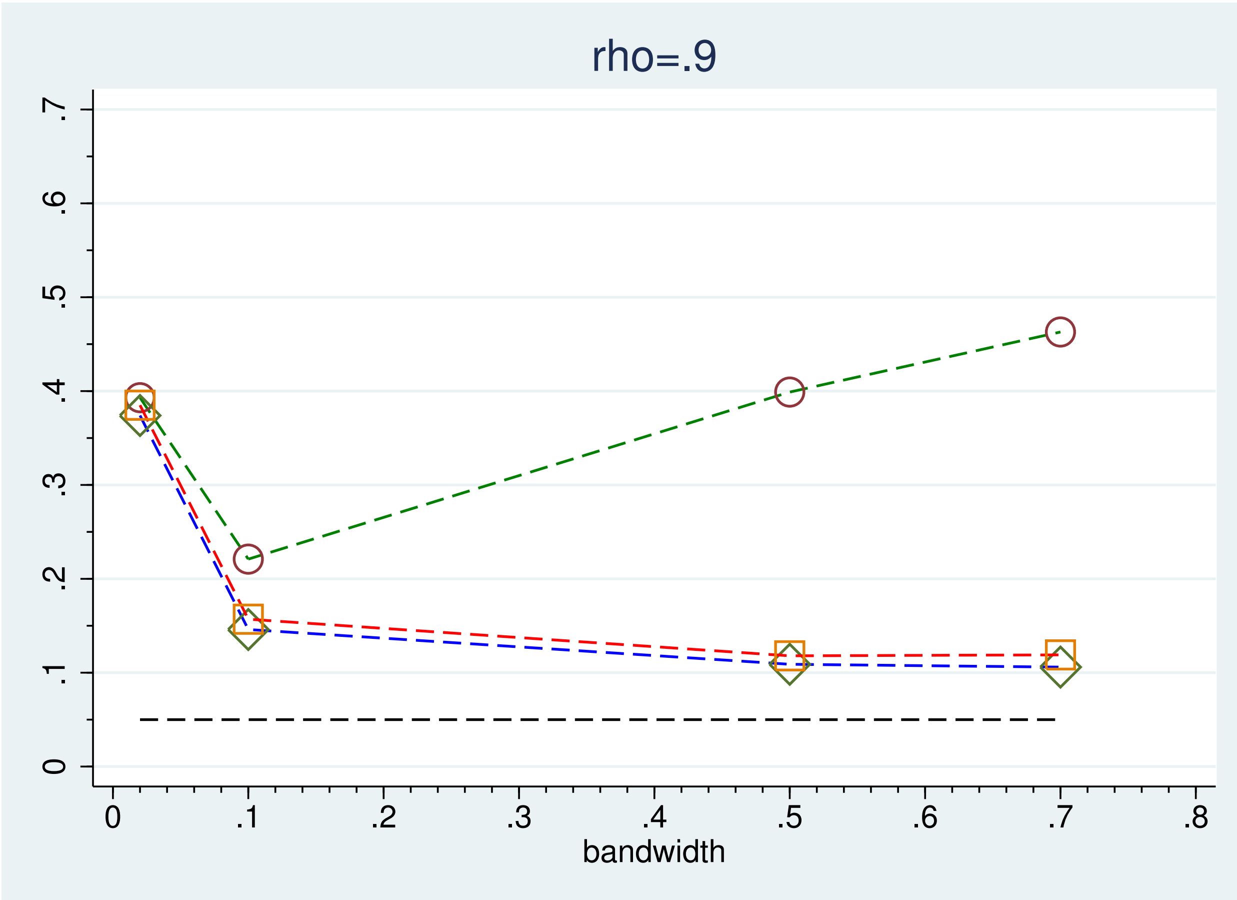
(d) N=250, T=250

Figure E.5: (cont'd)

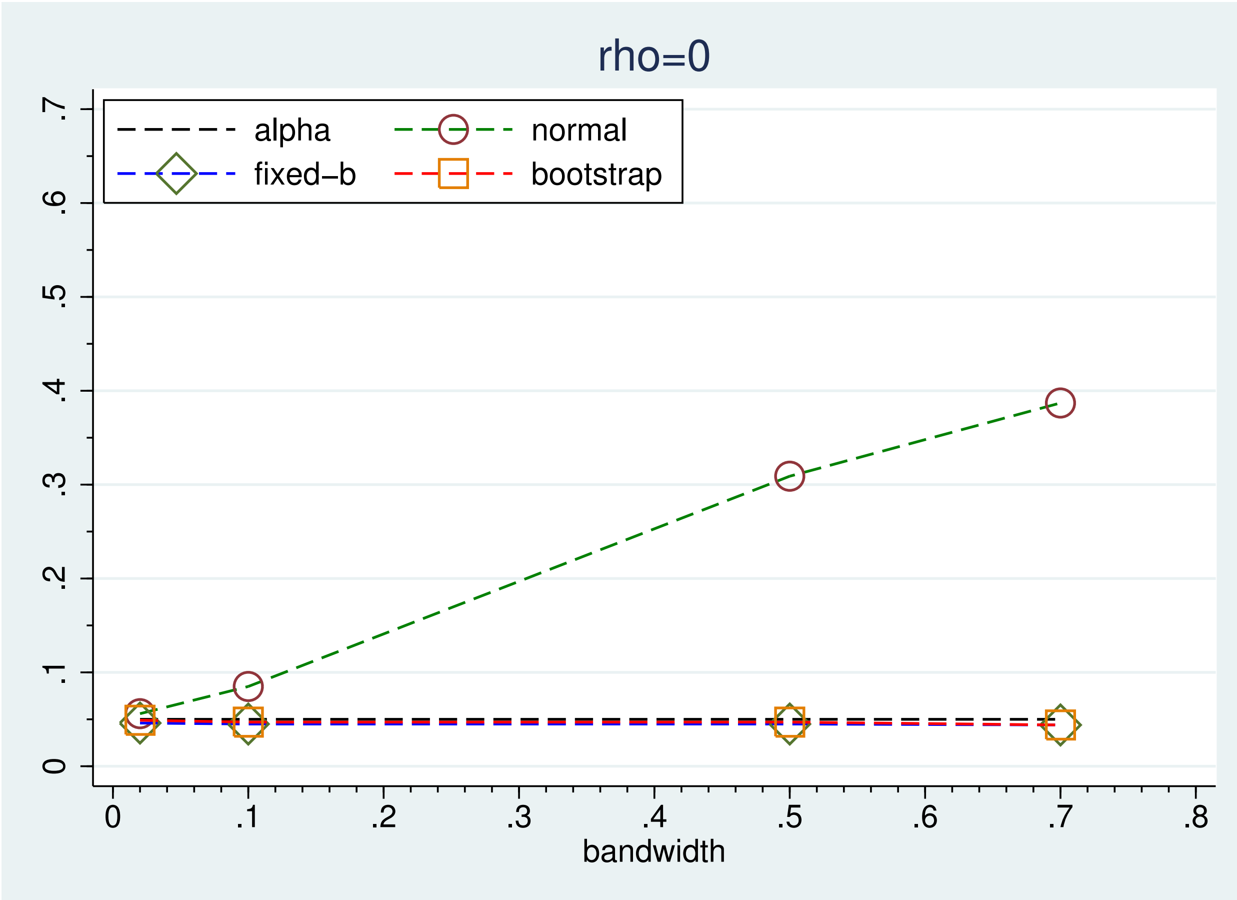


(e) N=250, T=250

Figure E.5: (cont'd)



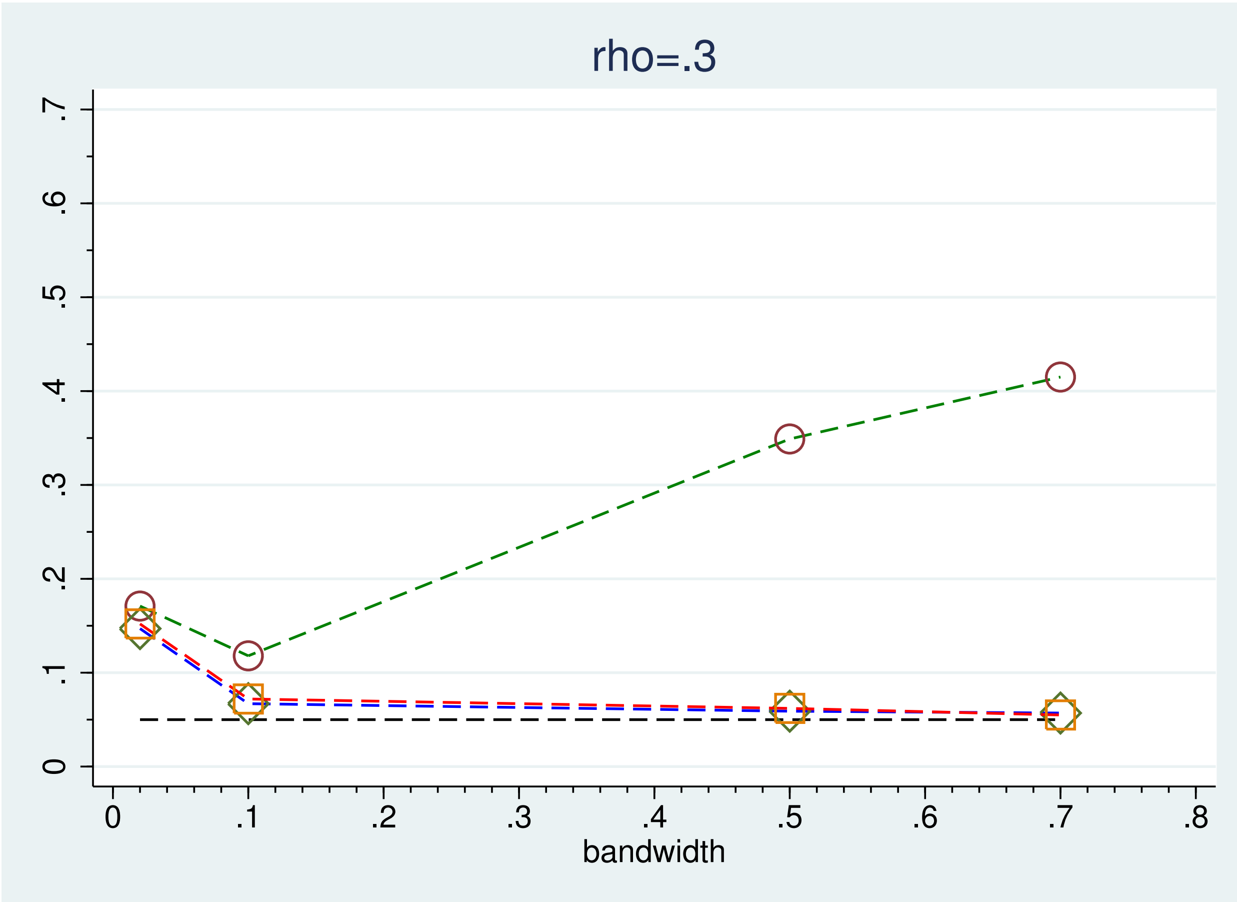
(f) N=250, T=250



(a) N=256, T=50

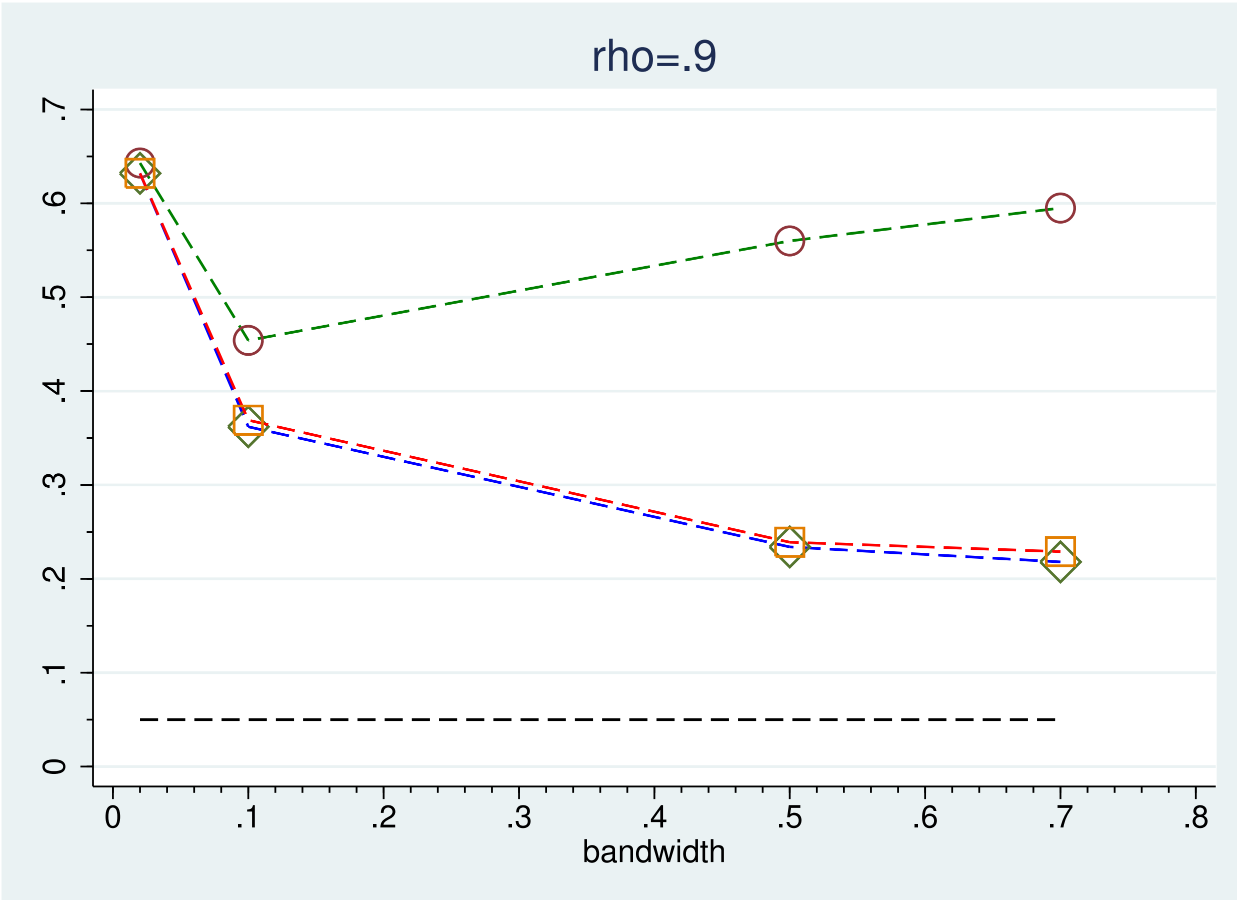
Figure E.6: Empirical null rejection probabilities, spatial $MA(2)$, Bartlett kernel, $N = 256$, $\lambda = 0.5$.

Figure E.6: (cont'd)



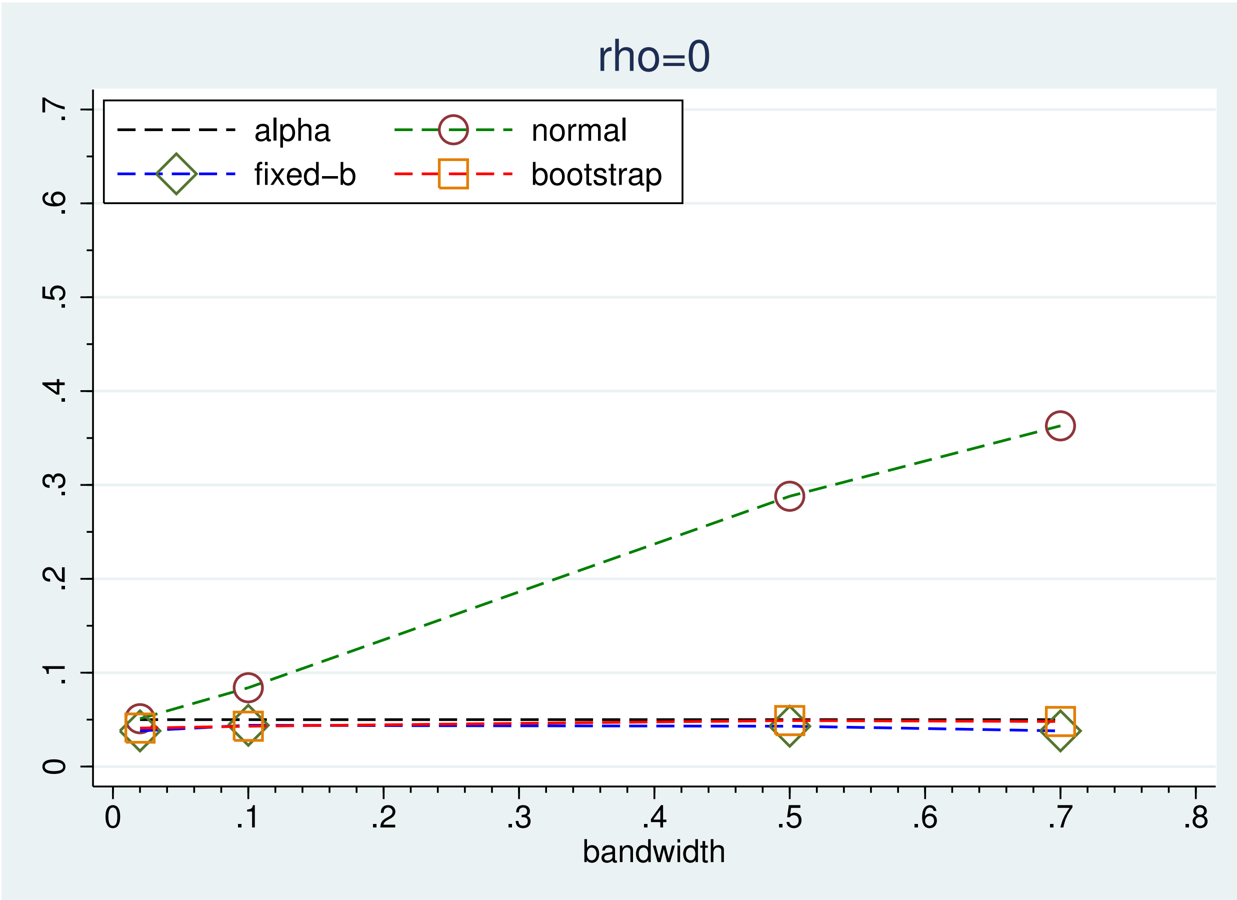
(b) N=256, T=50

Figure E.6: (cont'd)



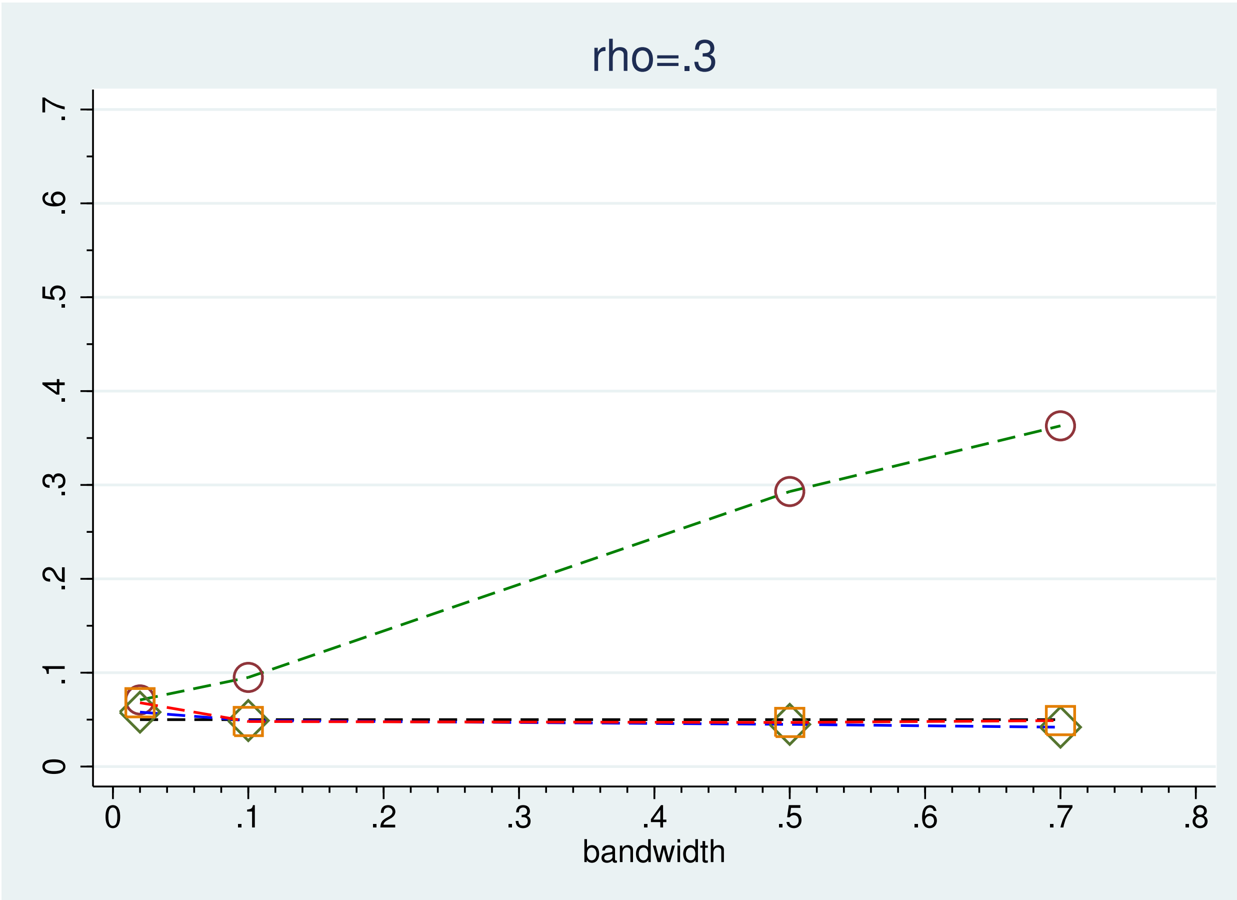
(c) N=256, T=50

Figure E.6: (cont'd)



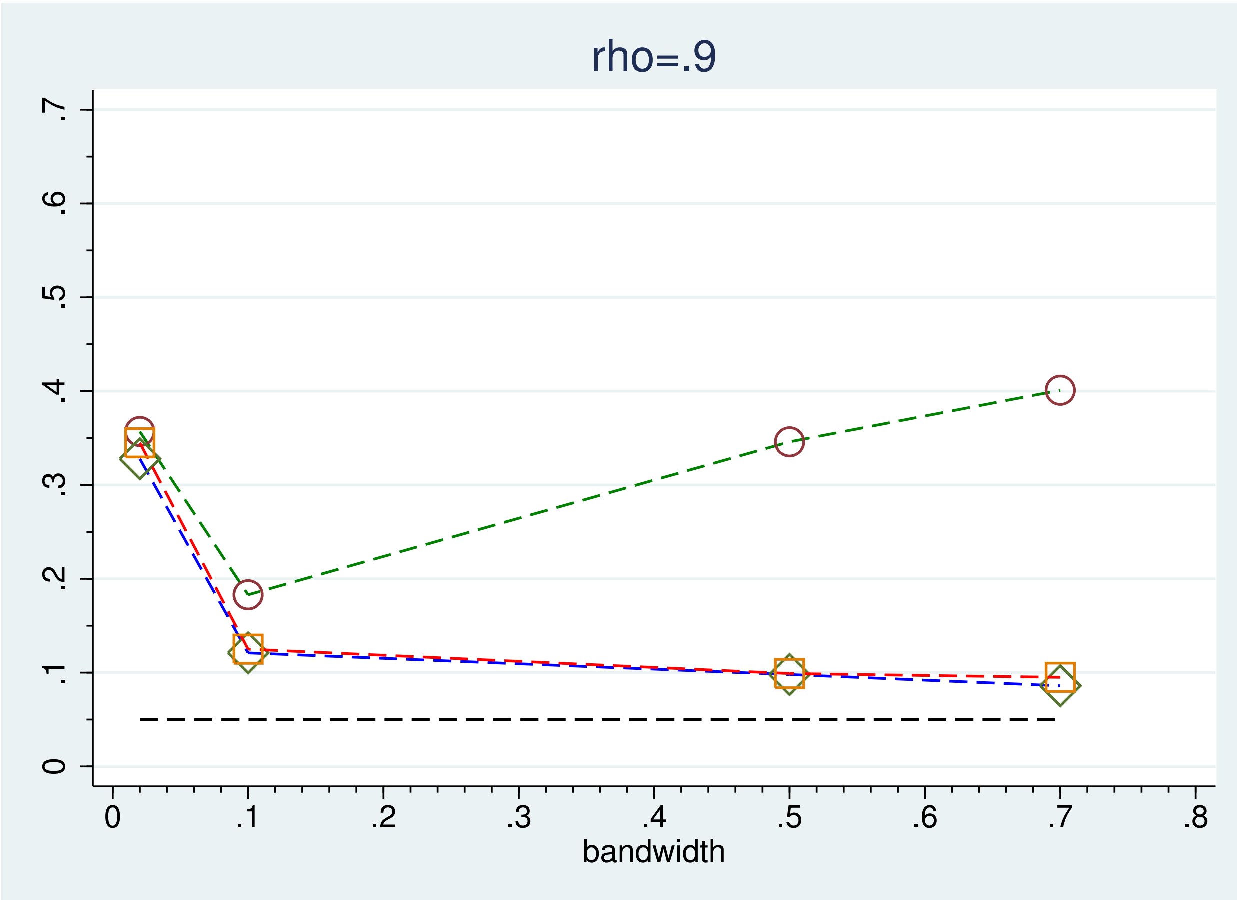
(d) N=256, T=250

Figure E.6: (cont'd)

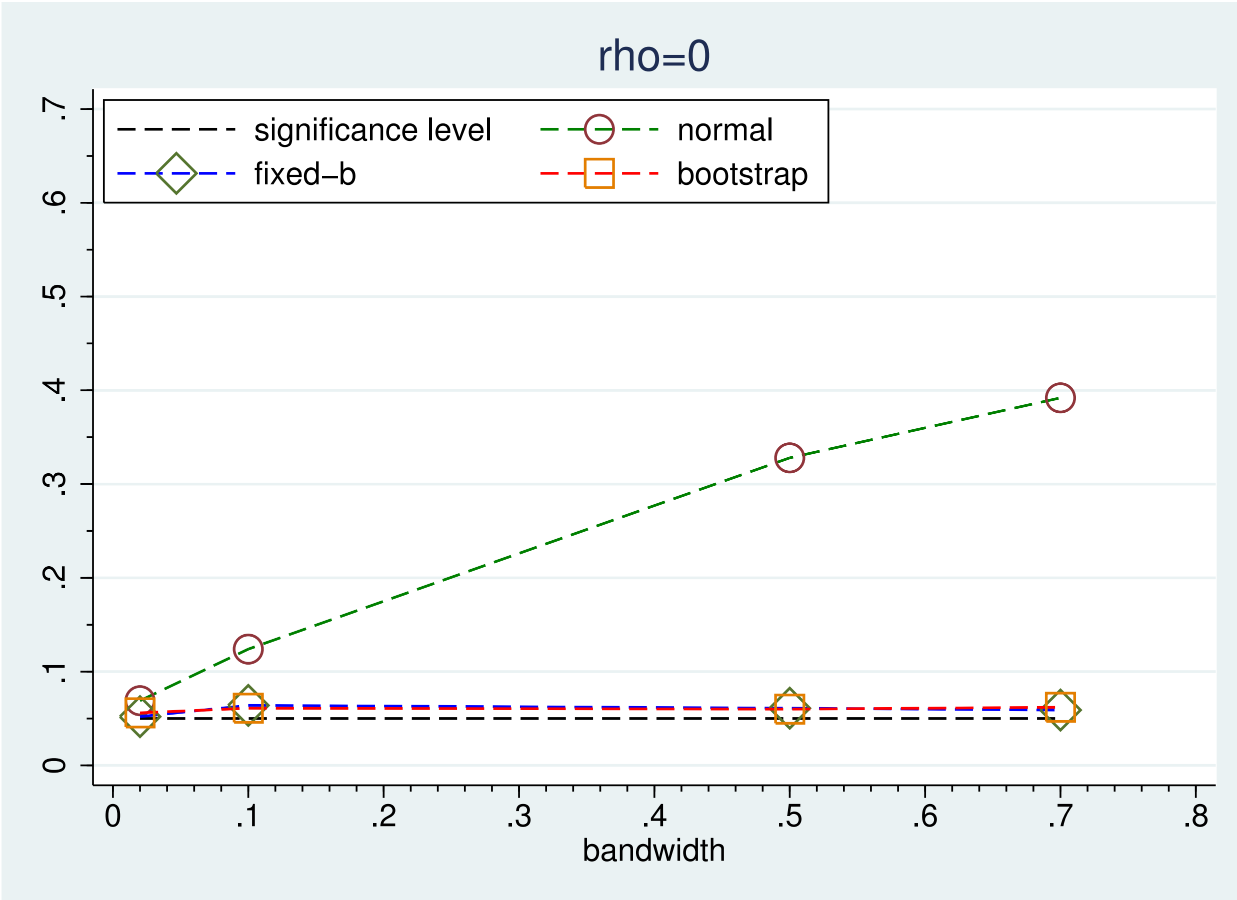


(e) N=256, T=250

Figure E.6: (cont'd)



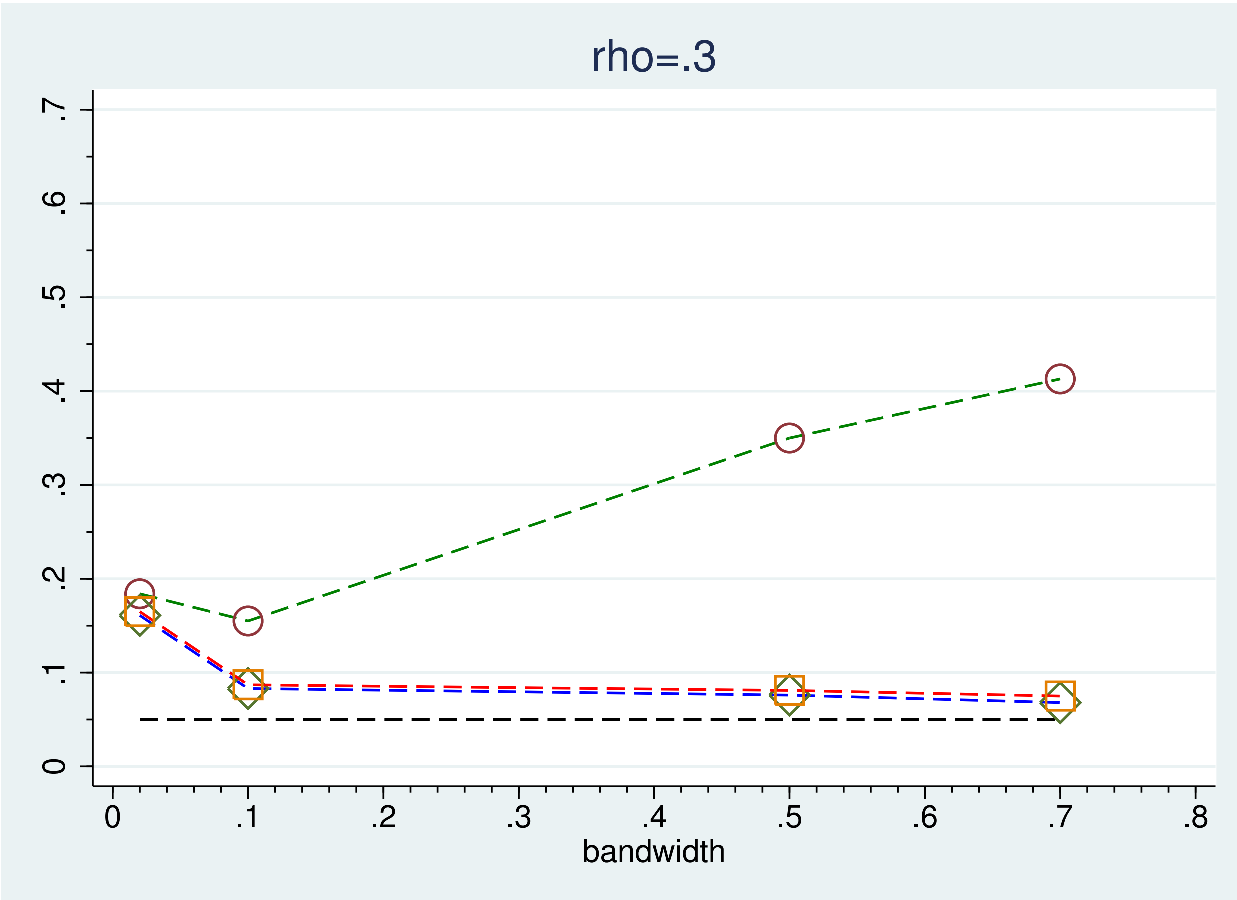
(f) N=256, T=250



(a) N=50, T=50

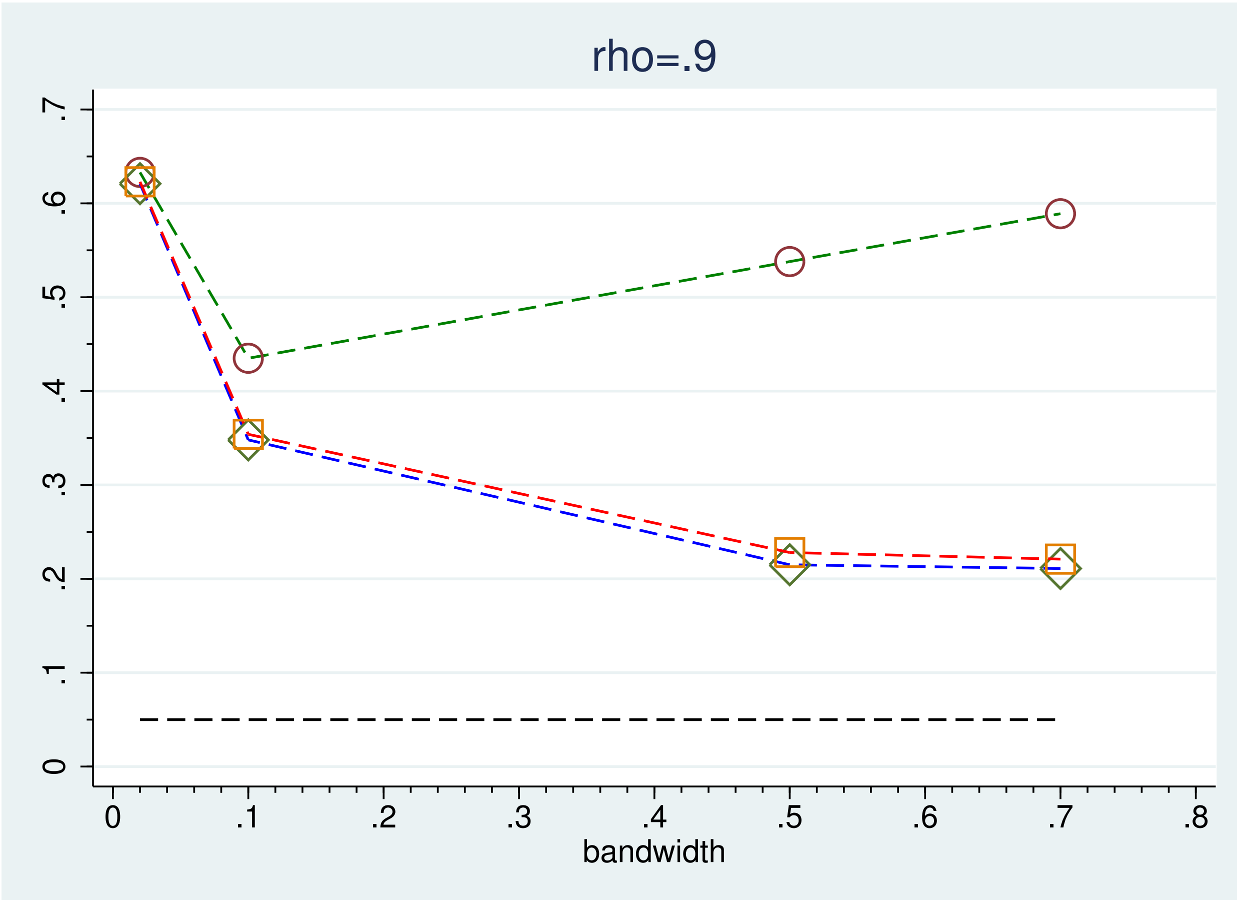
Figure E.7: Empirical null rejection probabilities, no spatial correlation, Bartlett kernel, $T = 50$, $\lambda = 0.5$.

Figure E.7: (cont'd)



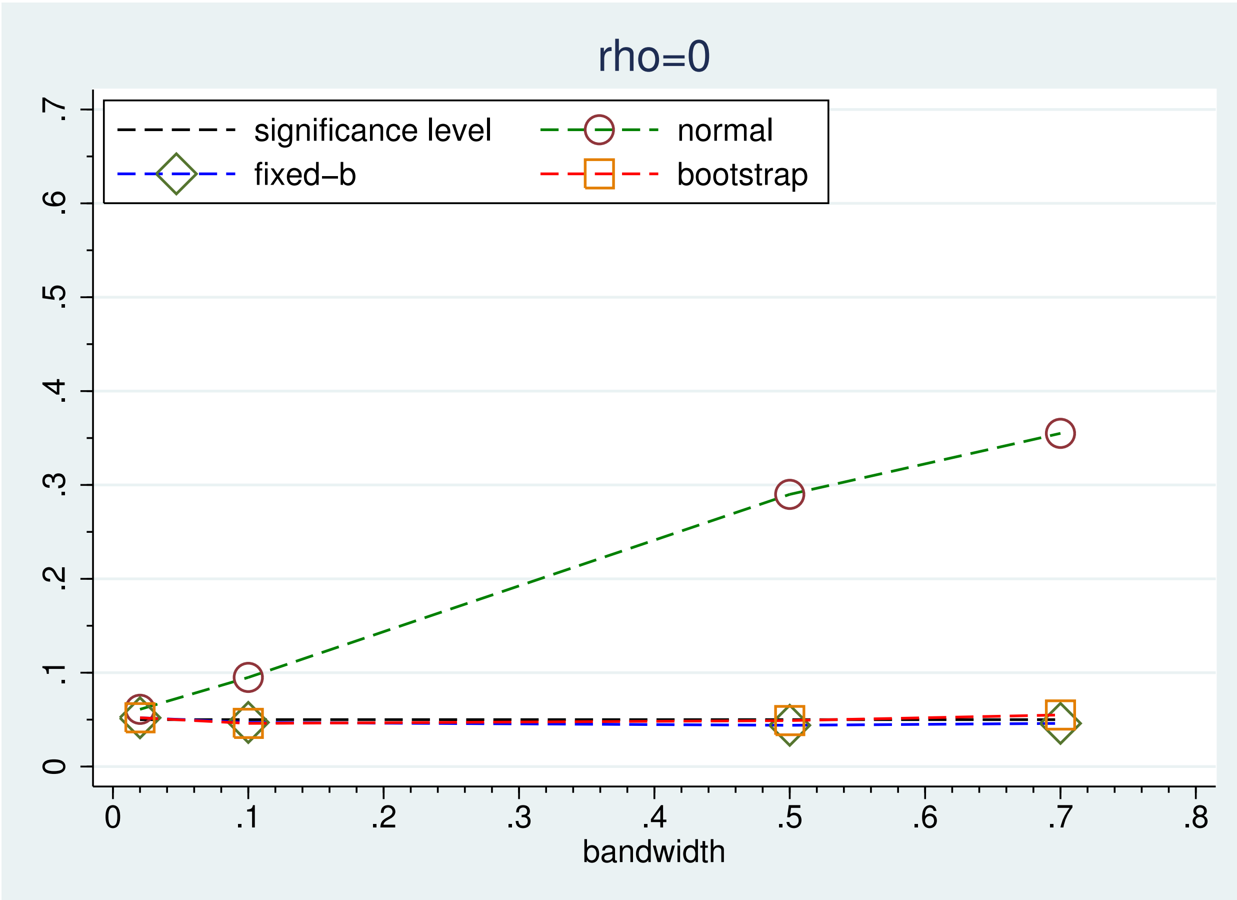
(b) N=50, T=50

Figure E.7: (cont'd)



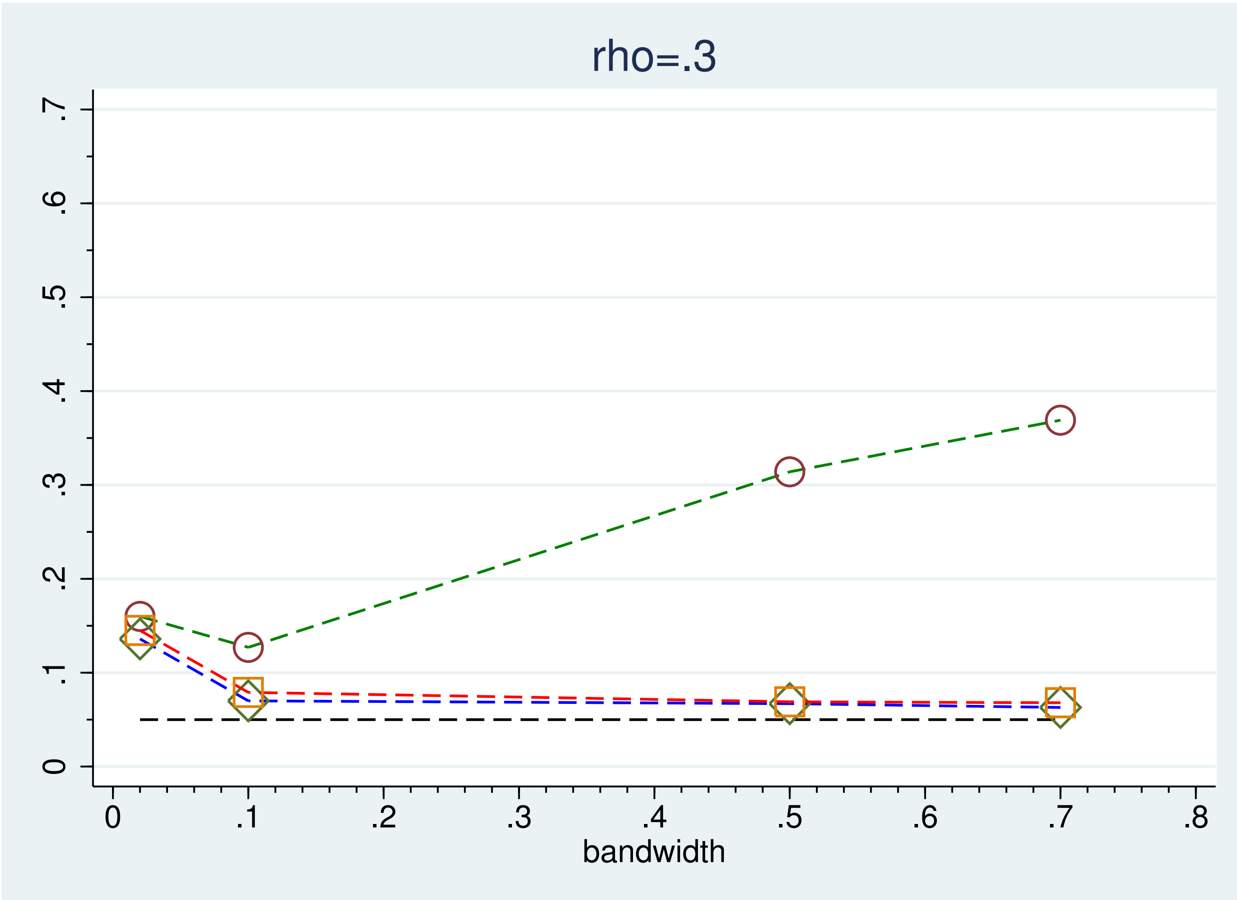
(c) N=50, T=50

Figure E.7: (cont'd)



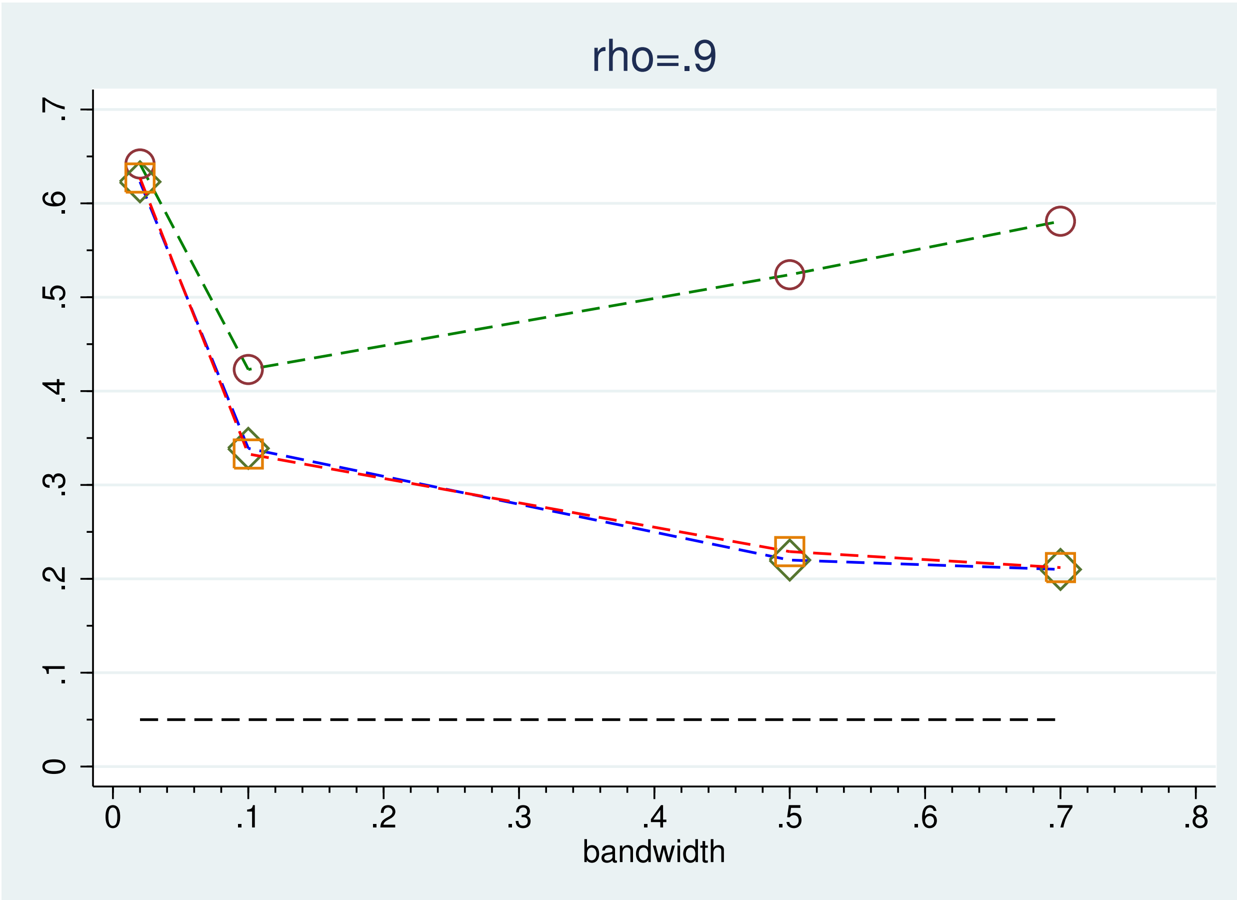
(d) N=250, T=50

Figure E.7: (cont'd)

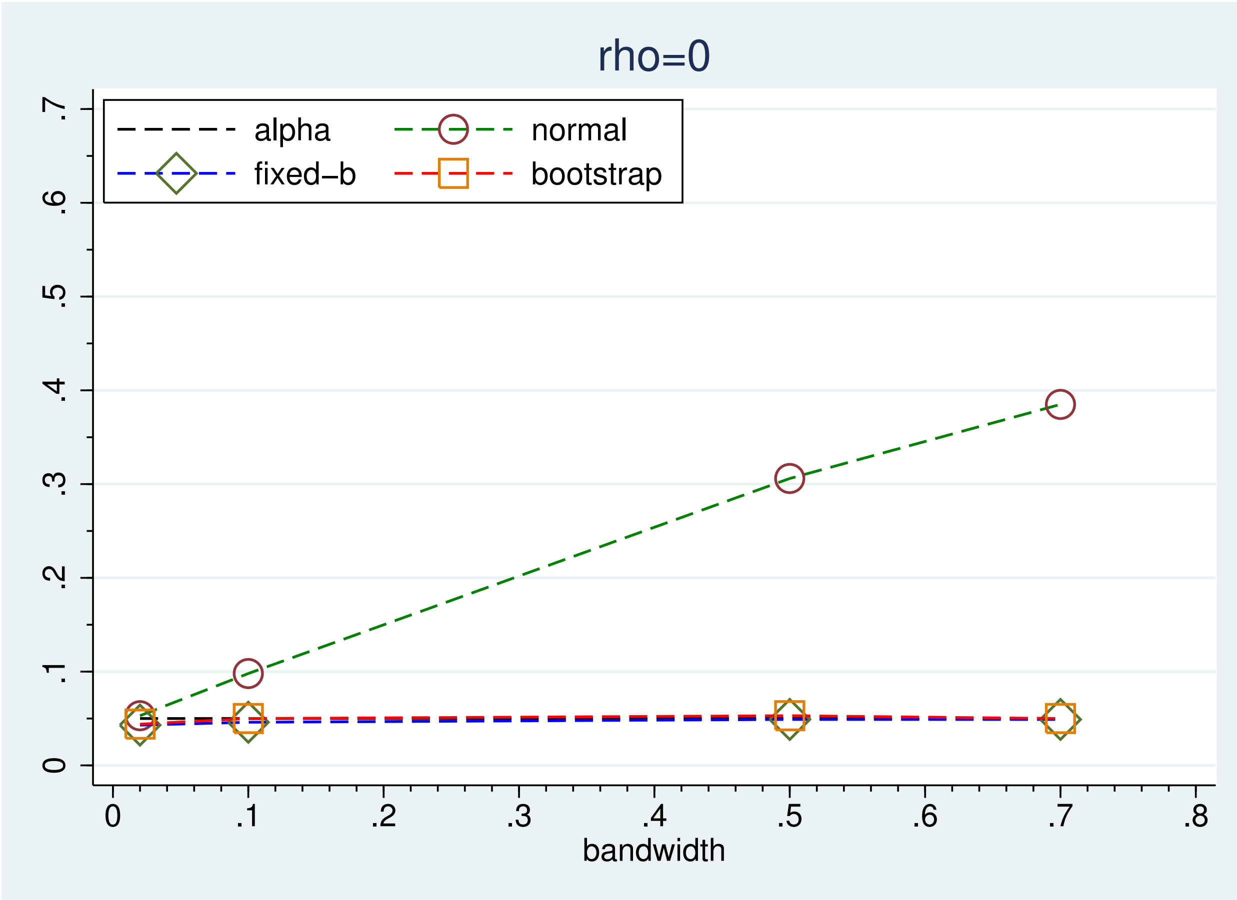


(e) N=250, T=50

Figure E.7: (cont'd)



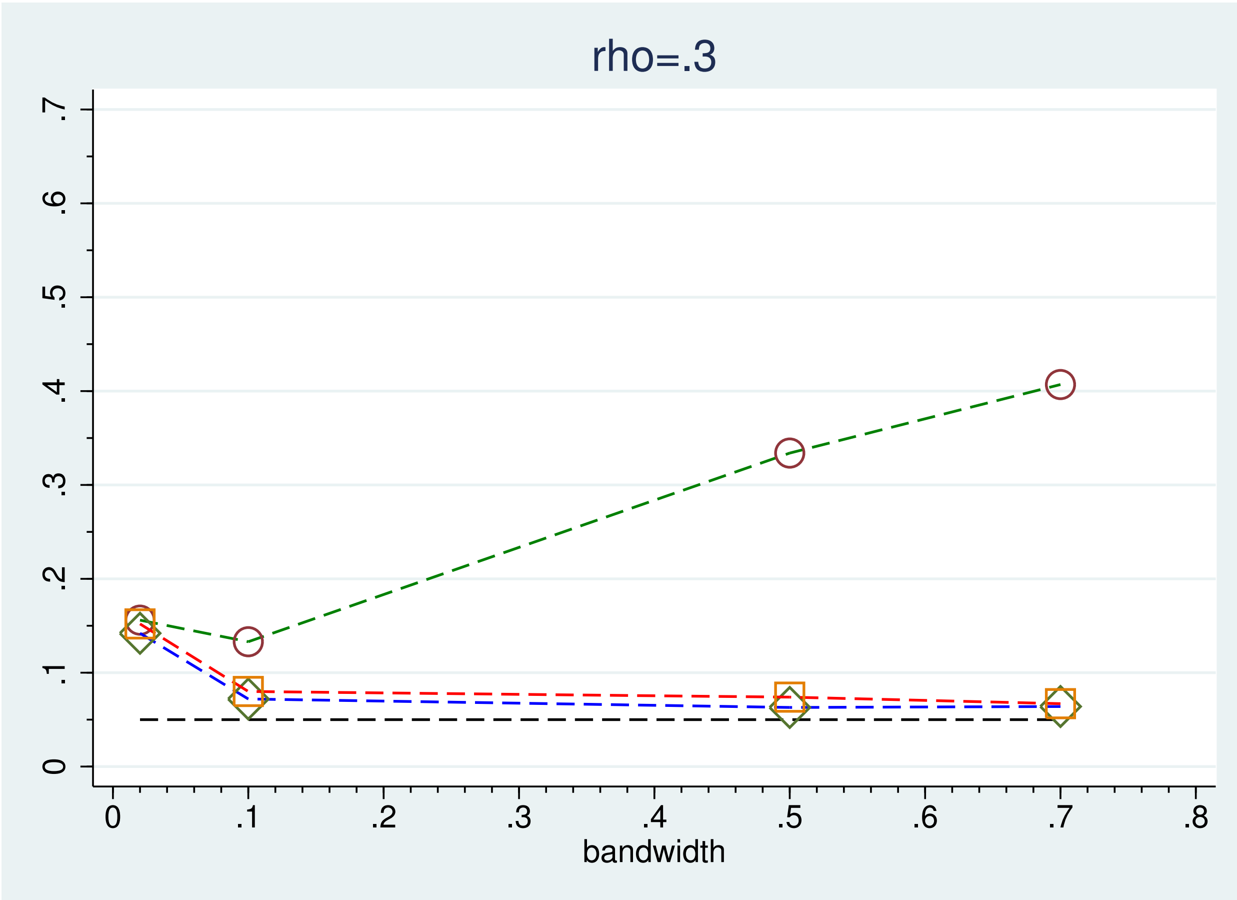
(f) N=250, T=50



(a) N=49, T=50

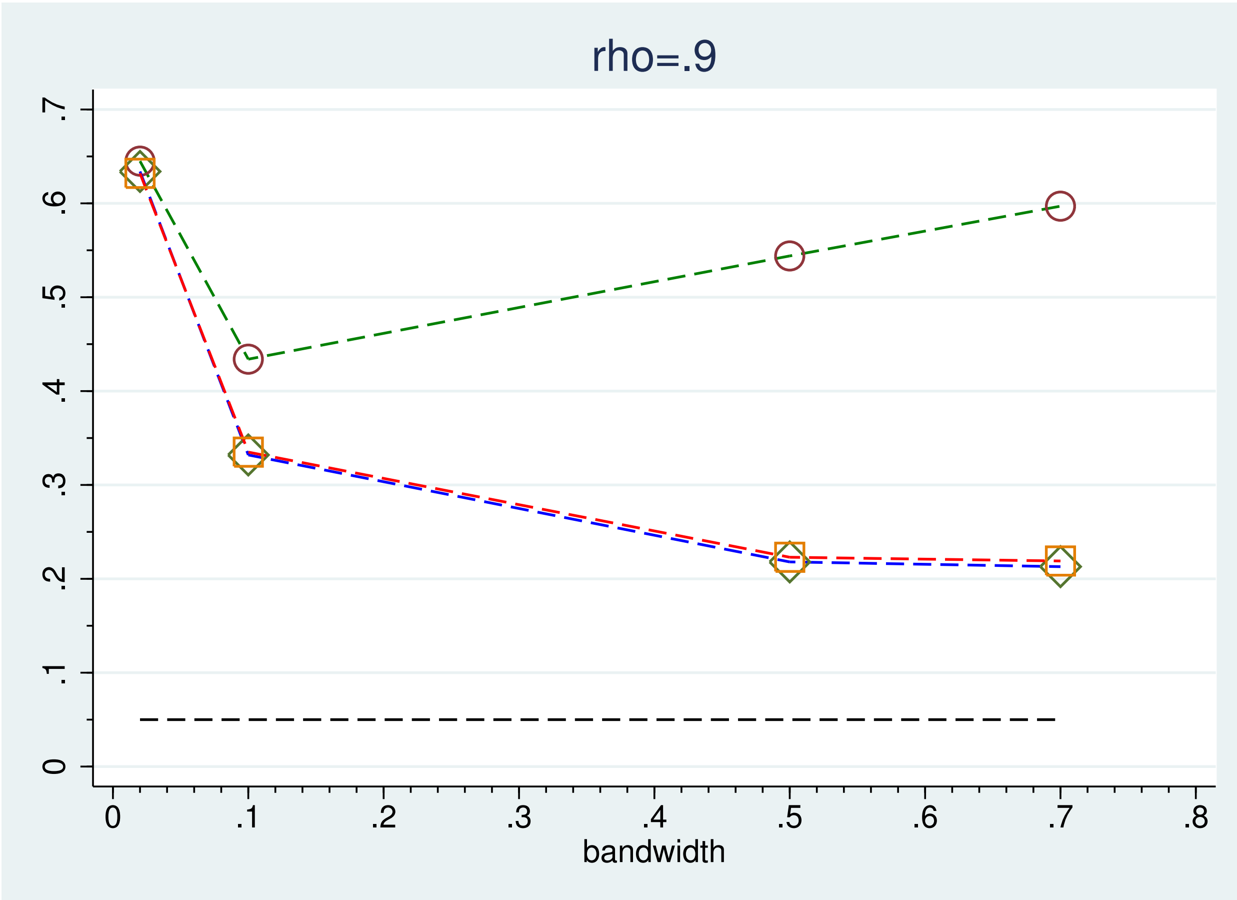
Figure E.8: Empirical null rejection probabilities, spatial $MA(2)$, Bartlett kernel, $T=49$, $\lambda=0.5$.

Figure E.8: (cont'd)



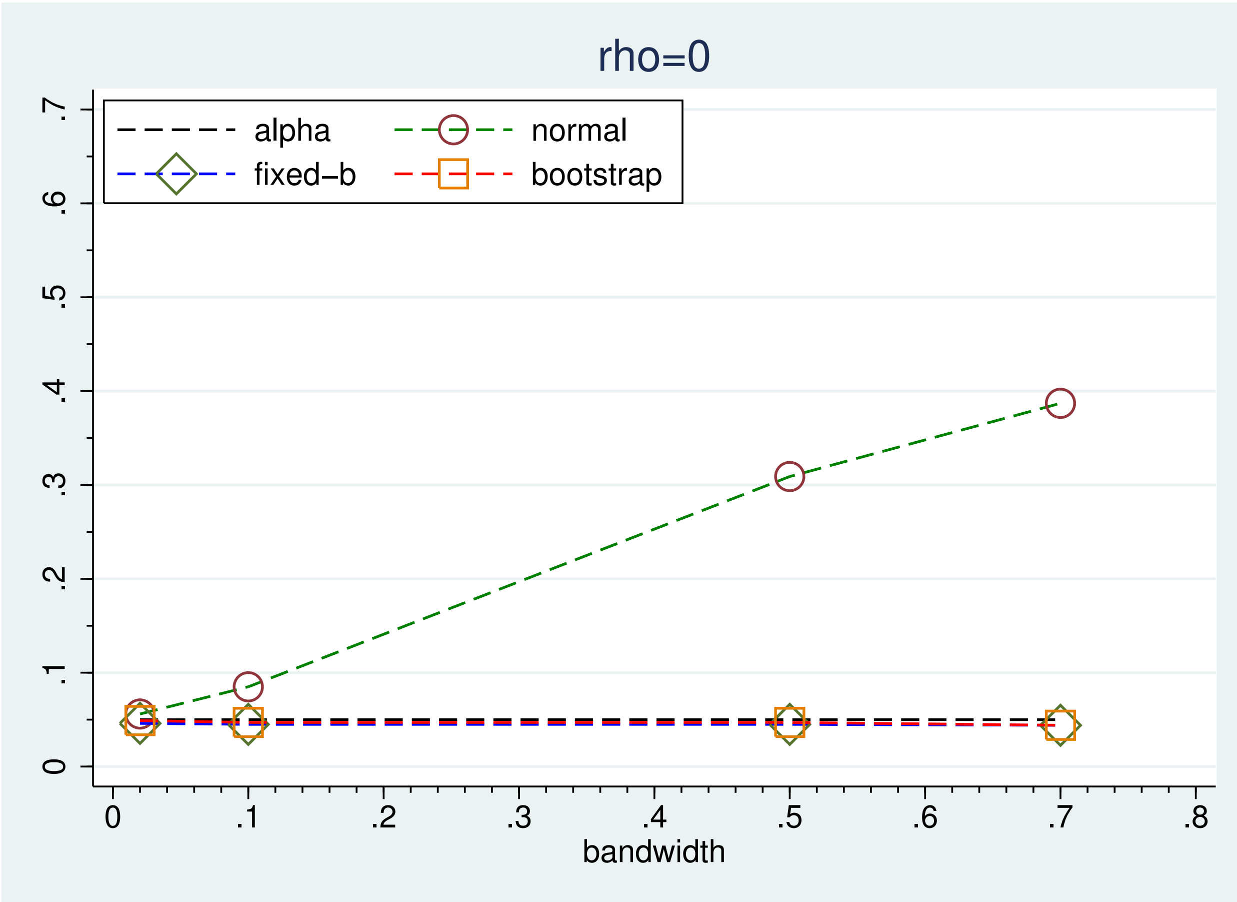
(b) N=49, T=50

Figure E.8: (cont'd)



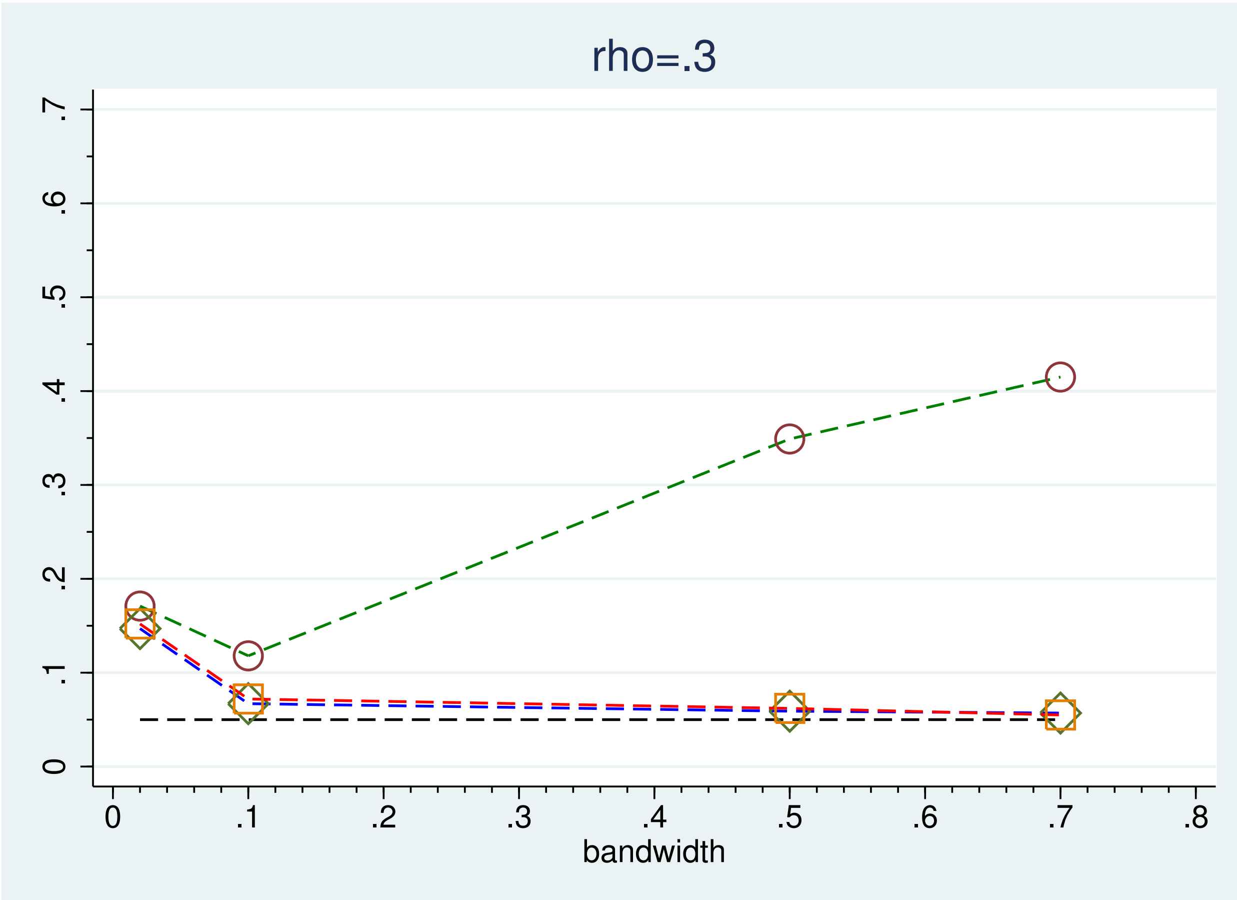
(c) N=49, T=50

Figure E.8: (cont'd)



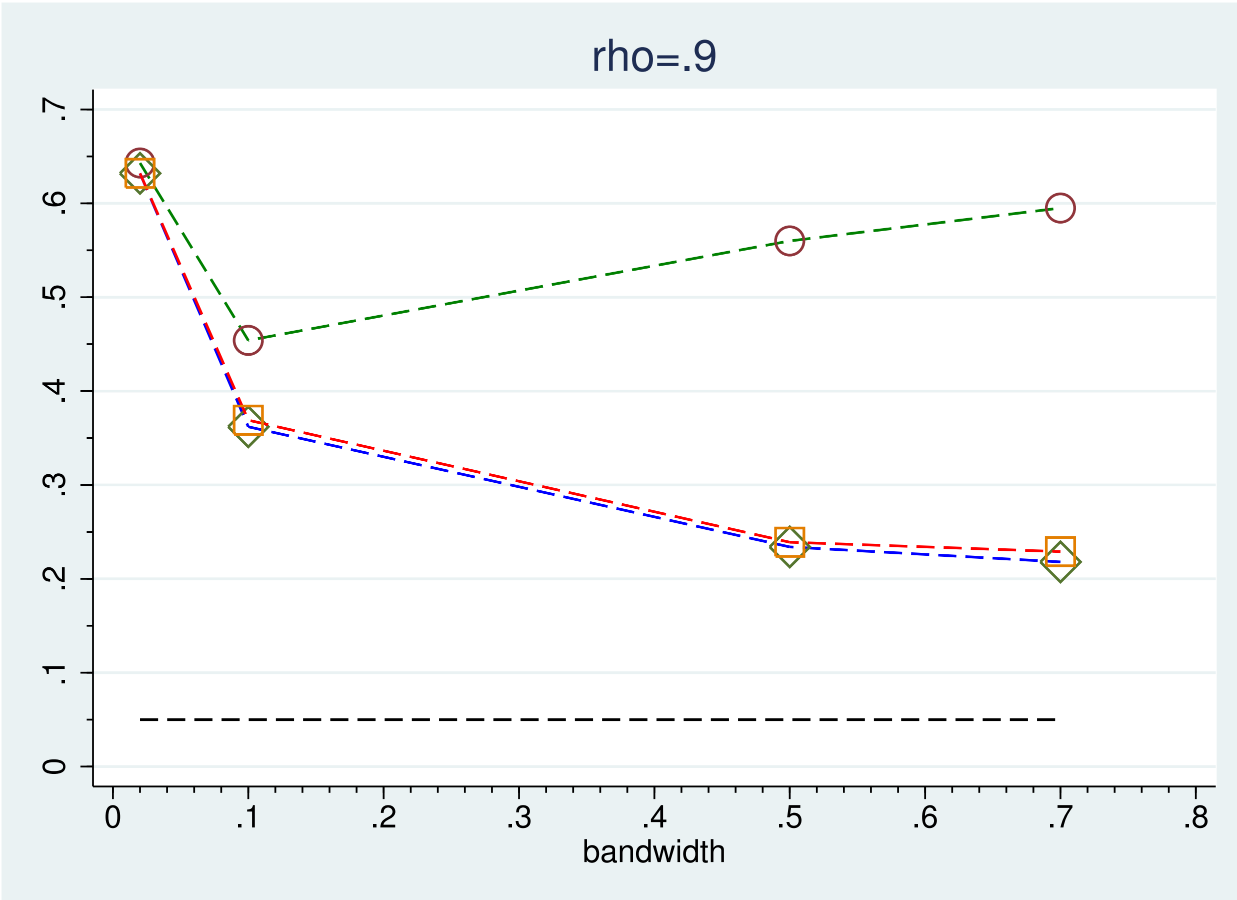
(d) N=256, T=50

Figure E.8: (cont'd)

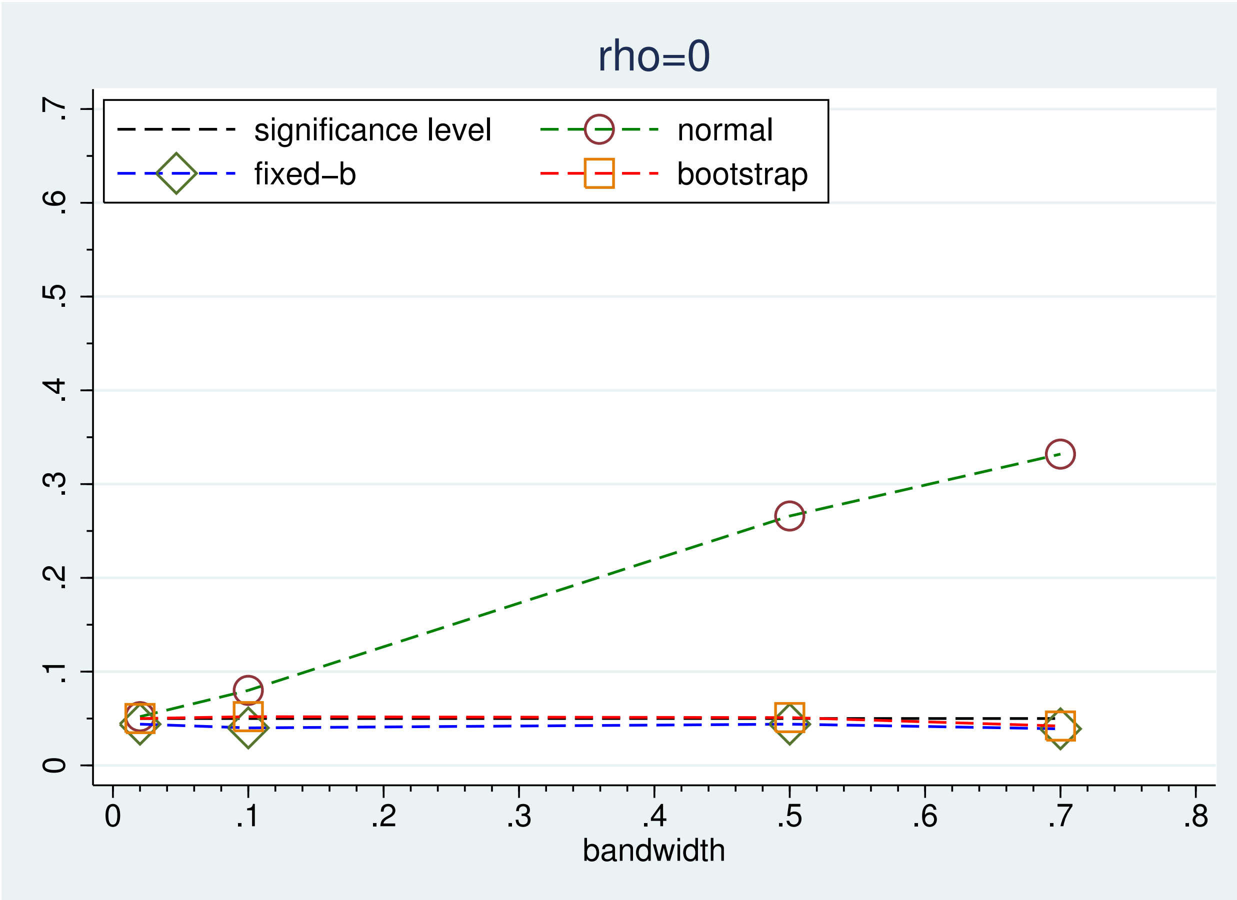


(e) N=256, T=50

Figure E.8: (cont'd)



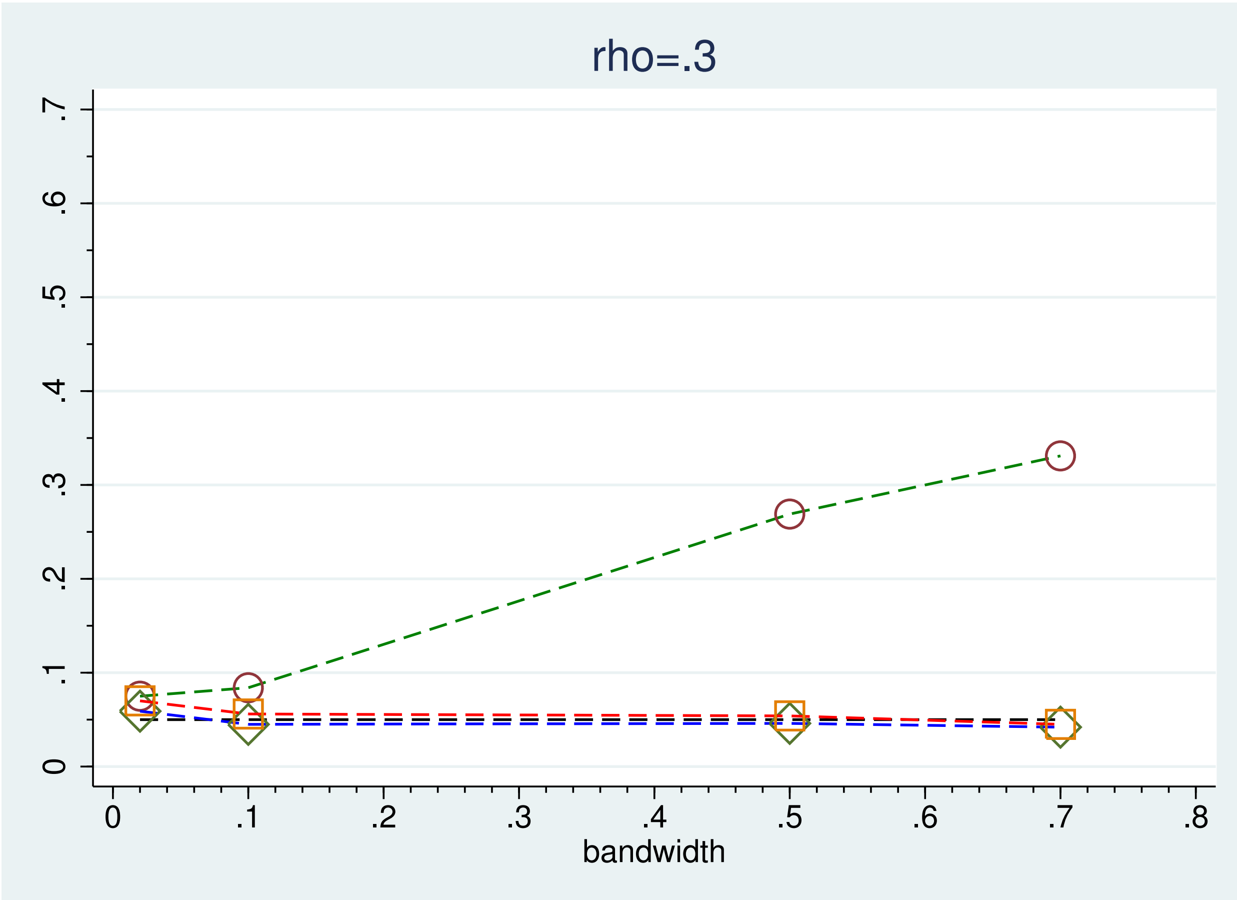
(f) N=256, T=50



(a) N=50, T=250

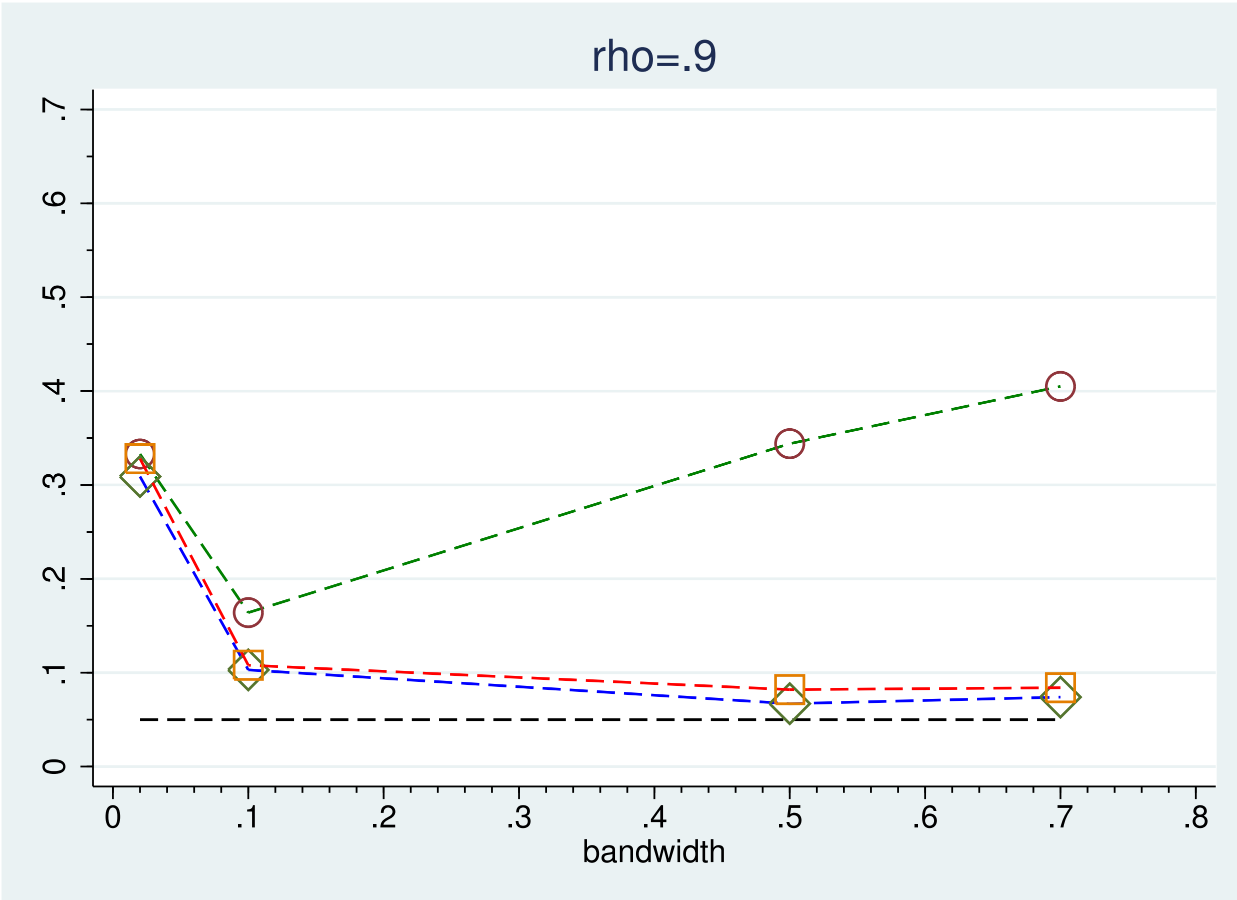
Figure E.9: Empirical null rejection probabilities, no spatial correlation, Bartlett kernel, $T = 250$, $\lambda = 0.5$.

Figure E.9: (cont'd)



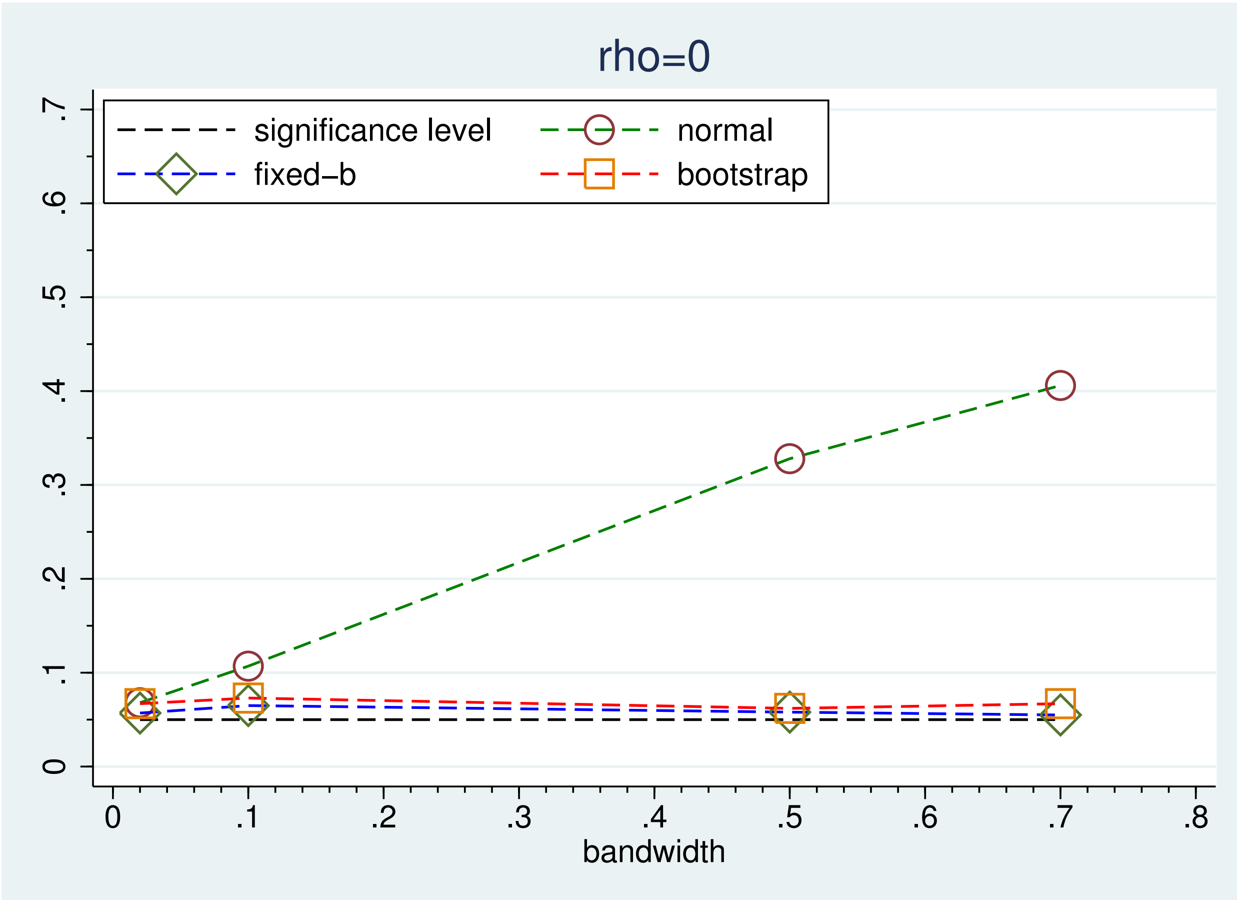
(b) N=50, T=250

Figure E.9: (cont'd)



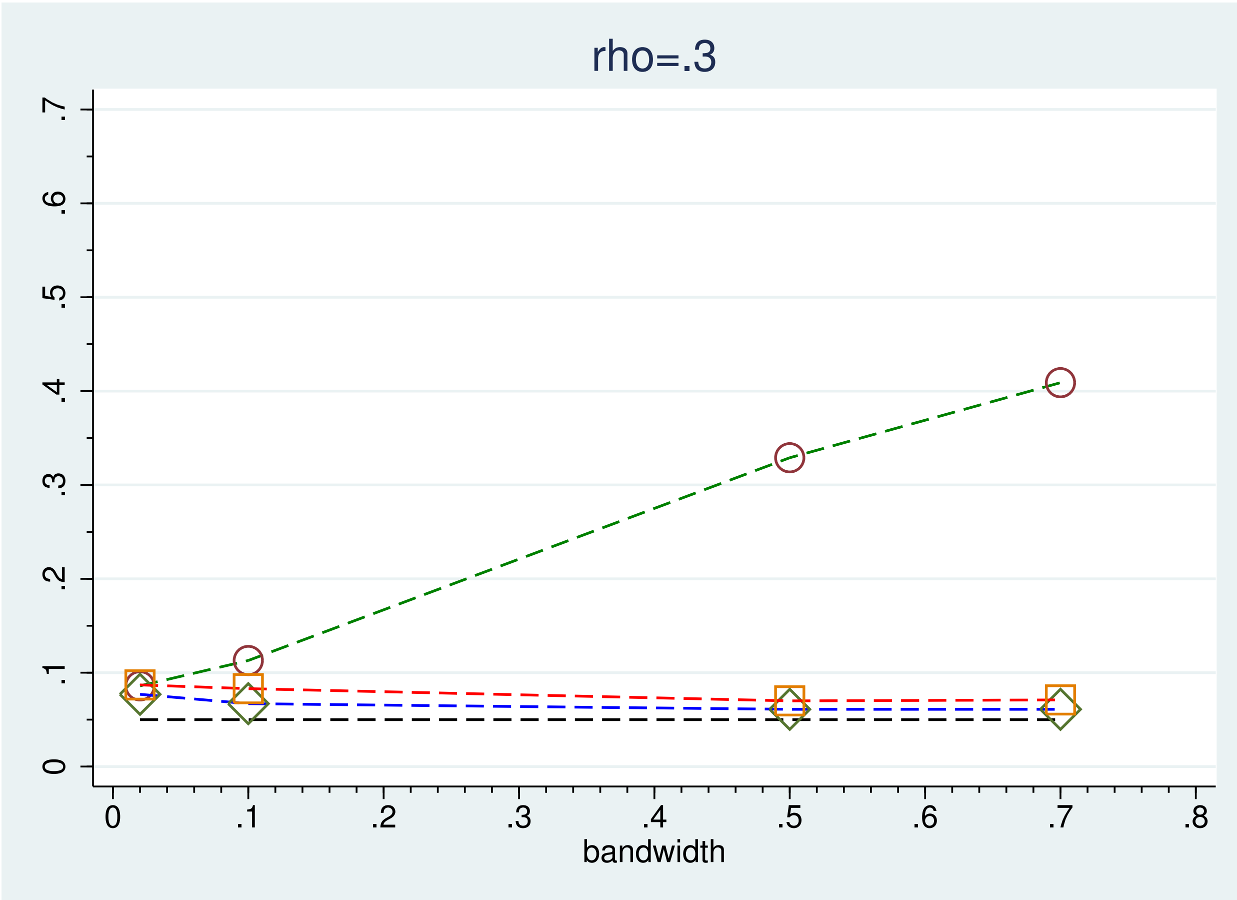
(c) N=50, T=250

Figure E.9: (cont'd)



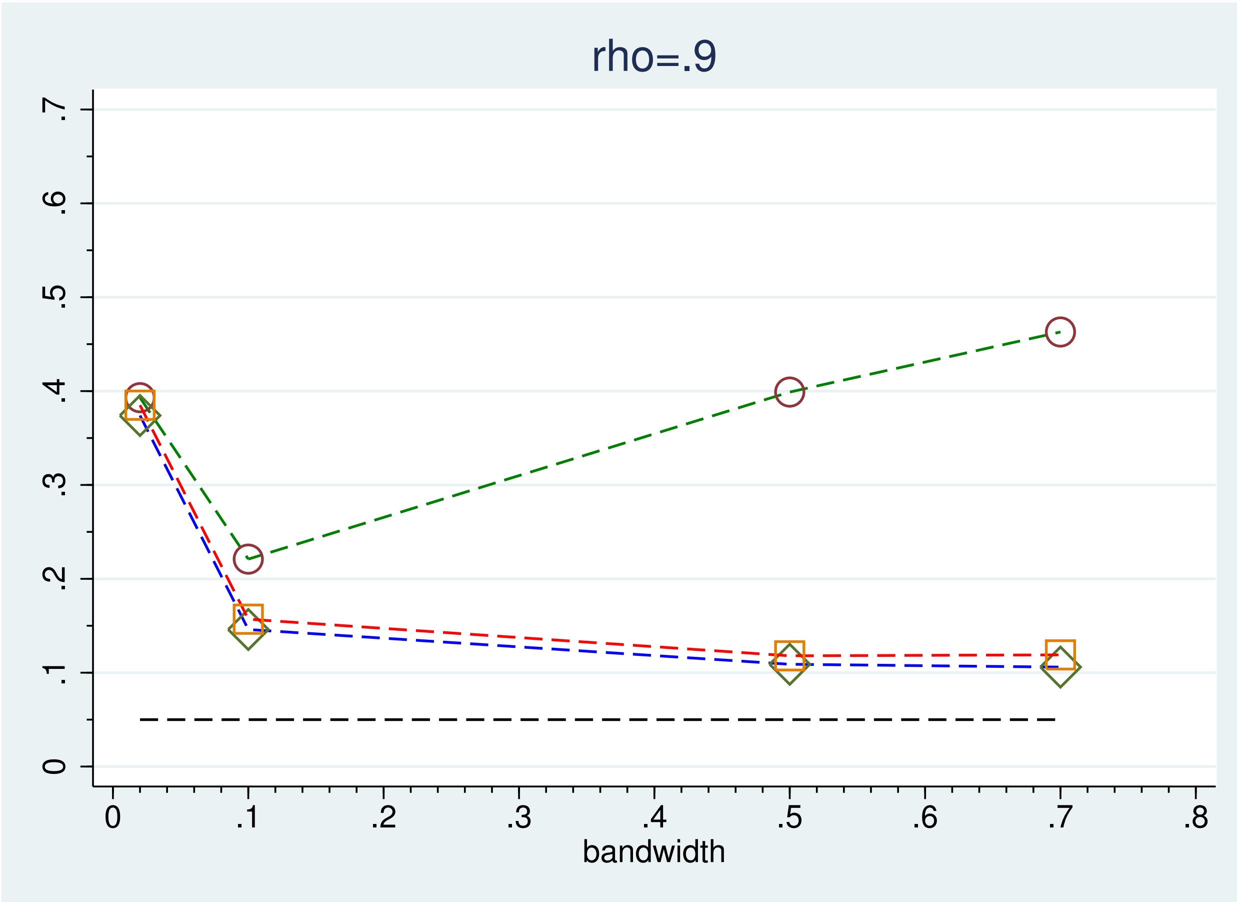
(d) N=250, T=250

Figure E.9: (cont'd)

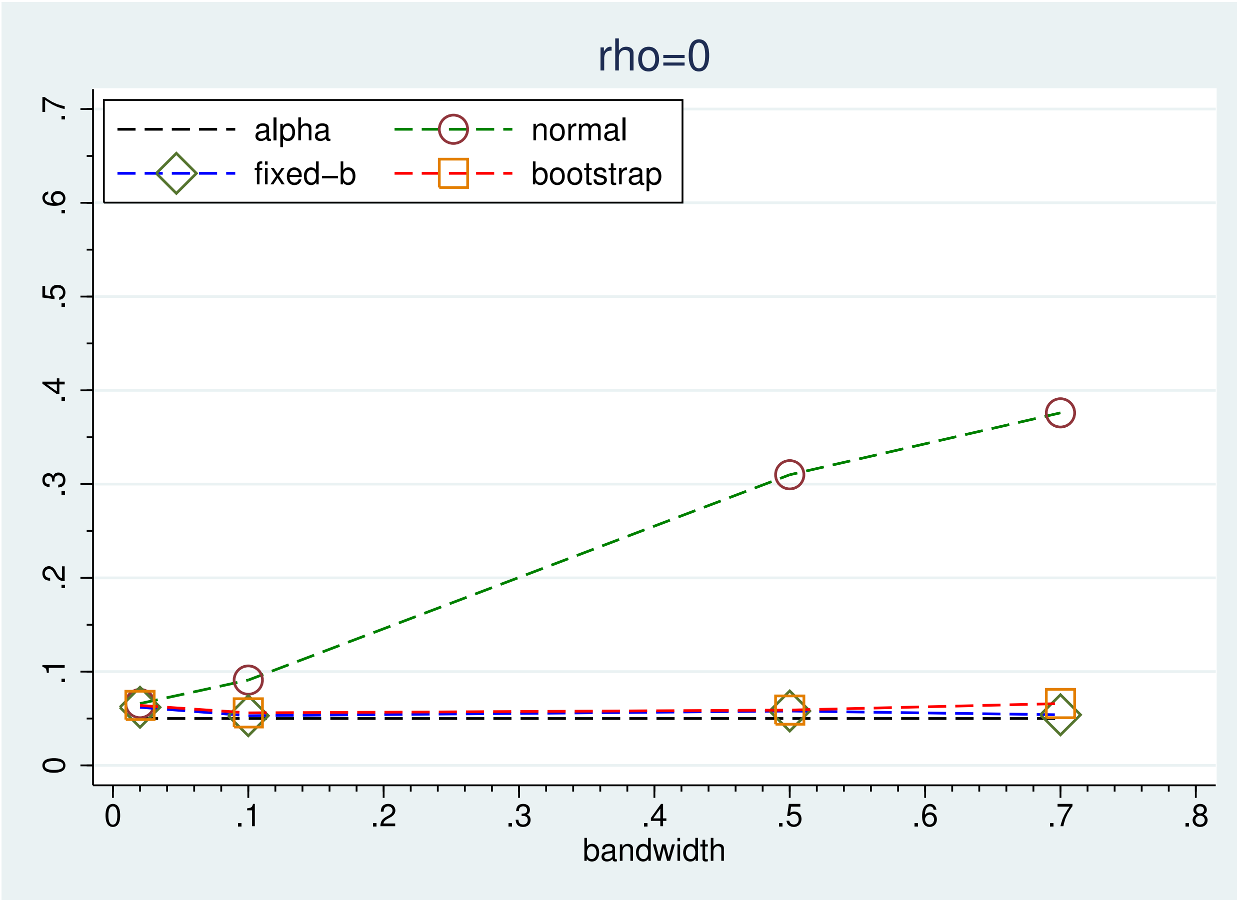


(e) N=250, T=250

Figure E.9: (cont'd)



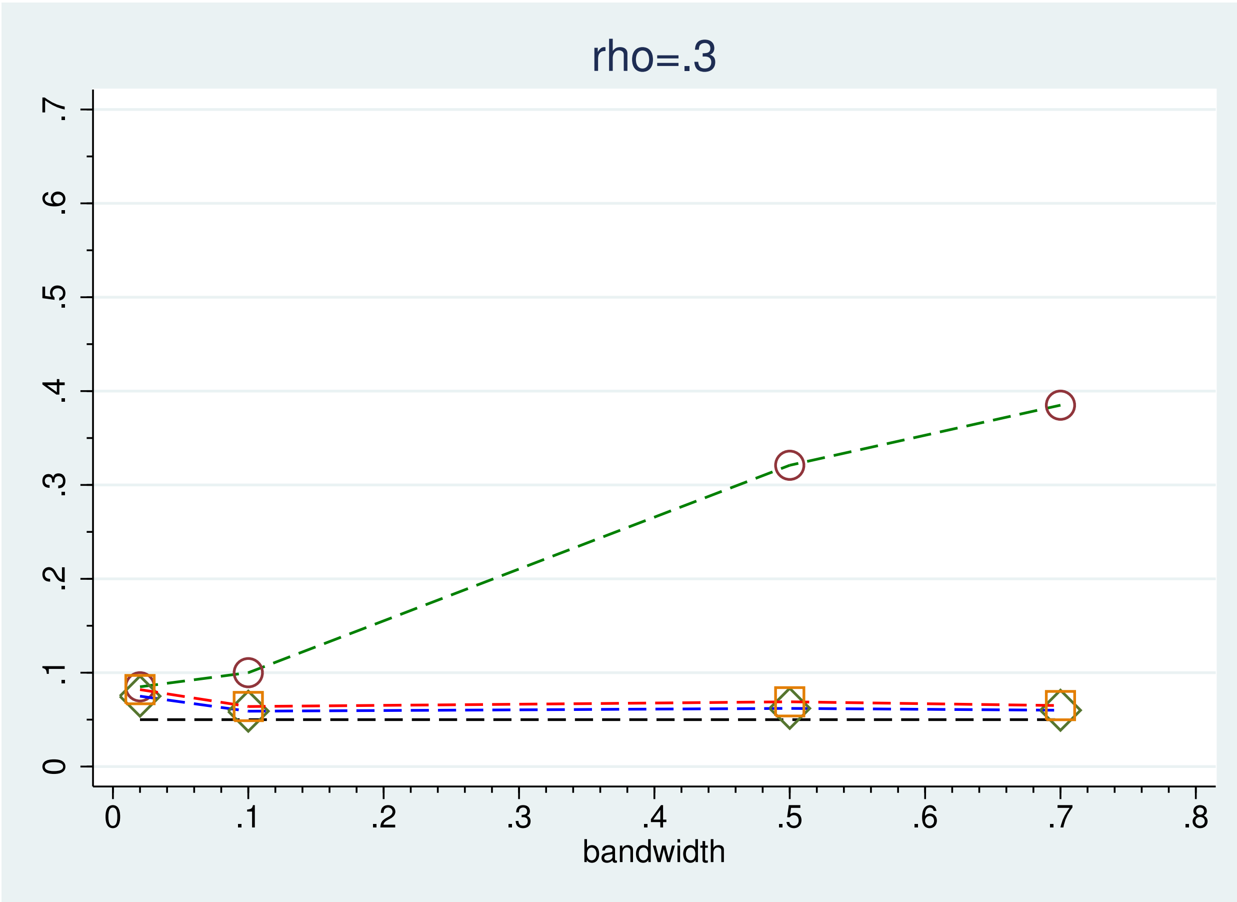
(f) N=250, T=250



(a) N=49, T=250

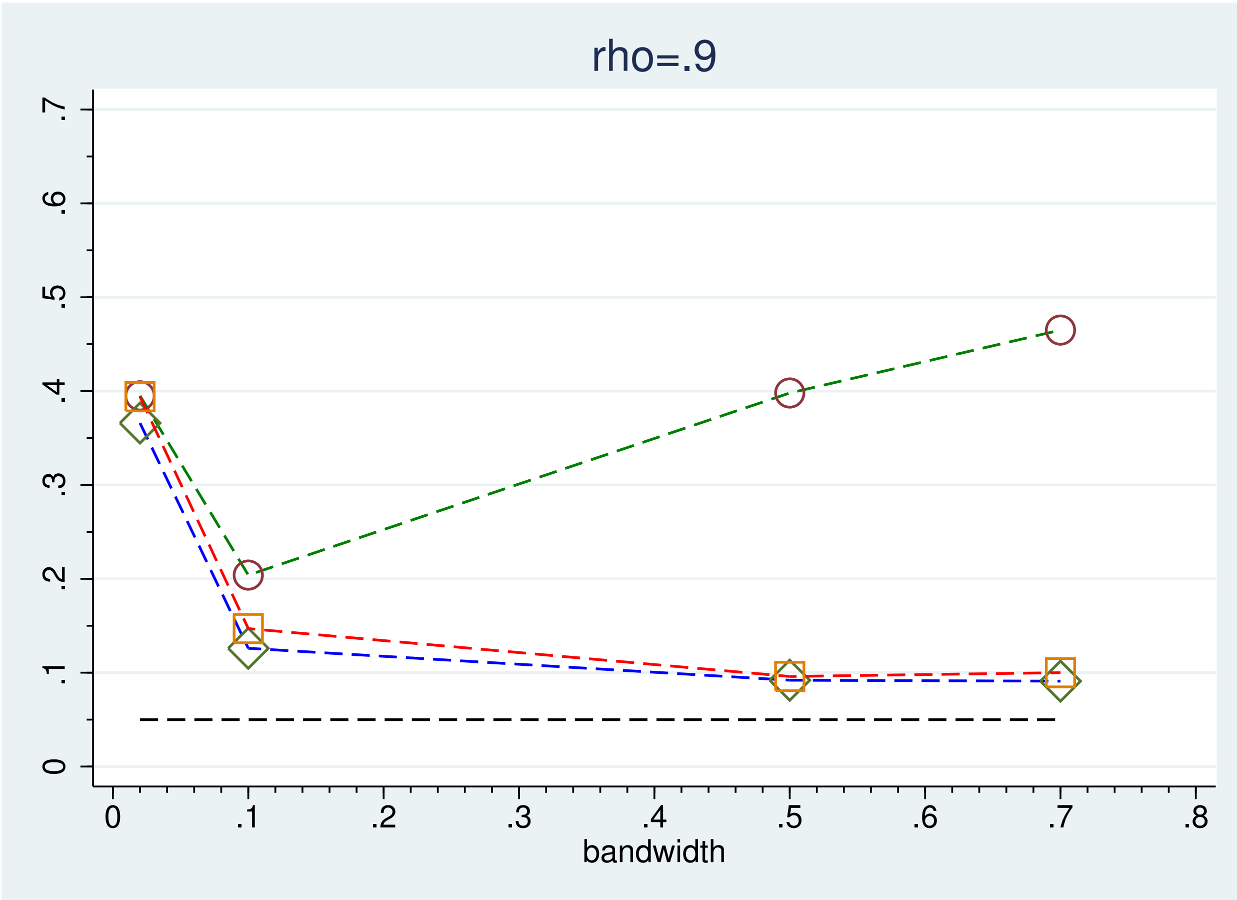
Figure E.10: Empirical null rejection probabilities, spatial $MA(2)$, Bartlett kernel, $T = 250$, $\lambda = 0.5$.

Figure E.10: (cont'd)



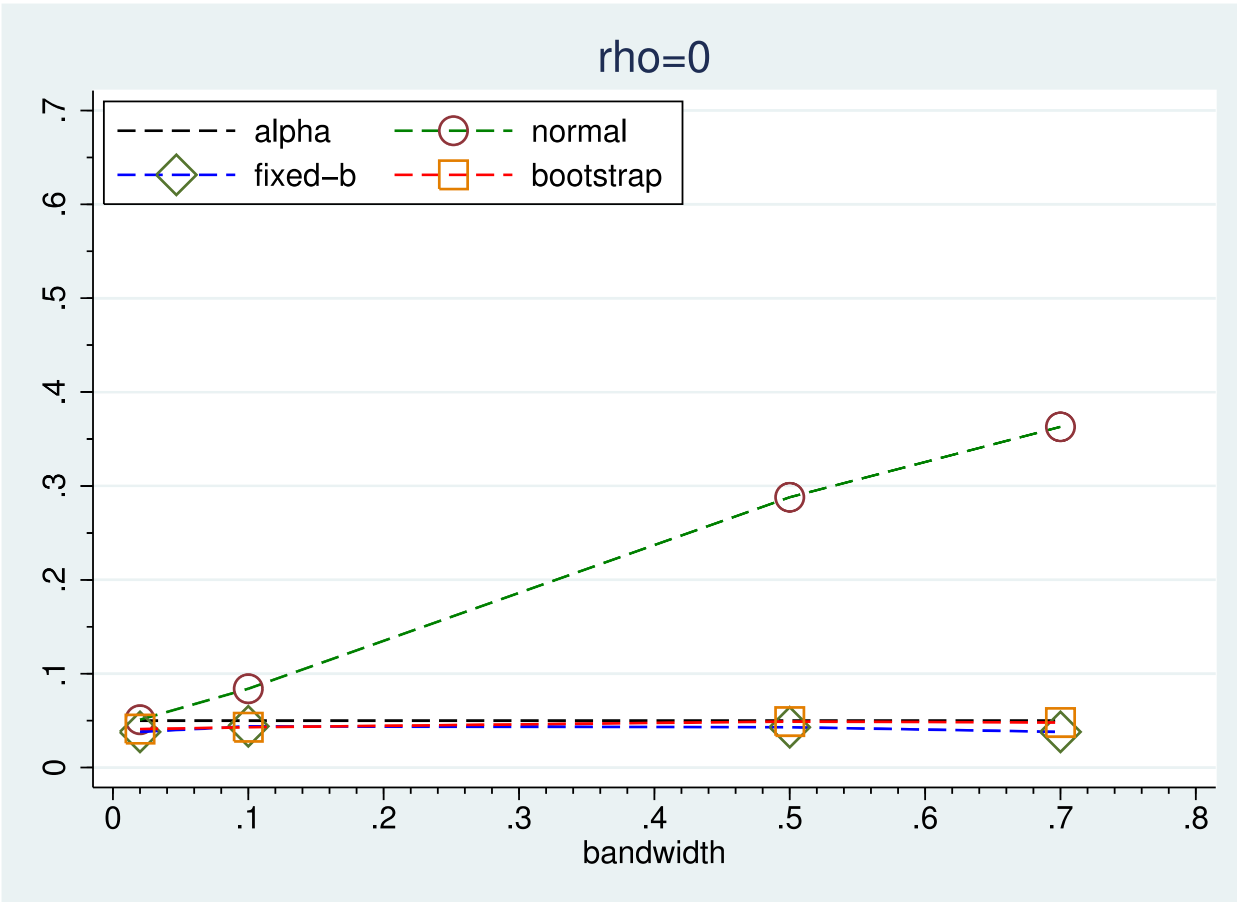
(b) N=49, T=250

Figure E.10: (cont'd)



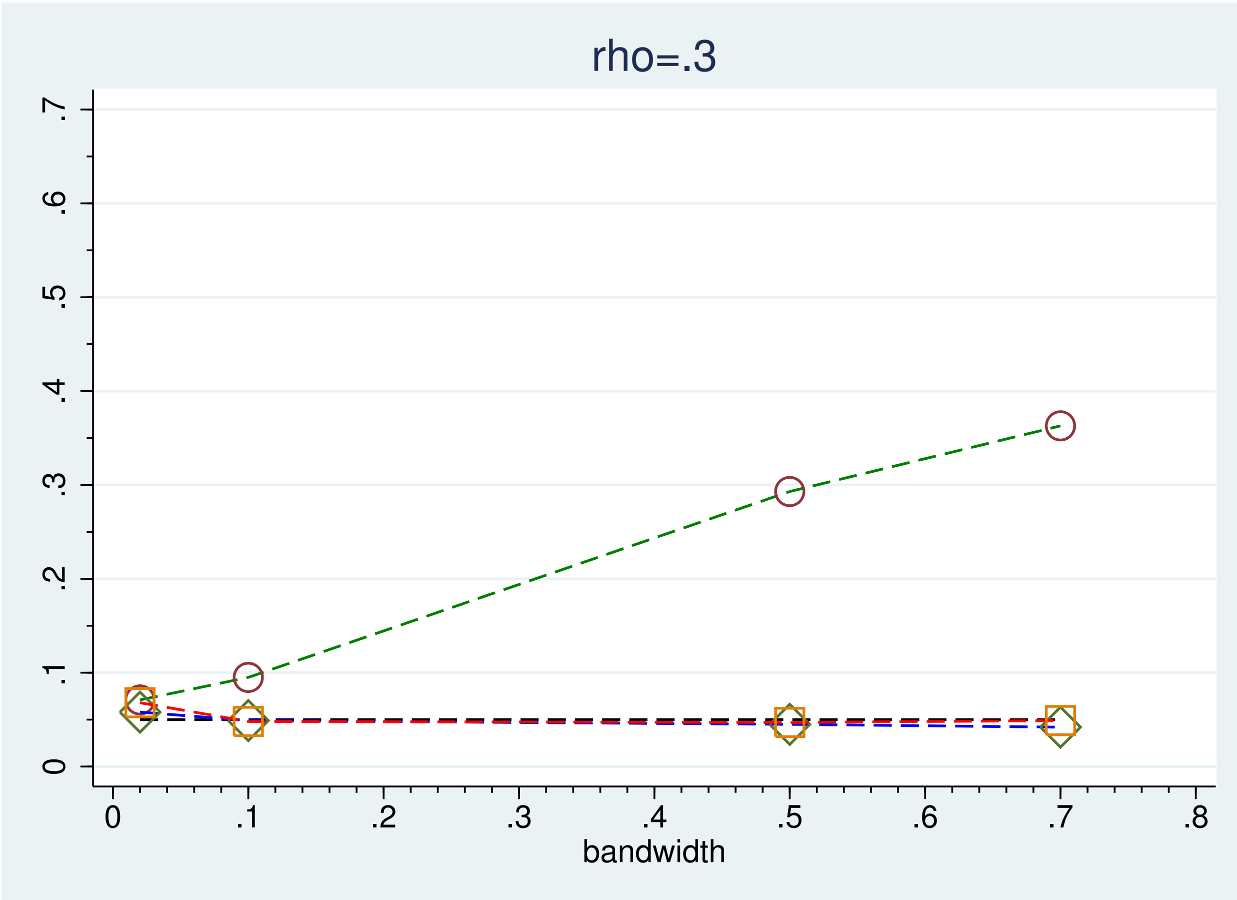
(c) N=49, T=250

Figure E.10: (cont'd)



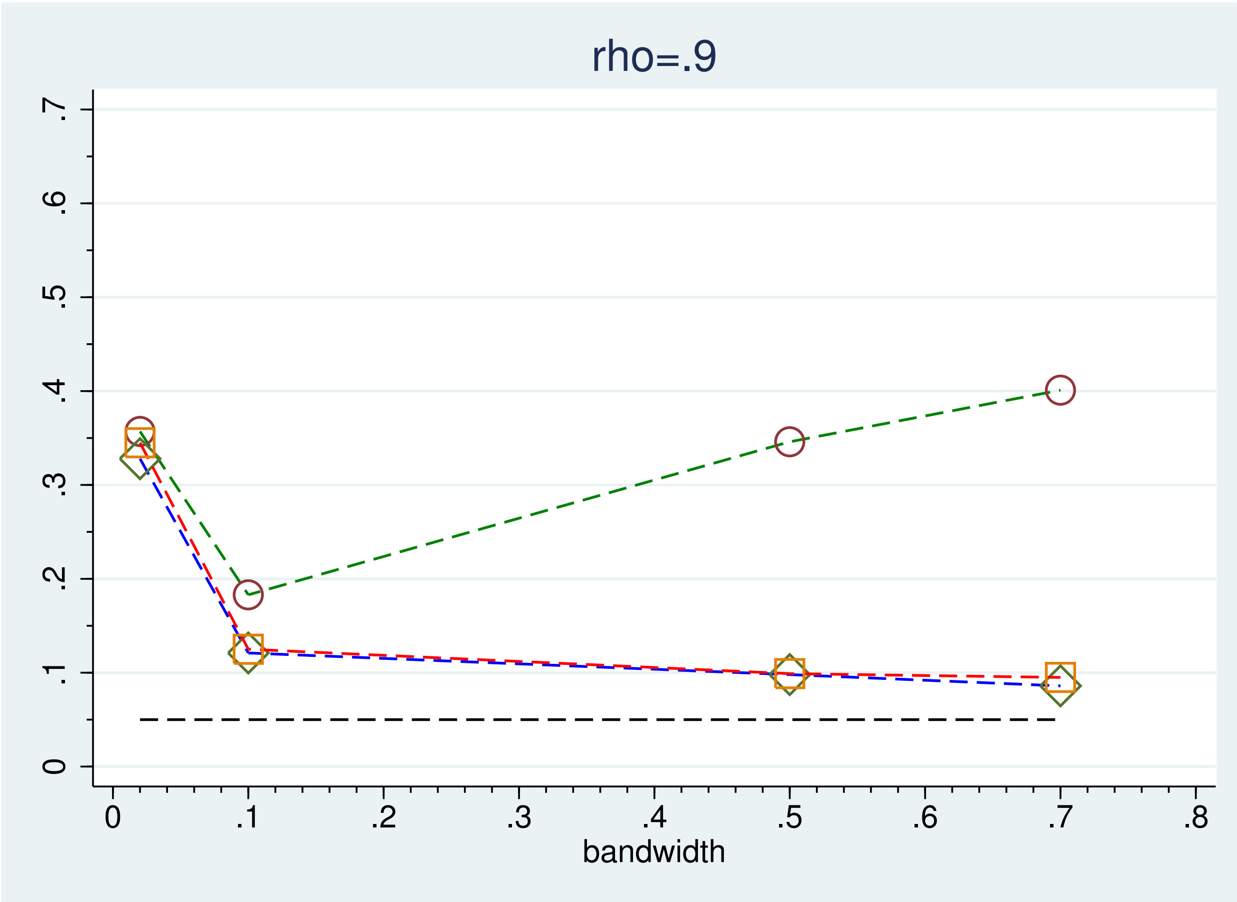
(d) N=256, T=250

Figure E.10: (cont'd)

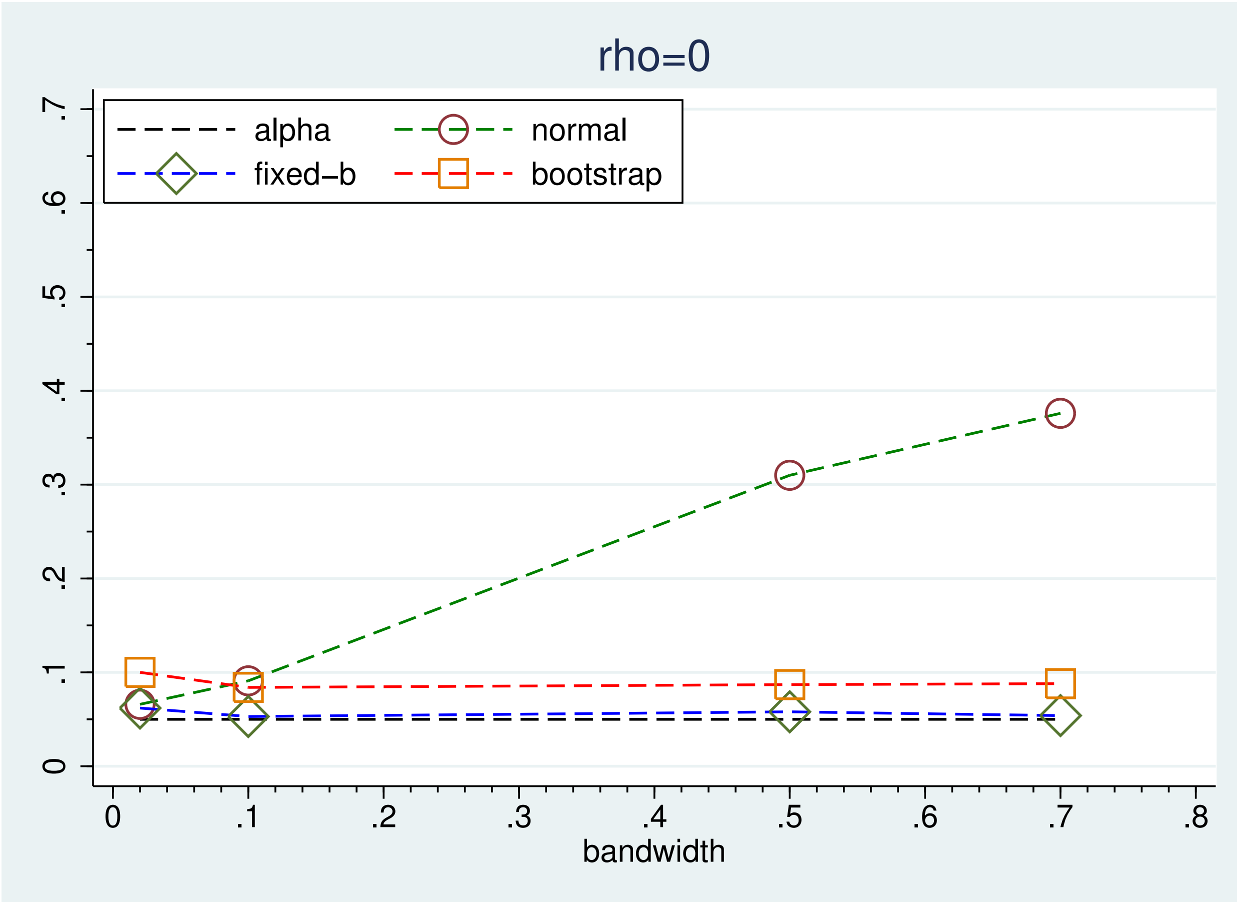


(e) N=256, T=250

Figure E.10: (cont'd)



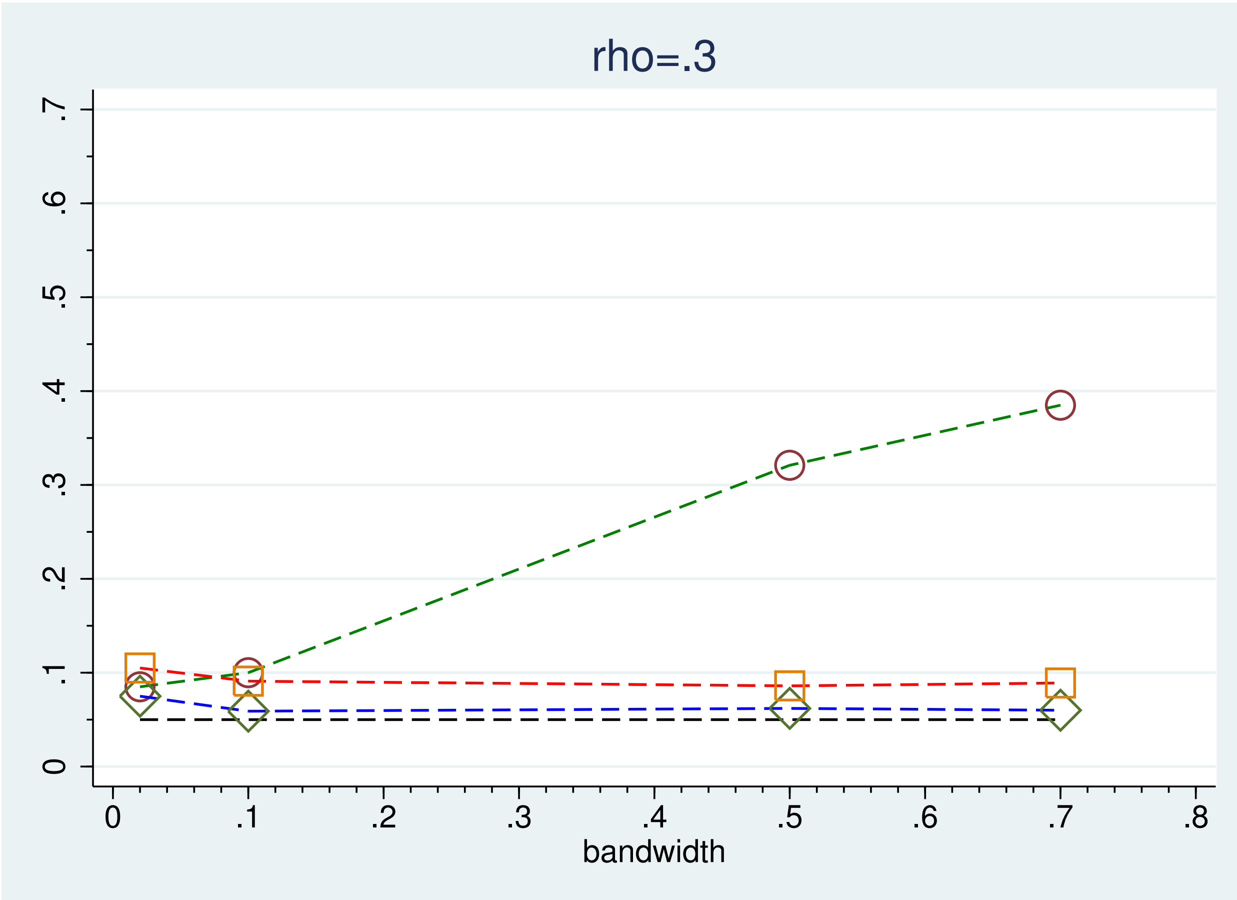
(f) N=256, T=250



(a) N=49, T=250, block length=25

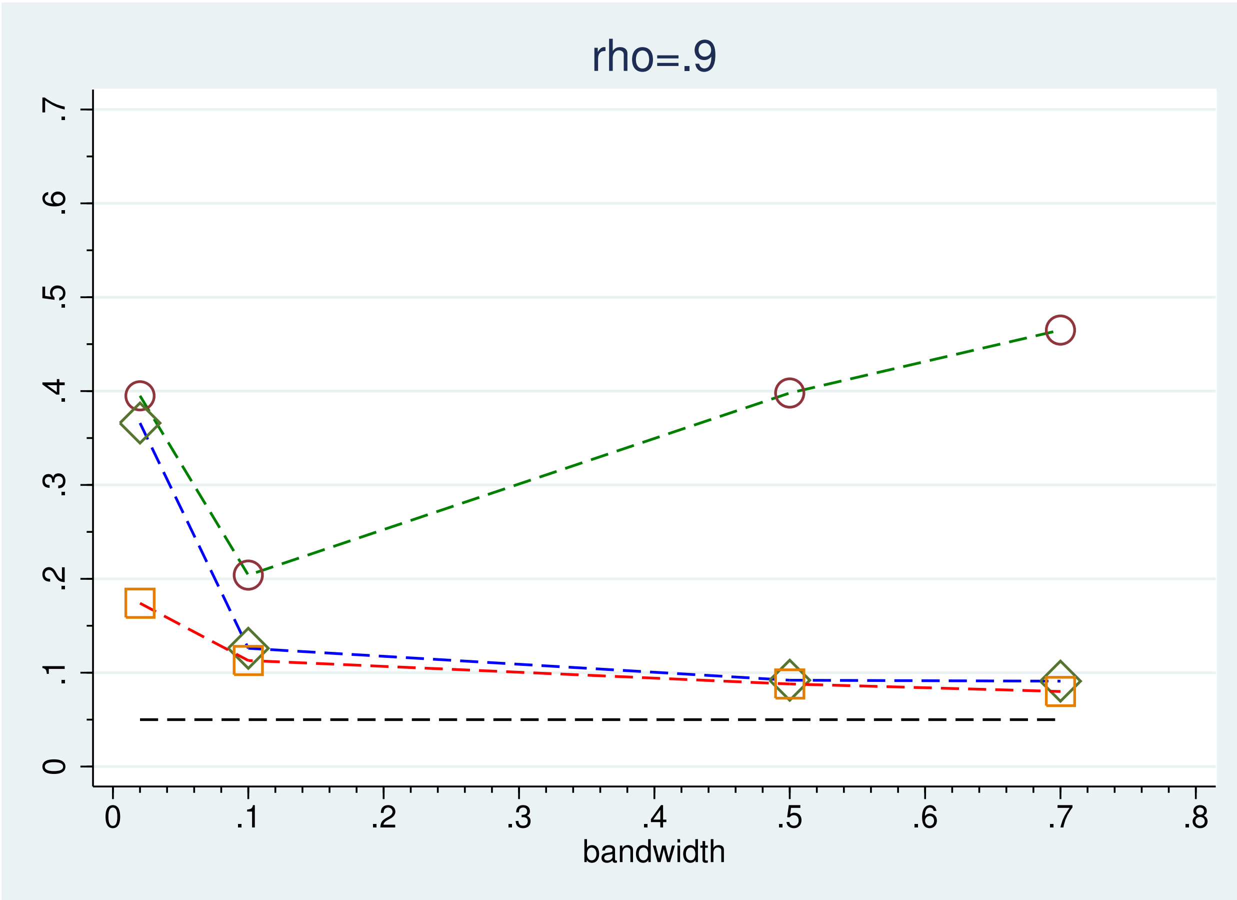
Figure E.11: Empirical null rejection probabilities, spatial $MA(2)$, Bartlett kernel, $N = 49$, $T = 250$, $\lambda = 0.5$.

Figure E.11: (cont'd)



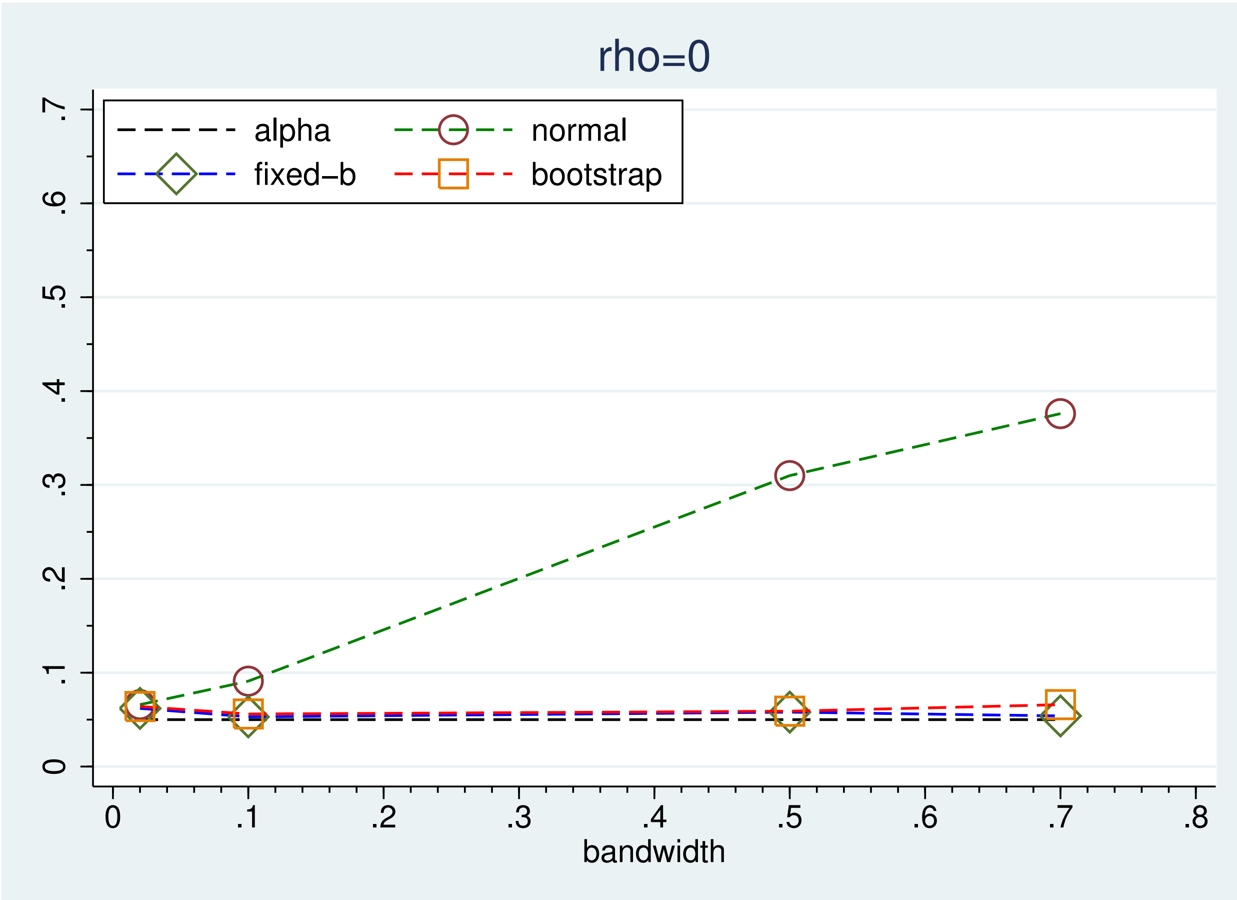
(b) N=49, T=250, block length=25

Figure E.11: (cont'd)



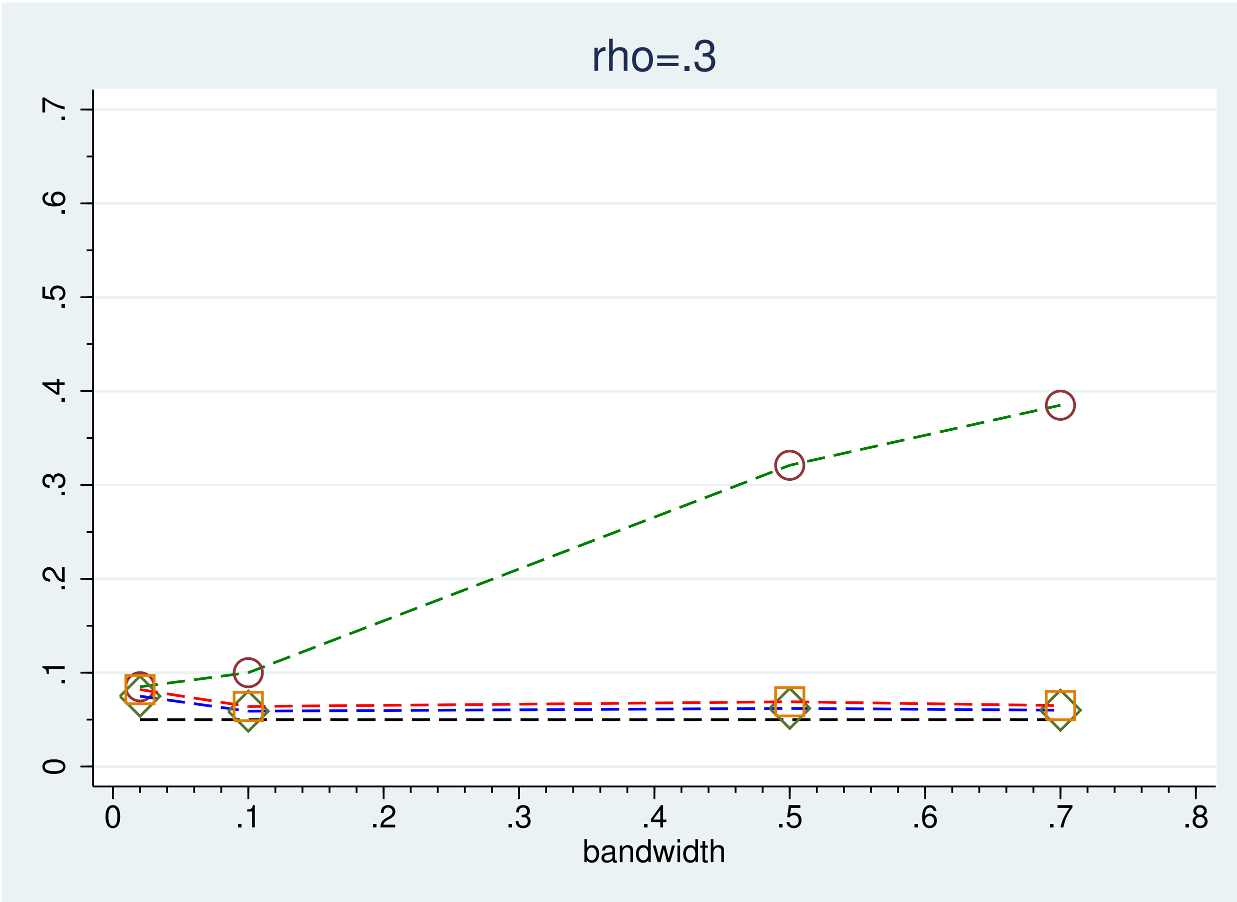
(c) N=49, T=250, block length=25

Figure E.11: (cont'd)



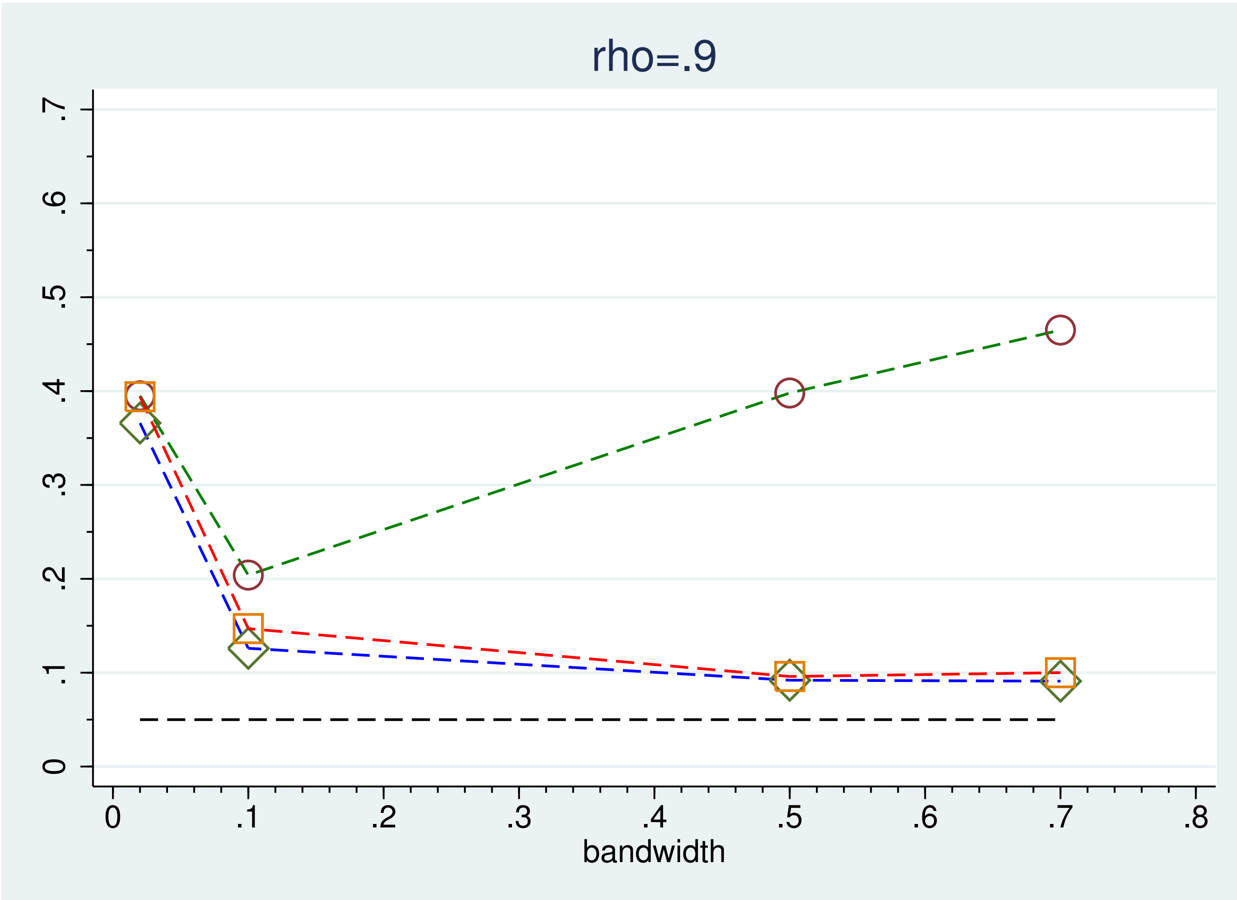
(d) N=49, T=250, block length=1

Figure E.11: (cont'd)

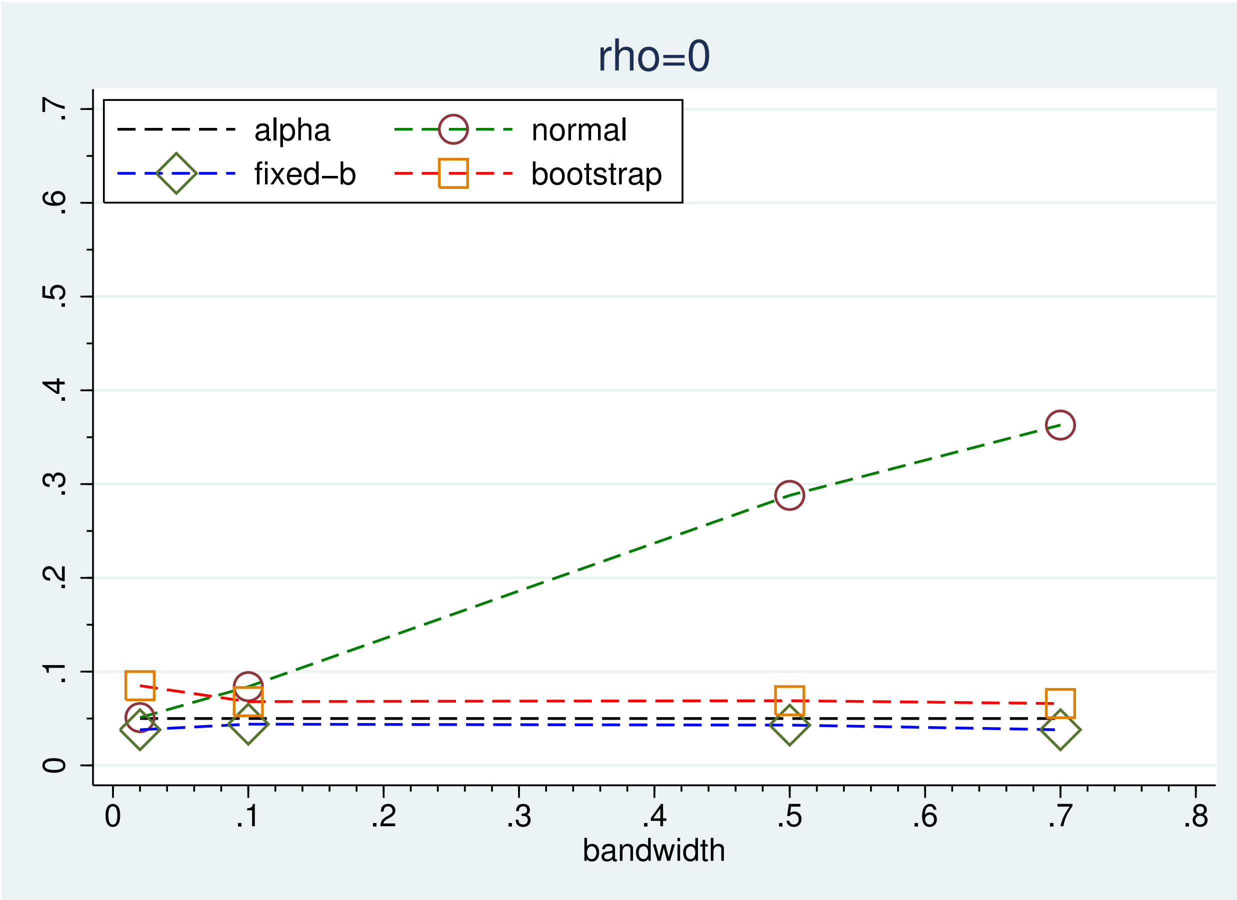


(e) N=49, T=250, block length=1

Figure E.11: (cont'd)



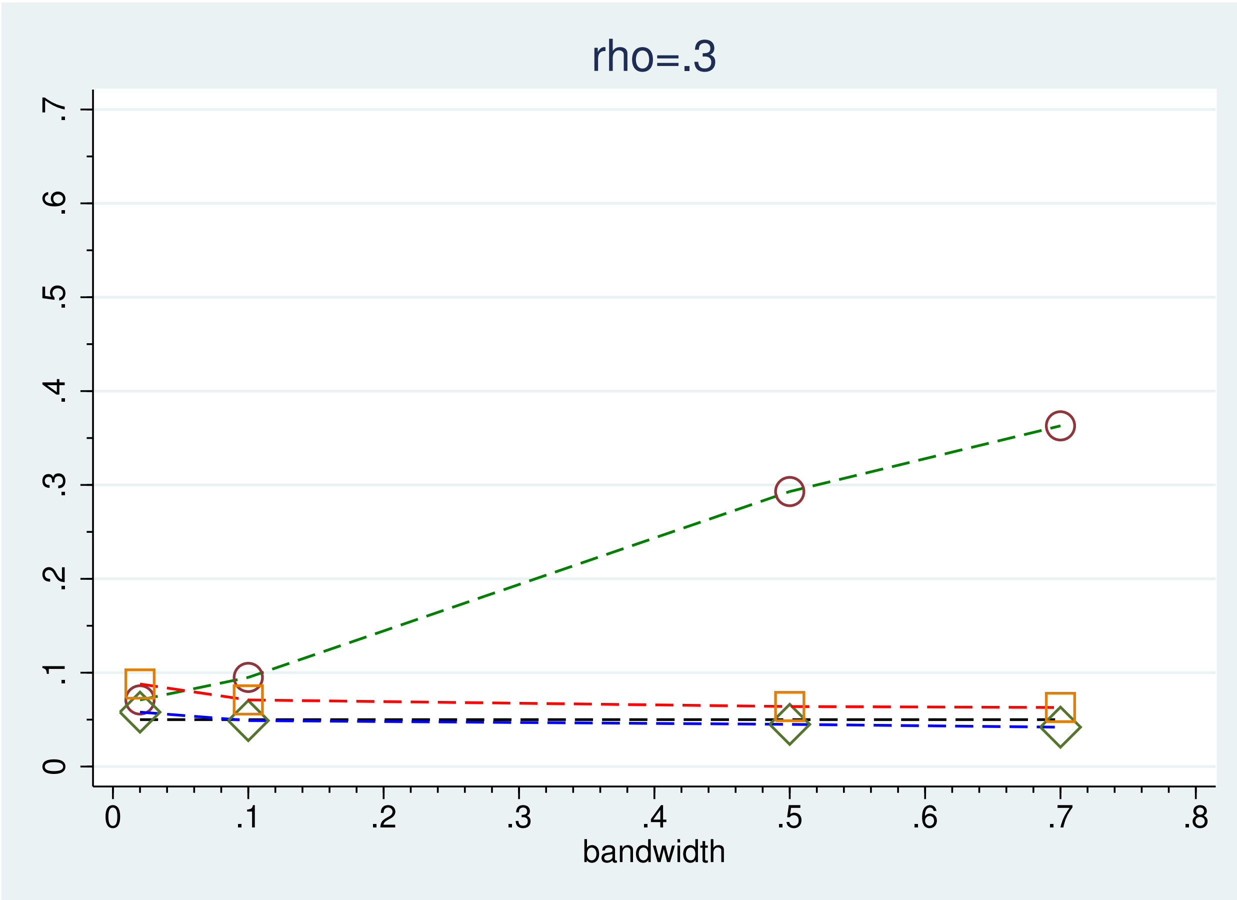
(f) N=49, T=250, block length=1



(a) N=256, T=250, block length=25

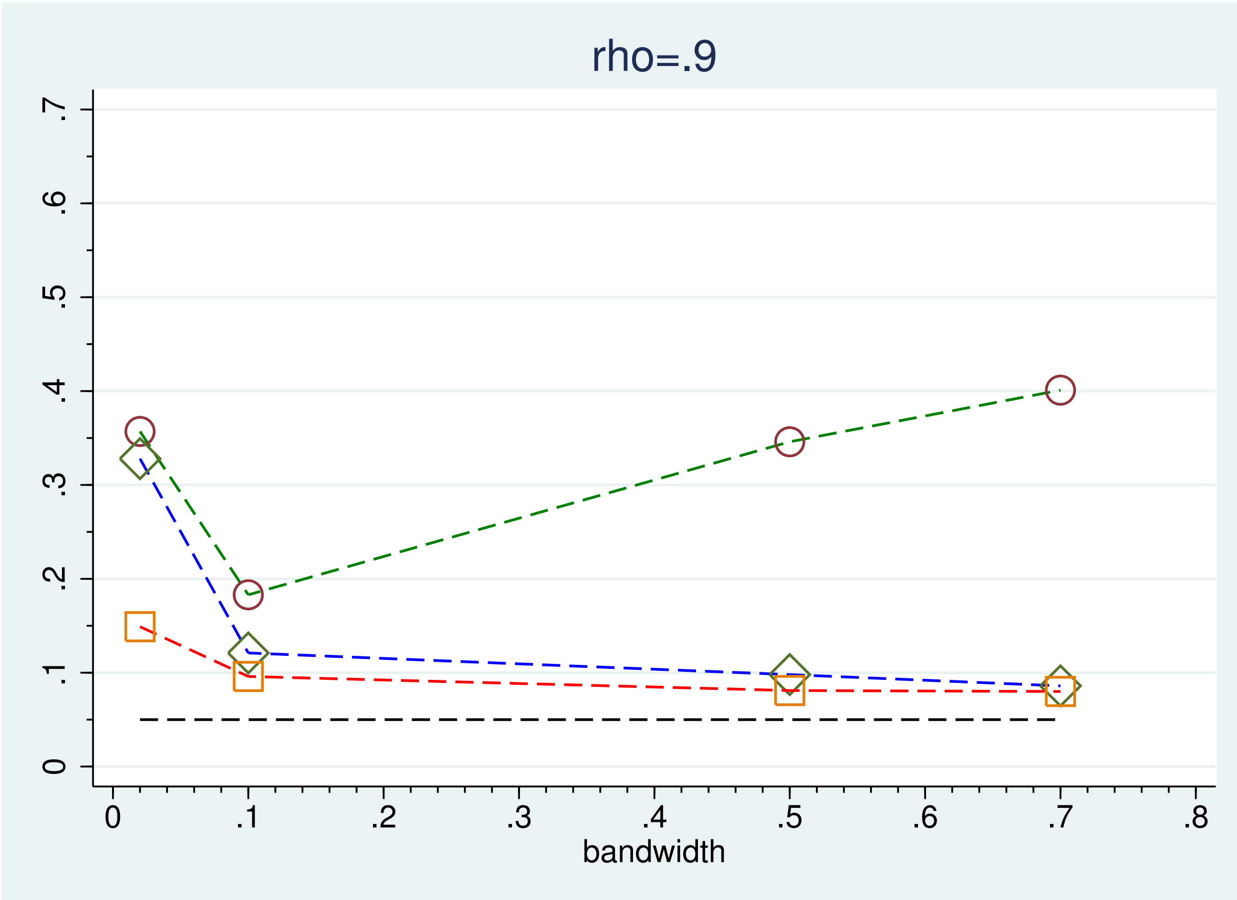
Figure E.12: Empirical null rejection probabilities, spatial $MA(2)$, Bartlett kernel, $N = 256$, $T = 250$, $\lambda = 0.5$.

Figure E.12: (cont'd)



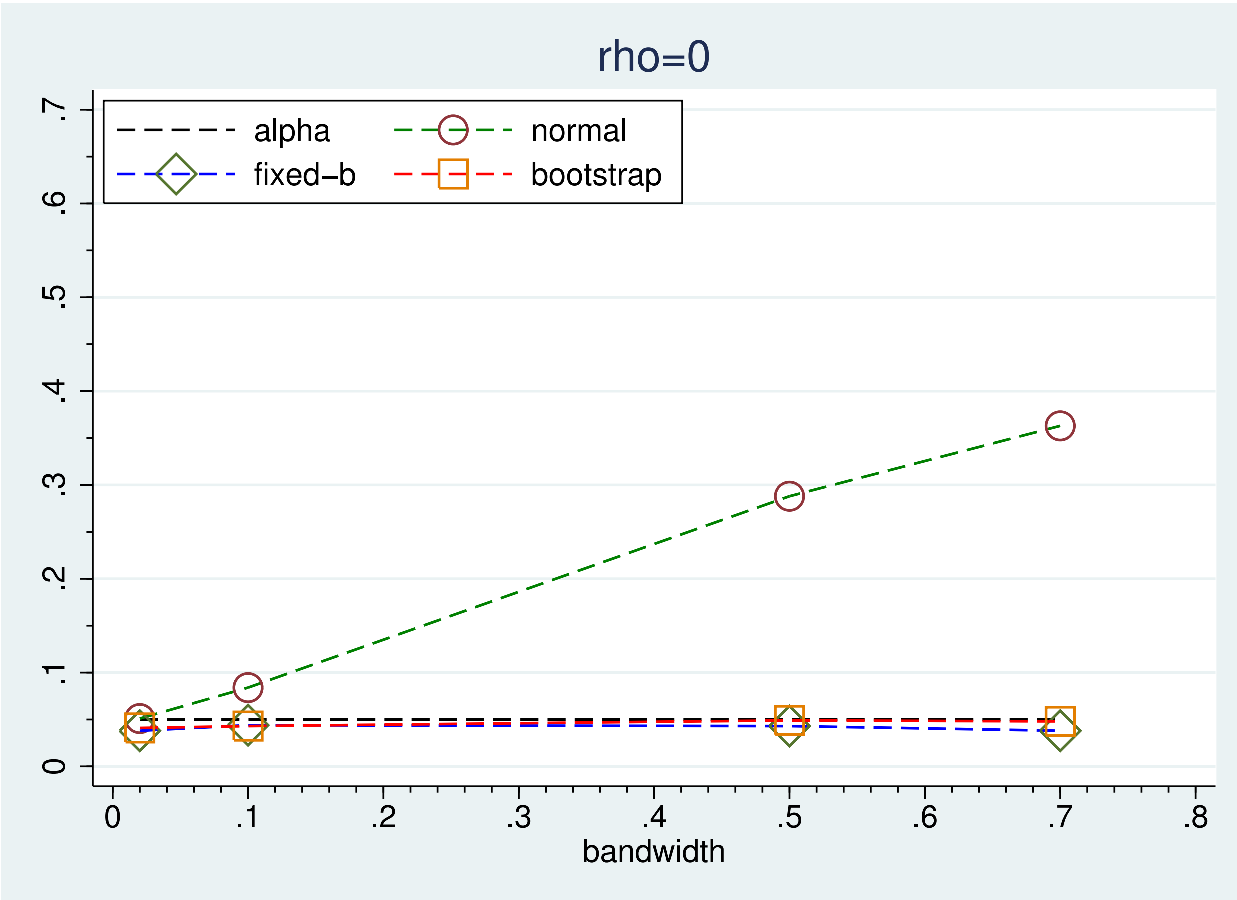
(b) N=256, T=250, block length=25

Figure E.12: (cont'd)



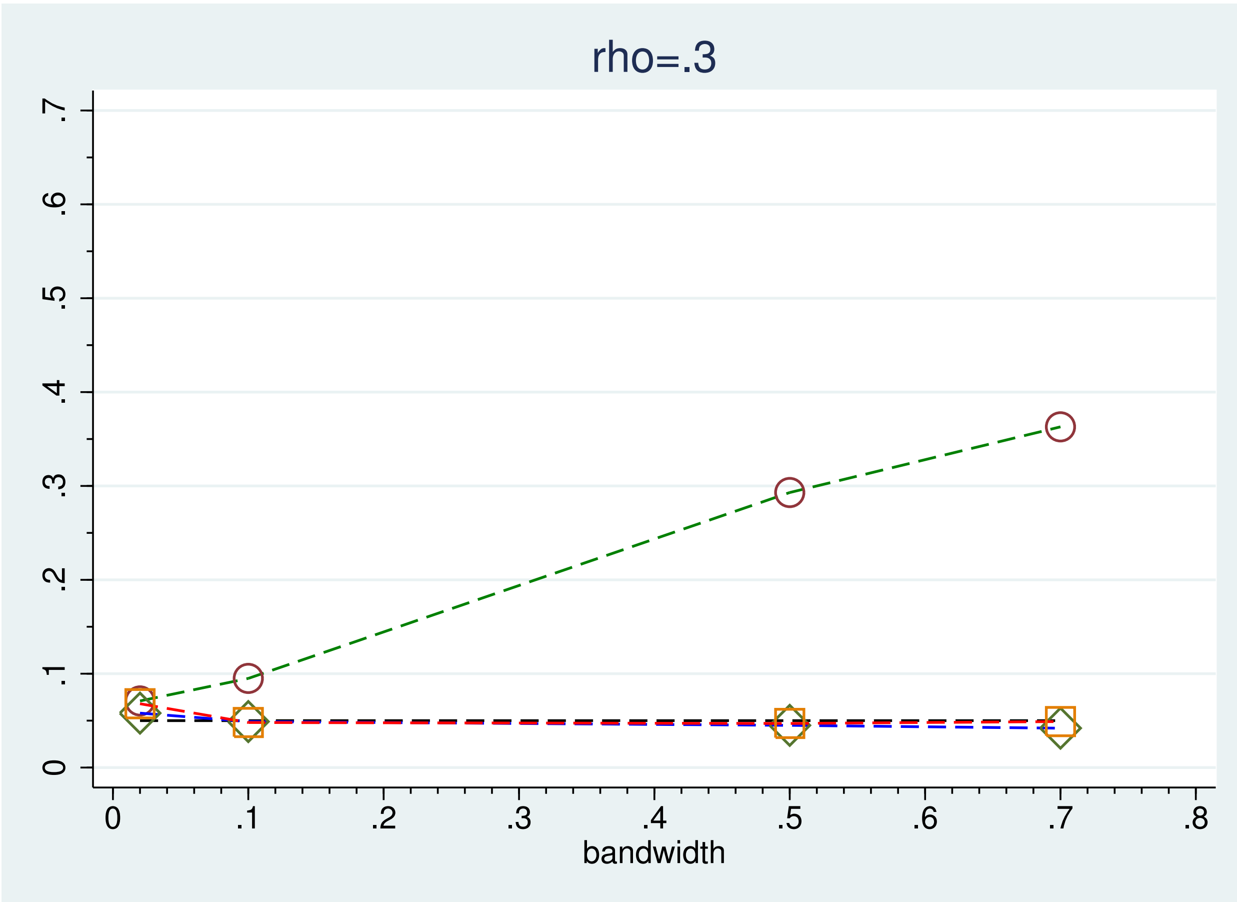
(c) N=256, T=250, block length=25

Figure E.12: (cont'd)



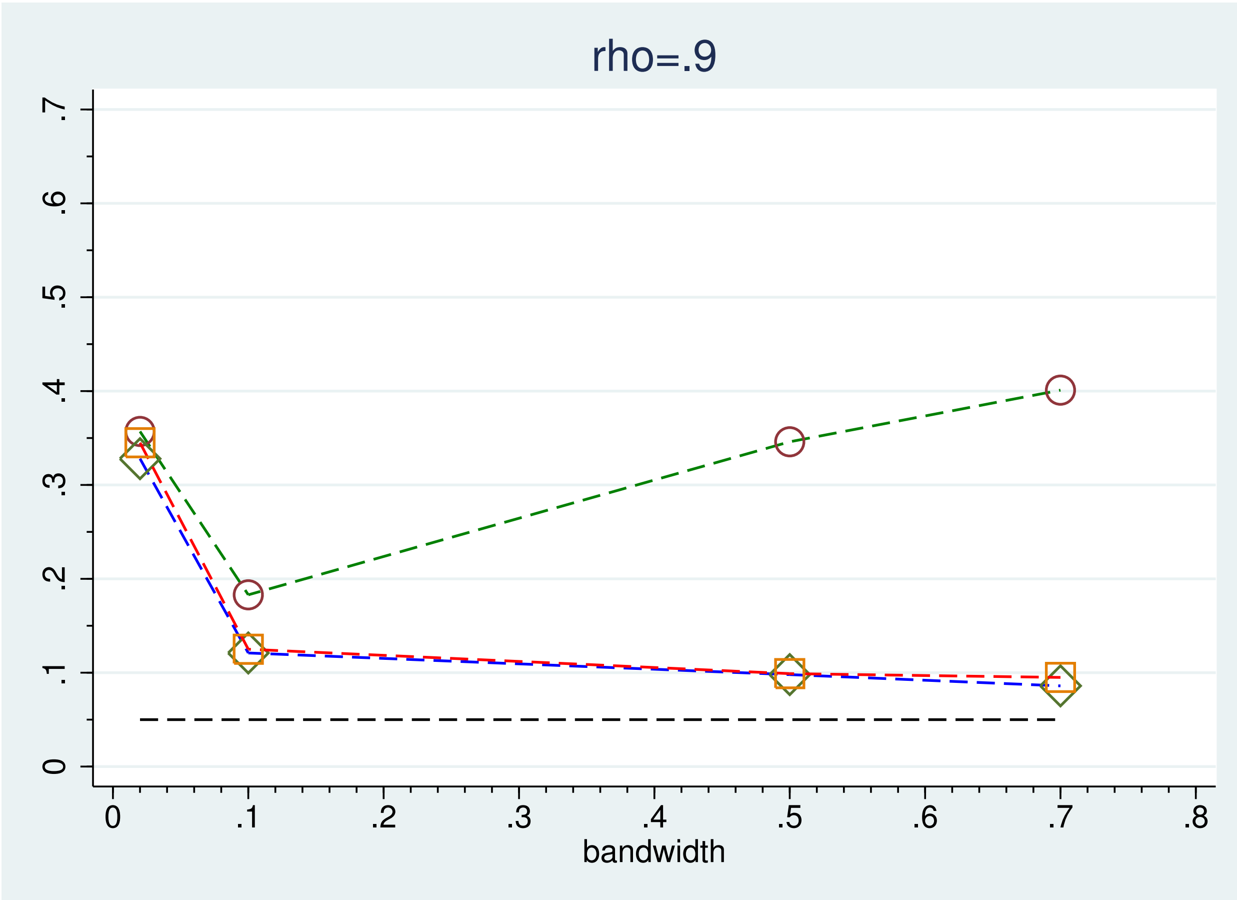
(d) N=256, T=250, block length=1

Figure E.12: (cont'd)

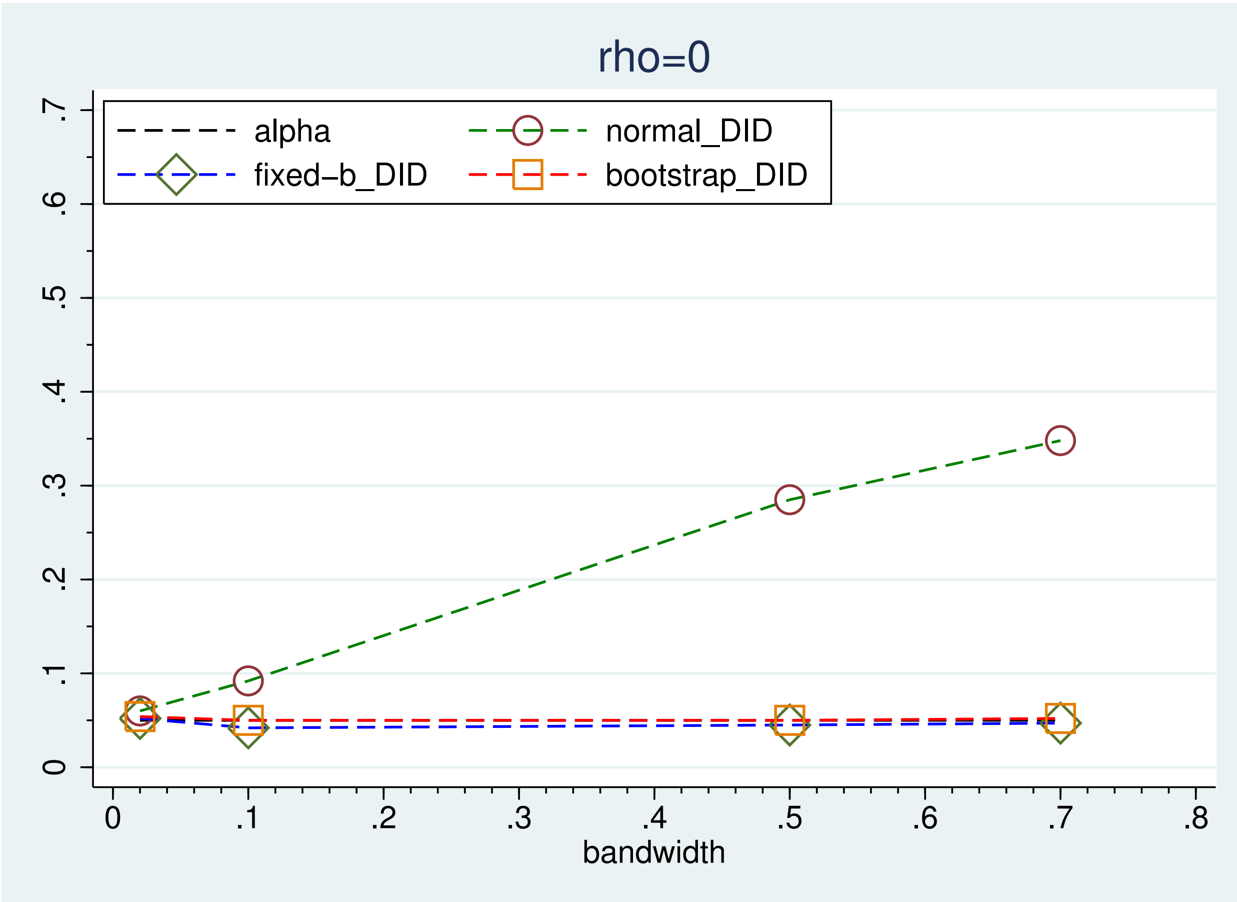


(e) N=256, T=250, block length=1

Figure E.12: (cont'd)



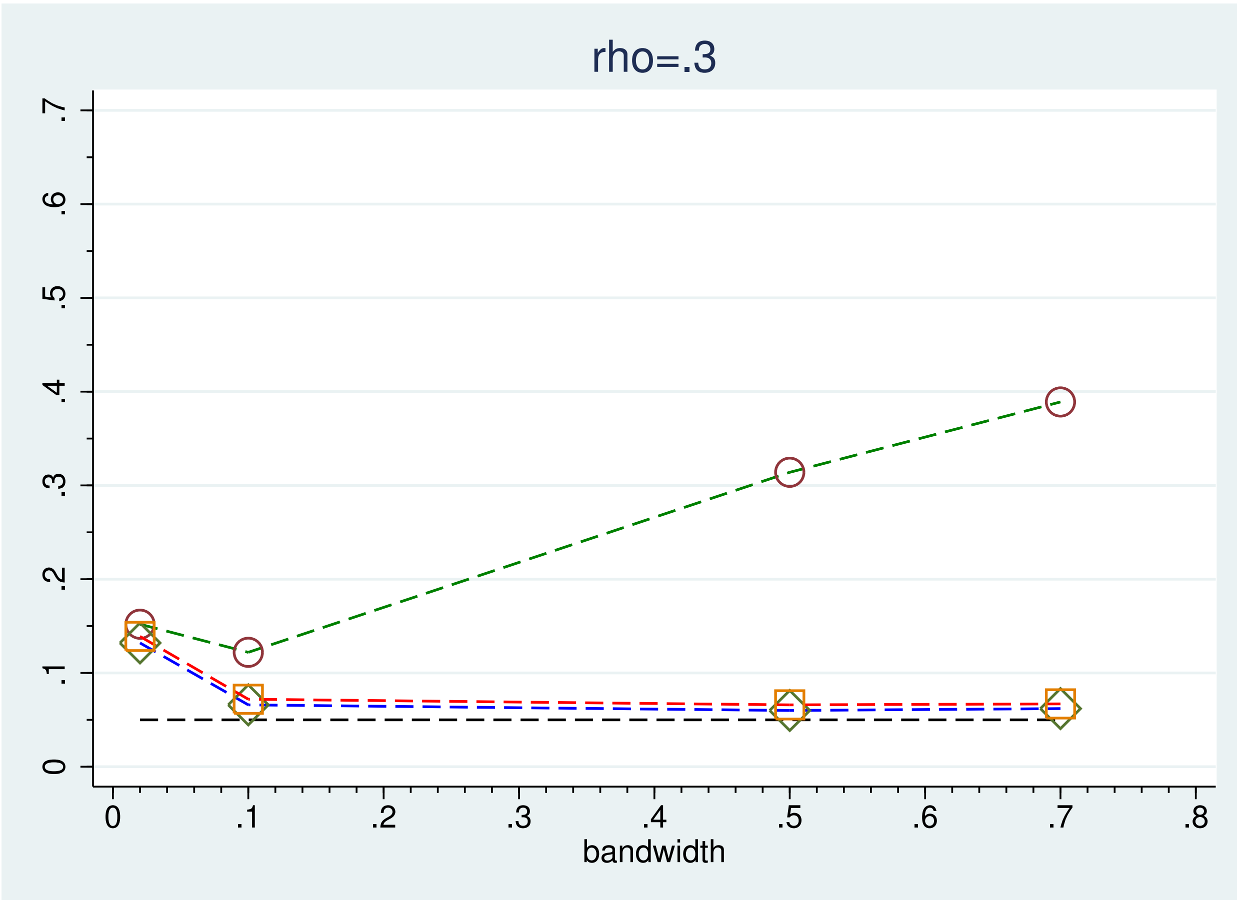
(f) N=256, T=250, block length=1



(a) N=49, T=50, DD parameter

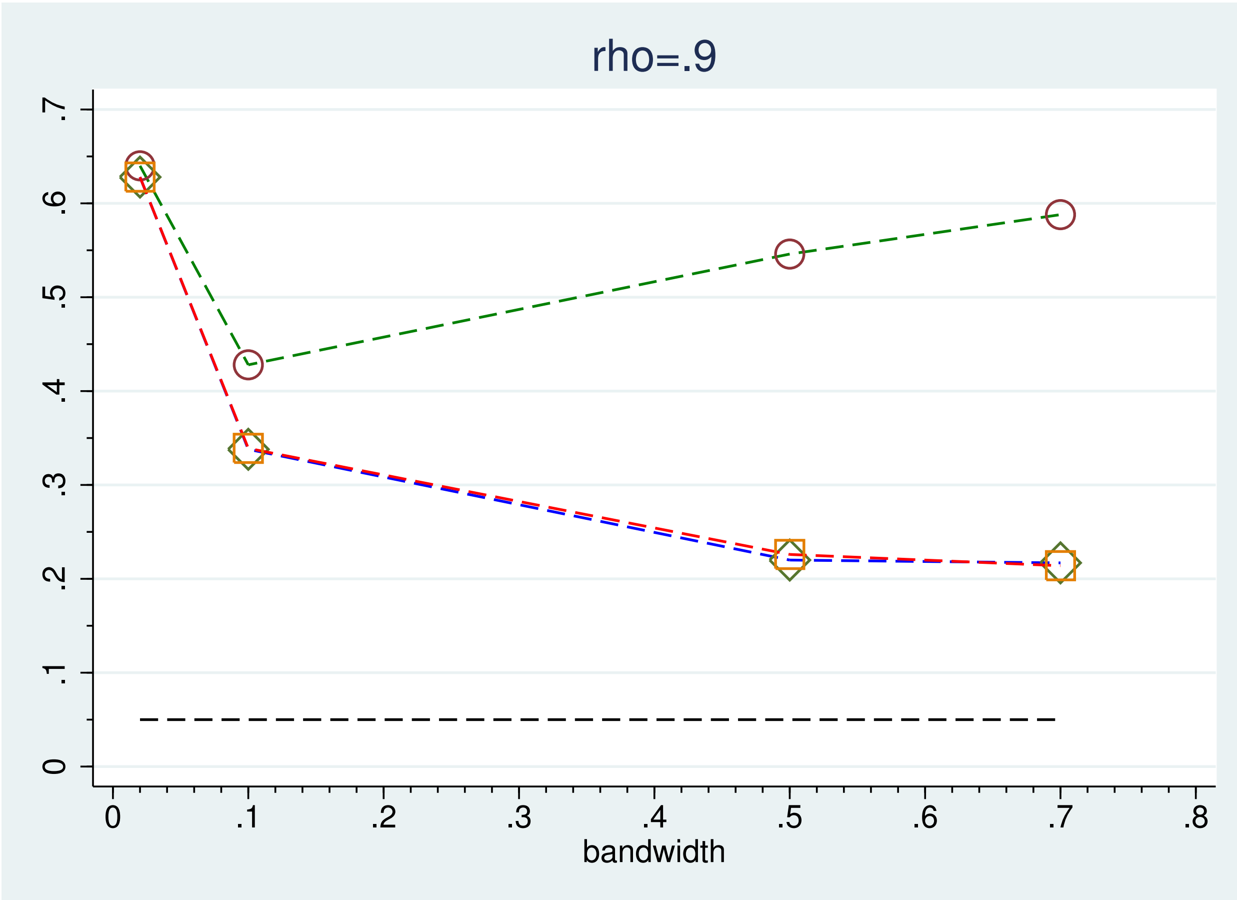
Figure E.13: Empirical null rejection probabilities, additional regressor, spatial $MA(2)$, Bartlett kernel, $N = 49$, $T = 50$, $\lambda = 0.5$.

Figure E.13: (cont'd)



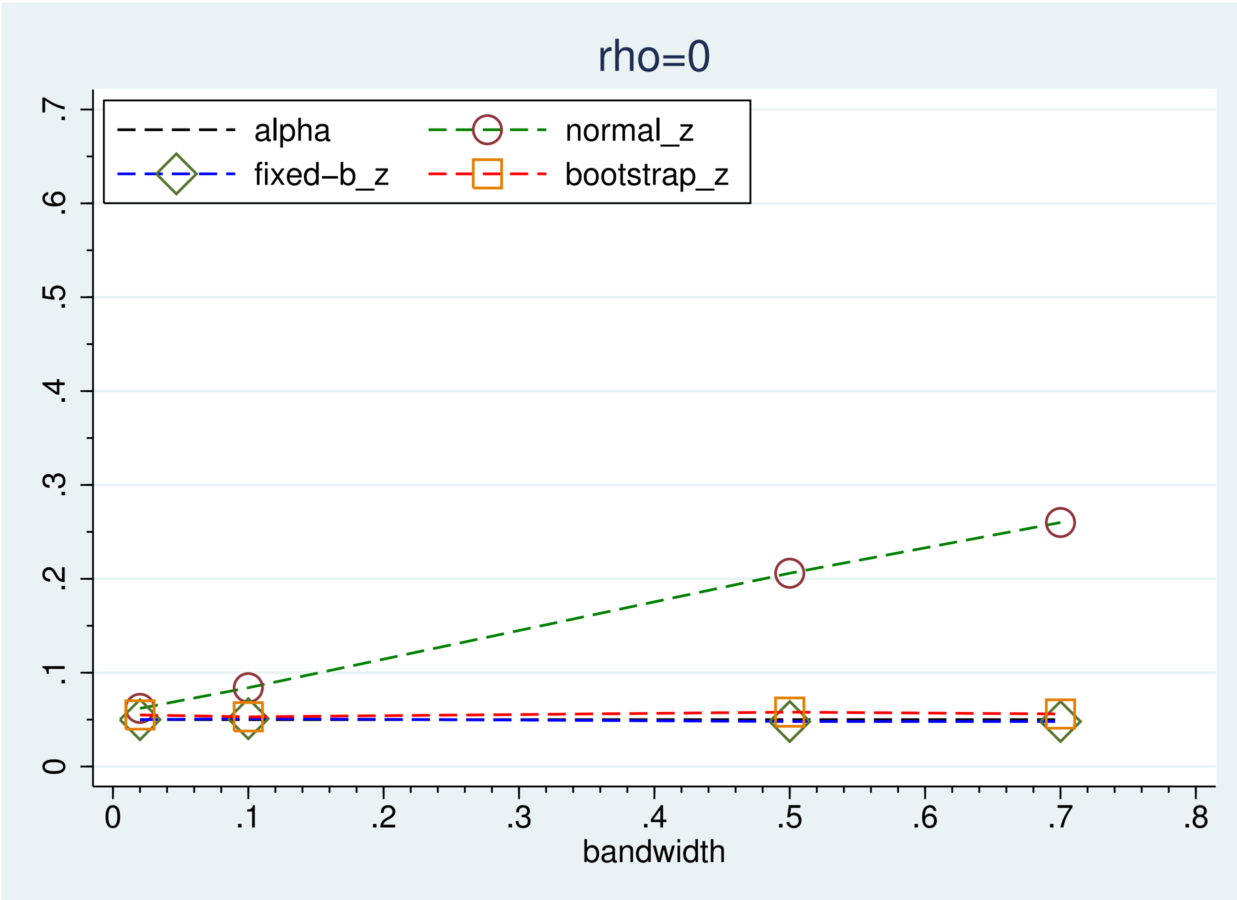
(b) N=49, T=50, DD parameter

Figure E.13: (cont'd)



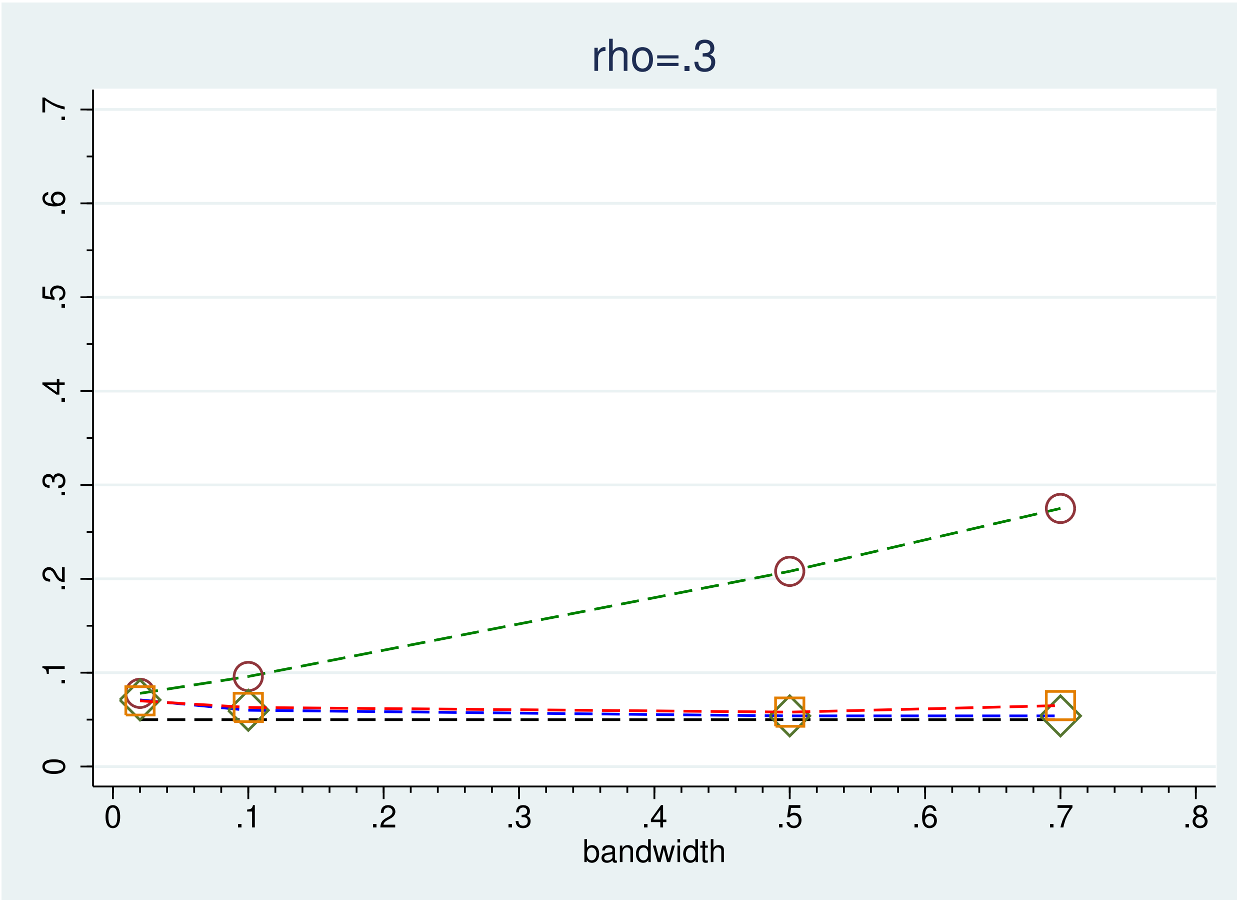
(c) N=49, T=50, DD parameter

Figure E.13: (cont'd)



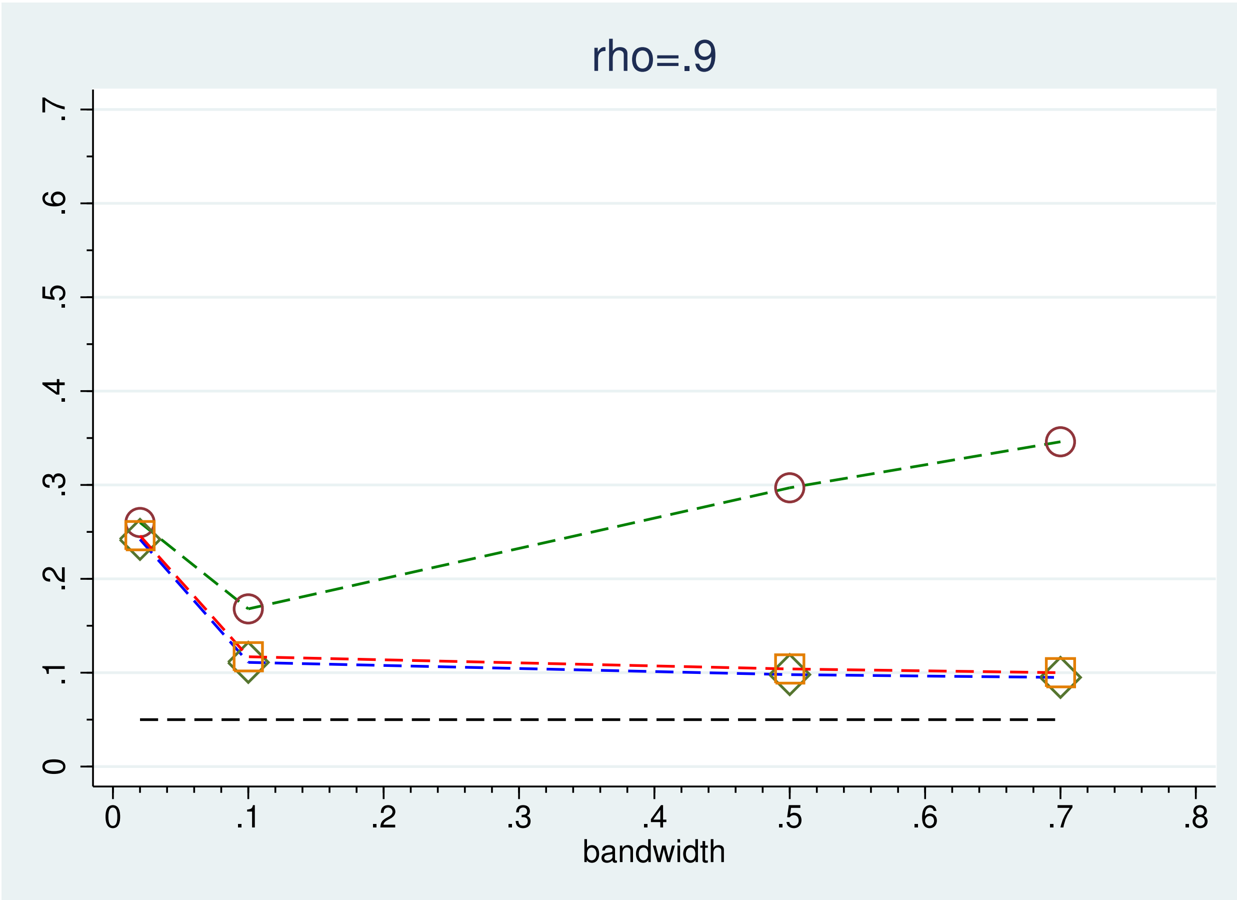
(d) N=49, T=50, z parameter

Figure E.13: (cont'd)

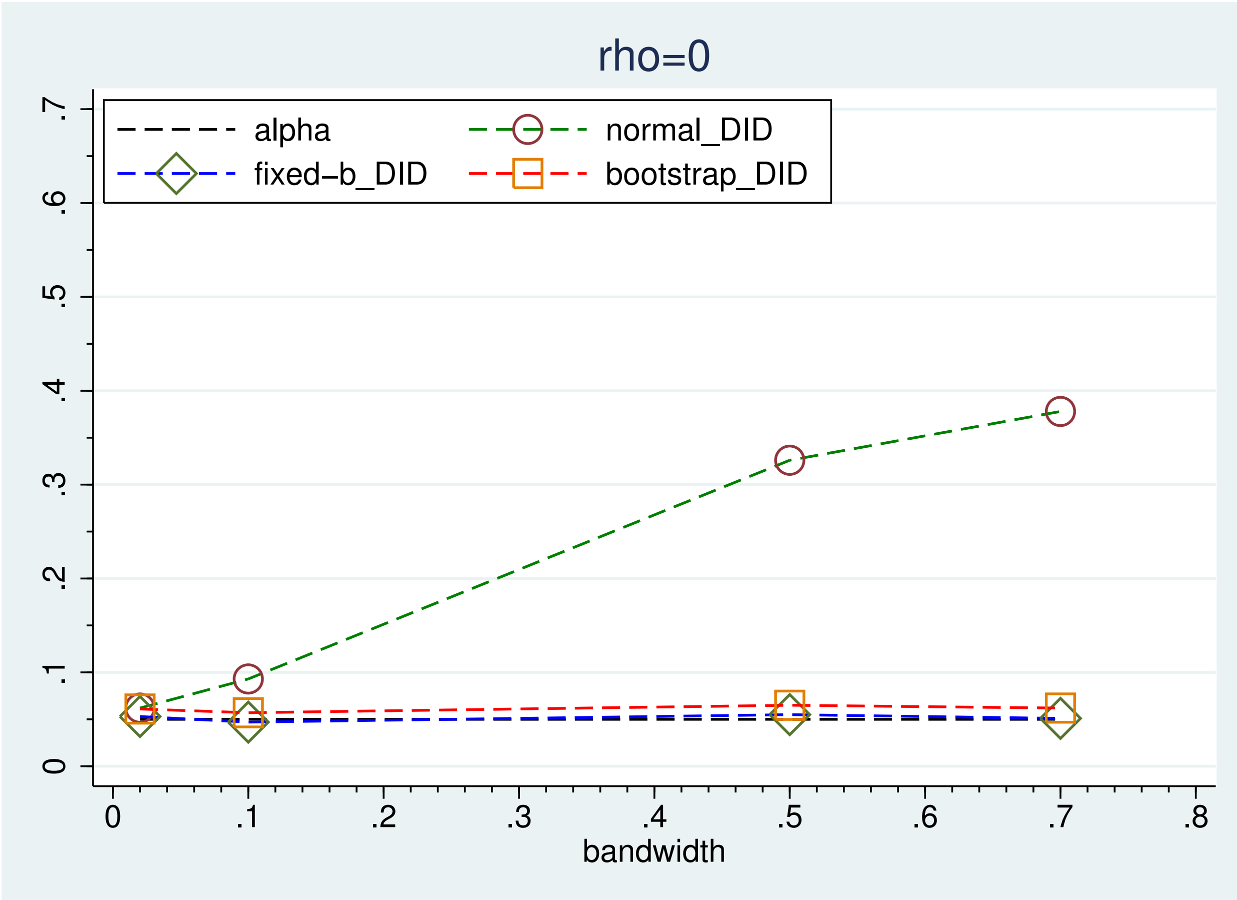


(e) N=49, T=50, z parameter

Figure E.13: (cont'd)



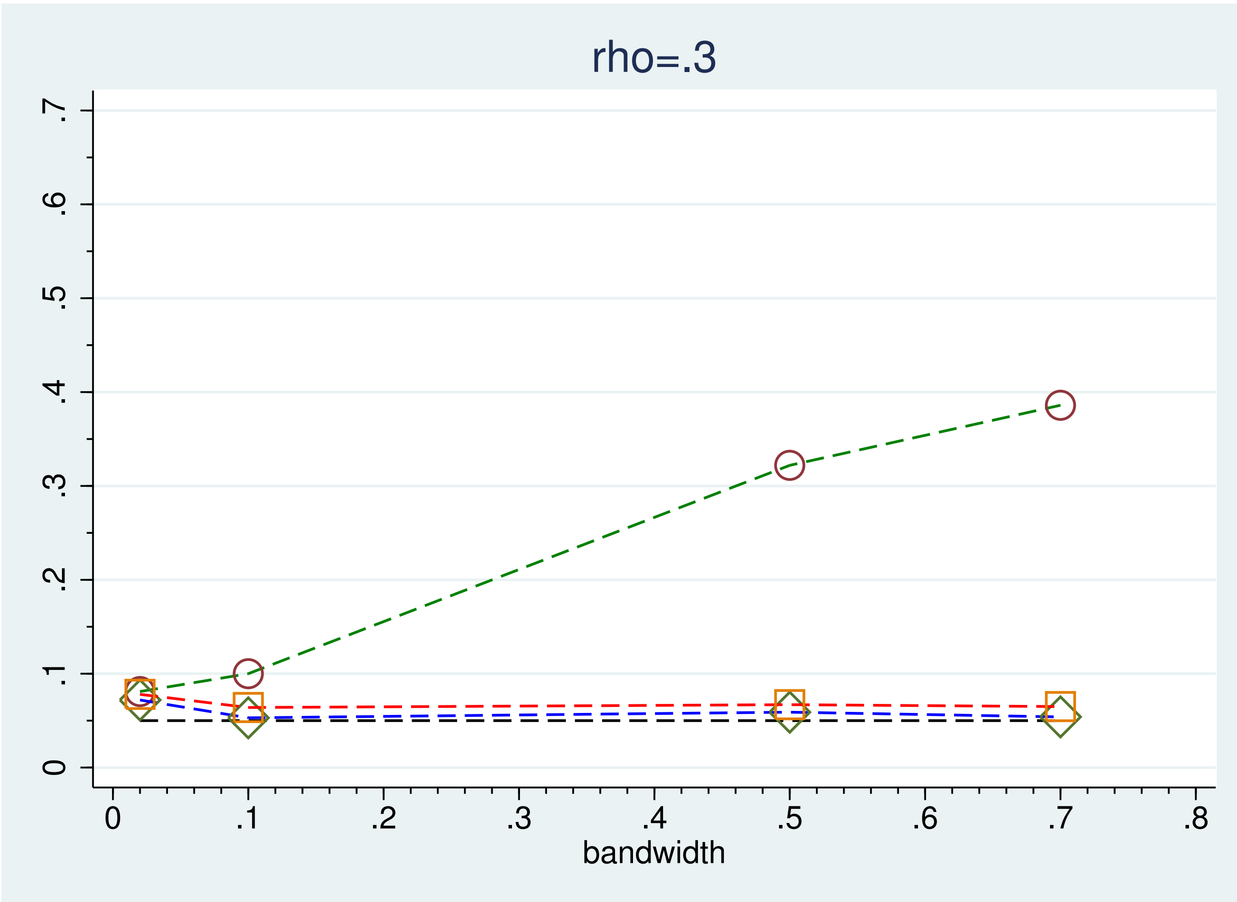
(f) N=49, T=50, z parameter



(a) N=49, T=250, DD parameter

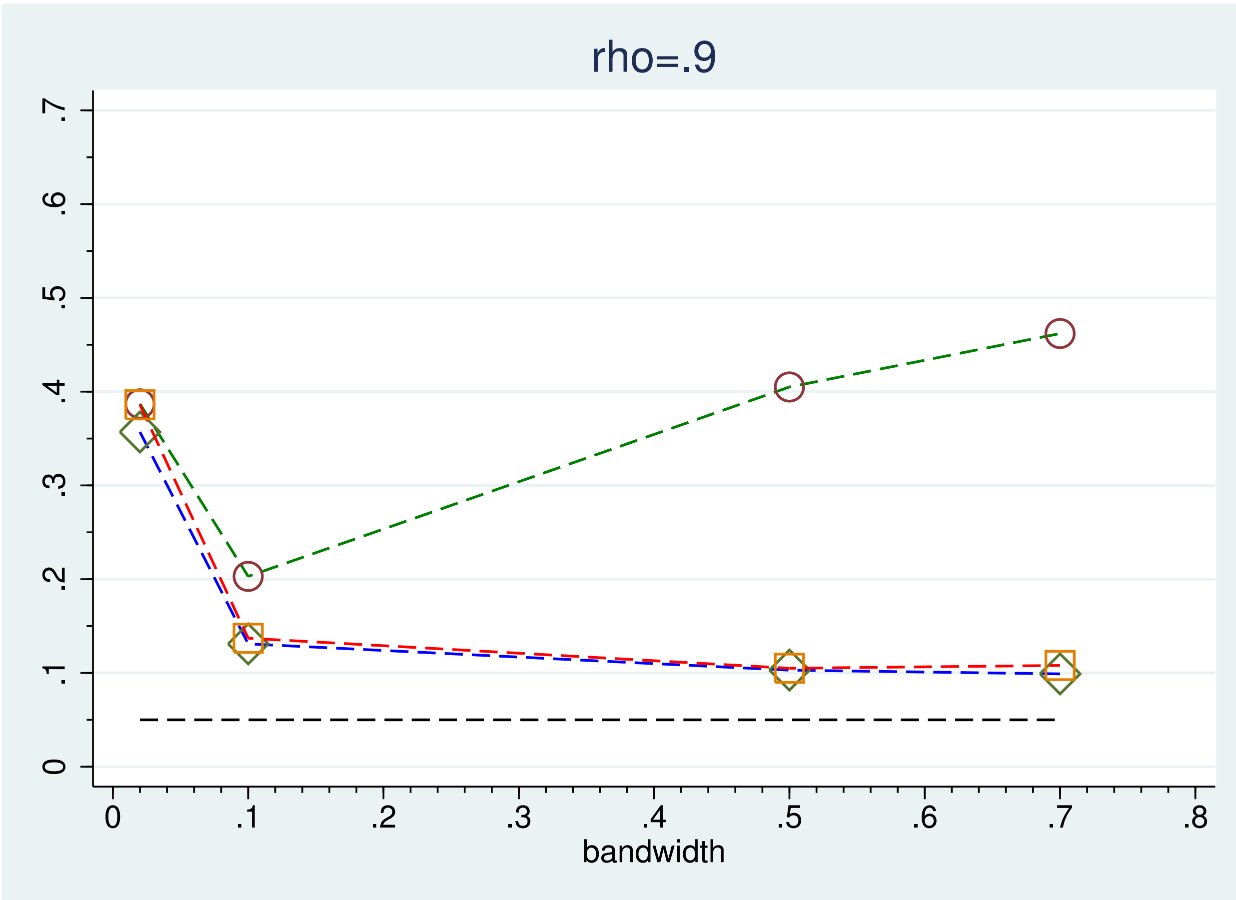
Figure E.14: Empirical null rejection probabilities, additional regressor, spatial $MA(2)$, Bartlett kernel, $N = 49$, $T = 250$, $\lambda = 0.5$.

Figure E.14: (cont'd)



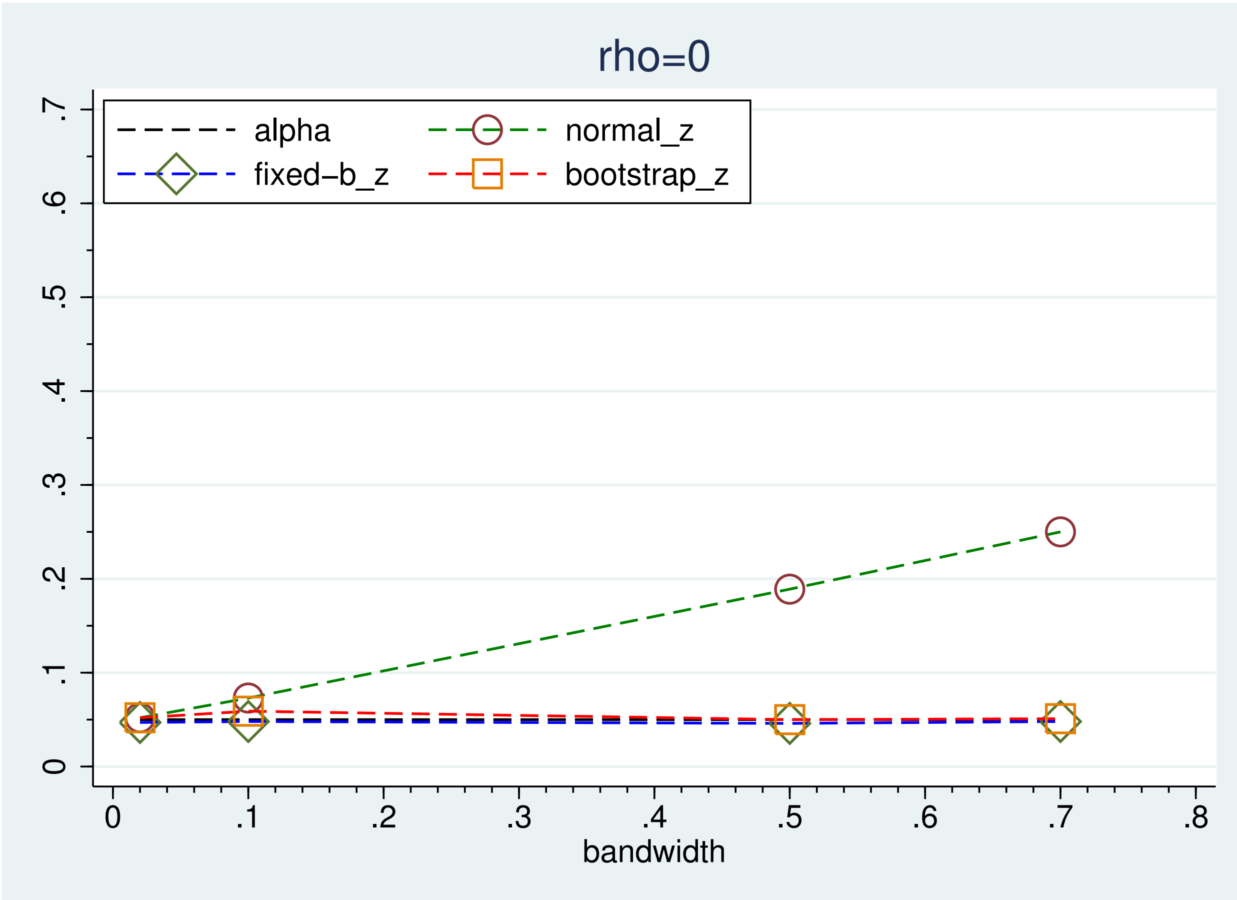
(b) N=49, T=250, DD parameter

Figure E.14: (cont'd)



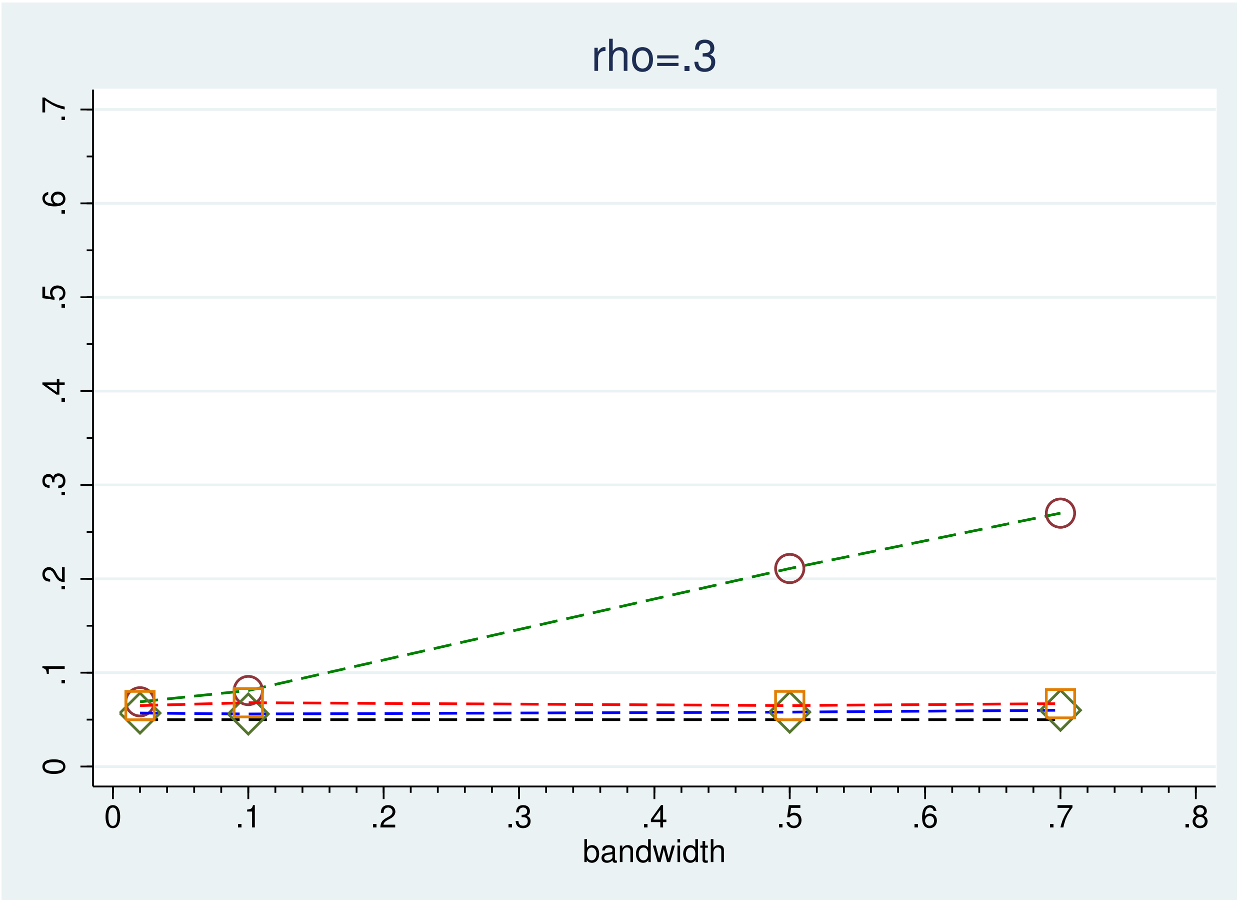
(c) N=49, T=250, DD parameter

Figure E.14: (cont'd)



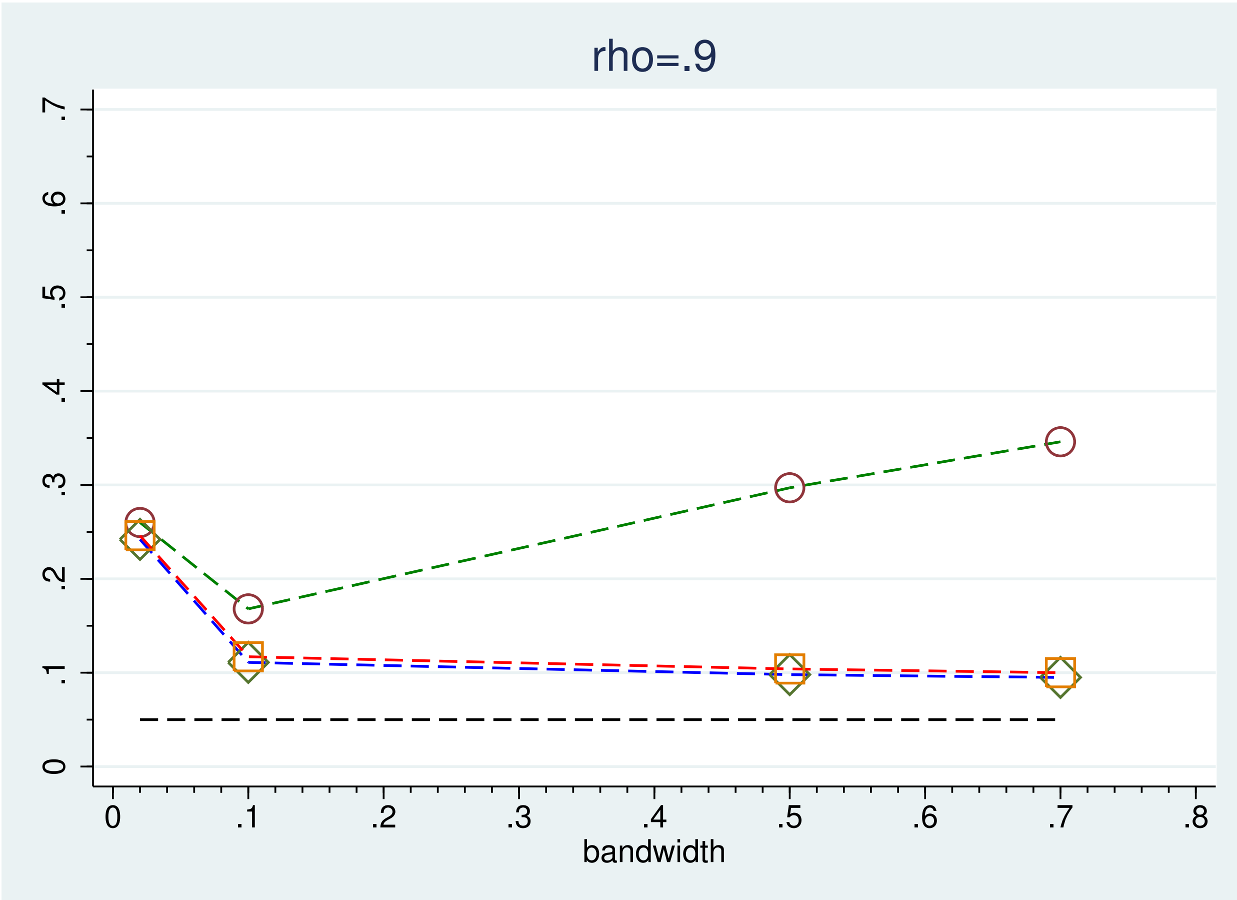
(d) N=49, T=250, z parameter

Figure E.14: (cont'd)

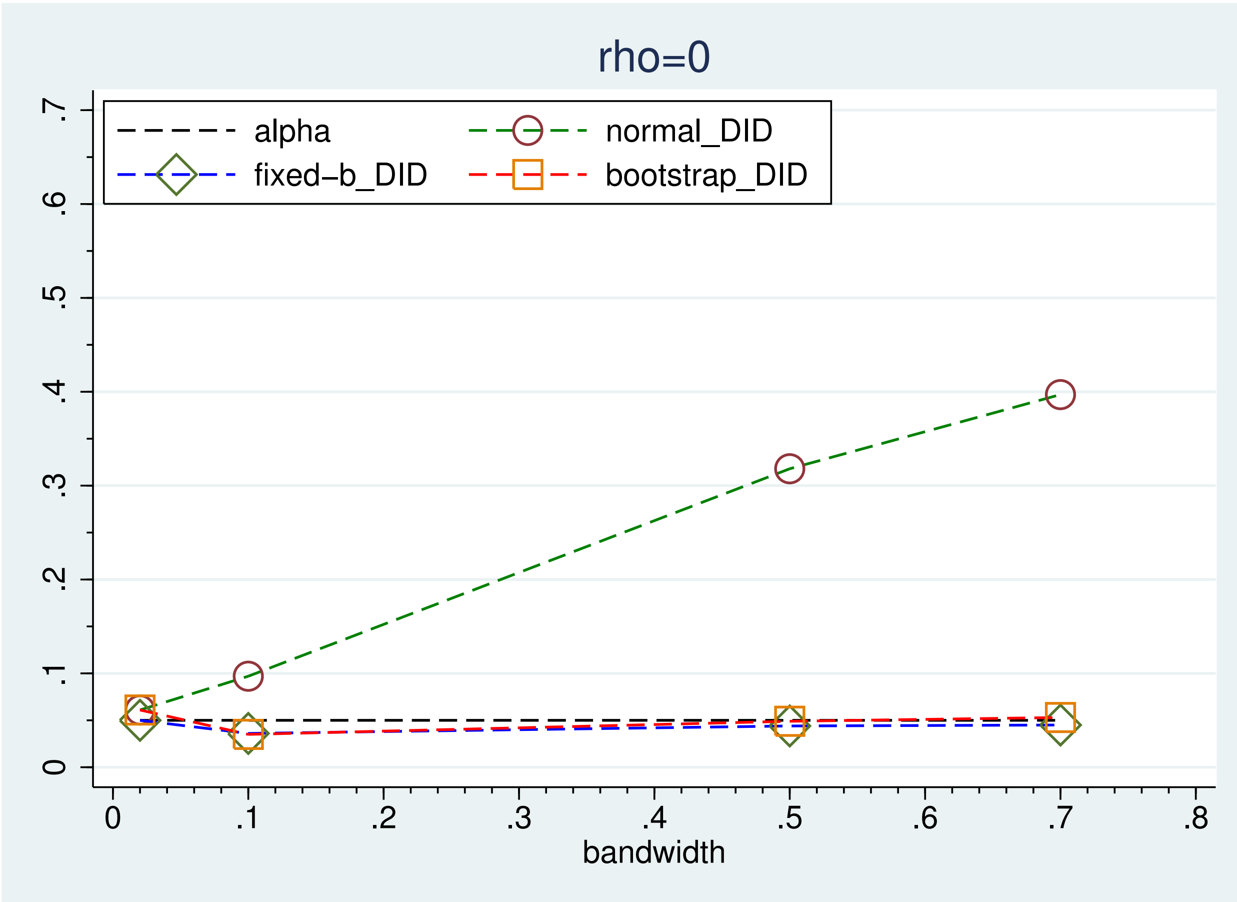


(e) N=49, T=250, z parameter

Figure E.14: (cont'd)



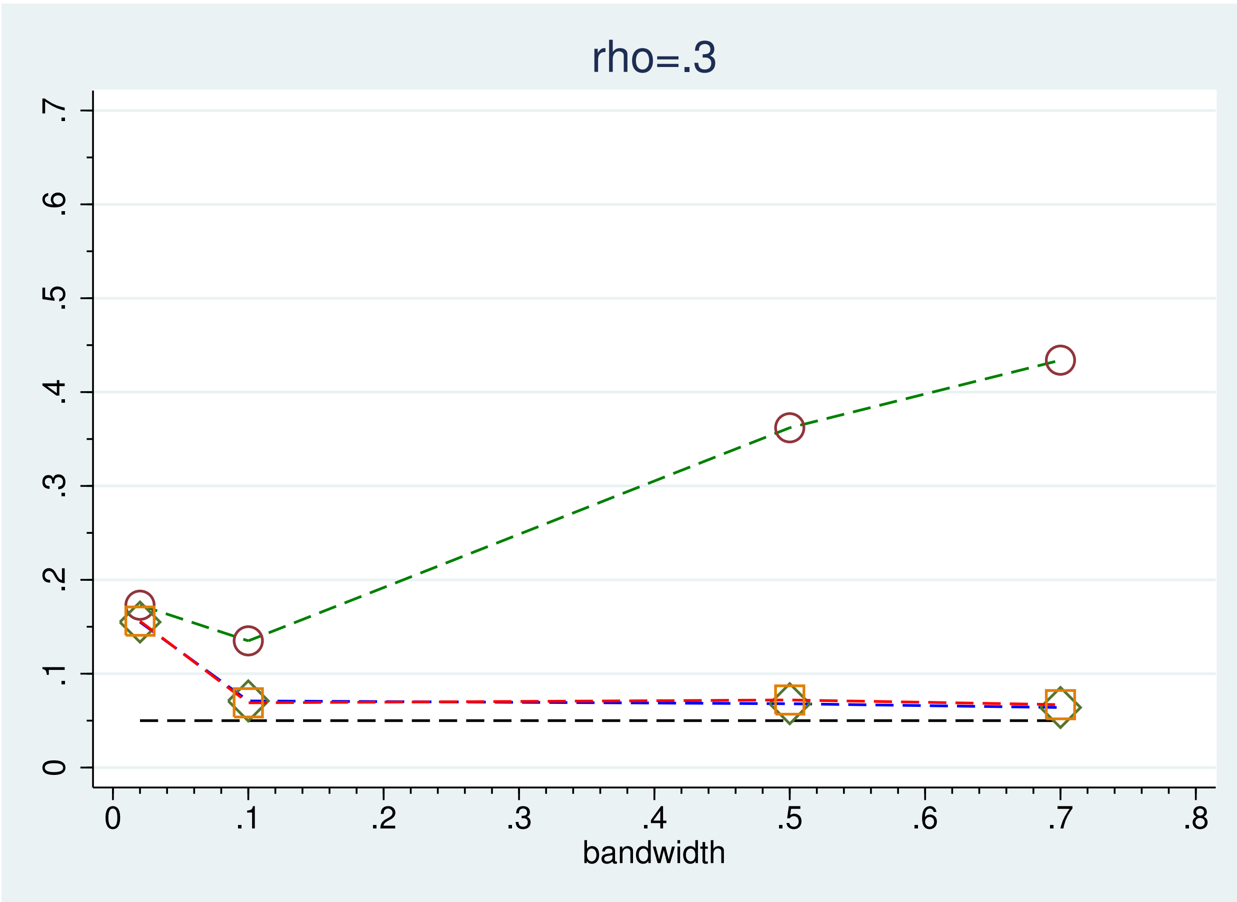
(f) N=49, T=250, z parameter



(a) N=256, T=50, DD parameter

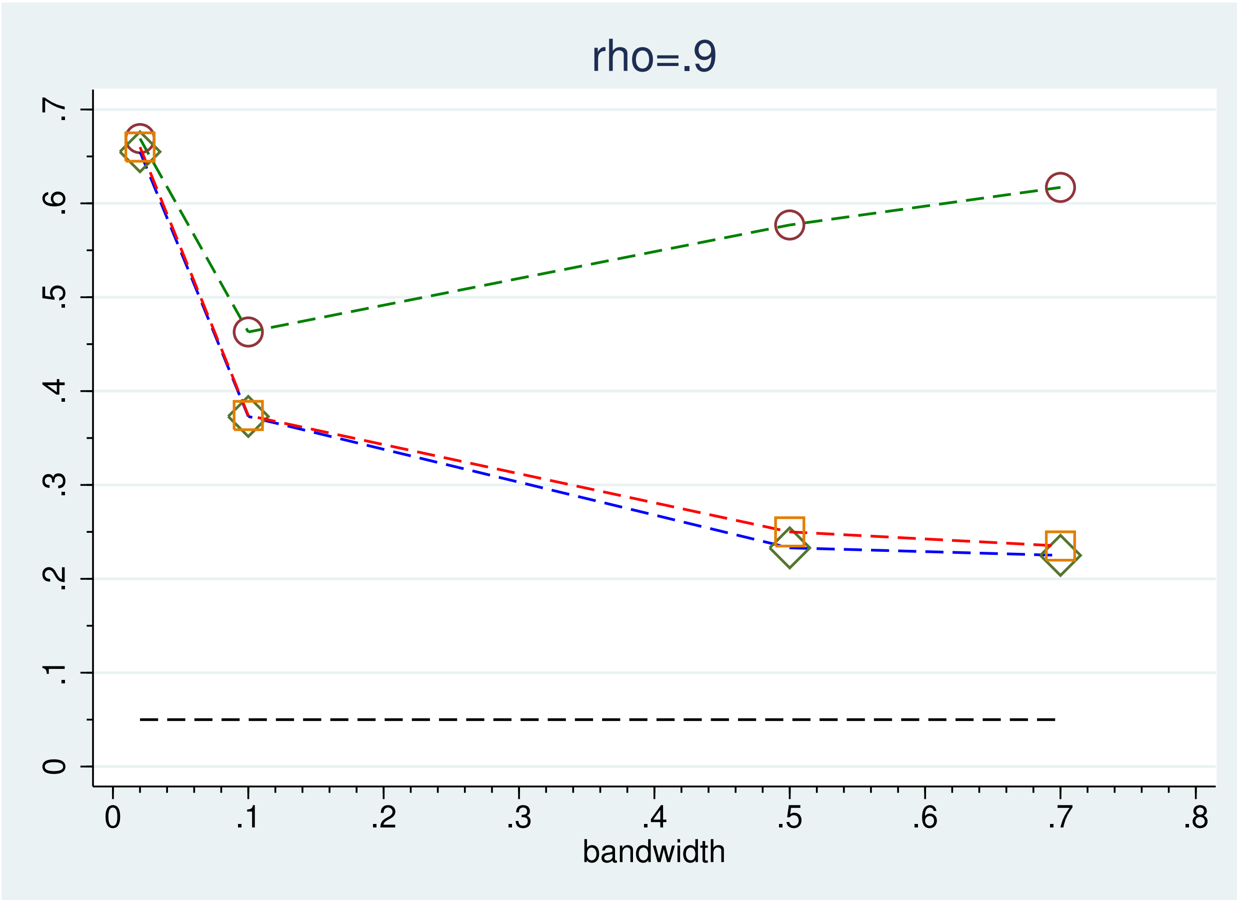
Figure E.15: Empirical null rejection probabilities, additional regressor, spatial $MA(2)$, Bartlett kernel, $N = 256$, $T = 50$, $\lambda = 0.5$.

Figure E.15: (cont'd)



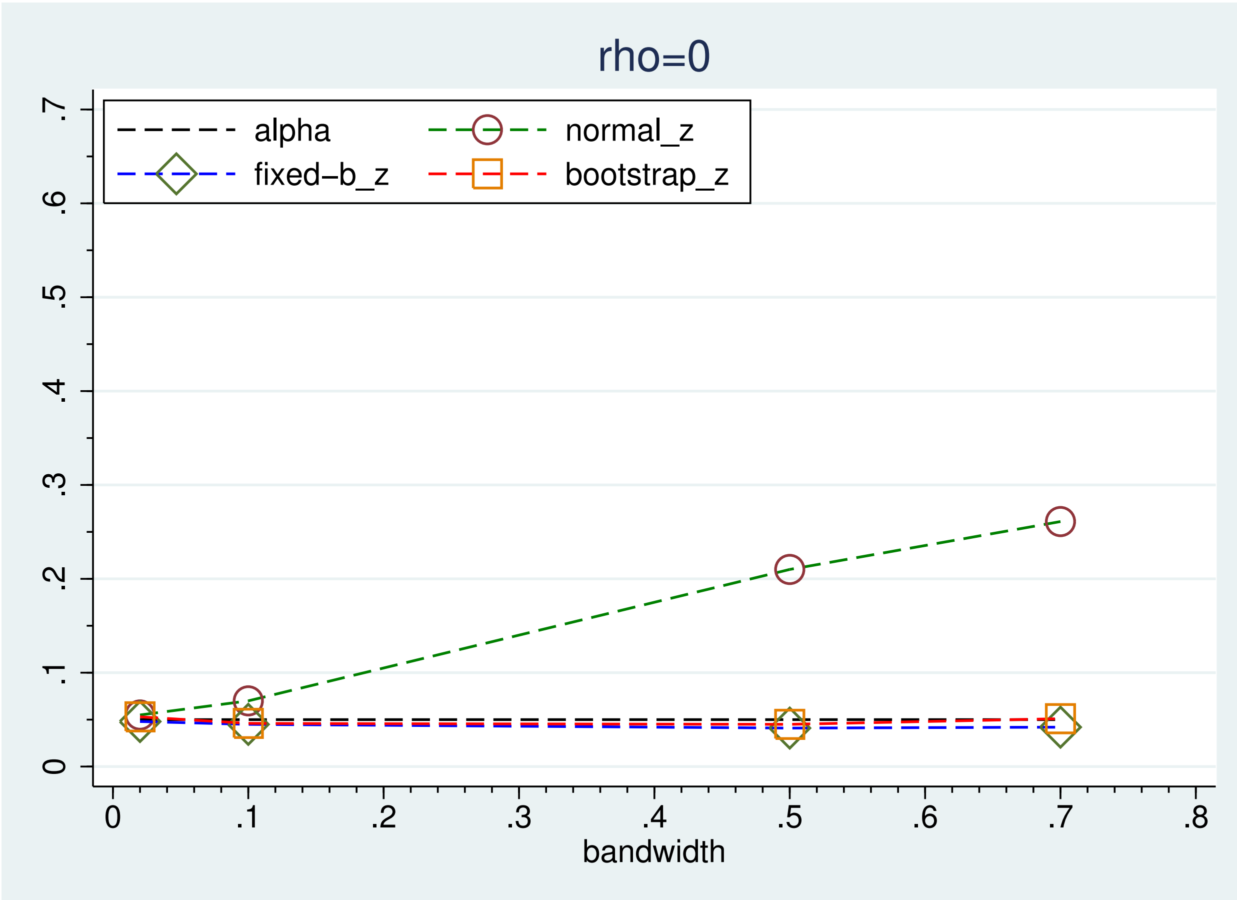
(b) N=256, T=50, DD parameter

Figure E.15: (cont'd)



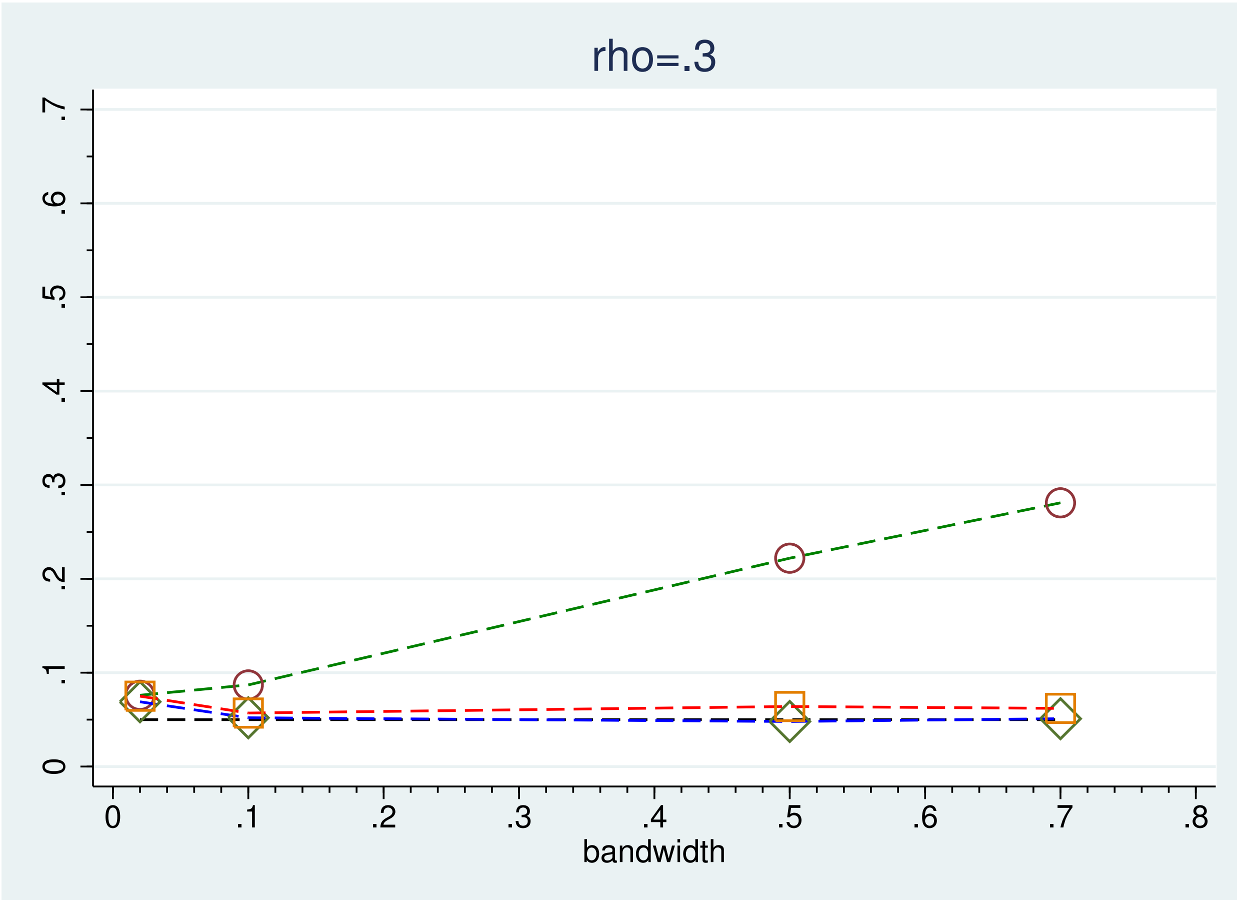
(c) N=256, T=50, DD parameter

Figure E.15: (cont'd)



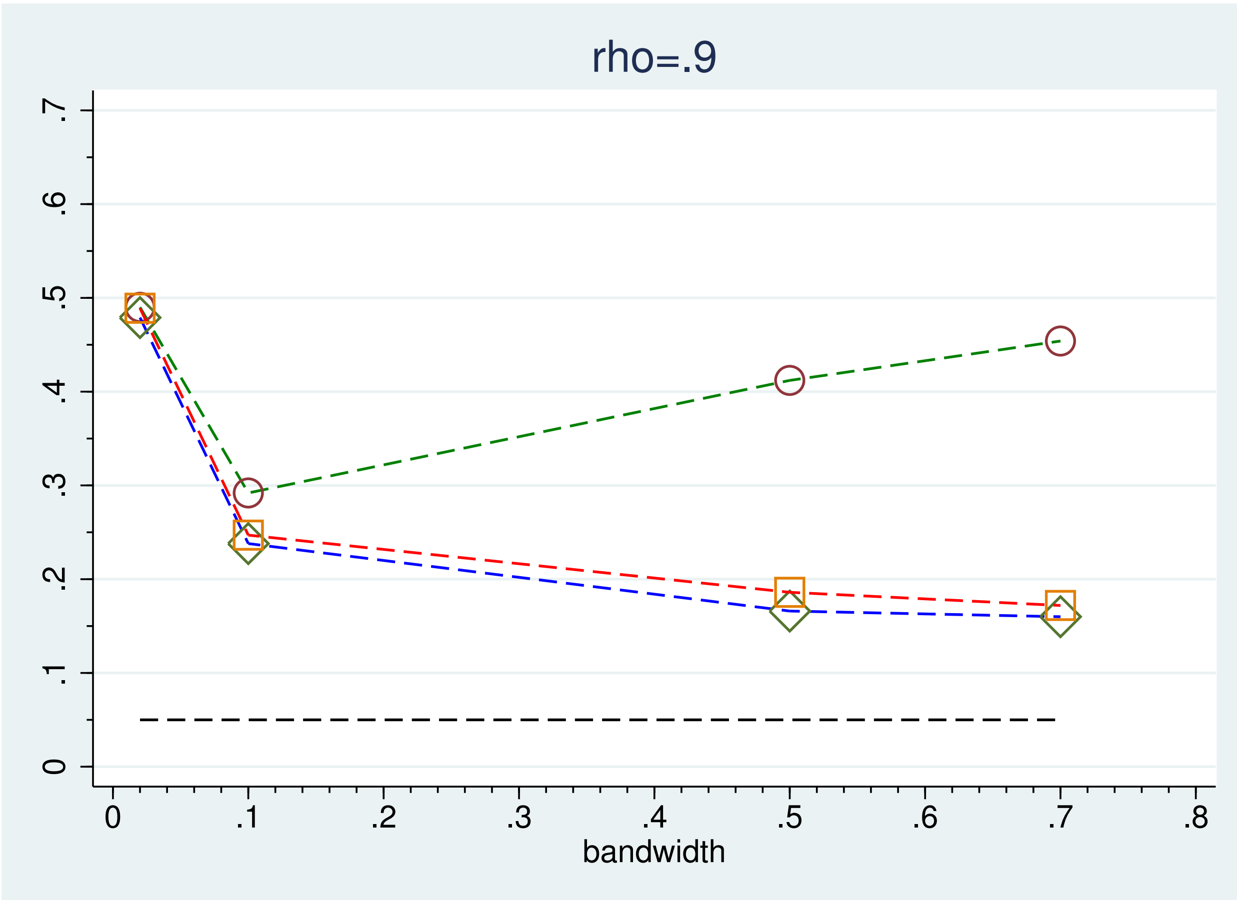
(d) N=256, T=50, z parameter

Figure E.15: (cont'd)

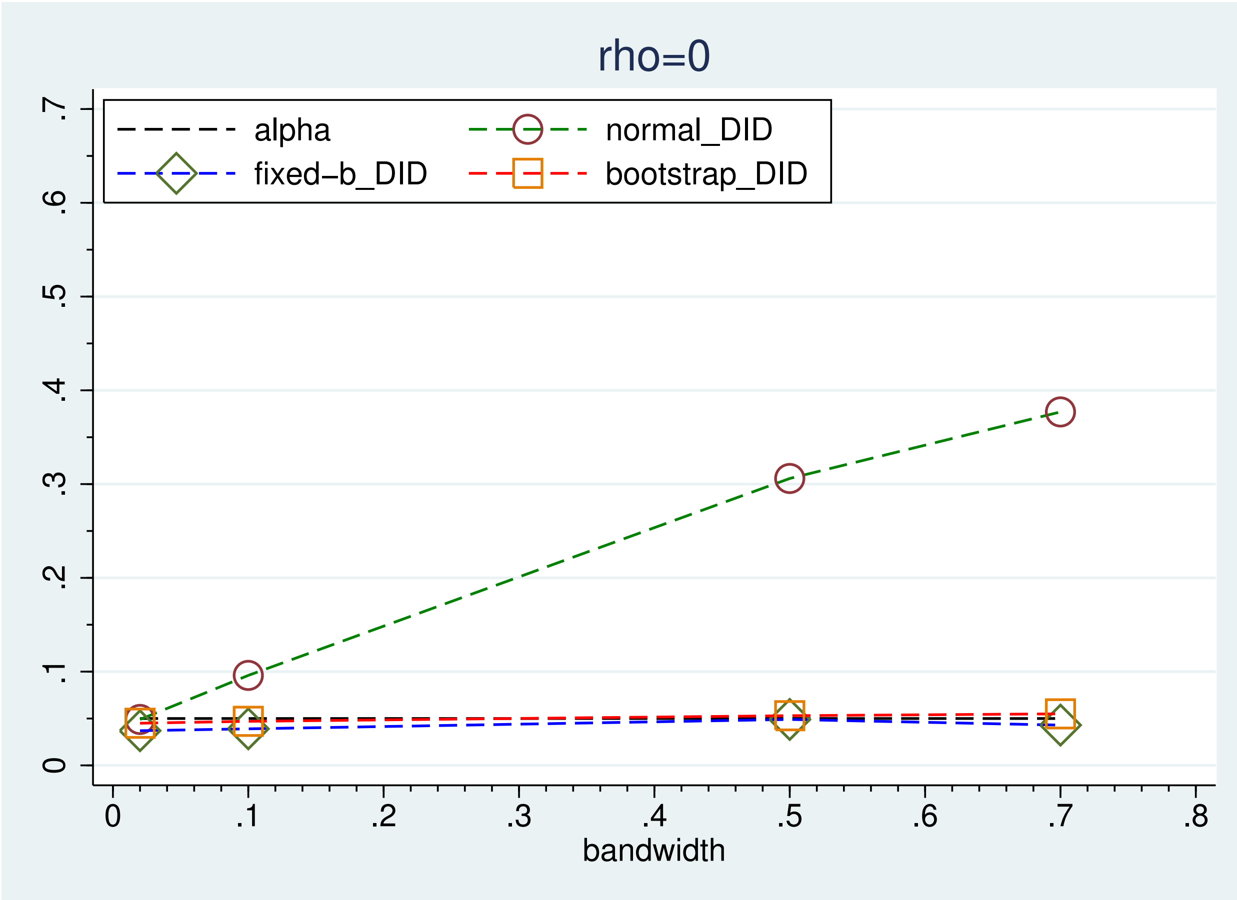


(e) N=256, T=50, z parameter

Figure E.15: (cont'd)



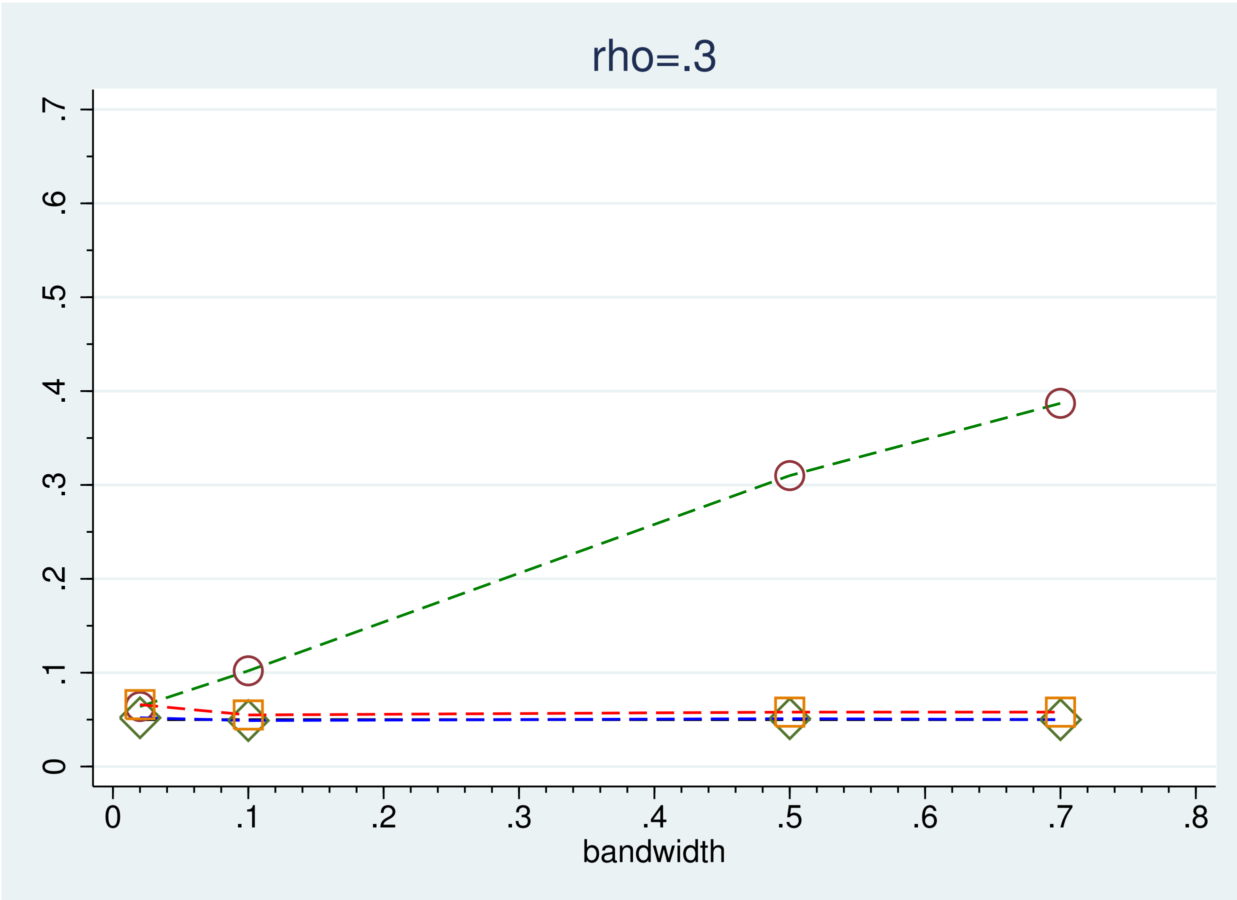
(f) N=256, T=50, z parameter



(a) N=256, T=250, DD parameter

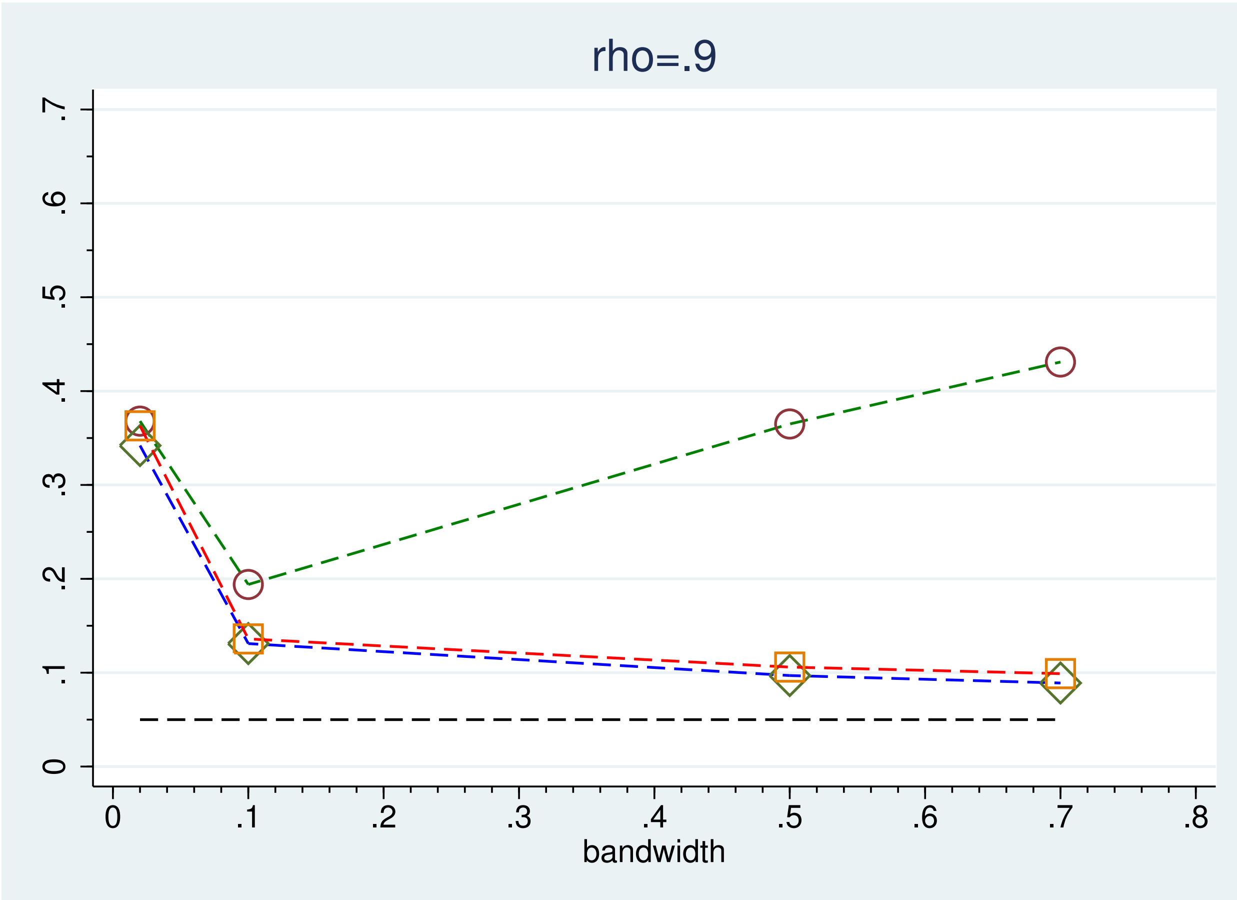
Figure E.16: Empirical null rejection probabilities, additional regressor, spatial $MA(2)$, Bartlett kernel, $N = 256$, $T = 250$, $\lambda = 0.5$.

Figure E.16: (cont'd)



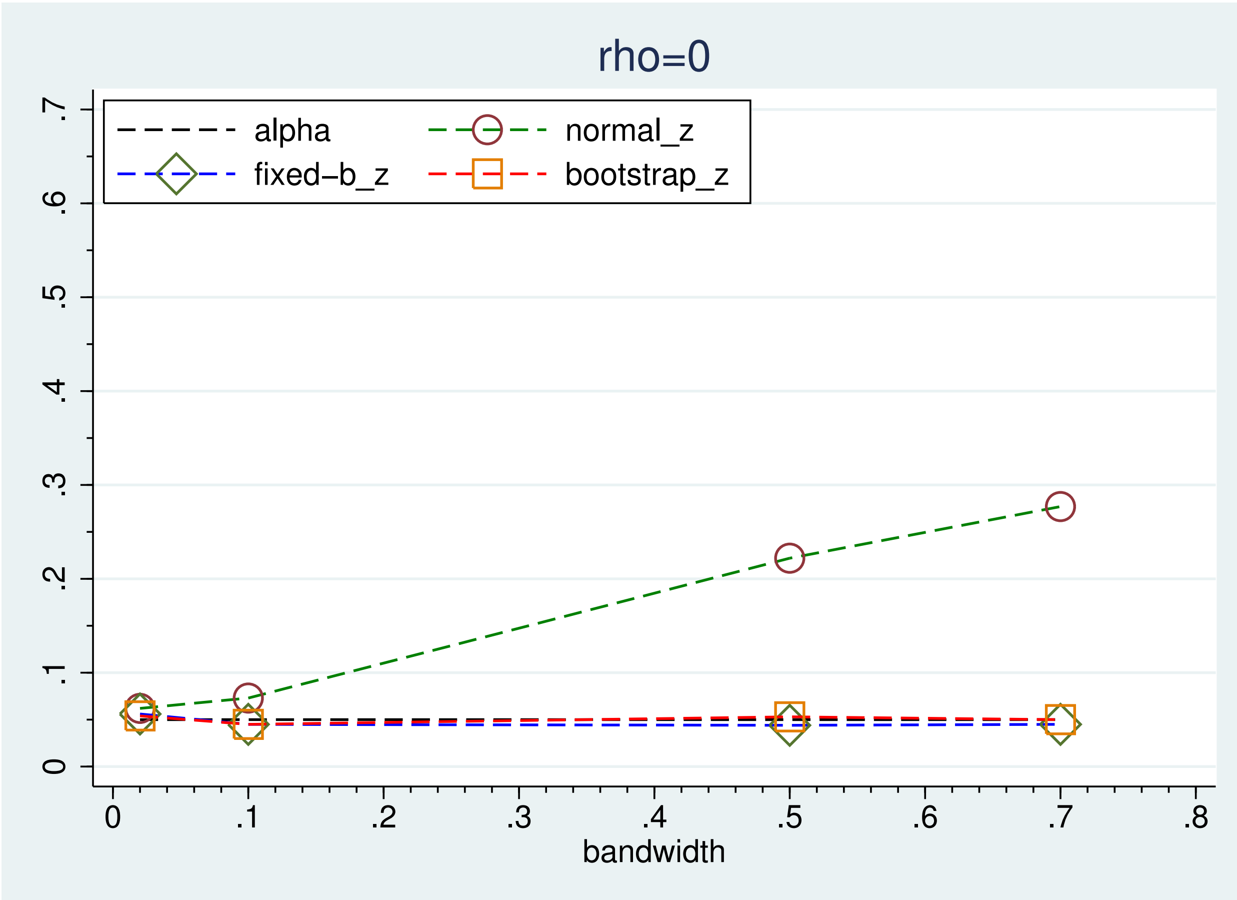
(b) N=256, T=250, DD parameter

Figure E.16: (cont'd)



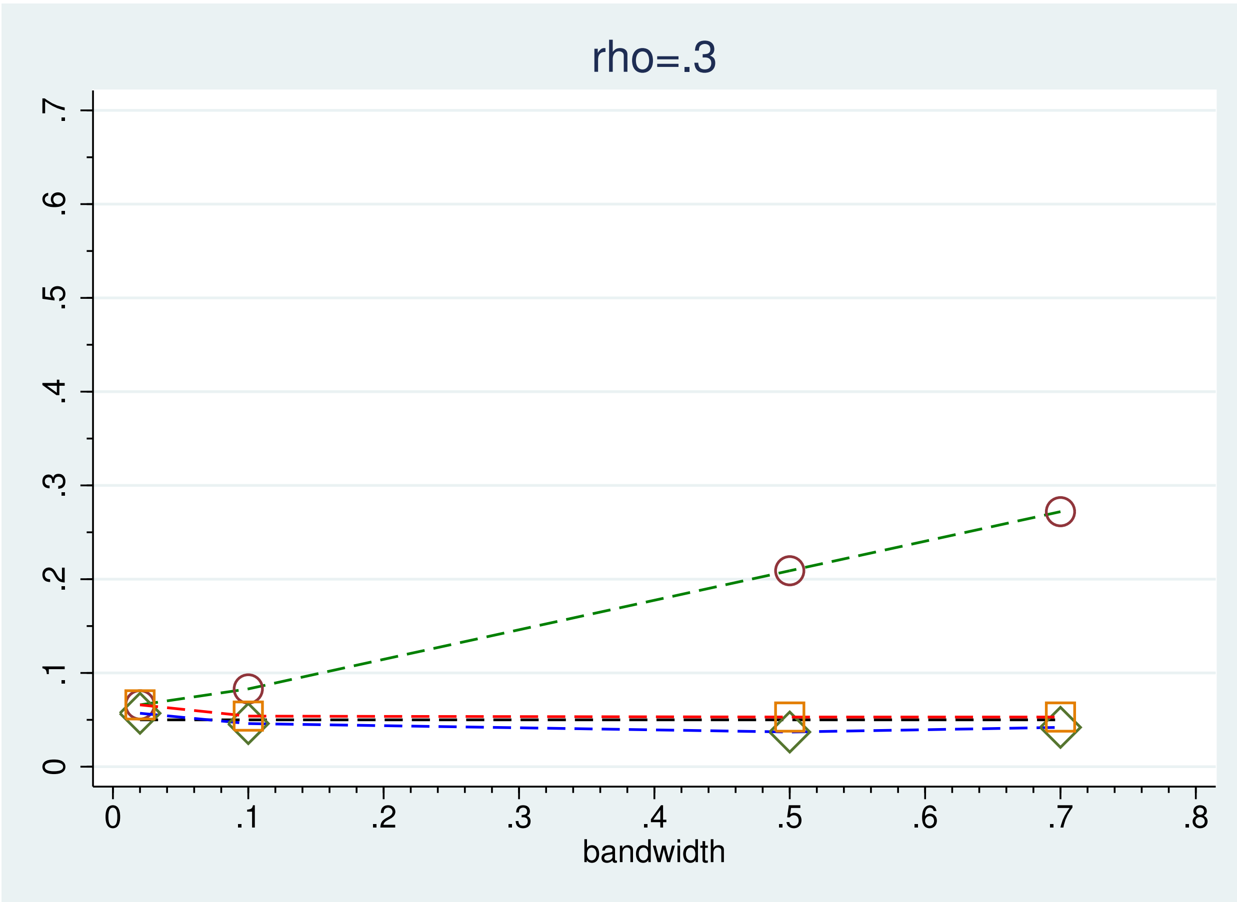
(c) N=256, T=250, DD parameter

Figure E.16: (cont'd)



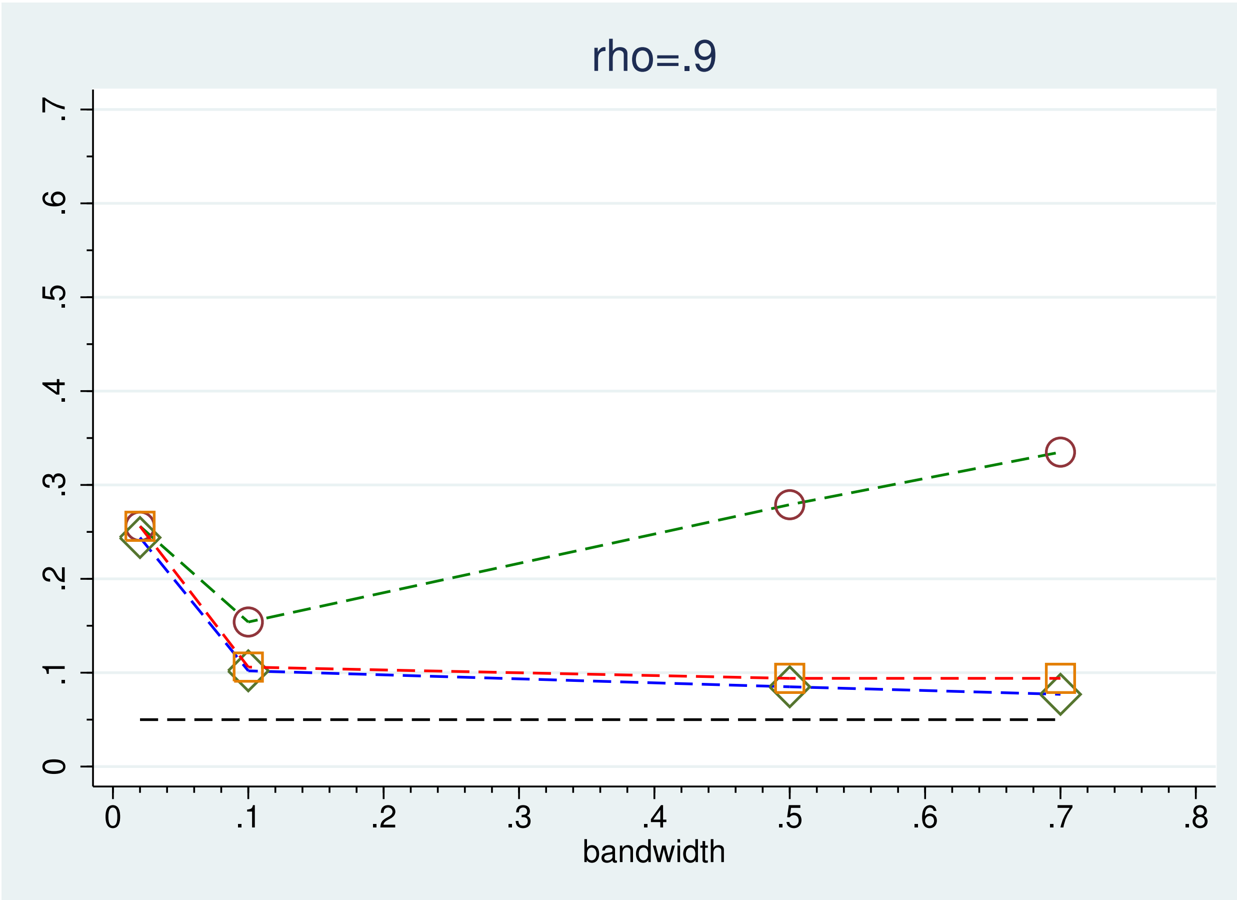
(d) N=256, T=250, z parameter

Figure E.16: (cont'd)

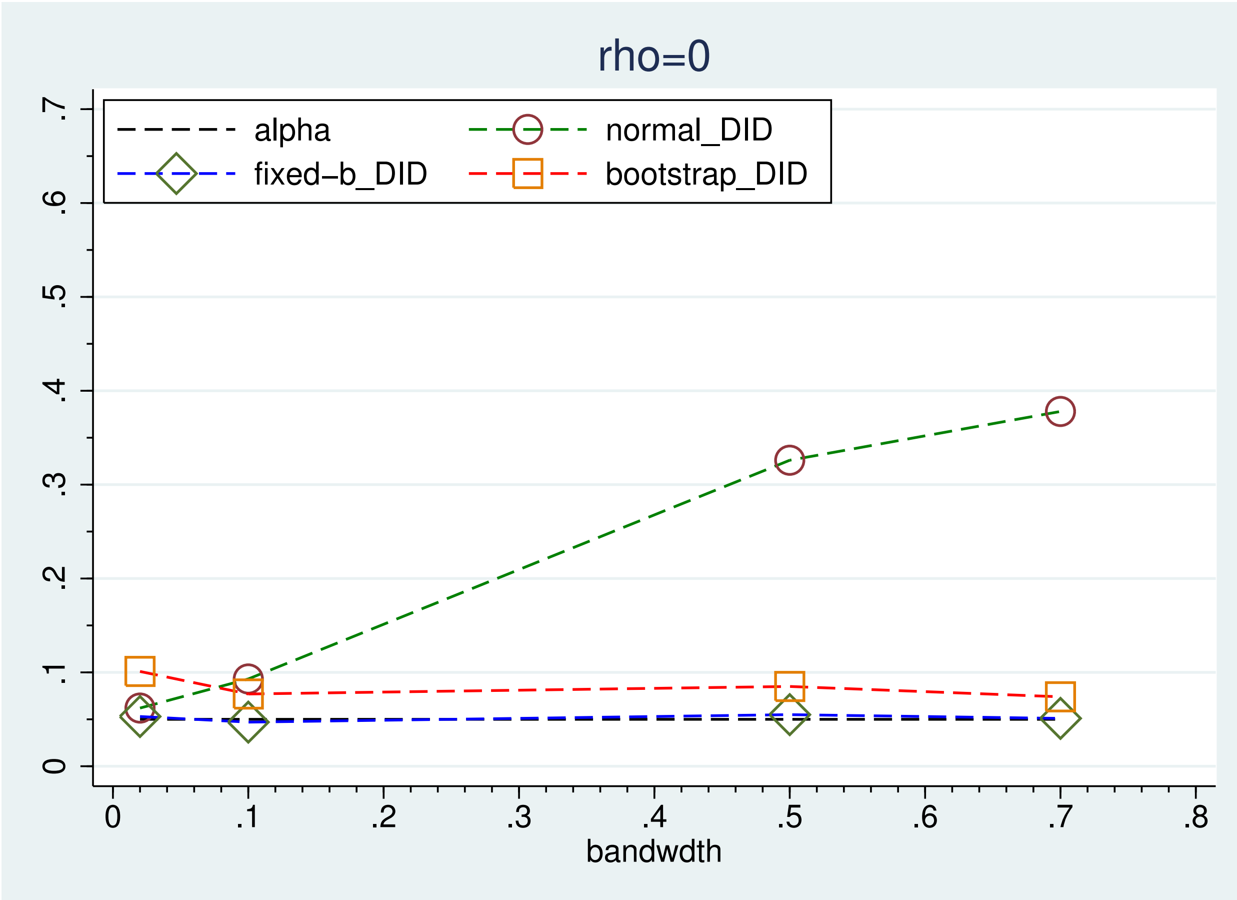


(e) N=256, T=250, z parameter

Figure E.16: (cont'd)



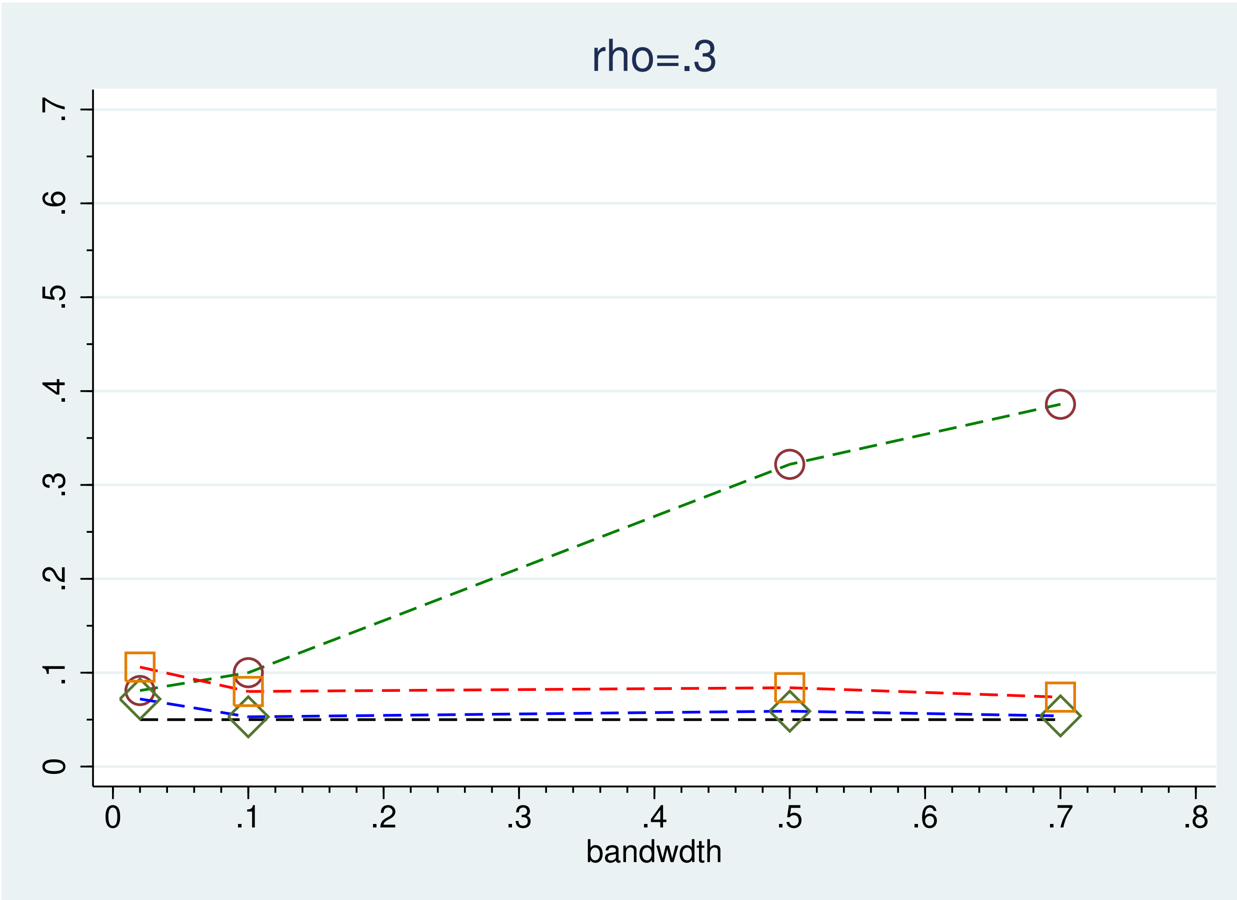
(f) N=256, T=250, z parameter



(a) N=49, T=250, l=25, DD parameter

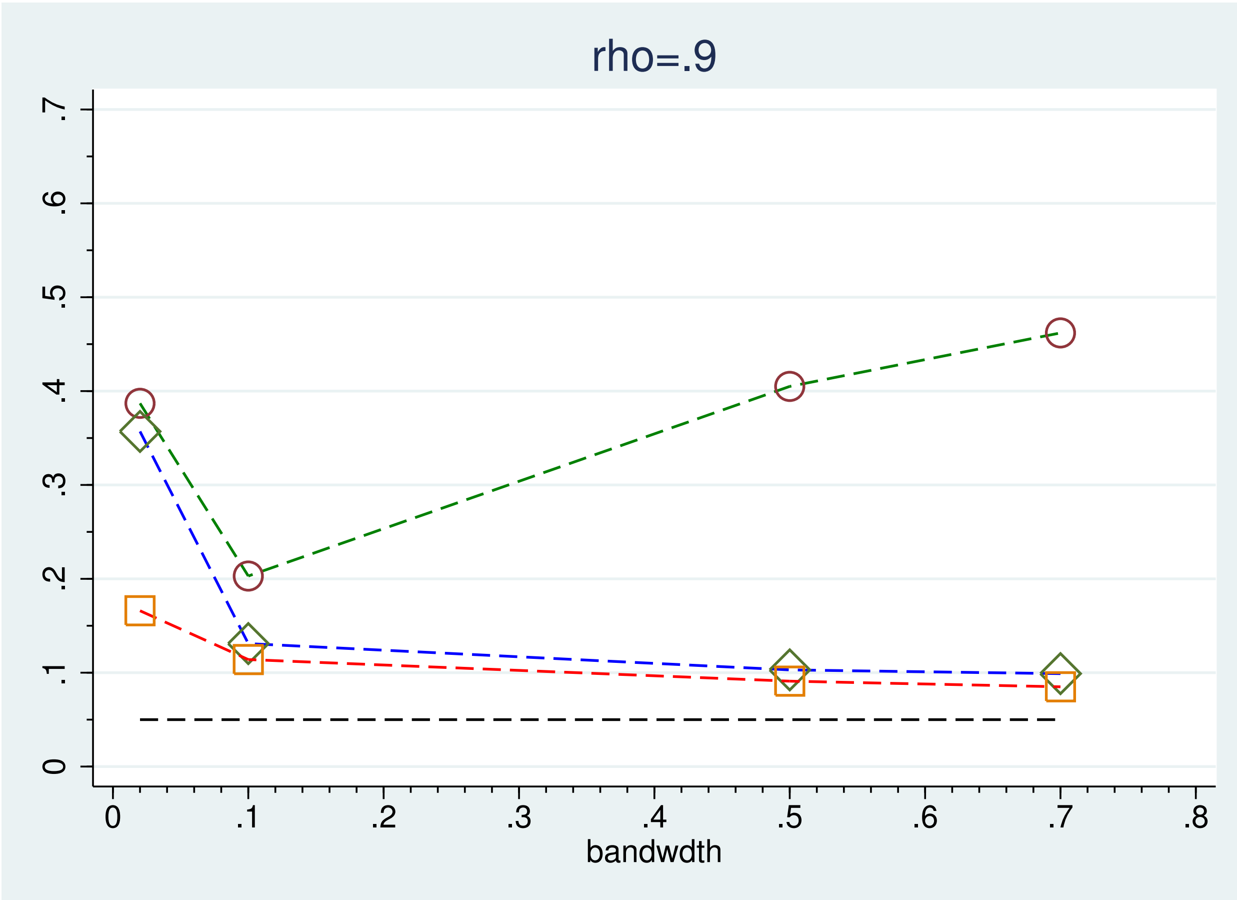
Figure E.17: Empirical null rejection probabilities for DD parameter, additional regressor, spatial $MA(2)$, Bartlett kernel, $N = 49$, $T = 250$, $\lambda = 0.5$.

Figure E.17: (cont'd)



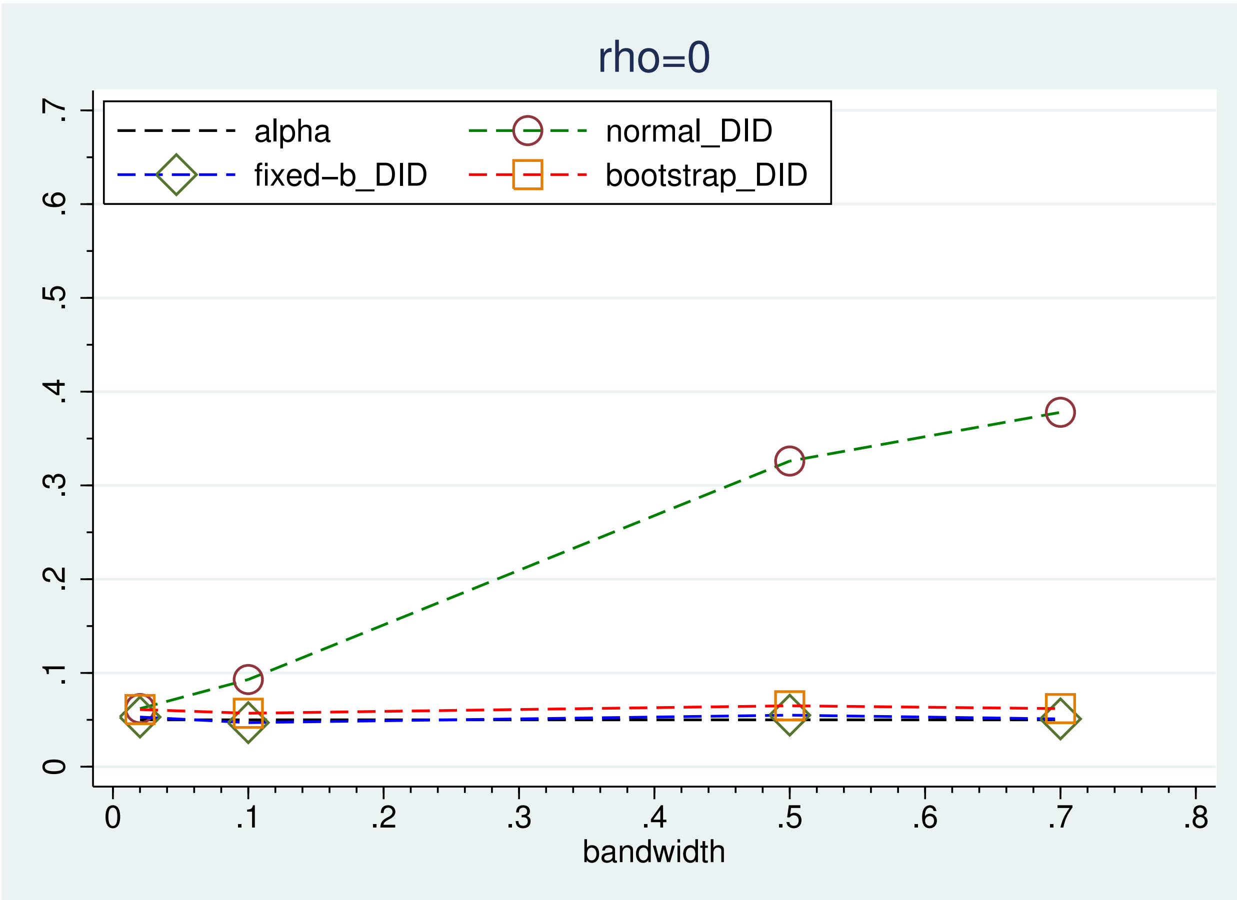
(b) N=49, T=250, l=25, DD parameter

Figure E.17: (cont'd)



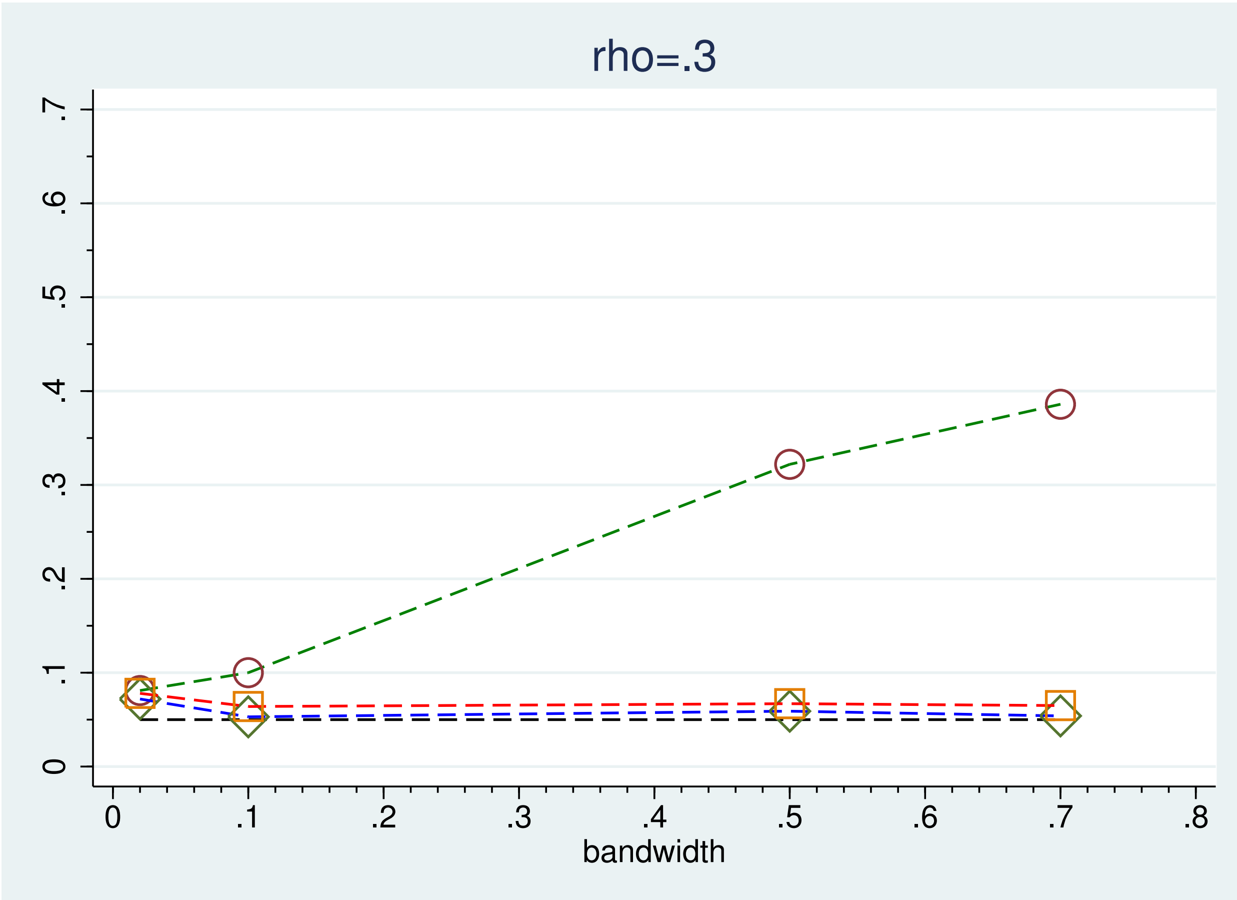
(c) N=49, T=250, l=25, DD parameter

Figure E.17: (cont'd)



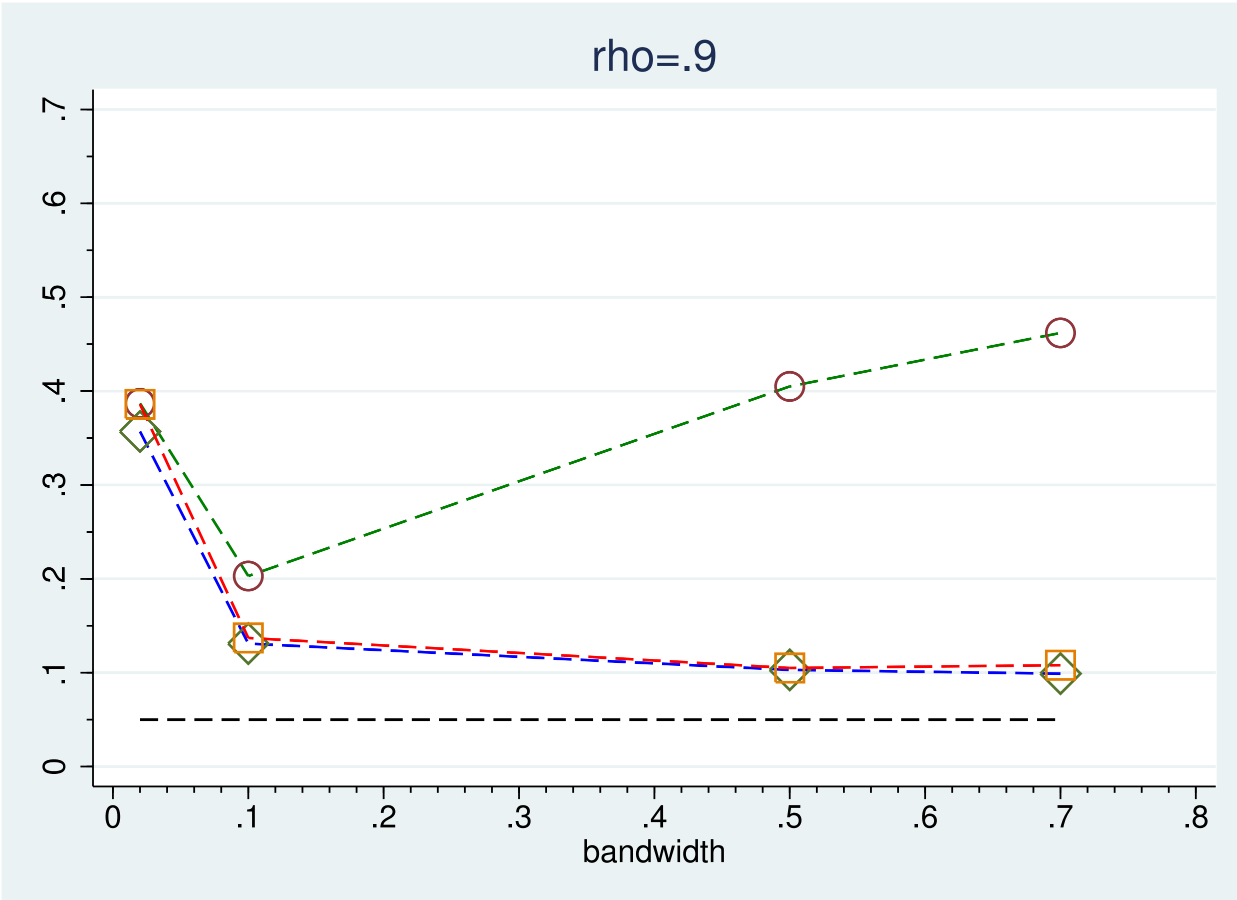
(d) N=49, T=250, l=1, DD parameter

Figure E.17: (cont'd)

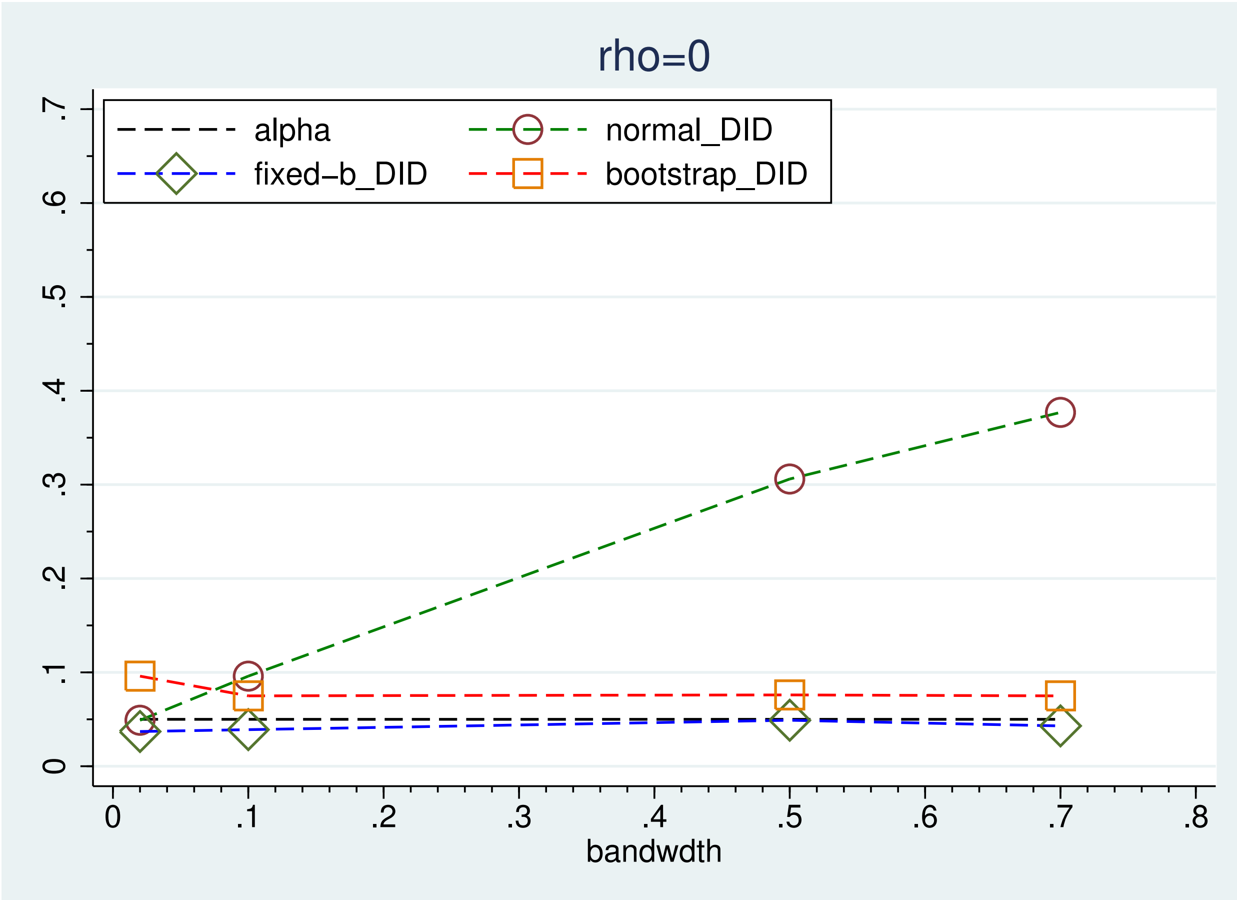


(e) N=49, T=250, l=1, DD parameter

Figure E.17: (cont'd)



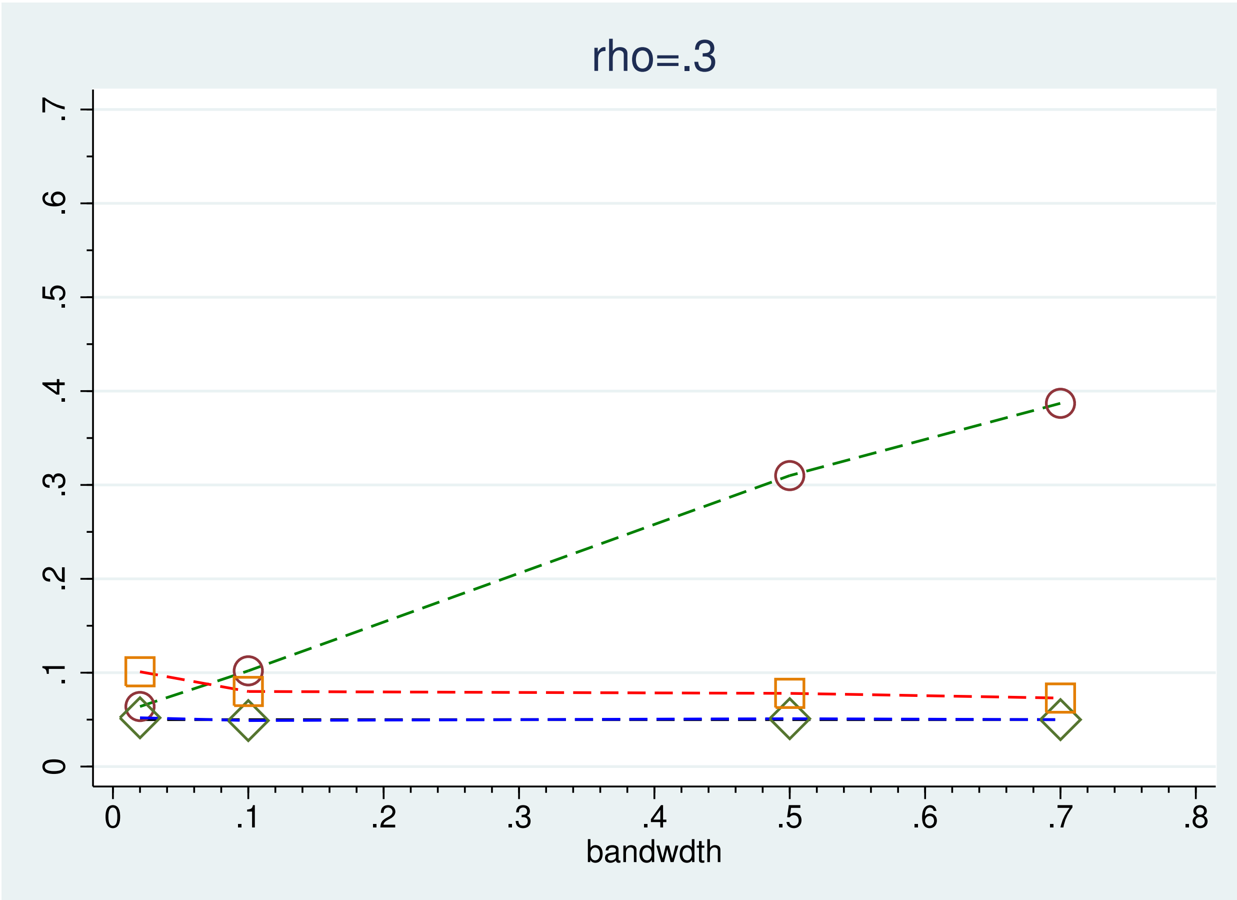
(f) N=49, T=250, l=1, DD parameter



(a) N=256, T=250, l=25, DD parameter

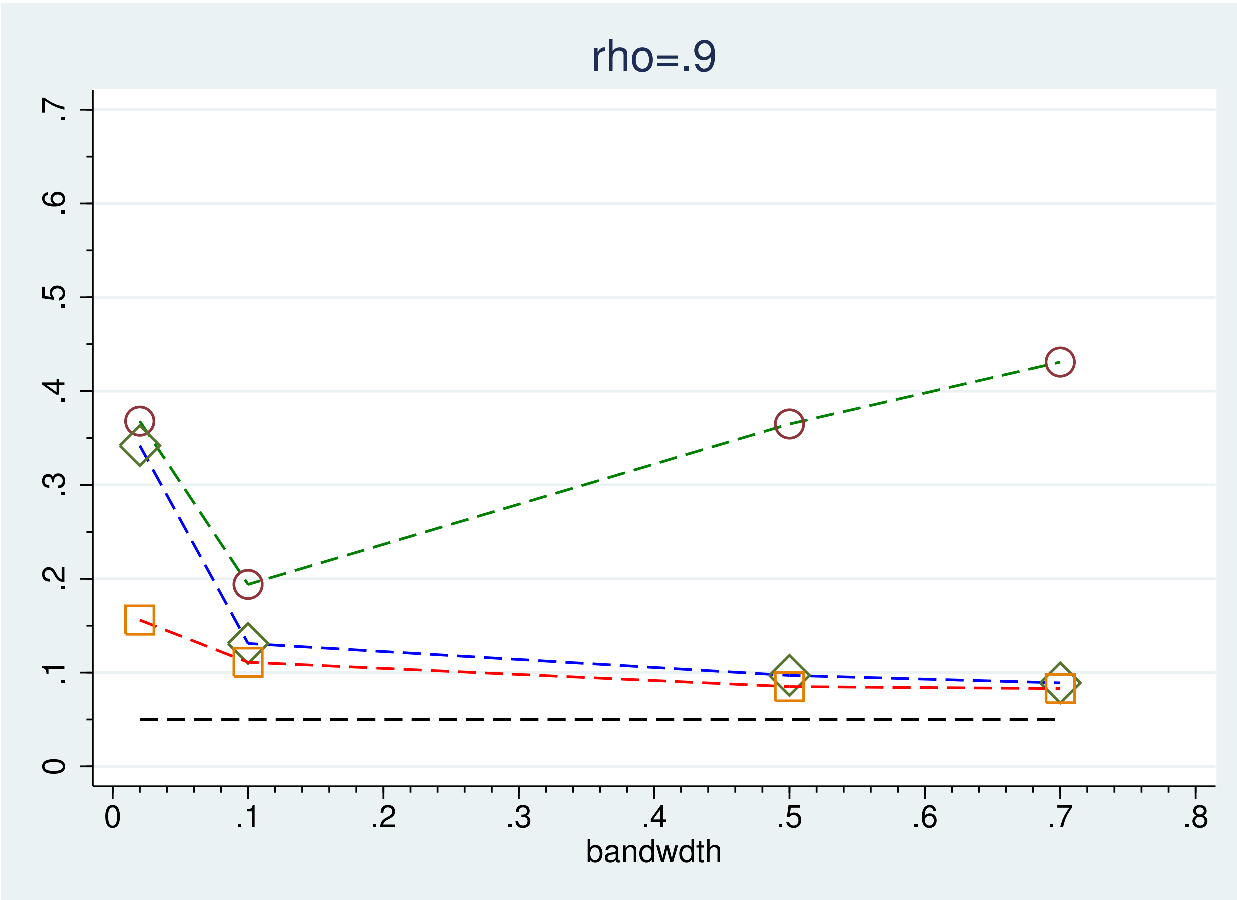
Figure E.18: Empirical null rejection probabilities for DD parameter, additional regressor, spatial $MA(2)$, Bartlett kernel, $N = 256$, $T = 250$, $\lambda = 0.5$.

Figure E.18: (cont'd)



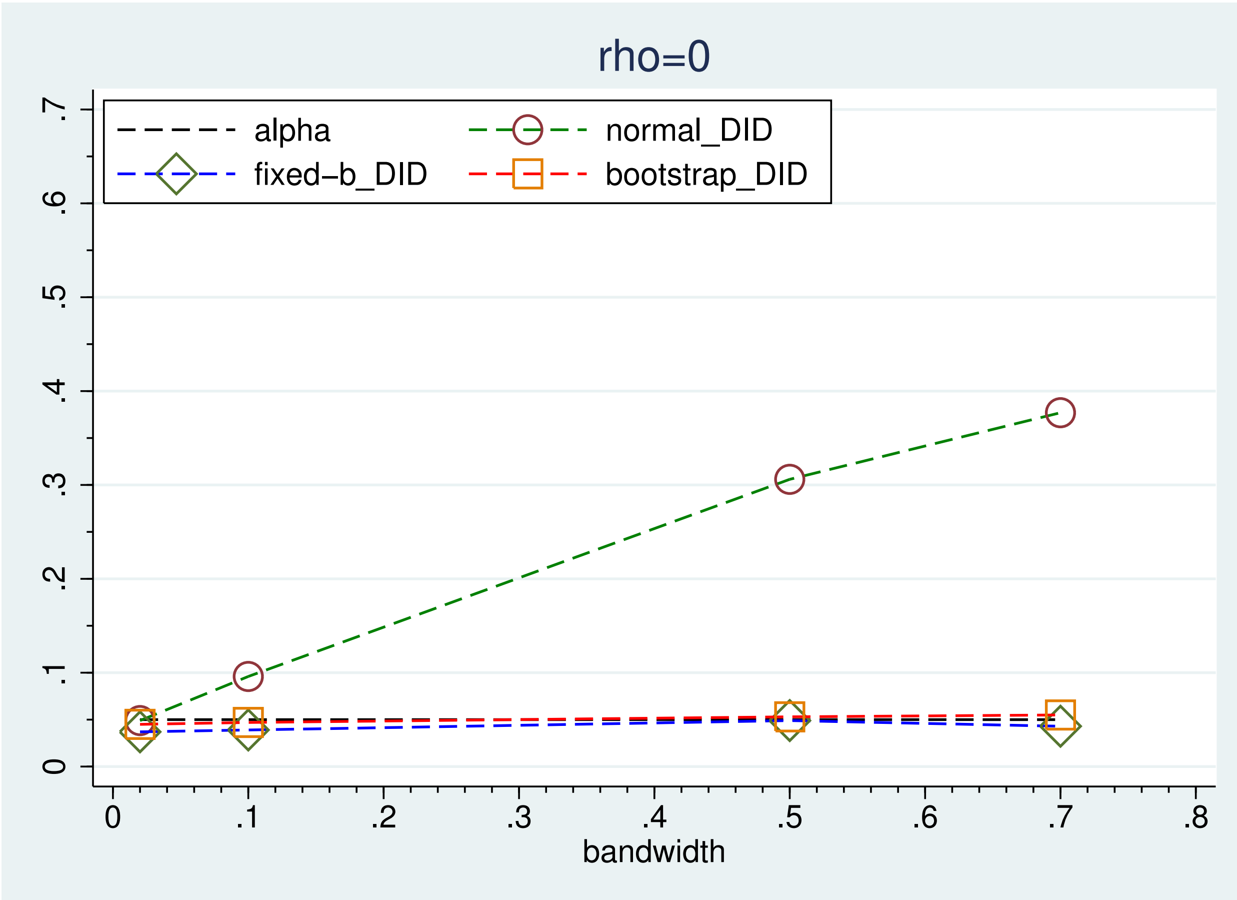
(b) N=256, T=250, l=25, DD parameter

Figure E.18: (cont'd)



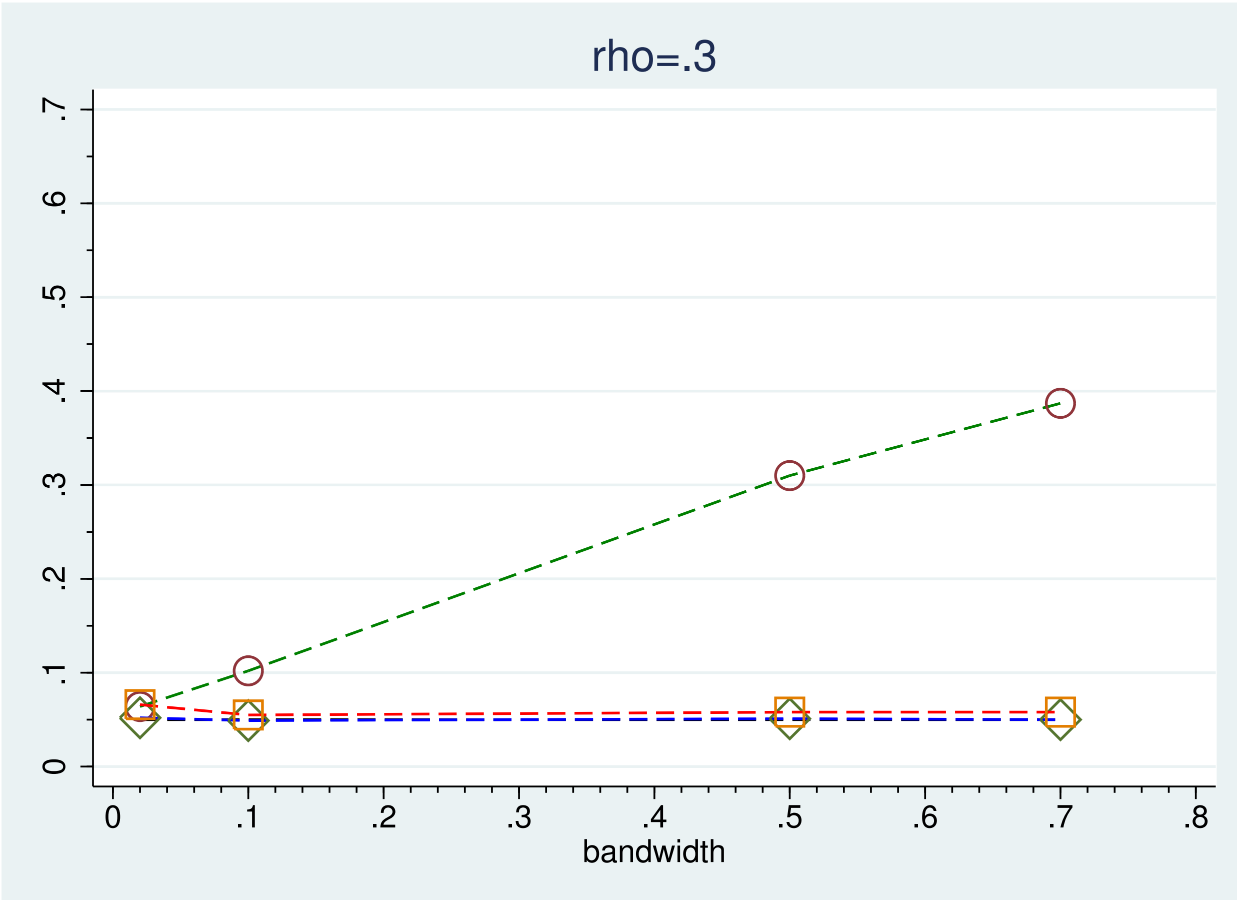
(c) N=256, T=250, l=25, DD parameter

Figure E.18: (cont'd)



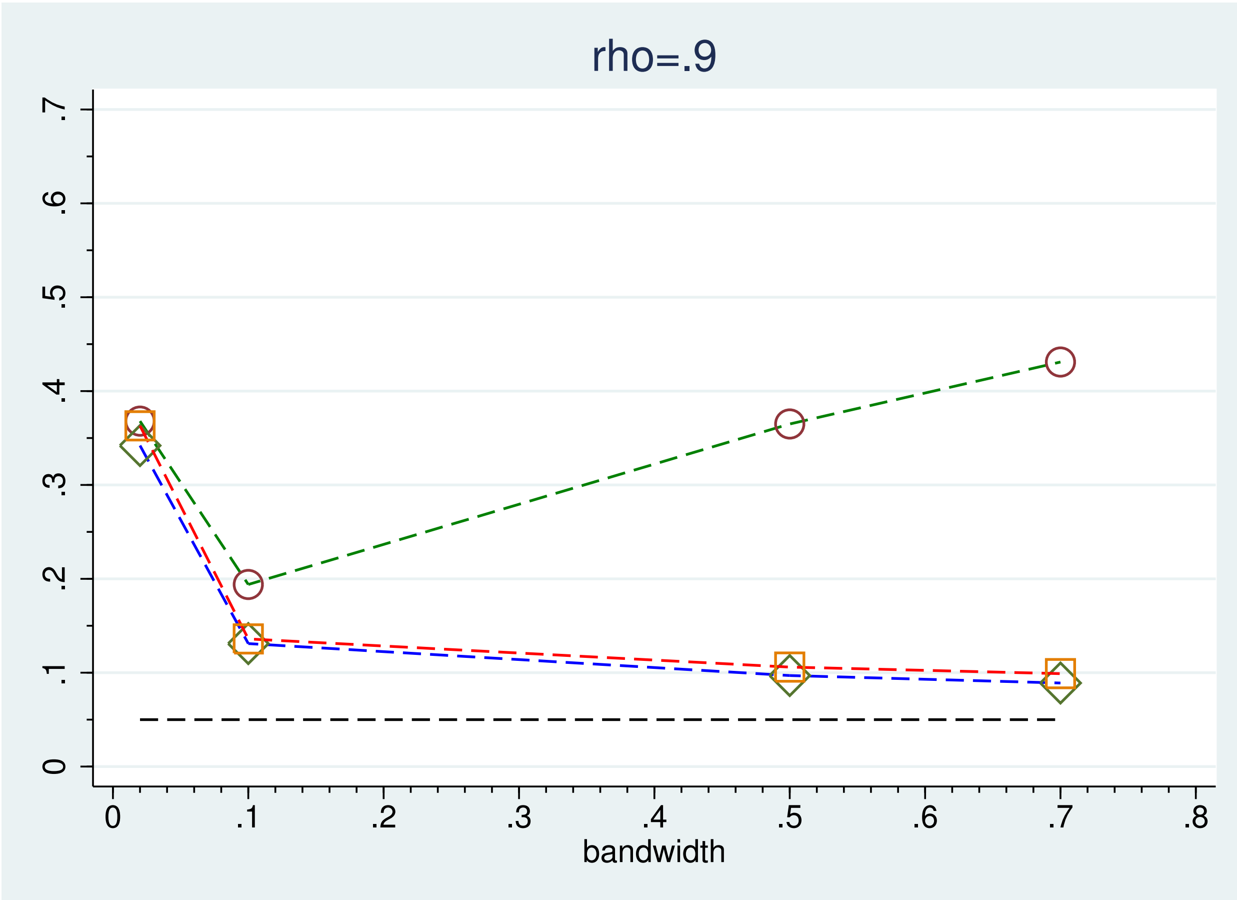
(d) N=256, T=250, l=1, DD parameter

Figure E.18: (cont'd)

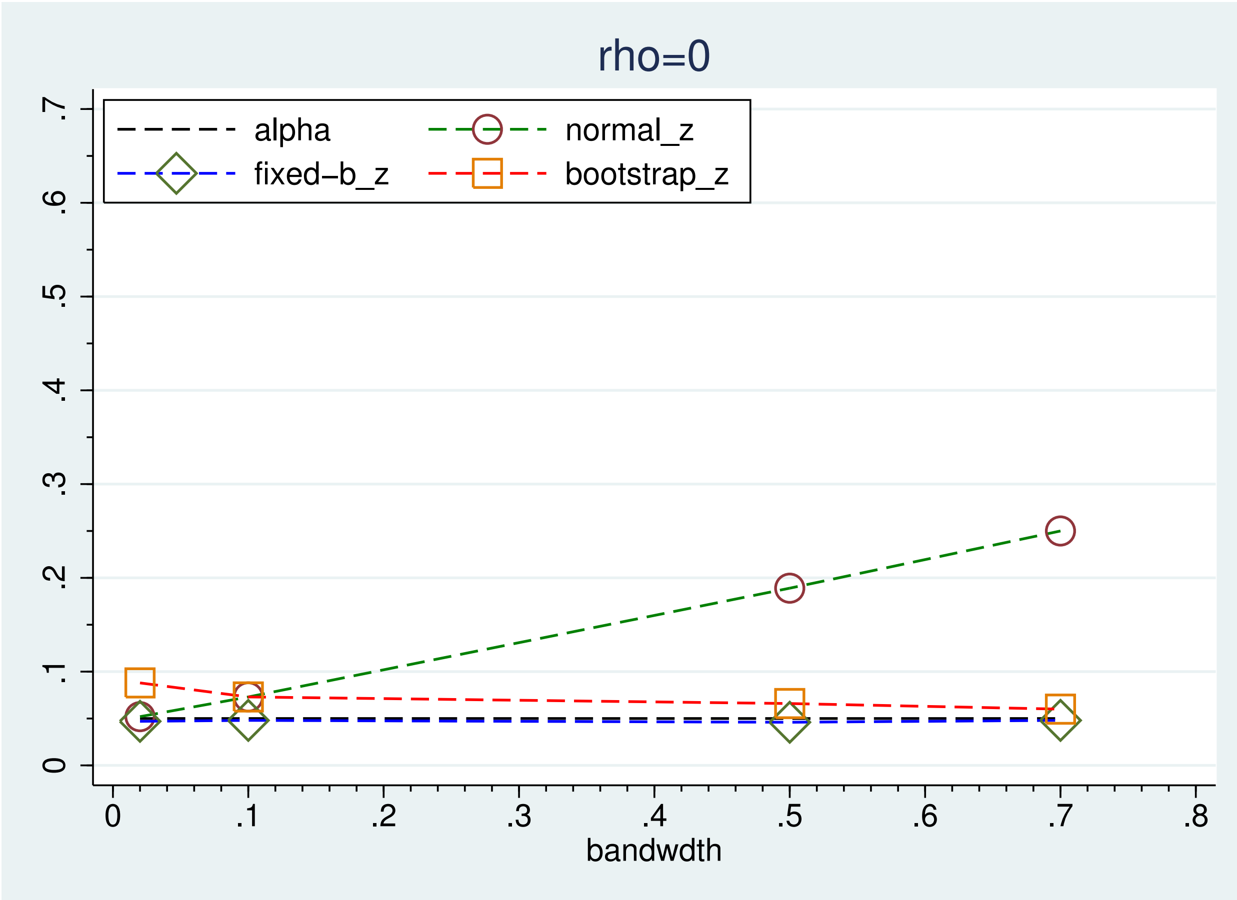


(e) N=256, T=250, l=1, DD parameter

Figure E.18: (cont'd)



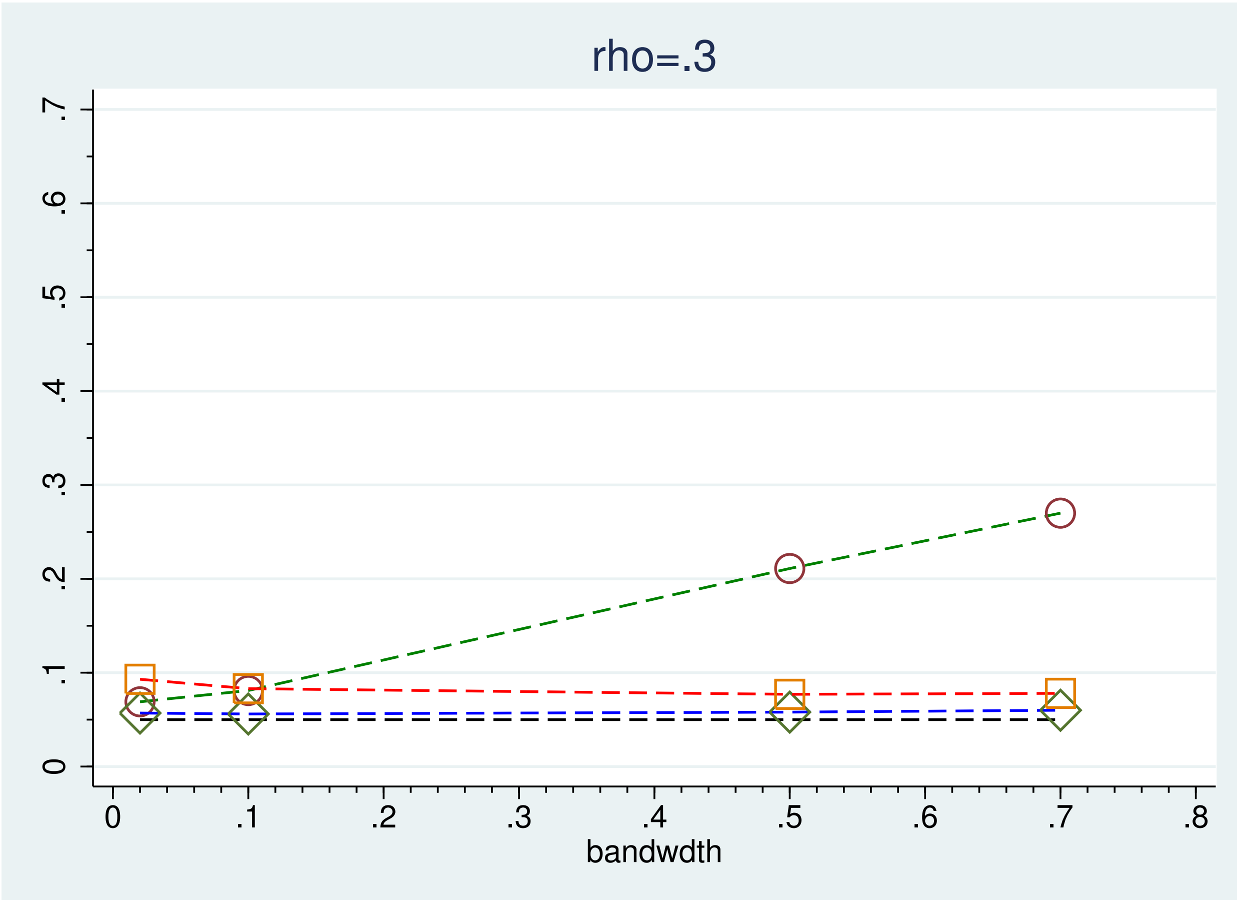
(f) N=256, T=250, l=1, DD parameter



(a) N=49, T=250, l=25, z parameter

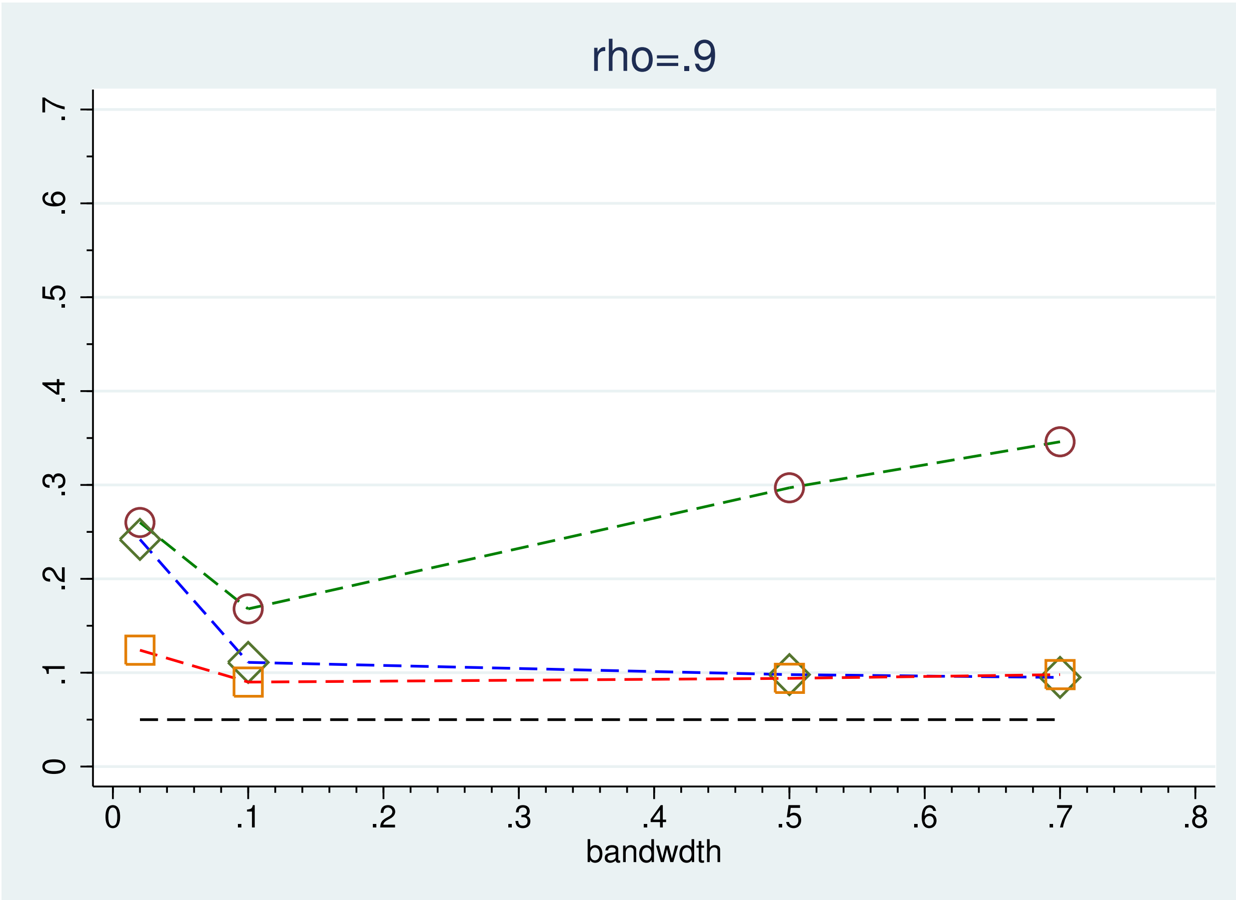
Figure E.19: Empirical null rejection probabilities for z parameter, additional regressor, spatial $MA(2)$, Bartlett kernel, $N = 49$, $T = 250$, $\lambda = 0.5$.

Figure E.19: (cont'd)



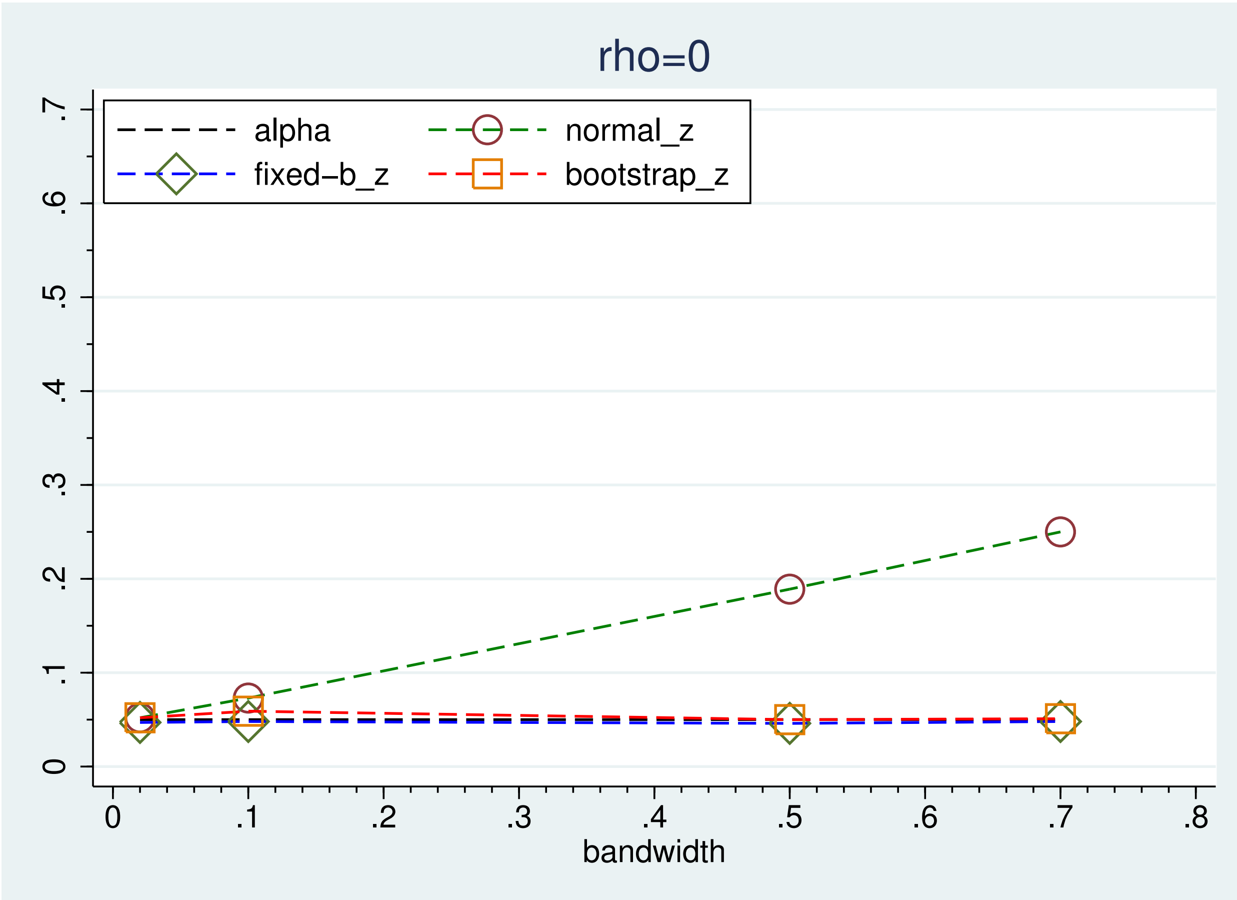
(b) N=49, T=250, l=25, z parameter

Figure E.19: (cont'd)



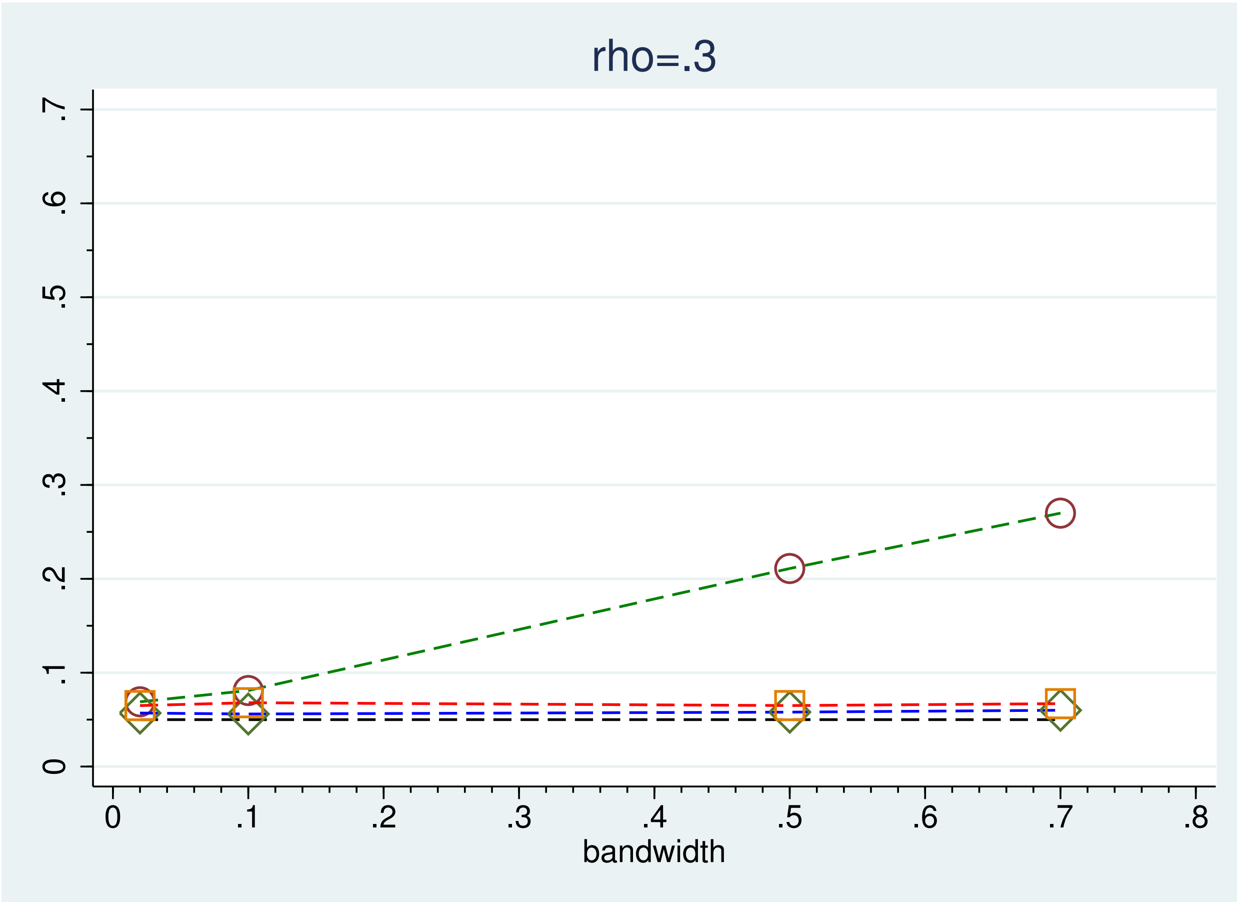
(c) N=49, T=250, l=25, z parameter

Figure E.19: (cont'd)



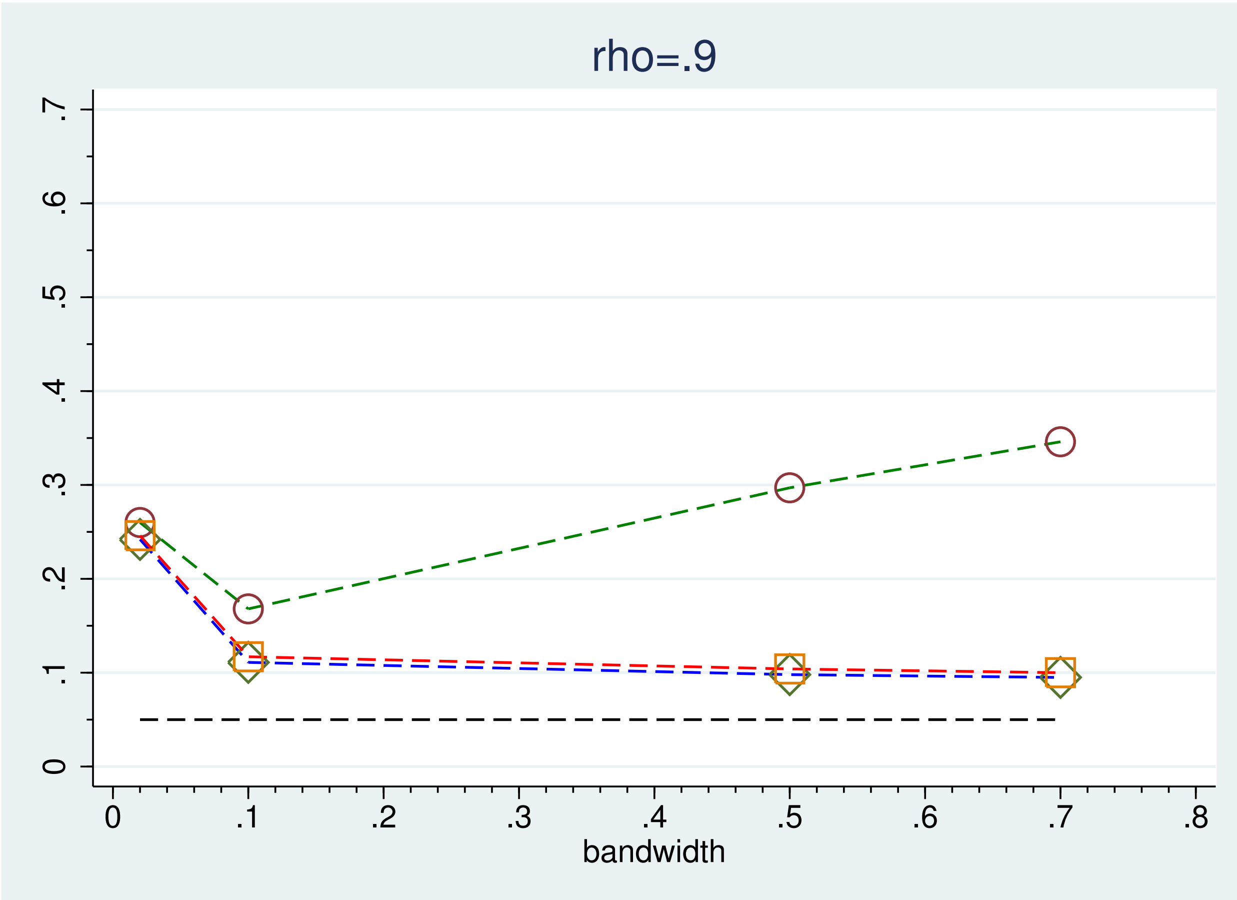
(d) N=49, T=250, l=1, z parameter

Figure E.19: (cont'd)

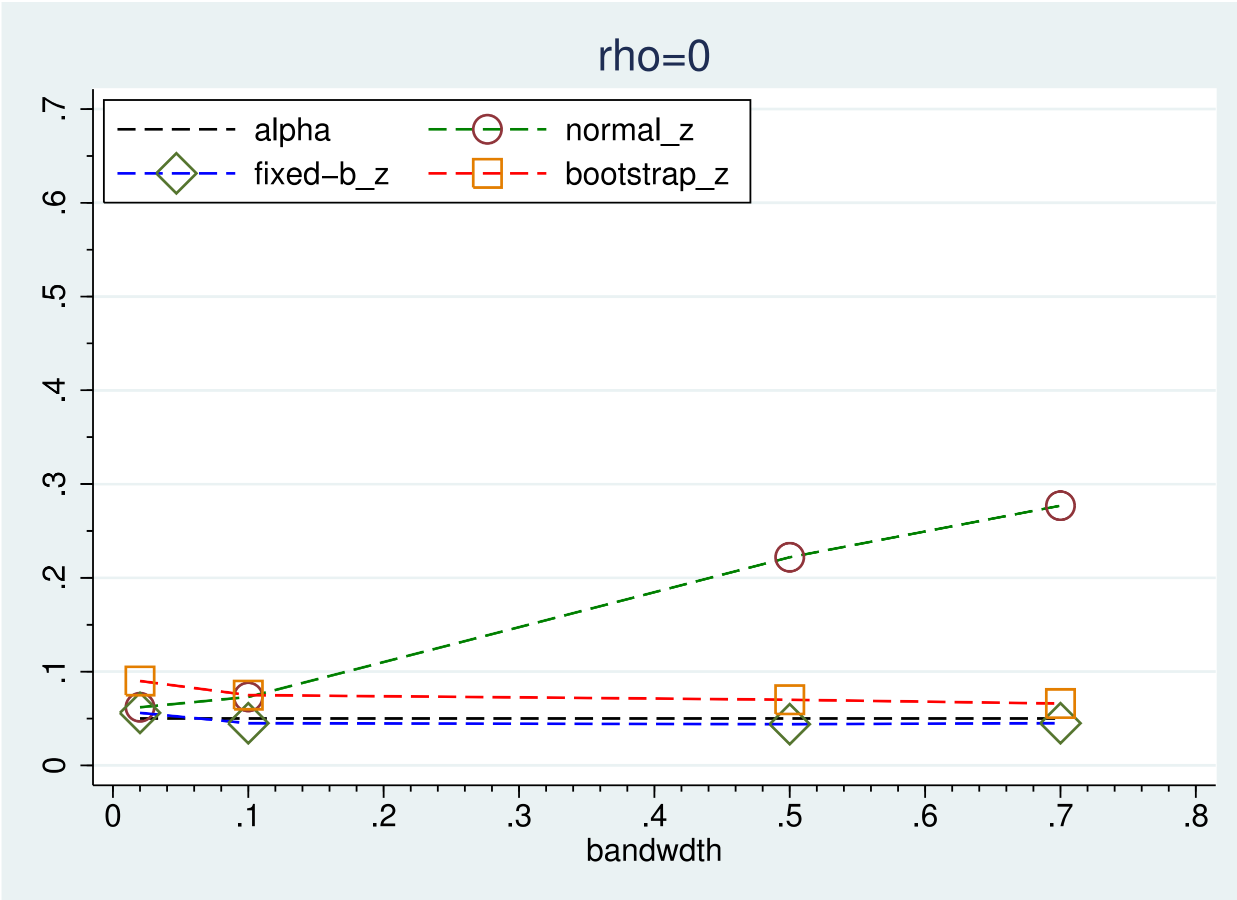


(e) N=49, T=250, l=1, z parameter

Figure E.19: (cont'd)



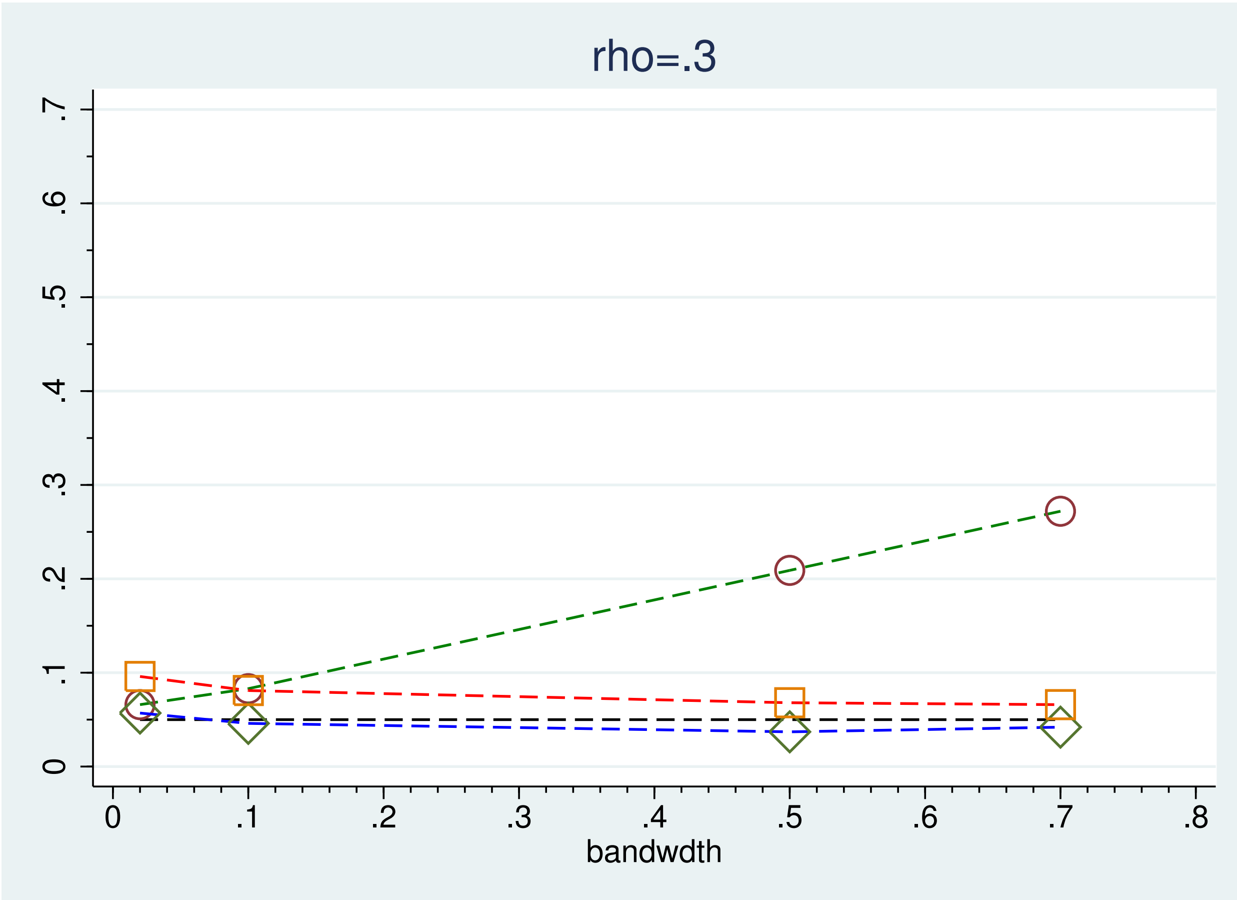
(f) N=49, T=250, l=1, z parameter



(a) N=256, T=250, l=25, z parameter

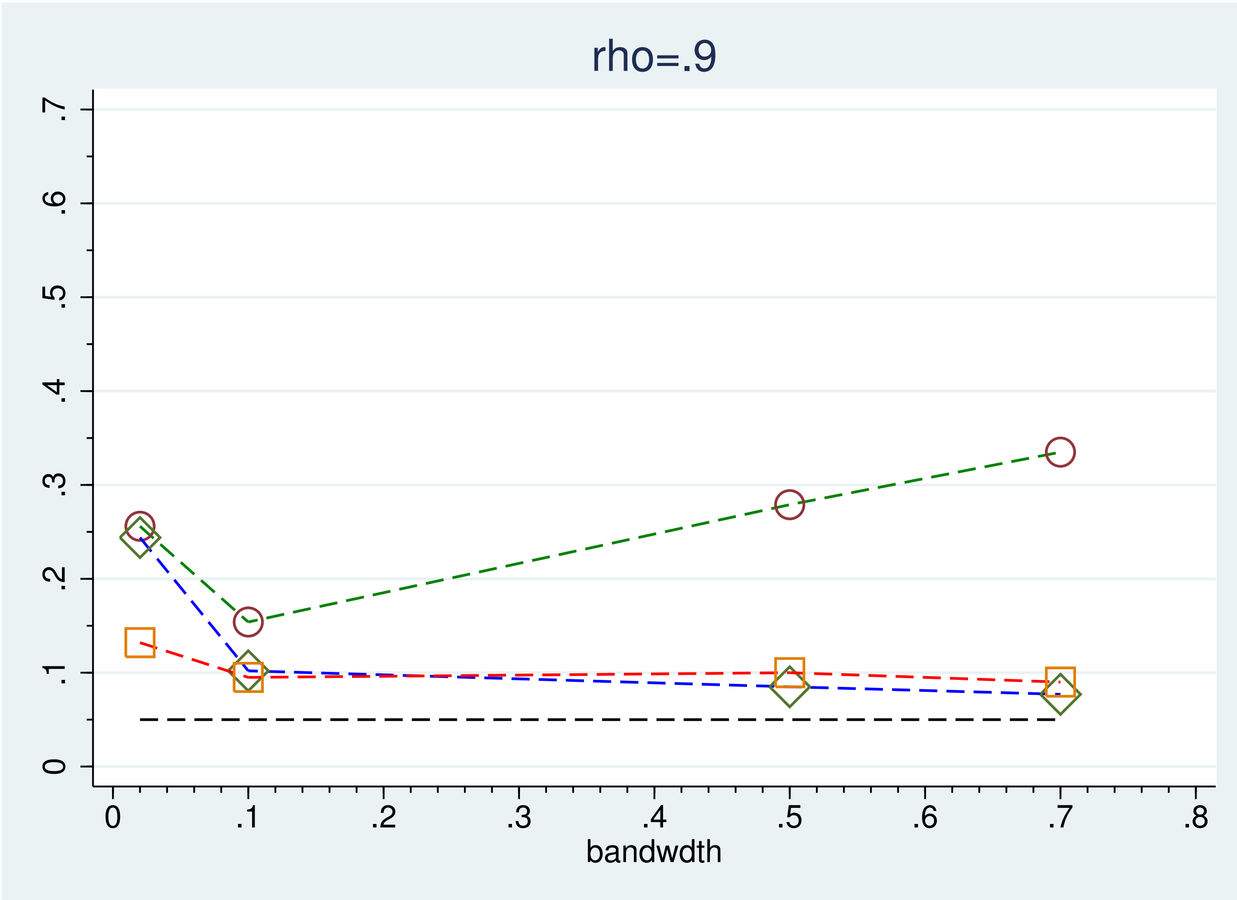
Figure E.20: Empirical null rejection probabilities for z parameter, additional regressor, spatial $MA(2)$, Bartlett kernel, $N = 256$, $T = 250$, $\lambda = 0.5$.

Figure E.20: (cont'd)



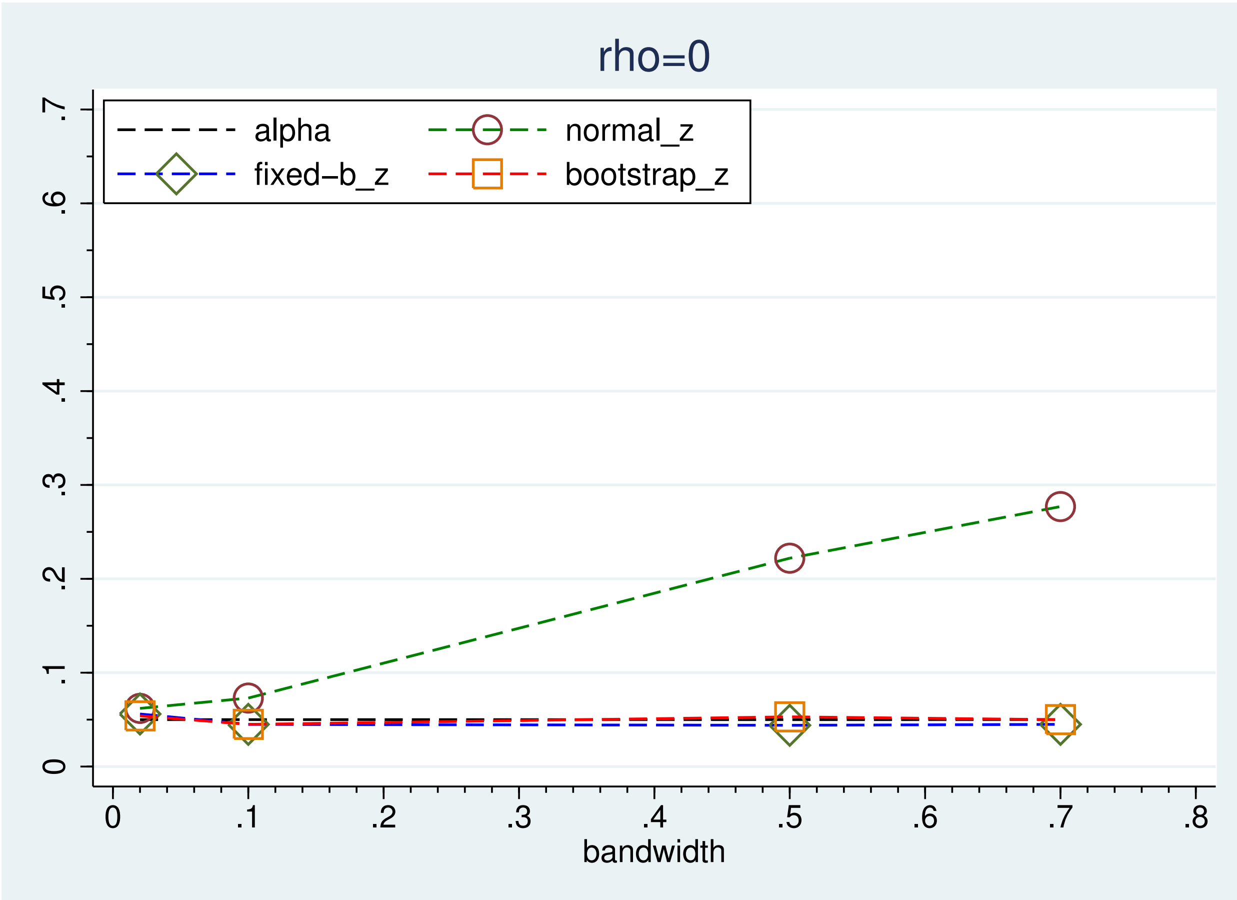
(b) N=256, T=250, l=25, z parameter

Figure E.20: (cont'd)



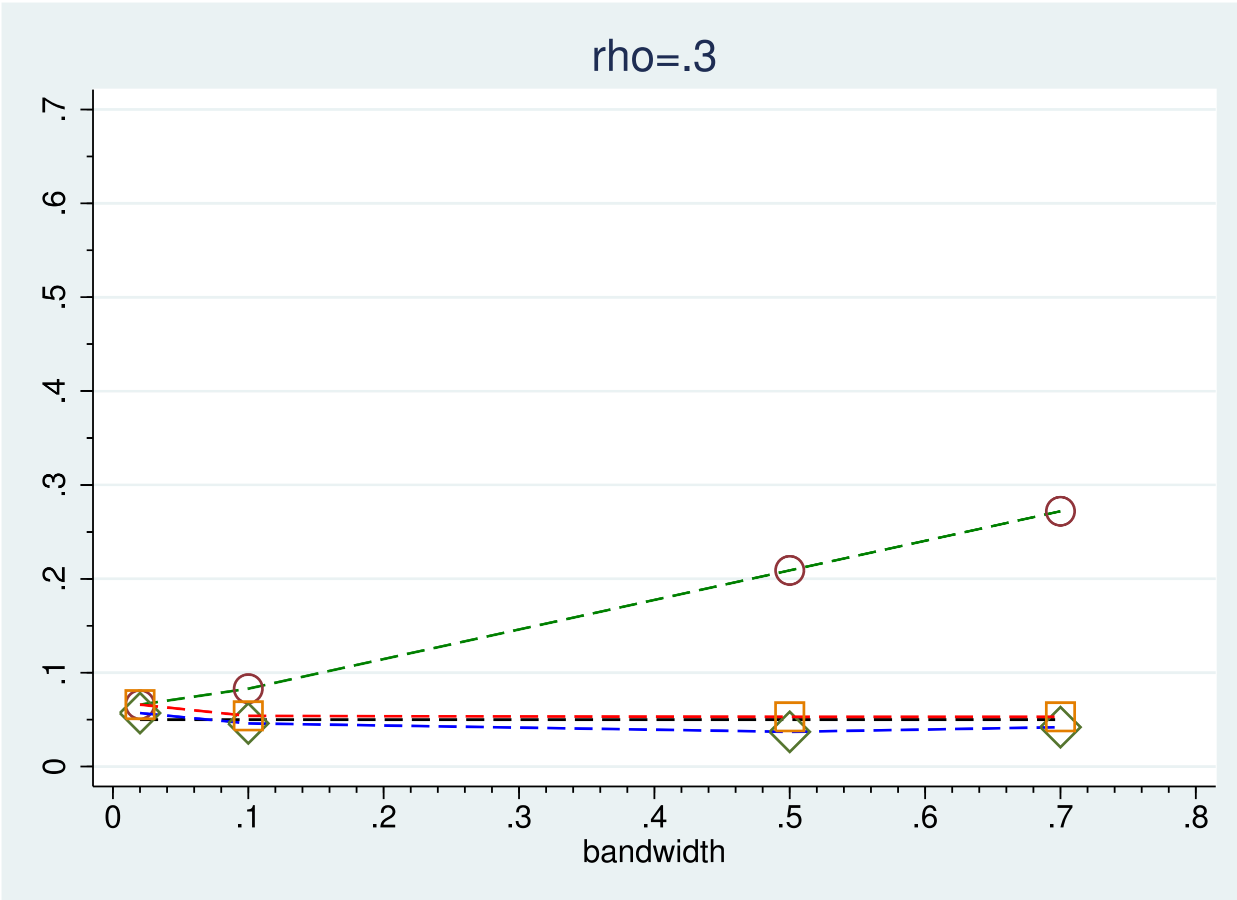
(c) N=256, T=250, l=25, z parameter

Figure E.20: (cont'd)



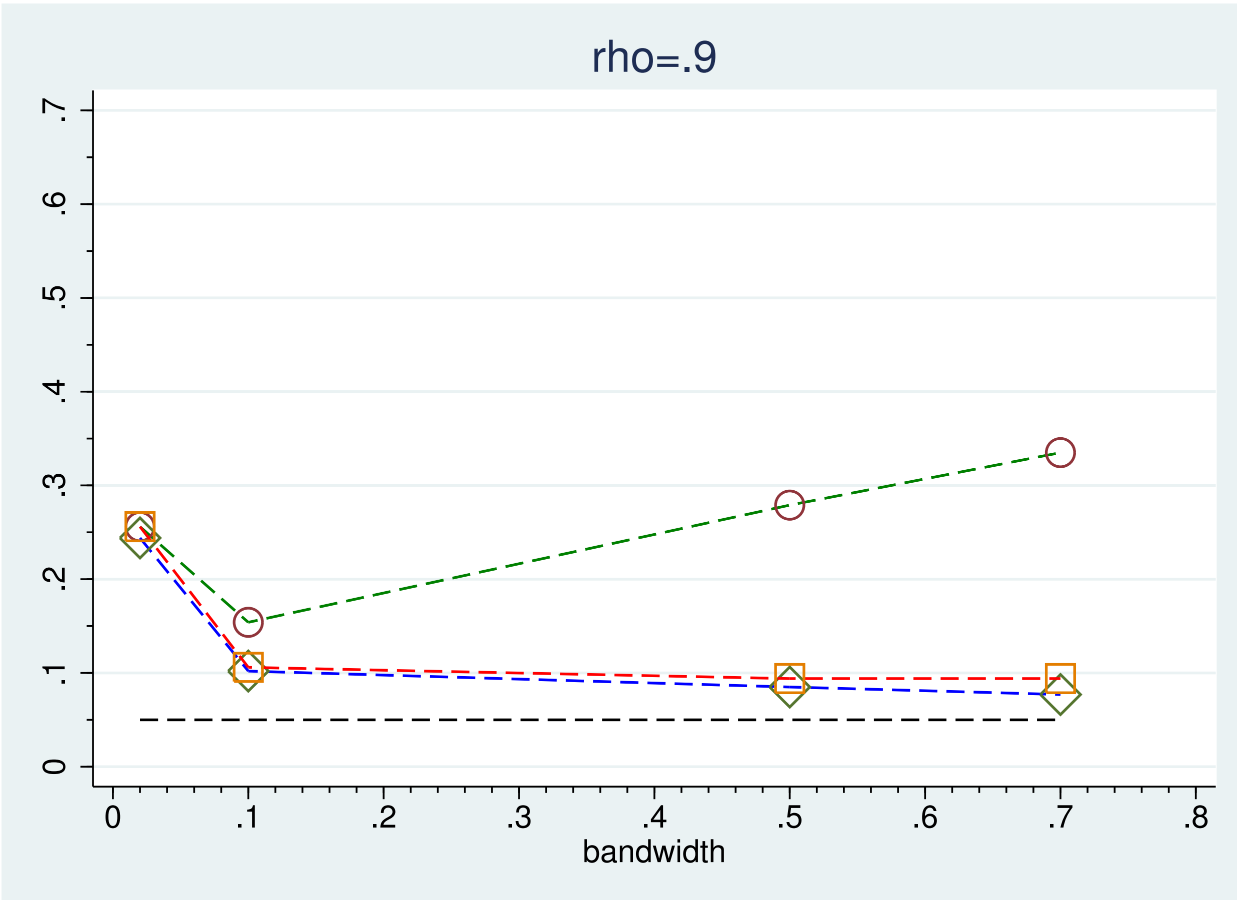
(d) N=256, T=250, l=1, z parameter

Figure E.20: (cont'd)



(e) N=256, T=250, l=1, z parameter

Figure E.20: (cont'd)



(f) N=256, T=250, l=1, z parameter

BIBLIOGRAPHY

BIBLIOGRAPHY

M. Arellano. Computing robust standard errors for within-groups estimators. *Oxford Bulletin of Economics and Statistics*, 49(4):431–434, 1987.

M. Bertrand, E. Duflo, and S. Mullainathan. How much should we trust differences-in-differences estimates? *Quarterly Journal of Economics*, 119:249–275, 2004.

A.C. Bester, T.C. Conley, C.B. Hansen, and T.J. Vogelsang. Fixed-*b* asymptotics for spatially dependent robust nonparametric covariance matrix estimators. Working Paper, Department of Economics, Michigan State University, 2008.

A.C. Bester, T.C. Conley, and C.B. Hansen. Inference with dependent data using cluster covariance estimators. *Journal of Econometrics*, 2011. doi:10.1016/j.jeconom.2011.01.007.

H. Bunzel and T. J. Vogelsang. Powerful trend function tests that are robust to strong serial correlation with an application to the prebisch-singer hypothesis. *Journal of Business and Economic Statistics*, 23:381–394, 2005.

A. Cameron, J. Gelbach, and D. Miller. Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statiscs*, 90:414–427, 2008.

A. Cameron, J. Gelbach, and D. Miller. Robust inference with multiway clustering. *Journal of Business and Economic Statistics*, 29:238–249, 2011.

C. K. Cho. Fixed b inference in a time series regression with a structural break. Working paper, Department of Economics, Michigan State University, 2012.

T. G. Conley. GMM estimation with cross sectional dependence. *Journal of Econometrics*, 92(1): 1–45, 1999.

J.C. Driscoll and A.C. Kraay. Consistent covariance matrix estimation with spatially dependent panel data. *Review of Economics and Statistics*, 80(4):549–560, 1998.

B. Efron. Bootstrap methods: Another look at the jackknife. *Annals of Statistics*, 7:1–26, 1979.

E.F. Fama and J.D. MacBeth. Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy*, 81(3):607–636, 1973.

S. Gonçalves. The moving blocks bootstrap for panel linear regression models with individual fixed-effects. *Econometric Theory*, 27:1048–1082, 2011.

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

I. Gow, G. Ormazabal, and D. Taylor. Correcting for cross-sectional and time-series dependence in accounting research. *The Accounting Review*, 85:483–512, 2010.

C.B. Hansen. Asymptotic properties of a robust variance matrix estimator for panel data when T is large. *Journal of Econometrics*, 141(2):597–620, 2007.

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

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

H.H. Kelejian and I.R. Prucha. HAC estimation in a spatial framework. *Journal of Econometrics*, 140(1):131–154, 2007.

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

M.S. Kim and Y. Sun. Spatial heteroskedasticity and autocorrelation consistent estimation of covariance matrix. *Journal of Econometrics*, 160:349–371, 2011a.

M.S. Kim and Y. Sun. Heteroskedasticity and spatiotemporal dependence robust inference for linear panel models with fixed effects. Working paper, Department of Economics, Ryerson University, 2011b.

H. R. Kunsch. The jackknife and the bootstrap for general stationary observations. *Annals of Statistics*, 17:1217–1241, 1989.

R.Y. Liu and K. Singh. Moving blocks jackknife and bootstrap capture weak dependence. In R. LePage and L. Billiard, editors, *Exploring the Limits of the Bootstrap*, New York, 1992. Wiley.

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

M. A. Petersen. Estimating standard errors in finance panel data sets: Comparing approaches. *Review of Financial Studies*, 22(1):435, 2009.

S. B. Thompson. Simple formulas for standard errors that cluster by both firm and time. *Journal of Financial Economics*, 99:1–10, 2011.

T. J. Vogelsang. Spectral analysis. In S. N. Durlauf and L. E. Blume, editors, *The New Palgrave Dictionary of Economics*. Palgrave Macmillan, 2008.

T. J. Vogelsang. Heteroskedasticity, autocorrelation, and spatial correlation robust inference in linear panel models with fixed-effects. *Journal of Econometrics*, 2012.

H. White. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48:817–38, 1980.

H. White. *Asymptotic Theory for Econometricians*. Academic Press, New York, 1984.

J. M. Wooldridge. *Analysis of Cross-sectional and Panel Data*. Cambridge, MA: MIT Press, 2002.

J. M. Wooldridge. Cluster-sample methods in applied econometrics. *American Economic Review Papers and Proceedings*, 93(2):133–138, 2003.

C. F. J. Wu. Jackknife, bootstrap and other resampling methods in regression analysis. *Annals of Statistics*, 14:1261–95, 1986.