THREE ESSAYS ON ECONOMETRICS

By

Seunghwa Rho

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 ECONOMETRICS

By

Seunghwa Rho

This dissertation consists of three chapters on econometrics. The first chapter, "*Are all firms inefficient?*", is related to the stochastic frontier model of Aigner et al. (1977). In the usual stochastic frontier model, all firms are inefficient, because inefficiency is non-negative and the probability that inefficiency is exactly zero equals zero. We modify this model by adding a parameter p which equals the probability that a firm is fully efficient. This model has also been considered by Kumbhakar et al. (2013). We extend their paper in several ways. We discuss some identification issues that arise if all firms are inefficient or no firms are inefficient. We show that the likelihood has a stationary point at parameters that indicate no inefficiency and that this point is a local maximum if the OLS residuals are positively skewed. We consider the case that a logit or probit model determines the probability of full efficiency in terms of some observable variables. Finally, we consider problems involved in testing the hypothesis that $p = 0$. We provide some simulations and an empirical example. The simulation results suggest that the proposed model appears to be useful when (i) it is reasonable to suppose that some firms are fully efficient, and (ii) the inefficiency levels of the inefficient firms are not small relative to statistical noise.

The focus of the second and third chapters lies on asymptotic theory for test statistics in time series that are robust to heteroskedasticity and autocorrelation (HAC) especially under the fixed-b asymptotic framework proposed by Kiefer and Vogelsang (2005). In the second chapter, "*Serial Correlation Robust Inference with Missing Data*", we investigate the properties of HAC robust test statistics when there is missing data. We characterize the time series with missing observations as amplitude modulated series following Parzen (1963). For estimation and inference this amounts to plugging in zeros for missing ob-

servations. We also investigate an alternative approach where the missing observations are simply ignored. There are three main theoretical findings. First, when the missing process is random and satisfies strong mixing conditions, HAC robust t and $Wald$ statistics computed from the amplitude modulated series follow the usual fixed-b limits as in Kiefer and Vogelsang (2005). Second, when the missing process is non-random, the fixed-b limits depend on the locations of missing observations but are otherwise pivotal. Third, when missing observations are ignored we obtain the surprising result that fixed-b limits of the robust t and $Wald$ statistics have the standard fixed-b limits whether the missing process is random or non-random. We discuss methods for obtaining fixed-b critical values with a focus on bootstrap methods. We find that the naive *i.i.d.* bootstrap is the most effective and practical way to obtain fixed-b critical values when data is missing especially when the bootstrap conditions on the locations of the missing data.

In the third chapter, "*Inference in time series models using smoothed clustered standard errors*", we propose a long run variance estimator for conducting inference in time series regression models that combines the traditional nonparametric kernel approach with a cluster approach. The basic idea is to divide the time periods into non-overlapping clusters and construct the long run variance estimator by first aggregating within clusters and then kernel smoothing across clusters. We derive asymptotic results holding the number of clusters fixed and also treating the clusters as increasing with the sample size. We find that the "fixed number of clusters"asymptotic approximation works well whether the number of clusters (G) is small or large. Also, we find that the naive *i.i.d.* bootstrap mimics the fixed number of clusters critical values regardless of G. Finite sample simulations suggest that clustering before kernel smoothing can reduce over-rejections caused by strong serial correlation without a great cost in terms of power.

ACKNOWLEDGEMENTS

There were so many days that I doubted if I was going the right way during the last five years. I would not have made this far without the help of many people.

First, I want to thank all of my committee members. Especially, I want to thank my main advisor, Peter Schmidt, who always surprised me with clear and concrete guidance for whatever question I had and whose detailed attention helped me stay on track. I also would like to thank my other advisor Timothy Vogelsang for his continuous support and inspiration. Hours of conversations with him were both helpful and memorable. They advised me from the very preliminary stage to the concluding level and they were patient to the very last moment. Without their support, this dissertation would have been literally impossible. I could not imagine having better advisors. I also would like to express special thanks to Christine Amsler for all the detailed and helpful comments in the completion of this dissertation.

I want to thank all of my friends at MSU, whom I miss already, including Cheolkeun Cho, Dooyeon Cho, Sukampon Chongwilaikasaem, Sunghoon Kang, Dowon Kwak, Jinyoung Lee, and Sun Yu.

Lastly, I also have been fortunate to have family and friends who somehow always believed in me in whatever crazy situations. I would not have survived the days of the Great Depression without their support. I thank my parents, my loving sister Seungmin, my dear friend Wiroy Shin and friend-like mentor Charles Becker.

I am very happy to finish my dissertation.

I made it!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

ARE ALL FIRMS INEFFICIENT?

1.1 INTRODUCTION

In the basic stochastic frontier model of Aigner et al. (1977) and Meeusen and van den Broeck (1977), all firms are inefficient to some degree. The one-sided error that represents technical inefficiency has a distribution (for example, half normal) for which zero is in the support, so that zero is a possible value, but it is still the case that the probability is zero that a draw from a half normal exactly equals zero. This may be restrictive empirically, since it is plausible, or at least possible, that an industry may contain a set of firms that are fully efficient.

In this chapter we allow the possibility that some firms are fully efficient. We introduce a parameter p which represents the probability that a firm is fully efficient. So the case of $p = 0$ corresponds to the usual stochastic frontier model and the case of $p = 1$ corresponds to the case of full efficiency (no one-sided error), while if $0 < p < 1$ a fraction p of the firms are fully efficient and a fraction $1 - p$ are inefficient. This may be important because if some of the firms actually are fully efficient, the usual stochastic frontier model is misspecified and can be expected to yield biased estimates of the technology and of firms' inefficiency levels.

This model is a special form of the latent class model considered by Caudill (2003), Orea and Kumbhakar (2004), Greene (2005) and others. It has the special feature that the frontier itself does not vary across the two classes of firms; only the existence or non-existence of inefficiency differs. Our model has previously been considered by Kumbhakar, Parmeter, and Tsionas (2013), hereafter KPT. See also Grassetti (2011). Our results were derived without knowledge of the KPT paper, but in this chapter we will naturally focus on our results which are not in their paper.

The plan of this chapter is as follows. In Section 1.2 we will present the model and give a brief summary of the basic results that are also in the KPT paper. These include the likelihood to be maximized, the form of the "posterior" probabilities of full efficiency for each firm, and the expression for the estimated inefficiencies for each firm. In Section 1.3 we provide some new results. We discuss identification issues. We give the generalization of the results of Waldman (1982), which establish that there is a stationary point of the likelihood at a point of full efficiency and that this point is a local maximum of the likelihood if the OLS residuals are positively skewed. We propose using logit or probit models to allow additional explanatory variables to affect the probability of a firm being fully efficient. We also discuss the problem of testing the hypothesis that $p = 0$. In Section 1.4 we present some simulations, and in Section 1.5 we give an empirical example. Finally, Section 1.6 gives our conclusions.

1.2 THE MODEL AND BASIC RESULTS

We begin with the standard stochastic frontier model of the form:

$$y_i = \mathbf{x}_i'\beta + \varepsilon_i,\ \varepsilon_i = v_i - u_i, \quad u_i \geq 0.$$

Here $i = 1, \ldots, n$ indexes firms. We have in mind a production frontier so that y is typically log output and $\mathbf{x}$ is a vector of functions of inputs. The v_i are iid $N\left(0, \sigma_v^2\right)$, the u_i are iid $N^+\left(0, \sigma_u^2\right)$ (i.e., half-normal), and $\mathbf{x}$, v, and u are mutually independent (so $\mathbf{x}$ can be treated as fixed). We will refer to this model as the basic stochastic frontier (or basic SF) model.

We now define some standard notation. Let ϕ be the standard normal density, and Φ

be the standard normal cdf. Let f_v and f_u represent the densities of v and u :

$$f_v(v) = \frac{1}{\sqrt{2\pi}\sigma_v}\exp(-\frac{v^2}{2\sigma_v^2}) = \frac{1}{\sigma_v}\phi\left(\frac{v}{\sigma_v}\right), \tag{1.1}$$

$$f_u(u) = \frac{2}{\sqrt{2\pi}\sigma_u}\exp(-\frac{u^2}{2\sigma_u^2}) = \frac{2}{\sigma_u}\phi\left(\frac{u}{\sigma_u}\right), \quad u \geq 0 \quad .$$

Also define $\lambda = \sigma_u/\sigma_v$ and $\sigma^2 = \sigma_u^2 + \sigma_v^2$. This implies that $\sigma_v^2 = \sigma^2/(1+\lambda^2)$ and $\sigma_u^2 = \sigma^2\lambda^2/(1+\lambda^2)$. Finally, we let f_ε represent the density of $\varepsilon = v - u$:

$$f_\varepsilon(\varepsilon) = \frac{2}{\sigma}\phi\left(\frac{\varepsilon}{\sigma}\right)\left[1 - \Phi\left(\frac{\varepsilon\lambda}{\sigma}\right)\right] \quad . \tag{1.2}$$

Now we define the model of this chapter. Suppose there is an *unobservable* variable z_i such that

$$z_i = 1\left(u_i = 0\right) = \begin{cases} 1 & \text{if } u_i = 0 \\ 0 & \text{if } u_i > 0 \quad . \end{cases}$$

Define $p = P\left(z_i = 1\right) = P\left(u_i = 0\right)$. We assume that $u_i \left| z_i = 0\right.$ is distributed as $N^+(0, \sigma_u^2)$, that is, half normal. Thus

$$u_i \quad = \quad \begin{cases} 0 & \text{with probability } \; p \\ N^+\left(0, \sigma_u^2\right) & \text{with probability } \; 1 - p \quad . \end{cases}$$

This model contains the parameters β, σ_u^2, σ_v^2, and p or β, λ, σ^2, and p.

We will follow the terminology of KPT and call this model the "zero-inefficiency stochastic frontier" (ZISF) model. The name refers to the fact that, in this model, the event $u_i = 0$ can occur with non-zero frequency. Note that

$$f\left(\varepsilon | z = 1\right) = f_v\left(\varepsilon\right),$$

$$f\left(\varepsilon | z = 0\right) = f_\varepsilon\left(\varepsilon\right),$$

(where f_v and f_ε are defined in (1.1) and (1.2) above) and so the marginal (unconditional) density of ε is

$$f_p\left(\varepsilon\right) = p f_v\left(\varepsilon\right) + \left(1 - p\right) f_\varepsilon\left(\varepsilon\right) \quad .$$

Using this density, we can form the (log) likelihood for the model:

$$\ln L\left(\beta,\sigma_u^2,\sigma_v^2,p\right) = \sum_{i=1}^{n} \ln fp\left(y_i-\mathbf{x}_i'\beta\right) \quad .$$

We will estimate the model by MLE; that is, by maximizing $\ln L$ with respect to β, σ_u^2, σ_v^2, and p. Or, alternatively, the model may be parameterized in terms of β, λ, σ^2, and p, with maximization over that set of parameters.

When we have estimated the model, we can obtain $\hat{\varepsilon}_i = y_i - \mathbf{x}_i'\hat{\beta}$, an estimate of $\varepsilon_i = y_i - \mathbf{x}_i'\beta$. Using Bayes rule, we can now update the probability that a particular firm is fully efficient, because ε_i is informative about that possibility. That is, we can calculate

$$P\left(z_i=1|\varepsilon_i\right) = \frac{P\left(z_i=1\right) f\left(\varepsilon_i|z_i=1\right)}{fp\left(\varepsilon_i\right)} = \frac{pfv\left(\varepsilon_i\right)}{fp\left(\varepsilon_i\right)} = \frac{pfv\left(\varepsilon_i\right)}{pfv\left(\varepsilon_i\right)+(1-p)f\varepsilon\left(\varepsilon_i\right)} \tag{1.3}$$

We will call this the "posterior" probability that firm i is fully efficient. It is evaluated at $\hat{p}$, $\hat{\varepsilon}_i$ and also $\hat{\sigma}_u^2$ and $\hat{\sigma}_v^2$, which enter into the densities of fv and $f\varepsilon$. We put quotes around "posterior" because it is not truly the posterior probability of $z_i = 1$ in a Bayesian sense. (A true Bayesian posterior would give $P\left(z_i=1|y_i,x_i\right)$ and would have started with a prior distribution for the parameters β, σ_u^2, σ_v^2, and p.)

We now wish to estimate (predict) u_i for each firm. Following the logic of Jondrow et al. (1982), we define $\hat{u}_i = E\left(u_i|\varepsilon_i\right)$. Now

$$\begin{aligned} E\left(u_i|\varepsilon_i\right) &= E_{z|\varepsilon}E\left(u_i|\varepsilon_i,z_i\right) \\ &= P\left(z_i=1|\varepsilon_i\right)E\left(u_i|\varepsilon_i,z_i=1\right)+P\left(z_i=0|\varepsilon_i\right)E\left(u_i|\varepsilon_i,z_i=0\right) \\ &= P\left(z_i=0|\varepsilon_i\right)E\left(u_i|\varepsilon_i,z_i=0\right) \end{aligned}$$

since $u_i \equiv 0$ when $z_i = 1$. But $E\left(u_i|\varepsilon_i,z_i=0\right)$ is the usual expression from Jondrow et al. (1982), and $P\left(z_i=0|\varepsilon_i\right) = 1 - P\left(z_i=1|\varepsilon_i\right)$ which can be evaluated using equation (1.3) above. Therefore,

$$\hat{u}_i = E\left(u_i|\varepsilon_i\right) = \frac{(1-p)\,f\varepsilon\left(\varepsilon_i\right)}{pfv\left(\varepsilon_i\right)+(1-p)\,f\varepsilon\left(\varepsilon_i\right)} \times \sigma_* \left[\frac{\phi\left(a_i\right)}{1-\Phi\left(a_i\right)} - a_i\right], \tag{1.4}$$

where $a_i = \varepsilon_i \lambda / \sigma$ and $\sigma_* = \sigma_u \sigma_v / \sigma = \lambda\sigma / (1+\lambda^2)$.

A slight extension of this result, which is not in KPT, is to follow Battese and Coelli (1988) and define technical efficiency as $TE = \exp(-u)$. Correspondingly technical inefficiency would be $1 - TE = 1 - \exp(-u)$, which is only approximately equal to u (for small u). They provide the expression for $E(TE|\varepsilon)$. Using our "posterior" probability $P(z_i = 1|\varepsilon_i)$ and their expression for $E(TE_i|\varepsilon_i, z_i = 0)$, we obtain

$$\begin{aligned}
\widehat{TE}_i &= E\left(e^{-u_i}|\varepsilon_i\right) \\
&= \frac{(1-p) f_\varepsilon(\varepsilon_i)}{p f_v(\varepsilon_i) + (1-p) f_\varepsilon(\varepsilon_i)} \times \frac{\Phi\left(\frac{\mu_i^*}{\sigma_*} - \sigma_*\right)}{\Phi\left(\frac{\mu_i^*}{\sigma_*}\right)} \exp\left(\frac{\sigma_*^2}{2} - \mu_i^*\right) \\
&\quad + \frac{p f_v(\varepsilon_i)}{p f_v(\varepsilon_i) + (1-p) f_\varepsilon(\varepsilon_i)}, \qquad (1.5)
\end{aligned}$$

where $\mu_i^* = -\varepsilon_i \sigma_u^2 / \sigma^2$, $\sigma_* = \sigma_u \sigma_v / \sigma$ (as above), and correspondingly $\mu_i^* / \sigma_* = -a_i$ where $a_i = \varepsilon_i \lambda / \sigma$ (as above).

As in Jondrow et al. (1982), the expression in either (1.4) or (1.5) would need to be evaluated at the estimated values of the parameters (p, σ_u^2, and σ_v^2) and at $\hat{\varepsilon}_i = y_i - \mathbf{x}_i' \hat{\beta}$.

1.3 EXTENSIONS OF THE BASIC MODEL

We now investigate some extensions of the basic results of the previous section. Most of the results in this section are not in KPT.

1.3.1 Identification Issues

Some of the parameters are not identified under certain circumstances. When $p = 1$, so that all firms are fully efficient, σ_u^2 is not identified. Conversely, when $\sigma_u^2 = 0$, p is not identified. In fact, the likelihood value is exactly the same when (i) $\sigma_u^2 = 0$, $p =$ anything as when (ii) $p = 1$, $\sigma_u^2 =$ anything. More generally, we might suppose that σ_u^2 and p will be estimated imprecisely when a data set contains little inefficiency, since it will be hard to determine whether there is little inefficiency because σ_u^2 is small or because p is close to one.

This issue of identification is relevant to the problem of testing the null hypothesis that $p = 1$ against the alternative that $p < 1$. This is a test of the null hypothesis that all firms are efficient against the alternative that some fraction (possibly all) of them are inefficient, and that is an economically interesting hypothesis. KPT suggest a likelihood ratio test of this hypothesis. As they note, the null distribution of their statistic is affected by the fact that the null hypothesis is on the boundary of the parameter space. They refer to Chen and Liang (2010), p. 608 to justify an asymptotic distribution of $1/2\chi_0^2 + 1/2\chi_1^2$ for the likelihood ratio statistic. However, it is not clear that this result applies, given that one of the parameters (σ_u^2) is not identified under the null that $p = 1$. Specifically, the argument of Chen and Liang (2010) depends on the existence and asymptotic normality of the estimator $\hat{\eta}(\gamma_0)$ [see p. 606, line 4] where γ_0 corresponds to $p_0 (= 1)$, and where η corresponds to the other parameters of our model, including σ_u^2.

A more relevant reference, which KPT note but do not pursue, is Andrews (2001). This chapter explicitly allows the case in which the parameter vector under the null may lie on the boundary of the maintained hypothesis *and* there may be a nuisance parameter that appears under the alternative hypothesis, but not under the null. See his Theorem 4, p. 707, for the relevant asymptotic distribution result, which unfortunately is considerably more complicated than the simple result (50-50 mixture of chi-squareds) of Chen and Liang (2010).

1.3.2 A Stationary Point for the Likelihood

For the basic stochastic frontier model, let the parameter vector be $\theta = (\beta', \lambda, \sigma^2)'$. Then Waldman (1982) established the following results. First, the log likelihood always has a stationary point at $\theta^* = (\hat{\beta}', 0, \hat{\sigma}^2)'$, where $\hat{\beta}$ =OLS and $\hat{\sigma}^2$ =(OLS sum of squared residuals)$/n$. Note that these parameter values correspond to $\hat{\sigma}_u^2 = 0$, that is, to full efficiency of each firm. Second, the Hessian matrix is singular at this point. It is negative semi-definite with one zero eigenvalue. Third, these parameter values are a local maximizer of the log likelihood if the OLS residuals are positively skewed. This is the so-called "wrong skew problem".

The log likelihood for the ZISF model has a stationary point very similar to that for the basic stochastic frontier model. This stationary point is also a local maximum of the log likelihood if the least squares residuals are positively skewed.

Theorem 1.1. *Let $\theta = (\beta', \lambda, \sigma^2, p)'$ and let $\theta^{**} = (\hat{\beta}', 0, \hat{\sigma}^2, \hat{p})'$, where $\hat{\beta}$ =OLS, $\hat{\sigma}^2$ =(OLS sum of squared residuals)$/n$, and where $\hat{p}$ is* any *value in $[0, 1]$. Then*

1. *θ^{**} is a stationary point of the log likelihood.*
2. *The Hessian matrix is singular at this point. It is negative semi-definite with two zero eigenvalues.*
3. *θ^{**} with $\hat{p} \in [0, 1)$ is a local maximizer of the log likelihood function if and only if $\sum_{i=1}^{n} \hat{\varepsilon}_i^3 > 0$, where $\hat{\varepsilon}_i = y_i - \mathbf{x}_i' \hat{\beta}$ is the OLS residual.*
4. *θ^{**} with $\hat{p} = 1$ is a local maximizer of the log likelihood function if $\sum_{i=1}^{n} \hat{\varepsilon}_i^3 > 0$.*

Proof. See Appendix A. □

As is typically done for the basic stochastic frontier model, we will presume that θ^{**} is the global maximizer of the log likelihood when the residuals have positive ("wrong") skew. Note that at θ^{**}, we have $\hat{\lambda} = 0$ or equivalently $\hat{\sigma}_u^2 = 0$, and p is not identified when $\sigma_u^2 = 0$. We get the same likelihood value for any value of p. In our simulations (in Section 1.4) we will set $\hat{p} = 1$ in the case of wrong skew, since $\hat{p} = 1$ is another way of reflecting full efficiency. However, for a given data set, the value of $\hat{p}$ does not matter when $\theta = \theta^{**}$.

Since for any $p \in [0, 1]$,

$$\operatorname*{plim}_{n\to\infty} \frac{1}{n}\sum \hat{\varepsilon}_i^3 = E\left(\varepsilon_i - E\left(\varepsilon_i\right)\right)^3$$
$$= \sigma_u^3\sqrt{2/\pi}(1-p)\left(-4p^2 + (8-3\pi)p + \pi - 4\right)/\pi \leq 0$$

as the number of observations increases, the probability of a positive third moment of the OLS residuals goes to zero asymptotically. In a finite sample, the probability of a positive third moment increases when λ is small and/or p is near 0 or 1. See Table 1.1. The entries in Table 1.1 are based on simulations with 100,000 replications, with $\sigma_u = 1$, $\lambda = \sigma_u/\sigma_v$, $\lambda \in \{0.5, 1, 2\}$, and $p \in \{0, 0.1, \ldots, 0.9\}$, for sample sizes 50, 100, 200, and 400.

1.3.3 Models for the Distribution of u_i

The ZISF model can be extended by allowing the distribution of u_i to depend on some observable variables w_i. For example, in our empirical analysis of Section 1.5, the w_i will include variables like the age and education of the farmer and the size of his household. These variables can be assumed to affect either $P(z_i = 1)$ or $f(u_i|z_i = 0)$ or both.

First consider the case in which we assume that w_i affects the distribution of u_i for the inefficient firms. A general assumption would be that the distribution of u_i conditional on w_i and on $z_i = 0$ is $N^+(\mu_i, \sigma_i^2)$ where μ_i and/or σ_i^2 depend on w_i. For example, in Section 1.5 we will assume the RSCFG model of Reifschneider and Stevenson (1991), Caudill and Ford (1993) and Caudill et al. (1995), under the specific assumptions that $\mu_i = 0$ and $\sigma_i^2 = \exp(w_i'\gamma)$. Another possible model is the KGMHLBC model of Kumbhakar et al. (1991), Huang and Liu (1994) and Battese and Coelli (1995), with $\sigma_i^2 = \sigma_u^2$ constant and with $\mu_i = w_i'\psi$ or $\mu_i = c\exp(w_i'\psi)$. Wang (2002) proposes parameterizing both μ_i and σ_i^2. See also Alvarez et al. (2006).

A second and more novel case is the one in which we assume that w_i affects $P(z_i = 1)$.

For example, we could assume a logit model:

$$P\left(z_i = 1 | w_i\right) = \frac{\exp\left(w_i' \delta\right)}{1 + \exp\left(w_i' \delta\right)} \quad . \tag{1.6}$$

A probit model would be another obvious possibility.

Finally, we can consider a more general model in which both $P(z_i = 1|w_i)$ and $f(u_i|z_i = 0, w_i)$ depend on w_i, as above. We will estimate such a model in our empirical section.

1.3.4 Testing the Hypothesis That $p = 0$

In this section, we discuss the problem of testing the null hypothesis $H_0 : p = 0$ against the alternative $H_A : p > 0$. The null hypothesis is that all firms are inefficient, so the basic stochastic frontier model applies. The alternative is that some firms are fully efficient and so the ZISF model is needed.

It is a standard result that, under certain regularity conditions, notably that the parameter value specified by the null hypothesis is an interior point of the parameter space, the likelihood (LR), Lagrange multiplier (LM), and Wald tests all have the same asymptotic χ^2 distribution. However, in our case p cannot be negative, and therefore the null hypothesis that $p = 0$ lies on the boundary of the parameter space. This is therefore a non-standard problem. Unlike the case of testing the hypothesis that $p = 1$, however, there is no problem with the identification of the other parameters (nuisance parameters) β, σ_u^2, and σ_v^2, or β, λ, and σ^2. We need to restrict $\sigma_u^2 > 0$ and $\sigma_v^2 > 0$ so that the nuisance parameters are in the interior of the parameter space, and also because p would not be identified if $\sigma_u^2 = 0$. However, with these modest restrictions, this is only a mildly non-standard problem, which has been discussed by Rogers (1986), Self and Liang (1987), and Gouriéroux and Monfort (1995) chapter 21, for example.

We consider five test statistics: the likelihood ratio (LR), Wald, Lagrange multiplier (LM), modified Lagrange multiplier (modified LM), and Kuhn-Tucker (KT) tests. All of these except the LM test will have asymptotic distributions that are different from the usual (χ_1^2) distribution.

We will assume that the likelihood function $L_n(\theta)$ satisfies the usual conditions,

$$\begin{aligned} \frac{1}{\sqrt{n}}\frac{\partial L_n(\theta_0)}{\partial\theta} &\xrightarrow{d} \mathcal{N}\left(0,\mathcal{I}_0\right), \\ \frac{1}{n}\frac{\partial^2 L_n(\theta_0)}{\partial\theta\partial\theta} &\xrightarrow{p} \mathcal{H}_0 = -\mathcal{I}_0, \end{aligned}$$

where $\theta = (\beta^{'}, \sigma_u, \sigma_v, p)'$, and the parameters other than p are away from the boundary of their parameter spaces. Define the restricted estimator ($\tilde{\theta}$) and the unrestricted estimator ($\hat{\theta}$):

$$\begin{aligned} \tilde{\theta} &= \underset{\sigma_u\geq 0,\sigma_v\geq 0,\, p=0}{\operatorname{argmax}} \ln L_n(\theta), \\ \hat{\theta} &= \underset{\sigma_u\geq 0,\sigma_v\geq 0,\, 0\leq p\leq 1}{\operatorname{argmax}} \ln L_n(\theta) . \end{aligned}$$

We also define $l_i = \ln f(\varepsilon_i)$, $\hat{s}_i = \partial l_i(\hat{\theta})/\partial\theta$, $\tilde{s}_i = \partial l_i(\tilde{\theta})/\partial\theta$, $\hat{h}_i = \partial^2 l_i(\hat{\theta})/\partial\theta\partial\theta'$, and $\tilde{h}_i = \partial^2 l_i(\tilde{\theta})/\partial\theta\partial\theta'$. Finally, we consider the "unconstrained" estimator ($\check{\theta}$) that ignores the logical restriction $0 \leq p \leq 1$:

$$\check{\theta} = \underset{\sigma_u\geq 0,\sigma_v\geq 0}{\operatorname{argmax}} \ln L_n(\theta) .$$

1.3.4.1 LR test

The LR statistic when testing $H_0 : p = 0$ is $\xi^{LR} = 2(\ln L_n(\hat{\theta}) - \ln L_n(\tilde{\theta}))$. Under standard regularity conditions, the asymptotic distribution of ξ^{LR} is a mixture of χ_0^2 and χ_1^2, with mixing weights $1/2$, where χ_0^2 is defined as the point mass distribution at zero. That is $\xi^{LR} \xrightarrow{d} 1/2\chi_0^2 + 1/2\chi_1^2$. This follows, for example, from Chen and Liang (2010), as cited by KPT.

1.3.4.2 Wald test

The Wald statistic for $H_0 : p = 0$ is

$$\xi^W = \frac{\hat{p}^2}{se\left(\hat{p}\right)^2}.$$

Note that $se(\hat{p})^2$ can be computed using the outer product of the score form of the variance matrix of $\hat{\theta}$, $[(\sum_{i=1}^n \hat{s}_i\hat{s}_i')^{-1}]$, the Hessian form, $[(\sum_{i=1}^n -\hat{h}_i)^{-1}]$, or the Robust form, $[(\sum_{i=1}^n -\hat{h}_i)^{-1}(\sum_{i=1}^n \hat{s}_i\hat{s}_i')(\sum_{i=1}^n -\hat{h}_i)^{-1}]$. As with the LR statistic, $\xi^W \xrightarrow{d} 1/2\chi_0^2 + 1/2\chi_1^2$. Note that the non-standard nature of this result means that the "significance" of an estimated $\hat{p}$ from the ZISF model cannot be assessed using standard results.

1.3.4.3 LM test

The LM statistic for $H_0 : p = 0$ is

$$\xi^{LM} = \left(\sum_{i=1}^n \tilde{s}_i\right)' \tilde{M}^{-1} \left(\sum_{i=1}^n \tilde{s}_i\right).$$

$\tilde{M}$ can be either $[(\sum_{i=1}^n \tilde{s}_i\tilde{s}_i')]$ or $[(\sum_{i=1}^n -\tilde{h}_i)]$, in either case evaluated at $\tilde{\theta}$. Unlike the other statistics considered here, the LM statistic has the usual χ_1^2 distribution. It ignores the one-sided nature of the alternative, because it rejects for a large (in absolute value) positive or *negative* value of $\tilde{s}_i$. As pointed out by Rogers (1986), this may result in a loss in power relative to tests that take the one-sided nature of the alternative into account.

1.3.4.4 Modified LM test

The LM statistic has the usual χ_1^2 distribution because it does not take account of the one-sided nature of the alternative. By taking account of the one-sided nature of the alternative, the LM test might have better power. The Modified LM statistic proposed by

Rogers (1986) is motivated by this point. The modified LM statistic is :

$$\xi^{\text{modified LM}} = \begin{cases} \xi^{\text{LM}}, & \text{if } \sum_{i=1}^{n} \tilde{s}_i > 0 \\ 0 \quad , & \text{otherwise} \end{cases} .$$

In the modified LM statistic, a positive score is taken as evidence against the null and in favor of the alternative $p > 0$, whereas a negative score is not. So a negative score is simply set to zero. The asymptotic distribution of $\xi^{\text{modified LM}}$ is $1/2\chi_0^2 + 1/2\chi_1^2$.

1.3.4.5 KT test

Another form of score test statistic that takes account of the one-sided nature of the alternative is the KT statistic proposed by Gouriéroux et al. (1982). The KT statistic for $H_0 : p = 0$ is

$$\xi^{KT} = \left(\sum_{i=1}^{n} \tilde{s}_i - \sum_{i=1}^{n} \hat{s}_i \right)' \tilde{M}^{-1} \left(\sum_{i=1}^{n} \tilde{s}_i - \sum_{i=1}^{n} \hat{s}_i \right),$$

where $\tilde{M}$ can be either $\sum_{i=1}^{n} \tilde{s}_i \tilde{s}_i'$ or $\sum_{i=1}^{n} -\tilde{h}_i$. When $\hat{p} = 0$, $\sum_{i=1}^{n} \tilde{s}_i = \sum_{i=1}^{n} \hat{s}_i$. Since $\hat{p} = 0$ when $\check{p} \leq 0$, the test statistic will have a degenerate distribution at zero when $\check{p} \leq 0$. Otherwise, $\sum_{i=1}^{n} \hat{s}_i = 0$ and the test statistic has the usual χ_1^2 distribution. Therefore, $\xi^{KT} \xrightarrow{d} 1/2\chi_0^2 + 1/2\chi_1^2$.

1.3.4.6 The Wrong Skew Problem, Revisited

When the OLS residuals are positively skewed ($\sum_{i=1}^{n} \hat{\varepsilon}_i^3 > 0$), we have $\hat{\sigma}_u^2 = 0$ (or equivalently, $\hat{\lambda} = 0$) and $\hat{p}$ is not well defined. Also the information matrix, whether evaluated at $\hat{\theta}$ or $\tilde{\theta}$, is singular. Specifically,

$$\sum_{i=1}^{n} \hat{s}_i \hat{s}_i' =$$

$$\begin{pmatrix} \frac{1}{\hat{\sigma}_v^2} \sum_{i=1}^{n} \hat{\varepsilon}_i^2 x_i x_i' & -(1-\hat{p})\sqrt{\frac{2}{\pi}} \frac{1}{\hat{\sigma}_v^2} \sum_{i=1}^{n} \hat{\varepsilon}_i^2 x_i & \frac{1}{\hat{\sigma}_v^5} \sum_{i=1}^{n} \hat{\varepsilon}_i^3 x_i & 0 \\ -(1-\hat{p})\sqrt{\frac{2}{\pi}} \frac{1}{\hat{\sigma}_v^2} \sum_{i=1}^{n} \hat{\varepsilon}_i^2 x_i' & \frac{2}{\pi}(1-\hat{p})^2 \frac{n}{\hat{\sigma}_v^2} & -(1-\hat{p})\sqrt{\frac{2}{\pi}} \frac{1}{\hat{\sigma}_v^5} \sum_{i=1}^{n} \hat{\varepsilon}_i^3 & 0 \\ \frac{1}{\hat{\sigma}_v^5} \sum_{i=1}^{n} \hat{\varepsilon}_i^3 x_i' & -(1-\hat{p})\sqrt{\frac{2}{\pi}} \frac{1}{\hat{\sigma}_v^5} \sum_{i=1}^{n} \hat{\varepsilon}_i^3 & \frac{1}{\hat{\sigma}_v^6} \sum_{i=1}^{n} \hat{\varepsilon}_i^4 - \frac{n}{\hat{\sigma}_v^2} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\sum_{i=1}^{n} \tilde{s}_i \tilde{s}_i' = \begin{pmatrix} \frac{1}{\tilde{\sigma}_v^2} \sum_{i=1}^{n} \tilde{\varepsilon}_i^2 x_i x_i' & -\sqrt{\frac{2}{\pi}} \frac{1}{\tilde{\sigma}_v^2} \sum_{i=1}^{n} \tilde{\varepsilon}_i^2 x_i & \frac{1}{\tilde{\sigma}_v^5} \sum_{i=1}^{n} \tilde{\varepsilon}_i^3 x_i & 0 \\ -\sqrt{\frac{2}{\pi}} \frac{1}{\tilde{\sigma}_v^2} \sum_{i=1}^{n} \tilde{\varepsilon}_i^2 x_i' & \frac{2}{\pi} \frac{n}{\tilde{\sigma}_v^2} & -\sqrt{\frac{2}{\pi}} \frac{1}{\tilde{\sigma}_v^5} \sum_{i=1}^{n} \tilde{\varepsilon}_i^3 & 0 \\ \frac{1}{\tilde{\sigma}_v^5} \sum_{i=1}^{n} \tilde{\varepsilon}_i^3 x_i' & -\sqrt{\frac{2}{\pi}} \frac{1}{\tilde{\sigma}_v^5} \sum_{i=1}^{n} \tilde{\varepsilon}_i^3 & \frac{1}{\tilde{\sigma}_v^6} \sum_{i=1}^{n} \tilde{\varepsilon}_i^4 - \frac{n}{\tilde{\sigma}_v^2} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\sum_{i=1}^{n} \hat{h}_i = \begin{pmatrix} -\sum_{i=1}^{n} \frac{x_i x_i'}{\hat{\sigma}_v^2} & (1-\hat{p})\sqrt{\frac{2}{\pi}} \frac{1}{\hat{\sigma}_v^2} \sum_{i=1}^{n} x_i & 0 & 0 \\ (1-\hat{p})\sqrt{\frac{2}{\pi}} \frac{1}{\hat{\sigma}_v^2} \sum_{i=1}^{n} x_i' & -(1-\hat{p})^2 \frac{2}{\pi} \frac{n}{\hat{\sigma}_v^2} & 0 & 0 \\ 0 & 0 & \frac{-2n}{\hat{\sigma}_v^2} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\sum_{i=1}^{n} \tilde{h}_i = \begin{pmatrix} -\sum_{i=1}^{n} \frac{x_i x_i'}{\tilde{\sigma}_v^2} & \sqrt{\frac{2}{\pi}} \frac{1}{\tilde{\sigma}_v^2} \sum_{i=1}^{n} x_i & 0 & 0 \\ \sqrt{\frac{2}{\pi}} \frac{1}{\tilde{\sigma}_v^2} \sum_{i=1}^{n} x_i' & -\frac{2}{\pi} \frac{n}{\tilde{\sigma}_v^2} & 0 & 0 \\ 0 & 0 & \frac{-2n}{\tilde{\sigma}_v^2} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

All the matrices above are singular for any $\hat{p} \in [0,1]$. Therefore, when the third moment of the OLS residuals is positive, only the LR statistic can be defined, and equals zero. It remains to decide if we should reject the null hypothesis or not when the OLS residuals have wrong skew. Clearly, the LR test will not reject the null hypothesis, since the statistic equals zero under wrong skew. But for the other tests, the statistic is undefined and it is not clear what to conclude. If we consider the wrong skew cases as indicating that all firms are efficient, then it would be reasonable to reject the null hypothesis. However, as a practical matter, whether we reject the null hypothesis or not does not affect anything, because the estimated model whether $p = 0$ or not collapses to the same model. It might be reasonable to simply say that p is not identified with incorrectly skewed OLS residuals. For a given data set, both the null and the alternative hypothesis would lead to same results.

Assuming that $\sigma_u^2 > 0$, the wrong skew problem occurs with a probability that goes to zero asymptotically. However, as shown in Table 1.1, it can occur with non-trivial probability in finite samples. Also, the discussion above may be relevant even when the data do not have the wrong skew problem. The log likelihood has a stationary point at θ^{**} regardless of the skew of the residuals. In the wrong skew case, the likelihood is perfectly flat in the p direction with $\hat{\beta}$ =OLS, $\hat{\lambda} = 0$, and $\hat{\sigma}^2 = 1/nSSE$. In the correct skew case, this is not true, but when λ is small, we expect that the partial of log likelihood with respect to p (evaluated at the MLE of the other parameters) would often be small in the vicinity of $p = 0$, so that the LM test and its variants might have low power. We will investigate this issue in the simulations of the next section.

1.4 SIMULATIONS

We conducted simulations in order to investigate the finite sample performance of the ZISF model, and to compare it to the performance of the basic stochastic frontier model. We are interested both in parameter estimation and in the performance of tests of the hypothesis $p = 0$.

We consider a very simple data generating process: $y_i = \beta + \varepsilon_i$, where as in Section 1.2 above, $\varepsilon_i = v_i - u_i$ and u_i is half-normal with probability $1 - p$ and $u_i = 0$ with probability p. We pick $n = 200$ and 500, $\beta = 1$, and $\sigma_u = 1$. We consider $p = 0$, 0.25, 0.5, and 0.75, and $\lambda = 1$, 2, and 5 (i.e., $\sigma_v = 1$, 0.5, and 0.2). Our simulations are based on 1000 replications. Because the MLE's were sensitive to the starting values used, we used several sets of starting values and chose the results with the highest maximized likelihood value.

Our experimental design was similar to that in KPT. They included a non-constant regressor, but in our experiments that made little difference. A more substantial difference is that we used $n = 200$ and 500 whereas they used $n = 500$ and 1000.

There were some technical problems related to the facts that σ_u^2 is not identified when $p = 1$, and p is not identified when $\sigma_u^2 = 0$. We define $\hat{p} = 1$ when $\hat{\sigma}_u = 0$ and $\hat{\sigma}_u = 0$ when $\hat{p} = 1$. This would imply that when the OLS residuals have incorrect skew, the MLE would be θ^{**} with $\hat{p} = 1$. It was very seldom the case that $\hat{\sigma}_u^2 = 0$ or $\hat{p} = 1$ other than in the wrong skew cases.

1.4.1 Parameter Estimation

Table 1.2 contains the mean, bias, and MSE of the various parameter estimates, for the basic stochastic frontier model and for the ZISF model, for the case that $n = 200$. We also present the mean, bias, and MSE of the technical inefficiency estimates, and the mean of the "posterior" probabilities of full efficiency.

Unsurprisingly, the basic stochastic frontier model performs poorly except when $p = 0$ (in which case it is correctly specified). This is true for all three values of λ. We overestimate technical inefficiency, because we act as if all firms are inefficient, whereas in fact they are not. This bias is naturally bigger when p is bigger.

For the ZISF model, the results depend strongly on the value of λ. When $\lambda = 1$, the results are not very good. Note in particular the mean values of $\hat{p}$, which are 0.53, 0.49, 0.51, and 0.57 for $p = 0$, 0.25, 0.50, and 0.75, respectively. It is disturbing that the mean estimate of p does not appear to depend on the true value of p.

These problems are less severe for larger values of λ. The mean value of $\hat{p}$ when $p = 0$

is 0.33 for $\lambda = 2$ and 0.16 for $\lambda = 5$. The estimates are considerably better for the other values of p. So basically the model performs reasonably well when λ is large enough and p is not too close to zero.

Table 1.3 is similar to Table 1.2 except that it reports the results only for the cases of correct skew (i.e., wrong skew cases are not included). This makes almost no difference for $\lambda = 2$ or 5, because there are very few wrong skew cases when $\lambda = 2$ or 5. For $\lambda = 1$, it matters more. However, the conclusions given above really do not change.

Table 1.4 contains the same information as Table 1.2, except that now we have $n = 500$ rather than $n = 200$. The results are better for $n = 500$ than for $n = 200$, but a larger sample size does not really solve the problems that the ZISF model has in estimating p when $p = 0$ and/or $\lambda = 1$. For example, when $p = 0$, the mean $\hat{p}$ for $\lambda = 1, 2, 5$ is $0.53, 0.33, 0.16$ when $n = 200$ and $0.50, 0.30, 0.12$ when $n = 500$. Reading the table in the other direction, when $\lambda = 1$, the mean $\hat{p}$ for $p = 0, 0.25, 0.5, 0.75$ is $0.53, 0.49, 0.51, 0.57$ when $n = 200$ and $0.50, 0.46, 0.48, 0.58$ when $n = 500$. So again there are problems in estimating p when $p = 0$ or when λ is small.

It is perhaps not surprising that we encounter problems when we estimate the ZISF model when the true value of p is zero. Essentially, we are estimating a latent-class model with more classes than there really are. It is true that the class with zero probability contains no new parameters. If it did, they would not be identified and the results would presumably be much worse.

These results do not always agree with the summary of the results in KPT. KPT concentrate on the technical inefficiency estimates, and the only results they show explicitly for the parameter estimates (their Figure 3) are for $n = 1000$, and $\lambda = 5$ and $p = 0.25$. We did successfully replicate their results, but $n = 1000$ and $\lambda = 5$ is a very favorable parameter configuration. In their Section 1.3.1, they say the following about the case when the true p equals zero: "The ML estimator from the ZISF model is found to perform quite well. ... Estimates of p were close to zero." It is not clear what parameter configuration this refers to, but in our simulations this is not true except when $\lambda = 5$. For smaller values of λ, the ZISF estimates of p when the true $p = 0$ are not very close to zero.

1.4.2 Testing the Hypothesis $p = 0$

We now turn to the results of our simulations that are designed to investigate the size and power properties of the tests of the hypothesis $p = 0$, as discussed in Section 1.3.4 above. This hypothesis is economically interesting, and it is also practically important to know whether $p = 0$, since our model does not appear to perform well in that case. We would like to be able to recognize cases when $p = 0$ and just use the basic SF model in these cases.

The data generating process and parameter values for these simulations are as discussed above (in the beginning of Section 1.4). Specifically, the simulations are for $n = 200$ and $n = 500$.

We begin with the likelihood ratio (LR) test, which is the test that we believed *ex ante* would be most reliable. The results for $n = 200$ are given in Table 1.5. For each value of λ and p, we give the mean of the statistic (over the full set of 1000 replications), the number of rejections and the frequency of rejection. The rejection rates in the rows corresponding to $p = 0$ are the size of the test, whereas the rejection rates in the rows corresponding to the positive values of p represent power.

Look first at the set of results for all replications. The size of the test is reasonable. It is undersized for $\lambda = 1$ and approximately correctly sized for $\lambda = 2$ and 5. However, the power is disappointing, except when λ is large. There is essentially no power, even against the alternative $p = 0.75$, when $\lambda = 1$. When $\lambda = 2$, power is 0.60 against $p = 0.75$, but only 0.24 against $p = 0.50$ and 0.06 against $p = 0.25$. Power is more reasonable when $\lambda = 5$.

Table 1.6 gives the same results for $n = 500$. Increasing n has little effect on the size of the test, but it improves the power. Power is still low when $\lambda = 1$ or when $\lambda = 2$ and p is not large.

In either case ($n = 200$ or 500), looking separately at the correct-skew cases does not change our conclusions.

In Tables 1.7 and 1.8, we give results for the Wald test, for $n = 200$ and 500, respectively. Since the Wald test is undefined in wrong-skew cases, we show the results only for the correct-skew cases. We consider separately the OPG, Hessian, and Robust forms of the test, as defined in Section 1.3.4 above. Regardless of which form of the test is used, the test is considerably over-sized. This is true for both sample sizes. The problem is worst for the Robust form and least serious for the OPG form, but there are serious size distortions in all three cases. Based on these results, the Wald test is not recommended.

In Tables 1.9 and 1.10, we give the results for the score-based tests (LM, modified LM, and KT). Once again the tests are undefined for wrong-skew cases so we report results only for the correct-skew cases. The (two-sided) LM test is the best of the three. It shows moderate size distortions and no power when $\lambda = 1$, but only modest size distortions when $\lambda = 2$ or 5. The modified LM test has bigger size distortions and less power when $\lambda = 2$ or 5. The KT test has the largest size distortions and is therefore not recommended.

Our results are easy to summarize. The likelihood ratio test is the best of the five tests we have considered, at least for these parameter values. It is the only one of the tests that does not over-reject the true null that $p = 0$. However, it does not have much power. That is, we will have trouble rejecting the hypothesis that the basic SF model is correctly specified, even if the ZISF model is needed and p is not close to zero. The exception to this pessimistic conclusion is the case when both p and λ are large, in which case the power of the test is satisfactory.

1.5 EMPIRICAL EXAMPLE

We apply the models defined in Sections 1.2 and 1.3 to the Philippine rice data used in the empirical examples of Coelli et al. (2005), chapters 8-9. The Philippine data are composed of 43 farmers over eight years and Coelli et al. (2005) estimate the basic stochastic frontier model with a trans-log production function, ignoring the panel nature of the observations. Their output variable is tonnes of freshly threshed rice, and the input variables are planted area (in hectares), labor, and fertilizer used (in kilograms). These variables are scaled to have unit means so the first-order coefficients of the trans-log function can be interpreted as elasticities of output with respect to inputs evaluated at the variable means. We follow the basic setup of Coelli et al. (2005) but estimate the extended models where some farms are allowed to be efficient, and the probability of farm *i* being efficient and/or the distribution of u_i depend on farm characteristics. Data on age of household head, education of household head, household size, number of adults in the household, and the percentage of area classified as bantog (upland) fields are used as farm characteristics that influence the probability of a farm begin fully efficient and/or the distribution of the inefficiency. See Coelli et al. (2005) Appendix 2 for a detailed description of the data.

1.5.1 Model

We consider models based on the following specification:

$$\begin{aligned}\ln y_i &= \beta_0 + \theta t + \beta_1 \ln area_i + \beta_2 \ln labor_i + \beta_3 \ln npk_i + \frac{1}{2}\beta_{11}(\ln area_i)^2 \\ &+ \beta_{12} \ln area_i \ln labor_i + \beta_{13} \ln area_i \ln npk_i + \frac{1}{2}\beta_{22}(\ln labor_i)^2 + \beta_{23} \ln labor_i \ln npk_i \\ &+ \frac{1}{2}\beta_{33}(\ln npk_i)^2 + v_i - u_i, \qquad (1.7)\end{aligned}$$

$$u_i \sim N^+(0, \sigma_i^2), \sigma_i^2 = \exp(\gamma_0 + age_i\gamma_1 + edyrs_i\gamma_2 + hhsize_i\gamma_3 + nadult_i\gamma_4 + banrat_i\gamma_5), \qquad (1.8)$$

$$P(z_i = 1|w_i) = \frac{\exp(\delta_0 + age_i\delta_1 + edyrs_i\delta_2 + hhsize_i\delta_3 + nadult_i\delta_4 + banrat_i\delta_5)}{1 + \exp(\delta_0 + age_i\delta_1 + edyrs_i\delta_2 + hhsize_i\delta_3 + nadult_i\delta_4 + banrat_i\delta_5)}, \qquad (1.9)$$

where $area_i$ is the size of planted area in hectares, $labor_i$ is a measure of labor, npk_i is fertilizer in kilograms, age_i is the age of household head, $edyrs_i$ is the years of education of the household head, $hhsize_i$ is the household size, $nadult_i$ is the number of adults in the household, and $banrat_i$ is the percentage of area classified as bantog (upland) fields.

We assume a trans-log production function with time trend as in (1.7). We estimate the following models:

[a] the basic stochastic frontier model, in which σ_i^2 is constant ($\equiv \sigma_u^2$) and $P(z_i = 1|w_i) = 0$;

[b] the ZISF model in which σ_u^2 is constant and $P(z_i = 1|w_i)$ is constant ($\equiv p$) but not necessarily equal to zero;

[c] the "heteroskedasticity" model in which $p = 0$ but σ_i^2 is as given in (1.8);

[d] the "logit" model in which σ_u^2 is constant but $P(z_i = 1|w_i)$ is as given in (1.9);

[e] the "logit+heteroskedasticity" model in which σ_i^2 is as given in (1.8) and $P(z_i = 1|w_i)$ is as given in (1.9).

1.5.2 The Estimates

The MLEs and their OPG standard errors are reported in Table 1.11.

Consider first the results for the basic stochastic frontier model (first column of results in the table). The inputs are productive and there are roughly constant returns to scale. Average technical efficiency is about 70%. The estimated value of λ is 2.75, and both that value and the sample size ($n = 344$) are big enough to feel confident about proceeding to the ZISF model and its extensions.

The next block of column of results is for the ZISF model. Here we have $\hat{p} = 0.58$, so a substantial fraction of the observations (farm - time period combinations) are characterized by full efficiency. The technology (effect of inputs on output) is not changed much from the basic SF model, but the intercept is lower and the level of technical efficiency is higher (between 85% and 90%). Based on our simulations, this is a predictable consequence of finding that a substantial number of observations are fully efficient.

The next block of columns of results is for the heteroskedasticity model in which all farms are inefficient but the level of inefficiency depends on farm characteristics. A number of farm characteristics (age of the farmer, education of the farmer and percentage of bantog fields) have significant effects on the level of inefficiency. In this parameterization, a positive coefficient indicates that an increase in the corresponding variable makes a farm more *in*efficient. The model implies that farms where the farmer is older and more educated, and where the percentage of bantog fields is lower, tend to be more inefficient (less efficient). Or, saying the same thing the other way around, farms are more efficient on average if the farmer is younger and less educated and the percentage of bantog fields is higher. The effect of education is perhaps surprising. Because this model does not allow any farms to be fully efficient, we once again have a low level of average technical efficiency, about 72%, which is similar to that for the basic SF model.

Next we consider the logit model in which the distribution of inefficiency is the same for all firms that are not fully efficient, but the probability of being fully efficient depends on farm characteristics according to a logit model. Now age of the farmer and percentage of bantog fields have significant effects on the probability of full efficiency, and the coefficient of household size is almost significant at the 5% level (t statistic=-1.93). The results indicate that farms with younger farmers, smaller household size, and a larger proportion of bantog fields are more likely to be fully efficient. The results for age of the farmer and percentage of bantog fields are similar in nature to those for the heteroskedasticity model. The average level of inefficiency is once again higher, about 86%, which is very similar to the result for the ZISF model with constant p.

Finally, the last set of results are for the logit+heteroskedasticity model in which farm characteristics influence both the probability of being fully efficient and the distribution of inefficiency for those farms that are not fully efficient. Now none of the farm characteristics considered have significant effects on the distribution of inefficiency for the inefficient farms, but three of them (age of the farmer, household size and proportion of bantog fields) do have significant effects on the probability of being fully efficient. The coefficients for these three variables have the same signs as in the logit model without heteroskedasticity. It is interesting that we can estimate a model this complicated and still get significant results. Also, we note that, because this model allows the probability of full efficiency, we are back to a high average level of technical inefficiency, between 85% and 90%.

1.5.3 Model Comparison and Selection

We will now test the restrictions that distinguish the various models we have estimated. Based on the results of our simulations, we will use the likelihood ratio (LR) test. We immediately encounter some difficulties because, to use the LR test (or the other tests we considered in Section 1.3.4), the hypotheses should be nested, whereas not all of our models are nested. There are two possible nested hierarchies of models: (a) basic SF$\subset$ ZISF$\subset$logit$\subset$logit-heteroskedasticity, and (b) basic SF$\subset$heteroskedasticity.

We begin with hierarchy (a). When we test the hypothesis that $p = 0$ in the ZISF model, we obtain LR$=$ 5.07, which exceeds the 5% critical value of 2.71 for the distribution $({}^1/{}_2\chi^2_0 + {}^1/{}_2\chi^2_1)$. So we reject the basic SF model in favor of the ZISF model. Next we test the ZISF model against the logit model. This is a standard test of the hypothesis that $\delta_1 = \delta_2 = \delta_3 = \delta_4 = \delta_5 = 0$ in the logit model. The LR statistic of 23.96 exceeds the 5% critical value for the χ^2_5 distribution (11.07), so we reject the ZISF model in favor of the logit model. Finally, we test the logit model against the logit-heteroskedasticity model. This is a standard test of the hypothesis that $\gamma_1 = \gamma_2 = \gamma_3 = \gamma_4 = \gamma_5 = 0$ in the logit-heteroskedasticity model. The LR test statistic of 11.12 very marginally exceeds the 5% critical value, so we reject the logit model in favor of the logit-heteroskedasticity model, but not overwhelmingly. Note that the logit model is rejected even though, in the logit-heteroskedasticity model, none of the individual γ_j in the heteroskedasticity portion of the model is individually significant.

Now consider hierarchy (b). We test the basic SF model against the heteroskedasticity model. This is a standard test of the hypothesis that $\gamma_1 = \gamma_2 = \gamma_3 = \gamma_4 = \gamma_5 = 0$ in the heteroskedasticity model. The LR statistic of 17.04 exceeds the 5% critical value, so we reject the basic SF model in favor of the heteroskedasticity model.

We cannot test the heteroskedasticity model against the logit-heteroskedasticity model, at least not by standard methods, since the restriction that would convert the logit-heteroskedasticity model into the heteroskedasticity model is $\delta_0 = -\infty$ and under this null the other δ_j are unidentified. Still, the difference in log-likelihoods, which is 11.56, would appear to argue in favor of the logit-heteroskedasticity model.

In order to compare the models in a slightly different way, and to amplify on the comment at the end of the preceding paragraph, we will also consider some standard model selection criteria. We consider AIC$= -2LF + 2d$ (Akaike (1974)), BIC$= -2LF + d \ln n$ (Schwarz (1978)) and HQIC$= -2LF + 2d \ln(\ln n)$ (Hannan and Quinn (1979)), where d is the number of estimated parameters, n is the number of observations, and LF is the log-likelihood value. Smaller values of these criteria indicate a "better" model. We note that all three criteria favor the logit model over the heteroskedasticity model, two of the three favor the logit-heteroskedasticity model over the heteroskedasticity model, and two of the three favor the logit model over the logit-heteroskedasticity model.

Based on the results of our hypothesis tests and the comparison of the model selection procedures, we conclude that a case could be made for either the logit model or the logit-heteroskedasticity model as the preferred model. As we saw above, the substantive conclusions from these two models were basically the same.

1.6 CONCLUDING REMARKS

In this chapter we considered a generalization of the usual stochastic frontier model. In this new "ZISF" model, there is a probability p that a firm is fully efficient. This model was proposed by Kumbhakar, Parmeter, and Tsionas (2013), who showed how to estimate the model by MLE, how to update the probability of a firm being fully efficient on the basis of the data, and how to estimate the inefficiency level of a specific firm.

We extend their analysis in a number of ways. We show that a result similar to that of Waldman (1982) holds in the ZISF model, namely, that there is always a stationary point of the likelihood at parameter values that indicate no inefficiency, and that this point is a local maximum if the OLS residuals are positively skewed. We propose a model in which the probability of a firm being fully efficient is not constant, but rather is determined by a logit or probit model based on observable characteristics. We show how to test the hypothesis that $p = 0$. We also provide a more comprehensive set of simulations than Kumbhakar, Parmeter, and Tsionas (2013) did, and we include an empirical example.

Let $\lambda = \sigma_u/\sigma_v$, a standard measure in the stochastic frontier literature of the relative size of technical inefficiency and statistical noise. The main practical implication of our simulations is that the ZISF model works well when neither λ nor p is small. However, we have trouble estimating p reliably, or testing whether it equals zero, when λ is small. And if the true p equals zero, we have trouble estimating it reliably unless λ is larger than is empirically plausible (e.g., $\lambda = 5$). Larger sample size obviously helps, but the above conclusions do not depend strongly on sample size in our simulations. Situations where the ZISF model may be useful therefore have the characteristics that (i) it is reasonable to suppose that some firms are fully efficient, and (ii) the inefficiency levels of the inefficient firms are not small relative to statistical noise. Such situations do not seem implausible, and it is an empirical question as to how common they are.

Table 1.1: Frequency of a positive third moment of the OLS residuals

	$n = 50$			$n = 100$			$n = 200$			$n = 400$		
	$\lambda = 0.5$	$\lambda = 1$	$\lambda = 2$	$\lambda = 0.5$	$\lambda = 1$	$\lambda = 2$	$\lambda = 0.5$	$\lambda = 1$	$\lambda = 2$	$\lambda = 0.5$	$\lambda = 1$	$\lambda = 2$
$p = 0$	0.476	0.363	0.114	0.463	0.300	0.038	0.447	0.224	0.006	0.421	0.139	0.000
$p = 0.1$	0.475	0.352	0.102	0.460	0.286	0.031	0.443	0.210	0.003	0.416	0.123	0.000
$p = 0.2$	0.472	0.338	0.080	0.456	0.266	0.019	0.438	0.185	0.001	0.407	0.101	0.000
$p = 0.3$	0.469	0.322	0.058	0.451	0.245	0.011	0.431	0.161	0.000	0.398	0.079	0.000
$p = 0.4$	0.466	0.308	0.042	0.447	0.226	0.006	0.424	0.141	0.000	0.391	0.062	0.000
$p = 0.5$	0.465	0.300	0.034	0.444	0.215	0.004	0.421	0.129	0.000	0.386	0.053	0.000
$p = 0.6$	0.466	0.299	0.033	0.445	0.215	0.004	0.420	0.128	0.000	0.387	0.052	0.000
$p = 0.7$	0.468	0.311	0.043	0.449	0.229	0.006	0.427	0.143	0.000	0.394	0.063	0.000
$p = 0.8$	0.474	0.342	0.076	0.458	0.268	0.017	0.439	0.185	0.001	0.414	0.098	0.000
$p = 0.9$	0.485	0.399	0.178	0.474	0.348	0.079	0.463	0.280	0.019	0.446	0.200	0.001

Table 1.2: Basic SF Model vs. ZISF Model, all replications : $n = 200$

	Basic SF Model			ZISF Model			Basic SF Model			ZISF Model		
	$\lambda = 1\,(\sigma_u = 1, \sigma_v = 1)\,, p = 0$						$\lambda = 1\,(\sigma_u = 1, \sigma_v = 1)\,, p = 0.25$					
	mean	bias	mse	mean	bias	mse	mean	bias	mse	mean	bias	mse
β_0	0.87	-0.13	0.20	0.60	-0.40	0.29	1.16	0.16	0.19	0.85	-0.15	0.15
σ_u	0.84	-0.16	0.30	0.91	-0.09	0.40	0.95	-0.05	0.26	1.00	0.00	0.33
σ_v	0.99	-0.01	0.02	1.02	0.02	0.02	0.97	-0.03	0.02	1.01	0.01	0.02
p		–		0.53	0.53	0.43		–		0.49	0.24	0.19
λ	0.93	-0.07	0.44	0.95	-0.05	0.45	1.07	0.07	0.45	1.06	0.06	0.43
σ	1.39	-0.02	0.04	1.47	0.06	0.12	1.44	0.03	0.04	1.51	0.09	0.11
$\log L$	-313.00			-312.85			-314.71			-314.55		
$\hat{E}\left[u_i\|\hat{\varepsilon}_i\right]$	0.67	-0.13	0.50	0.40	-0.40	0.60	0.76	0.16	0.51	0.45	-0.15	0.47
$\hat{p}\left(z_i = 1\|\hat{\varepsilon}_i\right)$				0.53						0.49		
	$\lambda = 1\,(\sigma_u = 1, \sigma_v = 1)\,, p = 0.5$						$\lambda = 1\,(\sigma_u = 1, \sigma_v = 1)\,, p = 0.75$					
	mean	bias	mse	mean	bias	mse	mean	bias	mse	mean	bias	mse
β_0	1.40	0.40	0.30	1.04	0.04	0.12	1.51	0.51	0.40	1.16	0.16	0.13
σ_u	1.00	0.00	0.22	1.07	0.07	0.31	0.89	-0.11	0.23	1.01	0.01	0.34
σ_v	0.93	-0.07	0.02	0.98	-0.02	0.02	0.90	-0.10	0.03	0.95	-0.05	0.02
p		–		0.51	0.01	0.12		–		0.57	-0.18	0.16
λ	1.16	0.16	0.45	1.15	0.15	0.43	1.08	0.08	0.43	1.12	0.12	0.44
σ	1.44	0.03	0.04	1.52	0.11	0.12	1.35	-0.07	0.04	1.47	0.06	0.13
$\log L$	-311.00			-310.77			-300.69			-300.40		
$\hat{E}\left[u_i\|\hat{\varepsilon}_i\right]$	0.80	0.40	0.59	0.44	0.05	0.40	0.71	0.51	0.62	0.36	0.16	0.33
$\hat{p}\left(z_i = 1\|\hat{\varepsilon}_i\right)$				0.51						0.57		

Table 1.2: (cont'd)

	Basic SF Model			ZISF Model			Basic SF Model			ZISF Model		
	$\lambda = 2\,(\sigma_u = 1, \sigma_v = 0.5)\,, p = 0$						$\lambda = 2\,(\sigma_u = 1, \sigma_v = 0.5)\,, p = 0.25$					
	mean	bias	mse	mean	bias	mse	mean	bias	mse	mean	bias	mse
β_0	0.98	−0.02	0.02	0.71	−0.29	0.14	1.26	0.26	0.08	0.98	−0.02	0.05
σ_u	0.98	−0.02	0.03	0.97	−0.03	0.05	1.08	0.08	0.03	1.04	0.04	0.03
σ_v	0.50	0.00	0.01	0.54	0.04	0.01	0.45	−0.05	0.01	0.50	0.00	0.01
p		–		0.33	0.33	0.19		–		0.30	0.05	0.06
λ	2.08	0.08	0.48	1.89	−0.11	0.41	2.53	0.53	0.88	2.21	0.21	0.55
σ	1.11	−0.01	0.01	1.12	0.00	0.03	1.18	0.06	0.01	1.16	0.05	0.02
$\log L$	−229.93			−229.65			−232.55			−232.22		
$\hat{E}\left[u_i \mid \hat{\varepsilon}_i\right]$	0.78	−0.02	0.17	0.51	−0.29	0.30	0.86	0.26	0.23	0.58	−0.02	0.20
$\hat{p}\left(z_i = 1 \mid \hat{\varepsilon}_i\right)$				0.33						0.30		
	$\lambda = 2\,(\sigma_u = 1, \sigma_v = 0.5)\,, p = 0.5$						$\lambda = 2\,(\sigma_u = 1, \sigma_v = 0.5)\,, p = 0.75$					
	mean	bias	mse	mean	bias	mse	mean	bias	mse	mean	bias	mse
β_0	1.45	0.45	0.21	1.07	0.07	0.05	1.52	0.52	0.28	1.05	0.05	0.03
σ_u	1.07	0.07	0.02	1.06	0.06	0.02	0.91	−0.09	0.02	1.03	0.03	0.04
σ_v	0.39	−0.11	0.02	0.47	−0.03	0.01	0.37	−0.13	0.02	0.48	−0.02	0.00
p		–		0.44	−0.06	0.06		–		0.67	−0.08	0.05
λ	2.83	0.83	1.29	2.34	0.34	0.57	2.54	0.54	0.72	2.18	0.18	0.28
σ	1.15	0.03	0.01	1.16	0.04	0.02	0.99	−0.13	0.03	1.14	0.02	0.03
$\log L$	−221.64			−220.71			−196.18			−193.77		
$\hat{E}\left[u_i \mid \hat{\varepsilon}_i\right]$	0.85	0.45	0.35	0.47	0.08	0.18	0.72	0.52	0.39	0.25	0.05	0.11
$\hat{p}\left(z_i = 1 \mid \hat{\varepsilon}_i\right)$				0.44						0.67		

Table 1.2: (cont'd)

	Basic SF Model			ZISF Model			Basic SF Model			ZISF Model		
	$\lambda = 5\,(\sigma_u = 1, \sigma_v = 0.2)\,, p = 0$						$\lambda = 5\,(\sigma_u = 1, \sigma_v = 0.2)\,, p = 0.25$					
	mean	bias	mse	mean	bias	mse	mean	bias	mse	mean	bias	mse
β_0	1.00	0.00	0.00	0.85	-0.15	0.04	1.23	0.23	0.05	1.03	0.03	0.01
σ_u	1.00	0.00	0.01	0.97	-0.03	0.01	1.05	0.05	0.01	1.02	0.02	0.00
σ_v	0.19	-0.01	0.00	0.23	0.03	0.00	0.14	-0.06	0.01	0.19	-0.01	0.00
p		–		0.16	0.16	0.05		–		0.23	-0.02	0.02
λ	5.61	0.61	7.04	4.79	-0.21	6.21	8.32	3.32	21.81	6.06	1.06	10.77
σ	1.02	0.00	0.00	1.00	-0.02	0.01	1.05	0.03	0.01	1.04	0.02	0.00
$\log L$	-176.43			-176.06			-174.14			-172.76		
$\hat{E}\left[u_i \mid \hat{\varepsilon}_i\right]$	0.80	0.00	0.04	0.65	-0.15	0.08	0.82	0.23	0.09	0.63	0.03	0.04
$\hat{p}\left(z_i = 1 \mid \hat{\varepsilon}_i\right)$				0.16						0.23		
	$\lambda = 5\,(\sigma_u = 1, \sigma_v = 0.2)\,, p = 0.5$						$\lambda = 5\,(\sigma_u = 1, \sigma_v = 0.2)\,, p = 0.75$					
	mean	bias	mse	mean	bias	mse	mean	bias	mse	mean	bias	mse
β_0	1.30	0.30	0.09	1.00	0.00	0.00	1.32	0.32	0.10	1.00	0.00	0.00
σ_u	0.93	-0.07	0.01	1.01	0.01	0.01	0.71	-0.29	0.09	1.00	0.00	0.02
σ_v	0.11	-0.09	0.01	0.20	0.00	0.00	0.11	-0.09	0.01	0.20	0.00	0.00
p		–		0.50	0.00	0.01		–		0.75	0.00	0.00
λ	8.96	3.96	36.70	5.17	0.17	0.61	6.96	1.96	19.52	5.05	0.06	0.43
σ	0.94	-0.08	0.01	1.03	0.01	0.01	0.72	-0.30	0.09	1.02	0.00	0.02
$\log L$	-148.06			-138.15			-99.12			-74.13		
$\hat{E}\left[u_i \mid \hat{\varepsilon}_i\right]$	0.70	0.30	0.13	0.40	0.00	0.03	0.52	0.32	0.13	0.20	0.00	0.02
$\hat{p}\left(z_i = 1 \mid \hat{\varepsilon}_i\right)$				0.50						0.75		

Table 1.3: Basic SF Model vs. ZISF Model, correct skew replications : $n = 200$

	Basic SF Model			ZISF Model			Basic SF Model			ZISF Model		
	$\lambda = 1\,(\sigma_u = 1, \sigma_v = 1)\,, p = 0$						$\lambda = 1\,(\sigma_u = 1, \sigma_v = 1)\,, p = 0.25$					
	mean	bias	mse	mean	bias	mse	mean	bias	mse	mean	bias	mse
β_0	1.06	0.06	0.07	0.71	−0.29	0.20	1.29	0.29	0.16	0.93	−0.07	0.12
σ_u	1.08	0.08	0.10	1.17	0.17	0.22	1.13	0.13	0.12	1.19	0.19	0.20
σ_v	0.94	−0.06	0.02	0.98	−0.02	0.02	0.94	−0.06	0.02	0.98	−0.02	0.02
p		–		0.40	0.40	0.26		–		0.39	0.14	0.12
λ	1.20	0.20	0.28	1.23	0.23	0.30	1.27	0.27	0.35	1.26	0.26	0.32
σ	1.46	0.05	0.03	1.56	0.15	0.13	1.50	0.08	0.03	1.57	0.16	0.11
$\log L$	−313.16			−312.97			−315.09			−314.90		
$\hat{E}\left[u_i \mid \hat{\varepsilon}_i\right]$	0.86	0.06	0.36	0.51	−0.29	0.49	0.90	0.30	0.47	0.54	−0.06	0.43
$\hat{p}\left(z_i = 1 \mid \hat{\varepsilon}_i\right)$				0.40						0.39		
	$\lambda = 1\,(\sigma_u = 1, \sigma_v = 1)\,, p = 0.5$						$\lambda = 1\,(\sigma_u = 1, \sigma_v = 1)\,, p = 0.75$					
	mean	bias	mse	mean	bias	mse	mean	bias	mse	mean	bias	mse
β_0	1.51	0.51	0.32	1.10	0.10	0.12	1.64	0.64	0.46	1.22	0.22	0.15
σ_u	1.14	0.14	0.11	1.22	0.22	0.21	1.05	0.05	0.09	1.19	0.19	0.23
σ_v	0.91	−0.09	0.02	0.96	−0.04	0.02	0.87	−0.13	0.03	0.93	−0.07	0.02
p		–		0.44	−0.06	0.11		–		0.49	−0.26	0.18
λ	1.33	0.33	0.38	1.31	0.31	0.35	1.27	0.27	0.33	1.31	0.31	0.35
σ	1.48	0.07	0.03	1.58	0.16	0.12	1.40	−0.02	0.03	1.54	0.13	0.13
$\log L$	−311.17			−310.91			−301.06			−300.72		
$\hat{E}\left[u_i \mid \hat{\varepsilon}_i\right]$	0.91	0.51	0.61	0.51	0.11	0.40	0.84	0.64	0.69	0.42	0.22	0.34
$\hat{p}\left(z_i = 1 \mid \hat{\varepsilon}_i\right)$				0.44						0.49		

Table 1.3: (cont'd)

	Basic SF Model			ZISF Model			Basic SF Model			ZISF Model		
	$\lambda = 2\,(\sigma_u = 1, \sigma_v = 0.5)\,, p = 0$						$\lambda = 2\,(\sigma_u = 1, \sigma_v = 0.5)\,, p = 0.25$					
	mean	bias	mse	mean	bias	mse	mean	bias	mse	mean	bias	mse
β_0	0.99	−0.01	0.02	0.72	−0.28	0.14	1.26	0.26	0.08	0.98	−0.02	0.05
σ_u	0.98	−0.02	0.03	0.97	−0.03	0.04	1.08	0.08	0.03	1.04	0.04	0.03
σ_v	0.49	−0.01	0.01	0.54	0.04	0.01	0.45	−0.05	0.01	0.50	0.00	0.01
p		–		0.32	0.32	0.18		–		0.30	0.05	0.06
λ	2.10	0.10	0.46	1.90	−0.10	0.39	2.53	0.53	0.88	2.21	0.21	0.55
σ	1.11	−0.01	0.01	1.12	0.00	0.03	1.18	0.06	0.01	1.16	0.05	0.02
$\log L$	−229.94			−229.66			−232.57			−232.24		
$\hat{E}\left[u_i \vert \hat{\varepsilon}_i\right]$	0.78	−0.01	0.16	0.52	−0.28	0.29	0.86	0.26	0.23	0.58	−0.02	0.20
$\hat{p}\left(z_i = 1 \vert \hat{\varepsilon}_i\right)$				0.32						0.30		
	$\lambda = 2\,(\sigma_u = 1, \sigma_v = 0.5)\,, p = 0.5$						$\lambda = 2\,(\sigma_u = 1, \sigma_v = 0.5)\,, p = 0.75$					
	mean	bias	mse	mean	bias	mse	mean	bias	mse	mean	bias	mse
β_0	1.45	0.45	0.21	1.07	0.07	0.05	1.52	0.52	0.28	1.05	0.05	0.03
σ_u	1.07	0.07	0.02	1.06	0.06	0.02	0.91	−0.09	0.02	1.03	0.03	0.04
σ_v	0.39	−0.11	0.02	0.47	−0.03	0.01	0.37	−0.13	0.02	0.48	−0.02	0.00
p		–		0.44	−0.06	0.06		–		0.67	−0.08	0.05
λ	2.83	0.83	1.29	2.34	0.34	0.57	2.54	0.54	0.72	2.19	0.19	0.28
σ	1.15	0.03	0.01	1.16	0.04	0.02	0.99	−0.13	0.03	1.14	0.02	0.03
$\log L$	−221.64			−220.71			−196.20			−193.79		
$\hat{E}\left[u_i \vert \hat{\varepsilon}_i\right]$	0.85	0.45	0.35	0.47	0.08	0.18	0.72	0.52	0.39	0.25	0.05	0.11
$\hat{p}\left(z_i = 1 \vert \hat{\varepsilon}_i\right)$				0.44						0.67		

Table 1.3: (cont'd)

	Basic SF Model			ZISF Model			Basic SF Model			ZISF Model		
	$\lambda = 5\,(\sigma_u = 1, \sigma_v = 0.2)\,, p = 0$						$\lambda = 5\,(\sigma_u = 1, \sigma_v = 0.2)\,, p = 0.25$					
	mean	bias	mse	mean	bias	mse	mean	bias	mse	mean	bias	mse
β_0	1.00	0.00	0.00	0.85	−0.15	0.04	1.23	0.23	0.05	1.03	0.03	0.01
σ_u	1.00	0.00	0.01	0.97	−0.03	0.01	1.04	0.04	0.01	1.02	0.02	0.00
σ_v	0.19	−0.01	0.00	0.23	0.03	0.00	0.14	−0.06	0.01	0.19	−0.01	0.00
p		–		0.16	0.16	0.05		–		0.23	−0.02	0.02
λ	5.61	0.61	7.04	4.79	−0.21	6.21	8.32	3.32	21.81	6.06	1.06	10.77
σ	1.02	0.00	0.00	1.00	−0.02	0.01	1.05	0.03	0.01	1.04	0.02	0.00
$\log L$	−176.43			−176.06			−174.14			−172.76		
$\hat{E}\left[u_i \mid \hat{\varepsilon}_i\right]$	0.80	0.00	0.04	0.65	−0.15	0.08	0.82	0.23	0.09	0.63	0.03	0.04
$\hat{p}\left(z_i = 1 \mid \hat{\varepsilon}_i\right)$				0.16						0.23		
	$\lambda = 5\,(\sigma_u = 1, \sigma_v = 0.2)\,, p = 0.5$						$\lambda = 5\,(\sigma_u = 1, \sigma_v = 0.2)\,, p = 0.75$					
	mean	bias	mse	mean	bias	mse	mean	bias	mse	mean	bias	mse
β_0	1.30	0.30	0.09	1.00	0.00	0.00	1.32	0.32	0.10	1.00	0.00	0.00
σ_u	0.93	−0.07	0.01	1.01	0.01	0.01	0.71	−0.29	0.09	1.00	0.00	0.02
σ_v	0.11	−0.09	0.01	0.20	0.00	0.00	0.11	−0.09	0.01	0.20	0.00	0.00
p		–		0.50	0.00	0.01		–		0.75	0.00	0.00
λ	8.96	3.96	36.70	5.17	0.17	0.61	6.96	1.96	19.52	5.05	0.06	0.43
σ	0.94	−0.08	0.01	1.03	0.01	0.01	0.72	−0.30	0.09	1.02	0.00	0.02
$\log L$	−148.06			−138.15			−99.12			−74.13		
$\hat{E}\left[u_i \mid \hat{\varepsilon}_i\right]$	0.70	0.30	0.13	0.40	0.00	0.03	0.52	0.32	0.13	0.20	0.00	0.02
$\hat{p}\left(z_i = 1 \mid \hat{\varepsilon}_i\right)$				0.50						0.75		

Table 1.4: Basic SF Model vs. ZISF Model, all replications : $n = 500$

	Basic SF Model			ZISF Model			Basic SF Model			ZISF Model		
	$\lambda = 1\,(\sigma_u = 1, \sigma_v = 1)\,, p = 0$						$\lambda = 1\,(\sigma_u = 1, \sigma_v = 1)\,, p = 0.25$					
	mean	bias	mse	mean	bias	mse	mean	bias	mse	mean	bias	mse
β_0	0.90	-0.10	0.11	0.59	-0.41	0.26	1.19	0.19	0.13	0.85	-0.15	0.12
σ_u	0.88	-0.12	0.17	0.94	-0.06	0.24	0.99	-0.01	0.13	1.03	0.03	0.19
σ_v	1.01	0.01	0.01	1.04	0.04	0.01	0.99	-0.01	0.01	1.03	0.03	0.01
p		–		0.50	0.50	0.38		–		0.46	0.21	0.16
λ	0.91	-0.09	0.21	0.93	-0.07	0.23	1.04	0.04	0.19	1.03	0.03	0.19
σ	1.39	-0.02	0.02	1.46	0.05	0.08	1.44	0.03	0.02	1.50	0.09	0.07
$\log L$	-785.23			-785.03			-790.14			-789.94		
$\hat{E}\left[u_i \vert \hat{\varepsilon}_i\right]$	0.70	-0.10	0.40	0.39	-0.41	0.55	0.79	0.19	0.42	0.45	-0.15	0.42
$\hat{p}\left(z_i = 1 \vert \hat{\varepsilon}_i\right)$				0.50						0.46		
	$\lambda = 1\,(\sigma_u = 1, \sigma_v = 1)\,, p = 0.5$						$\lambda = 1\,(\sigma_u = 1, \sigma_v = 1)\,, p = 0.75$					
	mean	bias	mse	mean	bias	mse	mean	bias	mse	mean	bias	mse
β_0	1.45	0.45	0.26	1.05	0.05	0.09	1.56	0.56	0.38	1.14	0.14	0.09
σ_u	1.06	0.06	0.10	1.10	0.10	0.14	0.95	-0.05	0.11	1.05	0.05	0.19
σ_v	0.94	-0.06	0.01	0.99	-0.01	0.01	0.91	-0.09	0.02	0.97	-0.03	0.01
p		–		0.48	-0.02	0.10		–		0.58	-0.17	0.14
λ	1.17	0.17	0.19	1.13	0.13	0.16	1.08	0.08	0.19	1.10	0.10	0.21
σ	1.45	0.04	0.02	1.51	0.10	0.06	1.35	-0.06	0.02	1.47	0.06	0.08
$\log L$	-779.58			-779.28			-754.00			-753.53		
$\hat{E}\left[u_i \vert \hat{\varepsilon}_i\right]$	0.85	0.45	0.53	0.45	0.05	0.36	0.76	0.56	0.58	0.34	0.14	0.27
$\hat{p}\left(z_i = 1 \vert \hat{\varepsilon}_i\right)$				0.48						0.58		

Table 1.4: (cont'd)

	Basic SF Model			ZISF Model			Basic SF Model			ZISF Model		
	$\lambda = 2\,(\sigma_u = 1, \sigma_v = 0.5)\,, p = 0$						$\lambda = 2\,(\sigma_u = 1, \sigma_v = 0.5)\,, p = 0.25$					
	mean	bias	mse	mean	bias	mse	mean	bias	mse	mean	bias	mse
β_0	0.99	−0.01	0.01	0.74	−0.26	0.11	1.26	0.26	0.07	1.00	0.00	0.04
σ_u	0.99	−0.01	0.01	0.95	−0.05	0.01	1.08	0.08	0.01	1.02	0.02	0.01
σ_v	0.50	0.00	0.00	0.54	0.04	0.01	0.46	−0.04	0.00	0.50	0.00	0.00
p		–		0.30	0.30	0.15		–		0.27	0.02	0.04
λ	2.02	0.02	0.15	1.79	−0.21	0.19	2.39	0.39	0.32	2.08	0.08	0.18
σ	1.11	0.00	0.00	1.10	−0.02	0.01	1.18	0.06	0.01	1.15	0.03	0.00
$\log L$	−577.71			−577.38			−584.97			−584.58		
$\hat{E}\left[u_i \vert \hat{\varepsilon}_i\right]$	0.79	−0.01	0.15	0.54	−0.26	0.26	0.86	0.26	0.22	0.60	0.00	0.18
$\hat{p}\left(z_i = 1 \vert \hat{\varepsilon}_i\right)$				0.30						0.27		
	$\lambda = 2\,(\sigma_u = 1, \sigma_v = 0.5)\,, p = 0.5$						$\lambda = 2\,(\sigma_u = 1, \sigma_v = 0.5)\,, p = 0.75$					
	mean	bias	mse	mean	bias	mse	mean	bias	mse	mean	bias	mse
β_0	1.45	0.45	0.21	1.05	0.05	0.03	1.52	0.52	0.27	1.02	0.02	0.01
σ_u	1.07	0.07	0.01	1.02	0.02	0.01	0.91	−0.09	0.01	1.00	0.00	0.02
σ_v	0.40	−0.10	0.01	0.49	−0.01	0.00	0.38	−0.12	0.02	0.49	−0.01	0.00
p		–		0.46	−0.04	0.04		–		0.72	−0.03	0.02
λ	2.72	0.72	0.68	2.14	0.14	0.15	2.43	0.43	0.32	2.04	0.04	0.07
σ	1.15	0.03	0.00	1.14	0.02	0.01	0.99	−0.13	0.02	1.12	0.00	0.01
$\log L$	−556.16			−554.57			−492.83			−487.45		
$\hat{E}\left[u_i \vert \hat{\varepsilon}_i\right]$	0.85	0.45	0.34	0.45	0.05	0.15	0.72	0.52	0.39	0.21	0.02	0.09
$\hat{p}\left(z_i = 1 \vert \hat{\varepsilon}_i\right)$				0.46						0.72		

Table 1.4: (cont'd)

	Basic SF Model			ZISF Model			Basic SF Model			ZISF Model		
	$\lambda = 5\,(\sigma_u = 1, \sigma_v = 0.2)\,, p = 0$						$\lambda = 5\,(\sigma_u = 1, \sigma_v = 0.2)\,, p = 0.25$					
β_0	1.00	0.00	0.00	0.88	−0.12	0.03	1.22	0.22	0.05	1.01	0.01	0.01
σ_u	1.00	0.00	0.00	0.97	−0.03	0.00	1.04	0.04	0.00	1.01	0.01	0.00
σ_v	0.20	0.00	0.00	0.22	0.02	0.00	0.14	−0.06	0.00	0.19	−0.01	0.00
p		–		0.12	0.12	0.03		–		0.24	−0.01	0.01
λ	5.18	0.18	0.99	4.55	−0.45	1.35	7.55	2.55	8.46	5.36	0.36	1.87
σ	1.02	0.00	0.00	0.99	−0.03	0.00	1.05	0.03	0.00	1.03	0.01	0.00
$\log L$	−442.32			−441.94			−436.65			−433.87		
$\hat{E}\left[u_i \vert \hat{\varepsilon}_i\right]$	0.80	0.00	0.03	0.68	−0.12	0.06	0.82	0.22	0.08	0.61	0.01	0.04
$\hat{p}\left(z_i = 1 \vert \hat{\varepsilon}_i\right)$				0.12						0.24		
	$\lambda = 5\,(\sigma_u = 1, \sigma_v = 0.2)\,, p = 0.5$						$\lambda = 5\,(\sigma_u = 1, \sigma_v = 0.2)\,, p = 0.75$					
	mean	bias	mse	mean	bias	mse	mean	bias	mse	mean	bias	mse
β_0	1.30	0.30	0.09	1.00	0.00	0.00	1.32	0.32	0.10	1.00	0.00	0.00
σ_u	0.93	−0.07	0.01	1.00	0.00	0.00	0.71	−0.29	0.09	1.00	0.00	0.01
σ_v	0.12	−0.08	0.01	0.20	0.00	0.00	0.11	−0.09	0.01	0.20	0.00	0.00
p		–		0.50	0.00	0.00		–		0.75	0.00	0.00
λ	8.01	3.01	10.75	5.03	0.03	0.17	6.36	1.36	2.77	5.00	0.00	0.16
σ	0.93	−0.09	0.01	1.02	0.00	0.00	0.72	−0.30	0.09	1.02	0.00	0.01
$\log L$	−371.51			−347.55			−250.31			−188.10		
$\hat{E}\left[u_i \vert \hat{\varepsilon}_i\right]$	0.70	0.30	0.12	0.40	0.00	0.02	0.52	0.32	0.13	0.20	0.00	0.01
$\hat{p}\left(z_i = 1 \vert \hat{\varepsilon}_i\right)$				0.50						0.75		

Table 1.5: Likelihood Ratio Test, $n = 200$

		All				Correct Skew				Incorrect Skew			
		Mean	Rejection		Total	Mean	Rejection		Total	Mean	Rejection		Total
$\lambda = 1$	$p = 0$	0.29	21	(0.02)	1000	0.38	21	(0.03)	776	0.00	0	(0.00)	224
	$p = 0.25$	0.32	20	(0.02)	1000	0.37	20	(0.02)	842	0.00	0	(0.00)	158
	$p = 0.5$	0.46	40	(0.04)	1000	0.53	40	(0.05)	878	0.00	0	(0.00)	122
	$p = 0.75$	0.57	53	(0.05)	1000	0.67	53	(0.06)	850	0.00	0	(0.00)	150
$\lambda = 2$	$p = 0$	0.56	42	(0.04)	1000	0.56	42	(0.04)	994	0.00	0	(0.00)	6
	$p = 0.25$	0.66	63	(0.06)	1000	0.66	63	(0.06)	999	0.00	0	(0.00)	1
	$p = 0.5$	1.87	244	(0.24)	1000	1.87	244	(0.24)	1000		–		0
	$p = 0.75$	4.81	596	(0.60)	1000	4.81	596	(0.60)	999	0.00	0	(0.00)	1
$\lambda = 5$	$p = 0$	0.73	60	(0.06)	999^{+}	0.73	61	(0.06)	999^{+}		–		0
	$p = 0.25$	2.76	393	(0.39)	996^{+}	2.76	395	(0.40)	996^{+}		–		0
	$p = 0.5$	19.82	988	(0.99)	997^{+}	19.82	988	(0.99)	997^{+}		–		0
	$p = 0.75$	49.98	997	(1.00)	997^{+}	49.98	997	(1.00)	997^{+}		–		0

1. $^{+}$ Some iterations dropped due to $\hat{\sigma}_v$ being too small such that $\hat{\lambda}$ is not well defined.

Table 1.6: Likelihood Ratio Test, $n = 500$

		All				Correct Skew				Incorrect Skew			
		Mean	Rejection		Total	Mean	Rejection		Total	Mean	Rejection		Total
$\lambda = 1$	$p = 0$	0.39	30	(0.03)	1000	0.45	30	(0.03)	879	0.00	0	(0.00)	121
	$p = 0.25$	0.41	28	(0.03)	1000	0.45	28	(0.03)	921	0.00	0	(0.00)	79
	$p = 0.5$	0.60	48	(0.05)	1000	0.62	48	(0.05)	964	0.00	0	(0.00)	36
	$p = 0.75$	0.95	102	(0.10)	1000	1.01	102	(0.11)	939	0.00	0	(0.00)	61
$\lambda = 2$	$p = 0$	0.66	56	(0.06)	1000	0.66	56	(0.06)	1000		–		0
	$p = 0.25$	0.77	63	(0.06)	1000	0.77	63	(0.06)	1000		–		0
	$p = 0.5$	3.19	461	(0.46)	1000	3.19	461	(0.46)	1000		–		0
	$p = 0.75$	10.76	911	(0.91)	1000	10.76	911	(0.91)	1000		–		0
$\lambda = 5$	$p = 0$	0.75	70	(0.07)	1000	0.75	70	(0.07)	1000		–		0
	$p = 0.25$	5.55	689	(0.69)	1000	5.55	689	(0.69)	1000		–		0
	$p = 0.5$	47.94	1000	(1.00)	1000	47.94	1000	(1.00)	1000		–		0
	$p = 0.75$	124.42	1000	(1.00)	1000	124.42	1000	(1.00)	1000		–		0

Table 1.7: Wald Test, $n = 200$

		OPG				Hessian				Robust			
		Mean	Rejection		Total	Mean	Rejection		Total	Mean	Rejection		Total
$\lambda = 1$	$p = 0$	5.92	128	(0.17)	773^*	57.97	189	(0.24)	776	143.75	546	(0.70)	776
	$p = 0.25$	4.42	147	(0.18)	838^*	36.90	215	(0.26)	842	104.61	592	(0.70)	842
	$p = 0.5$	6.98	179	(0.21)	873^*	40.61	270	(0.31)	878	135.57	629	(0.72)	878
	$p = 0.75$	9.74	247	(0.29)	849^*	63.53	334	(0.39)	850	146.11	607	(0.71)	850
$\lambda = 2$	$p = 0$	6.18	215	(0.22)	994	21.23	290	(0.29)	994	45.46	620	(0.62)	994
	$p = 0.25$	4.26	264	(0.26)	999	10.21	320	(0.32)	$997^\diamond$	19.28	618	(0.62)	999
	$p = 0.5$	10.91	580	(0.58)	1000	14.37	639	(0.64)	1000	19.46	735	(0.73)	1000
	$p = 0.75$	45.72	856	(0.86)	999	61.79	883	(0.88)	$998^\diamond$	80.71	906	(0.91)	999
$\lambda = 5$	$p = 0$	3.25	266	(0.27)	999^+	3.58	315	(0.32)	$996^{\diamond+}$	5.03	490	(0.49)	999^+
	$p = 0.25$	8.63	696	(0.70)	998^+	9.16	725	(0.73)	$997^{\diamond+}$	9.90	727	(0.73)	998^+
	$p = 0.5$	59.73	997	(1.00)	998^+	60.75	997	(1.00)	998^+	61.12	996	(1.00)	998^+
	$p = 0.75$	247.62	1000	(1.00)	1000	254.99	1000	(1.00)	1000	257.77	1000	(1.00)	1000

1. * Some iterations are dropped due to a singular OPG variance matrix.
2. $^\diamond$Some of the iterations where MLE is at the boundary ($\hat{p} = 0$) are dropped due to not negative definite Hessian.
3. $^+$ Some iterations dropped due to $\hat{\sigma}_v$ being too small such that $\hat{\lambda}$ is not well defined.

Table 1.8: Wald Test, $n = 500$

		OPG				Hessian				Robust			
		Mean	Rejection		Total	Mean	Rejection		Total	Mean	Rejection		Total
$\lambda = 1$	$p = 0$	12.05	201	(0.23)	878^*	112.28	250	(0.29)	$877^\diamond$	286.90	620	(0.71)	878°
	$p = 0.25$	10.34	203	(0.22)	921	94.59	264	(0.29)	921	215.21	639	(0.69)	921
	$p = 0.5$	12.69	275	(0.29)	963^*	47.99	347	(0.36)	964	121.96	661	(0.69)	964
	$p = 0.75$	24.74	368	(0.39)	938^*	120.86	447	(0.48)	$937^\diamond$	258.85	678	(0.72)	939
$\lambda = 2$	$p = 0$	5.32	262	(0.26)	1000	26.94	310	(0.31)	1000	47.13	618	(0.62)	1000
	$p = 0.25$	3.94	306	(0.31)	1000	0.47	373	(0.37)	$999^\diamond$	8.46	642	(0.64)	1000
	$p = 0.5$	17.10	800	(0.80)	1000	19.10	831	(0.83)	1000	22.49	837	(0.84)	1000
	$p = 0.75$	93.38	988	(0.99)	1000	105.45	987	(0.99)	1000	121.46	983	(0.98)	1000
$\lambda = 5$	$p = 0$	3.02	282	(0.28)	1000	3.24	311	(0.31)	1000	4.79	496	(0.50)	1000
	$p = 0.25$	17.87	890	(0.89)	1000	18.79	894	(0.89)	1000	20.02	893	(0.89)	1000
	$p = 0.5$	142.55	1000	(1.00)	1000	143.99	1000	(1.00)	1000	144.73	1000	(1.00)	1000
	$p = 0.75$	609.81	1000	(1.00)	1000	618.15	1000	(1.00)	1000	621.49	1000	(1.00)	1000

1. * Some iterations are dropped due to a singular OPG variance matrix.
2. $^\diamond$Some of the iterations where MLE is at the boundary ($\hat{p} = 0$ or $\hat{p} = 1$) are dropped due to not negative definite Hessian.
3. $^\circ$ One iteration dropped due to $\hat{p} = 1$

Table 1.9: Score-Based Tests, $n = 200$

		LM				Modified LM				KT			
		Mean	Rejection		Total	Mean	Rejection		Total	Mean	Rejection		Total
$\lambda = 1$	$p = 0$	1.83	99	(0.13)	776	1.47	122	(0.16)	776	1.82	146	(0.19)	776
	$p = 0.25$	1.83	101	(0.12)	840*	1.48	135	(0.16)	840*	1.83	160	(0.19)	840*
	$p = 0.5$	1.74	112	(0.13)	878	1.24	116	(0.13)	878	1.74	161	(0.18)	878
	$p = 0.75$	1.74	100	(0.12)	849*	1.07	88	(0.10)	849*	1.74	152	(0.18)	849*
$\lambda = 2$	$p = 0$	1.30	73	(0.07)	994	0.94	105	(0.11)	994	1.29	135	(0.14)	994
	$p = 0.25$	1.20	75	(0.08)	999	0.77	96	(0.10)	999	1.19	130	(0.13)	999
	$p = 0.5$	1.69	140	(0.14)	1000	0.18	16	(0.02)	1000	1.68	222	(0.22)	1000
	$p = 0.75$	3.82	422	(0.42)	999	0.04	3	(0.00)	999	3.82	531	(0.53)	999
$\lambda = 5$	$p = 0$	1.16	58	(0.06)	999+	0.71	79	(0.08)	999+	1.11	113	(0.11)	999+
	$p = 0.25$	1.78	139	(0.14)	996+	0.12	7	(0.01)	996+	1.73	216	(0.22)	996+
	$p = 0.5$	13.10	961	(0.96)	997+	0.00	0	(0.00)	997+	13.10	978	(0.98)	997+
	$p = 0.75$	29.46	996	(1.00)	997+	0.00	0	(0.00)	997+	29.46	996	(1.00)	997+

1. * Some iterations are dropped due to a singular OPG variance matrix.
2. + Some iterations dropped due to $\hat{\sigma}_v$ being too small such that $\hat{\lambda}$ is not well defined.

Table 1.10: Score-Based Tests, $n = 500$

		LM				Modified LM				KT			
		Mean	Rejection		Total	Mean	Rejection		Total	Mean	Rejection		Total
$\lambda = 1$	$p = 0$	1.39	84	(0.10)	878*	0.99	104	(0.12)	878*	1.39	130	(0.15)	878*
	$p = 0.25$	1.36	81	(0.09)	921	0.98	100	(0.11)	921	1.36	127	(0.14)	921
	$p = 0.5$	1.26	73	(0.08)	964	0.70	79	(0.08)	964	1.26	125	(0.13)	964
	$p = 0.75$	1.37	86	(0.09)	939	0.45	50	(0.05)	939	1.37	151	(0.16)	939
$\lambda = 2$	$p = 0$	1.21	76	(0.08)	1000	0.79	89	(0.09)	1000	1.20	127	(0.13)	1000
	$p = 0.25$	1.07	48	(0.05)	1000	0.62	68	(0.07)	1000	1.06	107	(0.11)	1000
	$p = 0.5$	2.59	249	(0.25)	1000	0.03	2	(0.00)	1000	2.57	370	(0.37)	1000
	$p = 0.75$	8.14	795	(0.80)	1000	0.00	0	(0.00)	1000	8.14	887	(0.89)	1000
$\lambda = 5$	$p = 0$	1.06	56	(0.06)	1000	0.62	69	(0.07)	1000	0.97	109	(0.11)	1000
	$p = 0.25$	2.97	280	(0.28)	1000	0.03	1	(0.00)	1000	2.92	415	(0.41)	1000
	$p = 0.5$	30.94	1000	(1.00)	1000	0.00	0	(0.00)	1000	30.94	1000	(1.00)	1000
	$p = 0.75$	69.77	1000	(1.00)	1000	0.00	0	(0.00)	1000	69.77	1000	(1.00)	1000

1. * Some iterations are dropped due to a singular OPG variance matrix.

Table 1.11: Model Comparison

	Basic SF			ZISF			Heteroskedasticity		
Variable	Coef	SE	t-stat	Coef	SE	t-stat	Coef	SE	t-stat
cons (β_0)	0.27	0.04	6.68	0.08	0.05	1.67	0.25	0.04	5.50
time period (θ)	0.02	0.01	2.27	0.01	0.01	2.19	0.02	0.01	2.60
area (β_1)	0.53	0.08	6.38	0.52	0.08	6.58	0.57	0.09	6.43
labor (β_2)	0.23	0.09	2.71	0.25	0.08	3.07	0.21	0.09	2.46
fertilizer (β_3)	0.20	0.05	3.95	0.21	0.05	4.54	0.19	0.05	3.83
β_{11}	−0.48	0.21	−2.27	−0.45	0.20	−2.26	−0.35	0.25	−1.39
β_{12}	0.61	0.17	3.60	0.55	0.17	3.21	0.58	0.20	2.87
β_{13}	0.06	0.15	0.43	0.08	0.13	0.62	0.01	0.15	0.07
β_{22}	−0.56	0.24	−2.33	−0.50	0.25	−2.04	−0.60	0.27	−2.21
β_{23}	−0.14	0.13	−1.04	−0.12	0.12	−1.02	−0.11	0.15	−0.78
β_{33}	−0.01	0.09	−0.08	−0.03	0.08	−0.45	0.01	0.09	0.15
σ_u	0.44	0.03	13.86	0.44	0.04	10.87			
σ_{ui}							0.42		
cons (γ_0)							−3.05	0.93	−3.27
age (γ_1)							0.03	0.01	2.06
edyrs (γ_2)							0.11	0.03	3.36
hhsize (γ_3)							0.08	0.09	0.99
nadult (γ_4)							−0.10	0.11	−0.98
banrat (γ_5)							−1.30	0.43	−2.98
σ_v	0.16	0.02	8.23	0.20	0.02	12.22	0.17	0.02	8.34
λ	2.75			2.18			2.43		
p				0.58	0.11	5.42			
p_i									
$\widehat{\text{pr}}\left[z_i = 1 \mid \widehat{\varepsilon}\right]$				0.58					
cons (δ_0)									
age (δ_1)									
edyrs (δ_2)									
hhsize (δ_3)									
nadult (δ_4)									
banrat (δ_5)									
$\widehat{E}\left[u_i \mid \widehat{\varepsilon}_i\right]$	0.35			0.15			0.33		
$\exp\left(-\widehat{E}\left[u_i \mid \widehat{\varepsilon}_i\right]\right)$	0.70			0.86			0.72		
$\widehat{E}\left[\exp\left(-u_i\right) \mid \widehat{\varepsilon}_i\right]$	0.73			0.89			0.74		
$\ln L$		−74.41			−71.88			−65.89	
# of parameter		13			14			18	
AIC		174.82			171.75			167.78	
BIC		224.75			225.52			236.91	
HQIC		194.70			193.17			195.31	

Table 1.11: (cont'd)

	Logit			Logit+Hetero.		
Variable	Coef	SE	t-stat	Coef	SE	t-stat
cons (β_0)	0.07	0.04	1.80	0.05	0.04	1.38
time period (θ)	0.02	0.01	2.57	0.02	0.01	2.60
area (β_1)	0.59	0.09	6.85	0.56	0.08	7.04
labor (β_2)	0.22	0.08	2.65	0.24	0.08	3.07
fertilizer (β_3)	0.18	0.05	3.82	0.18	0.04	3.96
β_{11}	−0.29	0.25	−1.13	−0.31	0.26	−1.19
β_{12}	0.53	0.21	2.60	0.50	0.21	2.35
β_{13}	0.00	0.14	0.02	0.02	0.14	0.16
β_{22}	−0.56	0.27	−2.08	−0.47	0.27	−1.73
β_{23}	−0.11	0.13	−0.80	−0.12	0.13	−0.91
β_{33}	0.00	0.09	−0.03	−0.01	0.09	−0.13
σ_u	0.42	0.04	10.43			
σ_{ui}				0.53		
cons (γ_0)				2.61	3.11	0.84
age (γ_1)				−0.07	0.04	−1.60
edyrs (γ_2)				−0.03	0.12	−0.28
hhsize (γ_3)				−0.28	0.22	−1.26
nadult (γ_4)				−0.02	0.23	−0.08
banrat (γ_5)				1.32	1.20	1.10
σ_v	0.20	0.01	14.15	0.21	0.01	15.84
λ	2.06			2.49		
p						
p_i	0.55			0.57		
$\widehat{\text{pr}}\left[z_i = 1 \vert \widehat{\varepsilon}\right]$	0.55			0.57		
cons (δ_0)	3.88	2.29	1.69	8.80	3.42	2.57
age (δ_1)	−0.10	0.03	−2.88	−0.16	0.05	−3.36
edyrs (δ_2)	−0.14	0.15	−0.93	−0.27	0.21	−1.30
hhsize (δ_3)	−0.44	0.23	−1.93	−0.81	0.32	−2.56
nadult (δ_4)	0.35	0.27	1.30	0.44	0.35	1.25
banrat (δ_5)	4.22	1.31	3.23	5.46	1.55	3.53
$\widehat{E}\left[u_i \vert \widehat{\varepsilon}_i\right]$	0.15			0.14		
$\exp\left(-\widehat{E}\left[u_i \vert \widehat{\varepsilon}_i\right]\right)$	0.86			0.87		
$\widehat{E}\left[\exp\left(-u_i\right) \vert \widehat{\varepsilon}_i\right]$	0.88			0.89		
$\ln L$		−59.90			−54.33	
# of parameter		19			24	
AIC		157.79			156.67	
BIC		230.76			248.84	
HQIC		186.86			193.38	

CHAPTER 2

HETEROSKEDASTICITY AUTOCORRELATION ROBUST INFERENCE IN TIME SERIES REGRESSIONS WITH MISSING DATA

2.1 INTRODUCTION

It is not unusual to encounter a time series data set with missing observations. Most of the times series literature in dealing with missing data focuses on the estimation of dynamic models where the goal is to forecast missing observations. However, in the relatively simple context of time series regression, there appears to be a sparsity of work related to missing data issues. In particular, little appears to be known about the impact of missing data on heteroskedasticity autocorrelation (HAC) robust tests in regression settings. This chapter attempts to fill this void by analyzing the impact of missing data on robust tests based on nonparametric kernel estimators of long run variances. Following Kiefer and Vogelsang (2005) we focus on obtaining fixed-b results for the robust tests. In addition to capturing the impact of the long run variance estimator's kernel and bandwidth on the robust test statistics, the fixed-b limits also capture the impact of the locations of the missing data on the robust test statistics when either the missing process is non-random or one conditions on the missing locations. In situations where the more traditional approach that seeks to obtain consistency results for variance estimators would be problematic, fixed-b theory delivers useful approximations.

Following Parzen (1963) we characterize missing observations as being driven by a missing process that is a 0-1 binary variable. In terms of a regression model, the Parzen (1963) approach amounts to plugging in zeros for missing observations. Time series with zeros in place of missing data have been labeled amplitude modulated series by Parzen (1963) which we adopt throughout this chapter. Because of the zeros, amplitude modulated series are intuitively sensible because the time distances between the observations remain preserved. This would seem particularly relevant for HAC robust testing based on nonparametric kernel estimator (Newey and West (1987) and Andrews (1991)) given that those estimators employ quadratic forms with weights that depend on the time distances of pairs of observations.

Soon after Parzen (1963) introduced the notion of modeling missing data with the amplitude modulated series approach, many authors investigated the impact of missing data on the consistent estimation of spectral density functions. For example, Scheinok (1965) and Bloomfield (1970) consider estimating a spectral density function of the observed process (with missing data) with independent Bernoulli and dependent Bernoulli missing processes respectively. Neave (1970) estimates a spectral density function with initially scarce data. Later work by Dunsmuir and Robinson (1981) investigated the consistent estimation of the spectral density of the underlying latent process. While HAC robust inference makes use of spectral estimation method, with the exception of a recent working paper by Datta and Du (2012), there appears to be no attempt in the literature to link this earlier literature on spectral density estimation with regression inference in the case of missing data.

Datta and Du (2012) used the amplitude modulated series approach to investigate robust inference in time series regression settings. Their approach is based on traditional asymptotic theory for HAC robust tests which appeals to the consistency of the HAC estimators. In the case of non-random missing locations, the traditional approach becomes complicated because of the need to consistently estimate the long run variance of the latent process. While this is possible using results in Dunsmuir and Robinson (1981), it is not clear how to obtain a positive definite variance estimator. In any case, given that it is now well established that fixed-b theory provides better approximations than the traditional approach (see Jansson (2002), Sun, Phillips, and Jin (2008), and Gonçalves and Vogelsang (2011)), obtaining fixed-b results for the missing data case is prudent.

There are three main theoretical findings in this chapter. First, when the missing process is random and satisfies strong mixing conditions, HAC robust t and $Wald$ statistics computed from the amplitude modulated series follow the usual fixed-b limits as in Kiefer and Vogelsang (2005). Second, when the missing process is non-random, the fixed-b limits depend on the locations of missing observations but are otherwise pivotal. Therefore, the fixed-b critical values that one would use in the amplitude modulated series approach depends on whether the missing process is best viewed as random or non-random. Third, a seemingly naive alternative to the amplitude modulated series approach is to simply ignore the missing data. One might reasonably conjecture that ignoring the missing data would be problematic for robust inference. Surprisingly we find that the fixed-b limits of the robust t and $Wald$ statistics have the standard fixed-b random variable whether the missing process is random or non-random. Here, ignoring the problem (missing data) has no negative consequences and generates the advantage of robustness to whether the missing process is random or non-random.

The rest of this chapter is organized as follows. Section 2.2 defines the model and the amplitude modulated series test statistics in the presence of missing data. Section 2.3 develops fixed-b asymptotic results for the amplitude modulated series test statistics for both random and non-random missing processes. Because the random and non-random missing processes require different regularity conditions they are treated separately. Simulation of the asymptotic critical values is discussed with a focus on bootstrap methods. Following Gonçalves and Vogelsang (2011), we find that the naive *i.i.d.* bootstrap is a particularly good option for obtaining valid fixed-b critical values. Finite sample performance of the amplitude modulated series tests for both random and non-random missing processes are examined in Section 2.4 by Monte Carlo simulations. Attention is focused on the relative performance of simulated asymptotic critical values with bootstrap critical values. Section 2.5 analyzes the approach of ignoring missing observations and makes some comparisons with the amplitude modulated series approach. Section 2.6 concludes and formal proofs are given in the Appendices B-D.

2.2 MODEL AND TEST STATISTICS

Consider a regression model without missing observations,

$$y_t^* = x_t^{*\prime} \beta + u_t^* \qquad (t = 1, 2, \ldots, T), \tag{2.1}$$

where β is a $(k \times 1)$ vector of regression parameters, x_t^* is a $(k \times 1)$ vector of regressors, and u_t^* is a mean zero random process. When there are missing observations, (2.1) is the underlying ***latent process***.

In the presence of missing observations, we characterize the missing process as a binary variable. Let $\{a_t\}$ be a ***missing process*** where

$$a_t = \begin{cases} 1 & \text{data is observed at time } t \\ 0 & \text{data is missing at time } t. \end{cases}$$

Whether we treat the missing process as non-random or random depends on the structure of the data and the reason why the observations are missing. We consider both stochastic and non-stochastic missing processes.

With the missing process, $\{a_t\}$, we define the regression model with missing observations as[1]

$$y_t = x_t' \beta + u_t, \quad y_t = a_t y_t^*, \, x_t = a_t x_t^*, \, u_t = a_t u_t^*, \quad (t = 1, 2, \ldots, T). \tag{2.2}$$

Characterizing the missing process as a 0-1 binary variable and constructing regression model as (2.2) is one of the standard approaches of treating missing observations in panel data regression models. In time series, Parzen (1963) first characterized time series with missing data using a dummy variable and modeled observed process as (2.2). However this approach has not become standard in time series which is surprising because (2.2) can be thought of as a natural way of formulating a regression model with missing observations when there is no particular interest in forecasting the missing data. Model (2.2) is

[1] For simplicity, we assume that the dependent variable and the independent variables are missing at the same time points.

intuitively sensible. Because the zeros are plugged in for missing observations, the true time distances between observations are preserved. At a conceptual level this would appear important when we are using nonparametric kernel covariance matrix estimators. Parzen (1963) labelled the time series in model (2.2) as *amplitude modulated (AM) series* because the original time series are amplitude modulated by the missing process $\{a_t\}$. We adopt the same language here.

Throughout our analysis we assume that the latent regression model satisfies exogeneity, $E(x_t^* u_t^*) = 0$, and we assume that the mechanism generating the missing data does not generate an endogeneity problem, i.e. we assume that $E(x_t u_t) = 0$ or equivalently $E(a_t x_t u_t^*) = 0$. This allows us to focus on the impact of missing data on robust inference assuming that β is identified by the observed data.

The focus of this chapter is on inference regarding β based on the ordinary least squares (OLS) estimator of β. Inference is carried out to be robust to the form of the heteroskedasticity and serial (auto) correlation. The OLS estimator of β is given by

$$\hat{\beta} = \left(\sum_{t=1}^{T} x_t x_t' \right)^{-1} \sum_{t=1}^{T} x_t y_t.$$

Plugging in for y_t gives the well known expression

$$\begin{aligned} \hat{\beta} - \beta &= \left(\sum_{t=1}^{T} x_t x_t' \right)^{-1} \sum_{t=1}^{T} x_t u_t \\ &= \left(\sum_{t=1}^{T} x_t x_t' \right)^{-1} \sum_{t=1}^{T} v_t \end{aligned}$$

where $v_t = x_t u_t$. The impact of serial correlation on $\hat{\beta}$ comes through v_t and robust standard errors can be obtained using a nonparametric kernel estimator of the asymptotic variance of $T^{-1/2} \sum_{t=1}^{T} v_t$ of the form

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

where $\hat{\Gamma}_j = T^{-1} \sum_{t=j+1}^{T} \hat{v}_t \hat{v}'_{t-j}$ are the sample autocovariances of $\hat{v}_t = x_t \hat{u}_t$ with $\hat{u}_t = y_t - x'_t \hat{\beta}$ the OLS residuals of the AM series, and $\hat{\Gamma}_j = \hat{\Gamma}'_{-j}$ for $j < 0$. Here, $k(x)$ is a kernel function such that $k(x) = k(-x)$, $k(0) = 1$, $|k(x)| \leq 1$, $k(x)$ is continuous at $x = 0$, $\int_{-\infty}^{\infty} k^2(x) < \infty$, and M is the bandwidth parameter. Notice that $\hat{\Omega}$ is the usual long run variance estimator that is obtained after simply setting $\hat{v}_t = 0$ for any dates for which data is missing. This can be seen mechanically by noting that $\hat{v}_t = x_t \hat{u}_t = a_t x_t^* \hat{u}_t$. Using well known algebra, we can rewrite $\hat{\Omega}$ as

$$\hat{\Omega} = T^{-1} \sum_{t=1}^{T} \sum_{s=1}^{T} k\left(\frac{t-s}{M}\right) \hat{v}_t \hat{v}'_s. \tag{2.3}$$

Because $\hat{\Omega}$ is computed using the AM series, the time distances, $|t-s|$, between observed data points are preserved which is conceptually sensible. In addition, $\hat{\Omega}$ will be positive definite with appropriate choices of the kernel function, e.g. the Bartlett, Parzen or quadratic spectral (QS) kernels.

Suppose we are interested in testing the null hypothesis, $H_0 : r(\beta_o) = 0$ against $H_A : r(\beta_o) \neq 0$, where $r(\beta)$ is a $q \times 1$ vector $(q \leq k)$ of continuously differentiable functions with a first derivative matrix $R(\beta) = \partial r(\beta) / \partial \beta'$. We analyze the following Wald statistic,

$$W_T = Tr(\hat{\beta})' \left[R(\hat{\beta}) \hat{Q}^{-1} \hat{\Omega} \hat{Q}^{-1} R(\hat{\beta})' \right]^{-1} r(\hat{\beta}),$$

where $\hat{Q} = T^{-1} \sum_{t=1}^{T} x_t x_t'$. The case where one restriction is being tested, $q = 1$, we can also use a t-statistic of the form

$$t_T = \frac{\sqrt{T} r(\hat{\beta})}{\sqrt{R(\hat{\beta}) \hat{Q}^{-1} \hat{\Omega} \hat{Q}^{-1} R(\hat{\beta})'}}.$$

2.3 ASSUMPTIONS AND ASYMPTOTIC THEORY

In this section we derive the asymptotic behavior of the OLS estimator, $\hat{\beta}$, the HAC estimator, $\hat{\Omega}$, and the HAC robust wald test, W_T, defined in Section 2.2 for the case of weakly dependent covariance stationary time series. We present results for both random and non-random missing processes. Results for random and non-random missing processes are treated separately as they require different regularity conditions. We first state results for the random missing process followed by results for the non-random missing process. Although we briefly discuss traditional asymptotic theory for HAC robust tests based on consistency of the HAC estimators, we are mainly interested in obtaining fixed-b asymptotic approximations as proposed by Kiefer and Vogelsang (2005). In the fixed-b asymptotic framework the bandwidth of the covariance matrix estimator is modeled as a fixed proportion, b, of the sample size. This is in contrast to the traditional approach where the bandwidth is modeled as increasing slower than the sample size. The advantage of the fixed-b approach is that the resulting asymptotic approximations for the test statistics depend on the choice of kernel and bandwidth. In the traditional approach the kernel and bandwidth choice do not appear in the asymptotic approximation. The fixed-b approach is therefore more accurate than the traditional approach because the fixed-b approach captures much of the impact of the sampling distribution of the HAC estimator on the test statistic. For theoretical evidence on the superior performance of the fixed-b approach see Jansson (2002), Sun, Phillips, and Jin (2008), and Gonçalves and Vogelsang (2011).

Because the fixed-b asymptotic distributions depend on the kernels used to compute the HAC estimators, some random matrices that appear in the asymptotic results need to be defined. Here we follow the notation of Kiefer and Vogelsang (2005).

Definition 1. *Let $h > 0$ be an integer and $B_h(r)$ denote a generic $h \times 1$ vector of stochastic processes. Let the random matrix, $P(b, B_h)$, be defined as follows for $b \in (0,1]$.*

Case (i) : if $k(x)$ is twice continuously differentiable everywhere,

$P(b, B_h) \equiv \int_0^1 \int_0^1 \frac{1}{b^2} k'' \left(\frac{r-s}{b} \right) B_h(r) B_h(s)' drds,$

Case (ii) : if $k(x)$ is continuous, $k(x) = 0$ for $|x| \geq 1$ and $k(x)$ is twice continuously differentiable everywhere except for $|x| = 1$,

$P\left(b,B_h\right)\equiv\int\int_{|r-s|<b}\frac{1}{b^2}k''\left(\frac{r-s}{b}\right)B_h\left(r\right)B_h\left(s\right)'drds$

$+\frac{k_-(1)'}{b}\int_0^{1-b}\left(B_h\left(r+b\right)B_h\left(r\right)'+B_h\left(r\right)B_h\left(r+b\right)'\right)dr,$

where $k_-\left(1\right)'=\lim_{h\to0}\left[\left(k\left(1\right)-k\left(1-h\right)\right)/h\right]$,

Case(iii) : if $k\left(x\right)$ *is the Bartlett kernel,*

$P\left(b,B_h\right)\equiv\frac{2}{b}\int_0^1 B_h\left(r\right)B_h\left(r\right)'dr-\frac{1}{b}\int_0^{1-b}\left(B_h\left(r+b\right)B_h\left(r\right)'+B_h\left(r\right)B_h\left(r+b\right)'\right)dr.$

Throughout, the symbol "$\Rightarrow$" denotes weak convergence of a sequence of stochastic processes to a limiting stochastic process and $\xrightarrow{p}$ denotes convergence in probability. We also use the following notation. Let $Q=E(x_tx_t')$ and $Q^*=E(x_t^*x_t^{*\prime})$. Let $v_t^*=x_t^*u_t^*$ and define $\Omega^*=\Lambda^*\Lambda^{*\prime}=\Gamma_0^*+\sum_{j=1}^{\infty}(\Gamma_j^*+\Gamma_j^{*\prime})$, where $\Gamma_j^*=E(v_t^*v_{t-j}^{*\prime})$ and Λ^* is the lower triangular matrix based on the Cholesky decomposition of Ω^*. Similarly, let and $v_t=a_tv_t^*$ and define $\Omega=\Lambda\Lambda'=\Gamma_0+\sum_{j=1}^{\infty}(\Gamma_j+\Gamma_j')$, where $\Gamma_j=E(v_tv_{t-j}')$ and Λ is the lower triangular matrix based on the Cholesky decomposition of Ω. The matrix Ω^* is the long run variance matrix of the latent vector v_t^* whereas Ω is the long run variance matrix of the AM series vector v_t.

We derive results under the assumptions that the latent processes are near epoch dependent (NED) on some underlying mixing process and that the missing process is strong mixing. We follow the definitions in Davidson (2002). Let the L_p norm of x be defined as $\|x\|_p=(E|x|^p)^{\frac{1}{p}}$. Also, let $|\bullet|$ denote the Euclidean norm of the corresponding vector or matrix. For a stochastic sequence $\{\varepsilon_t\}_{-\infty}^{\infty}$, on a probability space $(\Omega,\mathcal{F},P)$, let $\mathcal{F}_{t-m}^{t+m}=\sigma(\varepsilon_{t-m},\ldots,\varepsilon_{t+m})$, such that $\{\mathcal{F}_{t-m}^{t+m}\}_{m=0}^{\infty}$ is an increasing sequence of σ-fields. We say that a sequence of integrable random variables $\{w_t\}_{-\infty}^{\infty}$ is L$_p$-NED on $\{\varepsilon_t\}_{-\infty}^{\infty}$ if, for $p>0$, $\|w_t-E(w_t|\mathcal{F}_{t-m}^{t+m})\|_p<d_t\nu_m$, where $\nu_m\to0$ and $\{d_t\}_{-\infty}^{\infty}$ is a sequence of positive constants. For a sequence $\{a_t\}_{-\infty}^{\infty}$, let $\mathcal{F}_{-\infty}^{t}=\sigma(\ldots,a_{t-1},a_t)$, and similarly define $\mathcal{F}_{t+m}^{\infty}=\sigma(a_{t+m},a_{t+m+1},\ldots)$. The sequence is said to be α-mixing if $\lim_{m\to\infty}\alpha_m=0$, where $\alpha_m=\sup_t\sup_{G\in\mathcal{F}_{-\infty}^{t},H\in\mathcal{F}_{t+m}^{\infty}}|P(G\cap H)-P(G)P(H)|$. A sequence is α-mixing of size $-\psi_0$ if $\alpha_m=O(m^{\psi})$ for some $\psi>\psi_0$. Similarly, a sequence is L_p-NED of size $-\phi_0$ if $\nu_m = O(m^{\phi})$ for some $\phi > \phi_0$.

2.3.1 Random Missing Process

When the missing process is random, the asymptotic theory is driven by the observed AM series. If the AM series satisfies conditions required for fixed-b asymptotic theory, then the HAC estimator and the robust Wald test of the null hypothesis follow the usual fixed-b asymptotic limits as obtained by Kiefer and Vogelsang (2005).

The following high-level assumptions are sufficient for this purpose.

Assumption R.

1. $T^{-1}\sum_{t=1}^{[rT]} x_t x_t' \Rightarrow rQ, \ \forall r \in [0,1]$.
2. $T^{-1/2}\sum_{t=1}^{[rT]} v_t \Rightarrow \Lambda \mathcal{W}_k(r), \ \forall r \in [0,1]$.

Assumption R1 states that a uniform (in r) law of large numbers (LLN) holds for $\{x_t x_t'\}$. As long as $\{x_t\}$ is covariance stationary and weakly dependent, this assumption is a fairly general condition. Assumption R2 states that a functional central limit theorem (FCLT) holds for the normalized partial sum of the AM series $\{v_t\}$. Below Assumption R', which is in terms of the latent process and the missing process rather than the AM series itself, is sufficient for Assumption R.

Assumption R$'$.

1. *For some* $r > 2$, $\left\| x_t^* \right\|_{2r} \leq \Delta < \infty$ *for all* $t = 1, \ldots$.
2. $\left\{ x_t^* \right\}$ *is a weakly stationary sequence* L_2*−NED on* $\{\varepsilon_t\}$ *with NED coefficient of size* $-\frac{2(r-1)}{r-2}$.
3. $\left\| v_t^* \right\|_r \leq \Delta < \infty$, *and* $E(v_t^*) = 0$ *for all* $t = 1, 2, \ldots$.
4. $\left\{ v_t^* \right\}$ *is a mean zero weakly stationary sequence* L_2*-NED on* $\{\varepsilon_t\}$ *with NED coefficient of size* $-\frac{1}{2}$.

5. *$\{(a_t, \varepsilon_t)\}$ is a α-mixing sequence with α−mixing coefficient of size $-\frac{2r}{r-2}$.*

6. *$\{a_t\}$ is a weakly stationary process that is independent of $\left\{\left(x_t^*, u_t^*\right)\right\}$.*

7. *$\Omega = \lim_{T\to\infty} Var\left(T^{-1/2}\sum_{t=1}^{T} a_t v_t^*\right)$ is positive definite.*

Under Assumption R$'$, the latent process satisfies conditions sufficient for the fixed-b asymptotic theory to go through. In particular, under Assumption R', for all $r \in (0,1]$, $T^{-1}\sum_{t=1}^{[rT]} x_t^* x_t^{*\prime} \Rightarrow rQ^*$ and $T^{-1/2}\sum_{t=1}^{[rT]} v_t^* \Rightarrow \Lambda^* \mathcal{W}_k(r)$. In terms of the latent process, Assumptions R' are relatively weak. For example, Phillips and Durlauf (1986) states that if $\{v_t^*\}$ is $L_{2+\delta}$ bounded stationary process ($\delta > 0$) and strong mixing then the partial sums of $\{v_t^*\}$ satisfies the FCLT. The L_2-NED condition in Assumption R$'$4 is actually weaker than this condition. Hence, the presence of the missing observations generally does not require additional assumptions on the latent process. Assumption R$'$6 is relatively strong and states that the missing locations are not related to the latent process. This assumption is sufficient for $\hat{\beta}$ to be consistent for β because it implies if $E(x_t^* u_t^*) = 0$, then $E(x_t u_t) = E(a_t x_t^* u_t^*) = E(a_t)E(x_t^* u_t^*) = 0$. In addition, Assumptions R$'$5 and R$'$6 ensure that the LLN and FCLT that hold for the latent processes extend to the observed AM series, i.e. Assumptions R$'$5 and R$'$6 ensure that Assumptions R hold.

With these assumptions we can state our main result for the estimator and statistics based on the AM series when the missing process is random.

Theorem 2.1. *Under Assumption R' the following hold as $T \to \infty$.*

(a). (Asymptotic Behavior of OLS)

$$\sqrt{T}\left(\hat{\beta} - \beta\right) \Rightarrow Q^{-1}\Lambda\mathcal{W}_k(1) = N(0, Q^{-1}\Omega Q^{-1}).$$

(b). (Fixed-b approximation of HAC estimator) Let $\tilde{B}_k(r)$ denote a $k \times 1$ vector of stochastic processes defined as $\tilde{B}_k(r) \equiv \mathcal{W}_k(r) - r\mathcal{W}_k(1)$, for all $r \in (0,1]$. Assume $M = bT$ where $b \in (0,1]$ is fixed. Then,

$$\hat{\Omega} \Rightarrow \Lambda P\left(b, \tilde{B}_k\right)\Lambda',$$

where the form of $P(b, \tilde{B}_k)$ depends on the type of kernel via Definition 1.

(c). (Fixed-b asymptotic distribution of tests) Under H_0,

$$W_T \Rightarrow \mathcal{W}_q(1)' P\left(b, \tilde{B}_q\right)^{-1} \mathcal{W}_q(1)$$

and when $q = 1$,

$$t_T \Rightarrow \frac{\mathcal{W}_1(1)}{\sqrt{P\left(b, \tilde{B}_1\right)}}.$$

Because Assumptions R$'$ imply Assumptions R, Theorem 2.1 directly follows from Kiefer and Vogelsang (2005) although directly establishing Theorem 2.1 (a) is easy. If we plug in $y_t = x_t'\beta + u_t$ to the OLS estimator $\hat{\beta}$, then $\hat{\beta} = \beta + \left(\sum_{t=1}^{T} x_t x_t'\right)^{-1} \sum_{t=1}^{T} v_t$. Therefore we can write

$$\sqrt{T}(\hat{\beta} - \beta) = \left(T^{-1} \sum_{t=1}^{T} x_t x_t'\right)^{-1} T^{-1/2} \sum_{t=1}^{T} v_t.$$

and the limit is obtained by using Assumptions R with $r = 1$. While the fixed-b approximation is more useful than the traditional result that relies on a consistency result for $\hat{\Omega}$, one could easily obtain traditional results for the Wald and t-statistics under similar regularity conditions.[2]

Theorem 2.1 shows that when the missing process is random, one can simply plug in zeros for the missing observations and conduct standard fixed-b inference treating the zeros as though they were observed data. Given a particular sample with T time periods of data (including the zeros), rejections would be computed relative to fixed-b critical values obtained by Kiefer and Vogelsang (2005). The critical values are functions of the kernel and the value of $b = M/T$ where M is the bandwidth used to compute $\hat{\Omega}$.

The fixed-b asymptotic distributions in Kiefer and Vogelsang (2005) are non-standard. While it is relatively easy to simulate from the asymptotic distributions, more user-friendly

[2]Consistetncy of $\hat{\Omega}$ requires a slightly stronger assumption than Assumption R$'$. For example, Andrews (1991) requires $\{v_t^*\}$ to be a fourth order stationary process, and Hansen (1992) requires $\{v_t^*\}$ to be mixing with size $-(2+\delta)(r+\delta)/2(r-2)$. Assumption R$''$ in this section is sufficient for Hansen (1992). See Appendix B for the proof.

methods are available for the computation of critical values and p-values. For the case of the Bartlett kernel, Vogelsang (2012) has developed a numerical method for the easy computation of standard fixed-b critical values and p-values for any significance level. For other kernels comprehensive numerical approaches have not been developed. Kiefer and Vogelsang (2005) do provide critical value functions for popular significance levels but their functions do not allow the computation of p-values. A good alternative for the computation of fixed-b critical values and p-values is the bootstrap. Gonçalves and Vogelsang (2011) showed that the naive moving block bootstrap has the same limiting distribution as the fixed-b asymptotic distribution under regularity conditions similar to those used here. The bootstrap works with both a fixed block length (l) or a block length that increases with the sample size but at a slower rate ($l^2/T \to 0$). In particular, for the case of $l = 1$ the block bootstrap becomes an *i.i.d.* bootstrap. Therefore, the results of Gonçalves and Vogelsang (2011) indicate that valid fixed-b critical values can be obtained via a simple *i.i.d.* bootstrap method.

As shown in the next subsection, the fixed-b limit of the robust statistics becomes more complicated under the assumption that the missing locations are non-random. In this case the bootstrap becomes the ideal tool for obtaining fixed-b critical values on a case by case basis in practice. Therefore, it is useful to provide some details on the implementation of the bootstrap. Define the vector $\omega_t = (y_t, x_t')'$ that collects dependent and explanatory variables. Let $l \in \mathbb{N}(1 \leq l \leq T)$ be a 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. Draw $k_0 = T/l$ blocks randomly with replacement from the set of overlapping blocks $\{B_{1,l}, \ldots, B_{T-l+1,l}\}$ to obtain a bootstrap resample denoted as $\omega_t^\bullet = (y_t^\bullet, x_t^{\bullet\prime})'$, $t = 1, \ldots, T$. Notice that we are resampling from the AM series (the zeros are included). The bootstrap test statistics, $W_T^\bullet$ and $t_T^\bullet$, are defined as

$$W_T^\bullet = \left(r(\hat{\beta}^\bullet) - r(\hat{\beta})\right)' \left[TR(\hat{\beta}^\bullet)\hat{Q}^{\bullet-1}\hat{\Omega}^\bullet\hat{Q}^{\bullet-1}R(\hat{\beta}^\bullet)'\right]^{-1} \left(r(\hat{\beta}^\bullet) - r(\hat{\beta})\right)$$

and

$$t_T^\bullet = \frac{\sqrt{T}\left(r(\hat\beta^\bullet) - r(\hat\beta)\right)}{\sqrt{R(\hat\beta^\bullet)\hat Q^{\bullet-1}\hat\Omega^\bullet \hat Q^{\bullet-1} R(\hat\beta^\bullet)'}},$$

where

$$\hat Q^\bullet = T^{-1}\sum_{t=1}^{T} x_t^\bullet x_t^{\bullet\prime},$$
$$\hat\Omega^\bullet = T^{-1}\sum_{t=1}^{T}\sum_{t=1}^{T} k\left(t-s/M\right)\hat v_t^\bullet \hat v_s^{\bullet\prime},$$

and $\hat\beta^\bullet$ is the OLS estimate from the regression of $y_t^\bullet$ on $x_t^\bullet$, and $\hat v_t^\bullet = x_t^\bullet\left(y_t^\bullet - x_t^{\bullet\prime}\hat\beta^\bullet\right)$. Notice that bootstrap statistics use the same formula as W_T and t_T and this is what makes this bootstrap approach "naive". Let $p^\bullet$ denote the probability measure induced by the bootstrap resampling conditional on a realization of the original time series. If $T^{-1}\sum_{t=1}^{[rT]} x_t^\bullet x_t^{\bullet\prime} \Rightarrow^{p^\bullet} rQ^\bullet$ for some $Q^\bullet$ and $T^{-1/2}\sum_{t=1}^{[rT]} v_t^\bullet \Rightarrow^{p^\bullet} \Lambda^\bullet \mathcal{W}_k(r)$ for some $\Lambda^\bullet$, then because the fixed-b asymptotic distribution of the Wald test statistics is pivotal, the limiting distribution of $W_T^\bullet$ coincides with the limiting distribution of W_T, independently of $\Lambda^\bullet$ and $Q^\bullet$. We show that strengthening Assumption R'3-5 to R''3-5 is sufficient for this purpose.

Assumption R$''$.

3. $\left\|v_t^*\right\|_{r+\delta} < \infty, r > 2.$
4. $\left\{v_t^*\right\}$ *is a weakly stationary* $L_{2+\delta}-$*NED on* $\{\varepsilon_t\}$ *with* ν_m *of size* -1.
5. $\{(a_t, \varepsilon_t)\}$ *is a α-mixing sequence with* α_m *of size* $-\frac{(2+r)(r+\delta)}{r-2}$.

This strengthening is necessary for bootstrap resamples to satisfy conditions required for the FCLT. Also note that except for the assumptions related to the missing process, the other assumptions are identical to the those in Gonçalves and Vogelsang (2011), which implies that the existence of missing observations does not change the assumptions required for the latent process for the bootstrap to provide valid critical values. (See Gonçalves and

Vogelsang (2011, p764-766) for details.) Hence, in general, as long as the missing process satisfies the strong mixing condition, the naive moving block bootstrap provides valid critical values. We formally state the result below. Proofs are provided in the Appendix B.

Theorem 2.2. *Let $W_T^{\bullet}$ and $t_T^{\bullet}$ be naive bootstrap test statistics obtained from the moving block bootstrap resamples. Suppose that the block size l is either fixed as $T \to \infty$ or $l \to \infty$ as $T \to \infty$ such that $l^2/T \to 0$. Let $b \in (0, 1]$ be fixed and suppose $M = bT$. Then, under Assumption R' with Assumption R'3-5 strengthened to Assumption R''3-5, as $T \to \infty$,*

$$W_T^{\bullet} \overset{p^{\bullet}}{\Rightarrow} \mathcal{W}_q'(1) P(b, \tilde{B}_q(r))^{-1} \mathcal{W}_q(1)$$

and

$$t_T^{\bullet} \overset{p^{\bullet}}{\Rightarrow} \frac{\mathcal{W}_1(1)}{\sqrt{P(b, \tilde{B}_1(r))}}.$$

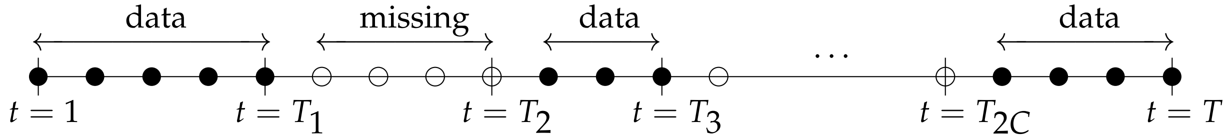
2.3.2 Non-random Missing Process

When the missing process is non-random, missing locations are fixed and hence the asymptotic behavior of the estimators and statistics depend on the locations of missing observations. We first define the structure of the timing of the missing observations.

Definition 2. *We characterize an arbitrary data set with missing observations as follows. From $t = 1$ to $t = T_1$ we observe data, from $t = T_1 + 1$ to $t = T_2$ data are missing, from $t = T_2 + 1$ to $t = T_3$ we observe data and so forth. Let the number of missing clusters be $C < \infty$. For simplicity, we assume that data are observed at $t = 1$ and $t = T$.*[3] *Thus, in general for $n = 1, \ldots, C$, from $t = T_{2n-1} + 1$ to $t = T_{2n}$ data are missing whereas from $t = T_{2n} + 1$ to $t = T_{2n+2}$ data are observed (see Figure 2.1). For notational purposes, let $T_0 = 0$ and $T_{2C+1} = T$.*

[3] This assumption is only for notational simplicity. The results of this chapter go through without this assumption.

Figure 2.1: Data with missing observations



When the missing process is non-random, the asymptotic theory is driven by the latent process. This is because the latent process is the only random process and what matters is whether the latent process satisfies conditions required for fixed-b asymptotic theory. The following assumptions are sufficient for us to obtain a fixed-b result.

Assumption NR.

1. *The missing/observed cutoffs satisfy* $\lim_{T\to\infty} \frac{T_n}{T} = \lambda_n$, $n = 0, 1, \ldots, 2C+1$, *where the number of cutoffs is non-random and finite, i.e.,* $C < \infty$.

2. $T \sum_{t=1}^{[rT]} x_t^* x_t^{*\prime} \Rightarrow rQ^*, \ \forall r \in [0,1]$.

3. $T^{-1/2} \sum_{t=1}^{[rT]} v_t^* \Rightarrow \Lambda^* \mathcal{W}_k(r), \ \forall r \in [0,1]$.

Assumption NR1 treats the number of observations in a missing or observed block as a fixed proportion of the sample size with the number of missing blocks also fixed. This is not meant to be a description of the way data is gathered but is simply a natural mathematical tool for obtaining approximations that depend on the locations of the missing and observed data. The total number of observed time periods is given by $\sum_{t=1}^{T} a_t$ and using Assumption NR1 we can quantify the proportion of the time periods that have observed data as

$$\lambda = \lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} a_t = \sum_{i=1}^{2C+1} (-1)^{i+1} \lambda_i. \tag{2.4}$$

Assumption NR2 states that a uniform (in r) LLN holds for $\{x_t^* x_t^{*\prime}\}$. Assumption NR3 states that the FCLT holds for the scaled partial sums of $\{v_t^*\}$. We now state more primitive conditions that are sufficient for Assumptions NR to hold:

Assumption NR′.

1. *For some* $r > 2$, $\left\|x_t^*\right\|_{2r} \leq \Delta < \infty$ *for all* $t = 1, \ldots$.

2. $\left\{x_t^*\right\}$ *is a weakly stationary sequence* L_2-*NED on* $\{\varepsilon_t\}$ *with NED coefficient of size* $-\frac{2(r-1)}{r-2}$.

3. $\left\|v_t^*\right\|_r \leq \Delta < \infty$, *and* $E(v_t^*) = 0$ *for all* $t = 1, 2, \ldots$.

4. $\left\{v_t^*\right\}$ *is a mean zero weakly stationary sequence* L_2*-NED on* $\{\varepsilon_t\}$ *with NED coefficient of size* $-\frac{1}{2}$.

5. $\{\varepsilon_t\}$ *is a* α*-mixing sequence with* $\alpha-$*mixing coefficient of size* $-\frac{2r}{r-2}$.

6. $\{a_t\}$ *is a non-random process.*

7. $\Omega^* = \lim_{T\to\infty} Var\left(T^{-1/2}\sum_{t=1}^{T} v_t^*\right)$ *is positive definite.*

Assumption NR$'$ is the same as Assumption R' except for the properties related to the missing process $\{a_t\}$. Recalling that in terms of the latent process, all that Assumption R$'$ required was the conditions sufficient for the latent process to satisfy fixed-b asymptotic theory, this is natural.

We now state our main results when the missing process is non-random. Note that for two numbers r and s, $r \wedge s$ denotes the minimum of r and s. The proof of Theorem 2.3 is given in Appendix C.

Theorem 2.3. *Let* $\overline{\mathcal{W}}_k \equiv \sum_{j=1}^{2C+1}(-1)^{j+1}\mathcal{W}_k\left(\lambda_j\right)$ *and let* $\breve{B}_k\left(r, \{\lambda_i\}\right)$ *be a* $k \times 1$ *vector of stochastic processes defined as* $\breve{B}_k\left(r, \{\lambda_i\}\right) \equiv \sum_{n=0}^{C}\mathbb{1}\left\{\lambda_{2n} < r \leq \lambda_{2(n+1)}\right\}\sum_{j=1}^{2n+1}(-1)^{j+1}\left(\mathcal{W}_k\left(r\wedge\lambda_j\right) - \left(r\wedge\lambda_j\right)\lambda^{-1}\overline{\mathcal{W}}_k\right)$, *for* $r \in (0,1]$. *Under Assumption NR*$'$, *as* $T \to \infty$,

(a). (Asymptotic Behavior of $\hat{\beta}$*)*

$$\sqrt{T}\left(\hat{\beta} - \beta\right) \Rightarrow \left(\lambda Q^*\right)^{-1}\Lambda^*\overline{\mathcal{W}}_k = N\left(0, \lambda^{-1}Q^{*-1}\Omega^*Q^{*-1}\right)$$

(b). (Fixed-b asymptotic approximation of $\hat{\Omega}$*) Assume* $M = bT$ *where* $b \in (0,1]$ *is fixed; then*

$$\hat{\Omega} \Rightarrow \Lambda^* P\left(b, \breve{B}_k(\{\lambda_i\})\right)\Lambda^{*\prime}.$$

(c). (Fixed-b asymptotic distribution of W_T) Under H_0,

$$W_T \Rightarrow \overline{\mathcal{W}}_q' P\left(b, \breve{B}_q(\{\lambda_i\})\right)^{-1} \overline{\mathcal{W}}_q$$

and when $q = 1$,

$$t_T \Rightarrow \frac{\overline{\mathcal{W}}_1}{\sqrt{P\left(b, \breve{B}_1(\{\lambda_i\})\right)}}.$$

Using the asymptotic normality result in Theorem 2.3 (a), one could pursue a traditional inference approach which would require a consistent estimator of the asymptotic variance. The challenge would be constructing a consistent estimator of the latent process long run variance matrix, Ω^*. Using results from Dunsmuir and Robinson (1981) a consistent estimator of Ω^* can be constructed as

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

where $\hat{\Gamma}_j^* = \sum_{t=j+1}^{T} \hat{v}_t \hat{v}_{t-j}' / \sum_{t=j+1}^{T} a_t a_{t-j}$. Because $\hat{\Gamma}_j^*$ is constructed using the effective sample size of the sequence $\{\widehat{v}_t \widehat{v}_{t-j}\}$ there is no guarantee that $\hat{\Omega}^*$ will be positive definite even if kernels like the Bartlett, Parzen and QS are used. Besides only providing a relatively crude approximation for test statistics, the difficulty in constructing a positive definite estimator of $\hat{\Omega}^*$ makes the traditional approach even less appealing in practice.

In contrast the fixed-b approach shows that one can simply use $\hat{\Omega}$ to construct valid test statistics because under fixed-b theory, $\hat{\Omega}$ is asymptotically proportional to Ω^* when the locations of missing data are non-random. Even though $\hat{\Omega}$ is not an estimator of Ω^*, it can still be used to construct test statistics because fixed-b theory shows that $\hat{\Omega}$ scales out Ω^*. Looking closely at the result given by Theorem 2.3 (b) we see that the fixed-b limit of $\hat{\Omega}$ is similar but is noticeably different from the limit obtained for the case of missing at random. The stochastic process $\breve{B}_k\left(r, \{\lambda_i\}\right)$ is different than the Brownian bridge $\tilde{B}(r)$ and depends on the locations of the missing/observed data. Therefore, critical values for the limiting random variables given by Theorem 2.3 (c) are different from the critical values given by the standard fixed-b limits given by Theorem 2.1 (c).

Given the locations of the missing data, the non-standard distribution in Theorem 2.3 (c) can be computed by simulation methods because the limiting distributions are still functions of Brownian motions. Although this method is feasible, it can be practically inconvenient because asymptotic critical values would have to be simulated on a case by case basis depending on the locations of the missing data. In this situation the bootstrap is a more convenient method for obtaining fixed-b critical values. Because the locations of missing data are treated as non-random, we need the bootstrap resampling scheme to preserve the missing locations. This means that blocking is not practical because blocks will shuffle the locations of the missing data upon resampling. Instead the *i.i.d.* bootstrap is more appropriate where bootstrap samples are created by first sampling with replacement from the observed data and creating a bootstrap sample with the same missing locations as the original data.

Specific details are as follows. Define $\omega_t = (y_t, x_t')'$, $t = 1, \ldots, T$, that collects the dependent and independent variables of the AM series. Among those T observations collect only the observed data, which we denote $\tilde{\omega}_t$, $t = 1, \ldots, \tilde{T}$, $\tilde{T} = \sum_{t=1}^{T} a_t$. Resample $\tilde{T}$ observations with replacement from $\tilde{\omega}_t$ and get bootstrap resample which we denote $\tilde{\omega}_t^{\bullet}$, $t = 1, \ldots, \tilde{T}$. Fill in the observed locations with resampled data $\tilde{\omega}_t^{\bullet}$ and leave the missing locations as zeros. This way we construct an *i.i.d.* resample with missing locations fixed. Denote this *i.i.d.* resample as $\omega_t^{\bullet} = (y_t^{\bullet}, x_t^{\bullet\prime})'$, $t = 1, \ldots, T$. The naive bootstrap test statistics $W_T^{\bullet}$ and $t_T^{\bullet}$ are computed as

$$W_T^{\bullet} = T\left(r(\hat{\beta}^{\bullet}) - r(\hat{\beta})\right)' [R(\hat{\beta}^{\bullet})\hat{Q}^{\bullet -1}\hat{\Omega}^{\bullet}\hat{Q}^{\bullet -1}R(\hat{\beta}^{\bullet})']^{-1}\left(r(\hat{\beta}^{\bullet}) - r(\hat{\beta})\right)$$

and

$$t_T^{\bullet} = \frac{\sqrt{T}\left(r(\hat{\beta}^{\bullet}) - r(\hat{\beta})\right)}{\sqrt{R(\hat{\beta}^{\bullet})\hat{Q}^{\bullet -1}\hat{\Omega}^{\bullet}\hat{Q}^{\bullet -1}R(\hat{\beta}^{\bullet})'}},$$

where $\hat{\beta}^{\bullet}$ is the OLS estimate from the regression of $y_t^{\bullet}$ on $x_t^{\bullet}$, $\hat{Q}^{\bullet} = {}^{1}/_{T}\sum_{t=1}^{T} x_t^{\bullet} x_t^{\bullet\prime}$, and $\hat{\Omega}^{\bullet} = {}^{1}/_{T}\sum_{t=1}^{T}\sum_{t=1}^{T} k\left({}^{t-s}/_{M}\right)\hat{v}_t^{\bullet}\hat{v}_s^{\bullet\prime}$, where $\hat{v}_t^{\bullet} = x_t^{\bullet}(y_t^{\bullet} - x_t^{\bullet\prime}\hat{\beta}^{\bullet})$.

Because we resample from observed time periods only, this resampling can be thought of as resampling from the latent process $\omega_t^* \equiv (y_t^*, x_t^{*\prime})'$. We do not know the value of ω_t^* when $a_t = 0$ and thus we are resampling from $\tilde{T}$ observations not the full number of time periods T. However, because the resampling is based on *i.i.d.* draws, this bootstrap resample has essentially the same properties as an *i.i.d.* resample of the latent process. We could take another $T - \tilde{T}$ independent draws from $\tilde{\omega}_t$ and fill in the missing locations of $\tilde{\omega}_t^\bullet$. Call this "filled-in" resample $\omega_t^{*\bullet}$. Then by construction $\omega_t^\bullet = a_t \omega_t^{*\bullet}$ where $\omega_t^{*\bullet}$ can be viewed as a sample from the latent process given the *i.i.d.* resampling. If the bootstrap process, $\omega_t^{*\bullet}$, satisfies (a) $T^{-1} \sum_{t=1}^T x_t^{*\bullet} x_t^{*\bullet\prime} \Rightarrow rQ^{*\bullet}$ for some $Q^{*\bullet}$ and (b) $T^{-1/2} \sum_{t=1}^T v_t^{*\bullet} \Rightarrow \Lambda^{*\bullet} \mathcal{W}_k(r)$ for some $\Lambda^{*\bullet}$, then using Theorem 2.3 (c) it follows that

$$W_T^\bullet \overset{p^\bullet}{\Rightarrow} \overline{\mathcal{W}}_q^{\bullet\prime} P(b, \breve{B}_q(\{\lambda_i^\bullet\})) \overline{\mathcal{W}}_q^\bullet,$$

with $\overline{\mathcal{W}}_q^\bullet = \sum_{j=1}^{2C^\bullet+1} (-1)^{j+1} \mathcal{W}_q(\lambda_j^\bullet)$ where $\{\lambda_m^\bullet\}_{m=0}^{2C^\bullet+1}$ are the missing locations in the bootstrap resample, $C^\bullet$ is the number of missing clusters in the bootstrap resample, and $p^\bullet$ denotes the probability measure induced by the bootstrap resampling conditional on a realization of the original time series. Because the missing locations of the bootstrap resamples are configured to be identical to the missing locations of the data, it follows that $\lambda_j^\bullet = \lambda_j$ and $C^\bullet = C$. Therefore,

$$W_T \overset{p^\bullet}{\Rightarrow} \overline{\mathcal{W}}_q' P(b, \breve{B}_q(\{\lambda_i\})) \overline{\mathcal{W}}_q,$$

which is the same fixed-b limit of W_T as in Theorem 2.3 (c). This asymptotic equivalence is mainly due to the fact that the limiting distribution in Theorem 2.3 (c) is pivotal with respect to Λ^* and Q^* so that $W_T^\bullet$ has an asymptotic distribution equivalent to W_T even though $\Lambda^{*\bullet}$ and $Q^{*\bullet}$ are potentially different from Λ^* and Q^*. Obviously, $t_T^\bullet$ and t_T have equivalent asymptotic approximations as well. Strengthening Assumption NR$'$3-5 to NR$''$3-5 is sufficient for $\omega_t^{*\bullet}$ to satisfy conditions (a) and (b) above.

Assumption NR$''$.

3. $\left\|v_t^*\right\|_{r+\delta} < \infty, r > 2$.

4. $\left\{v_t^*\right\}$ *is a weakly stationary* $L_{2+\delta}-NED$ *on* $\{\varepsilon_t\}$ *with* ν_m *of size* -1.

5. $\{\varepsilon_t\}$ *is a* α*-mixing sequence with* α_m *of size* $-\frac{(2+r)(r+\delta)}{r-2}$.

See Gonçalves and Vogelsang (2011) for the proofs. Here the result of Gonçalves and Vogelsang (2011) directly applies because these assumptions are made about the latent process which has nothing to do with the missing process when the missing locations are non-random. A formal statement of this bootstrap result is in the following theorem.

Theorem 2.4. *Let* $W_T^\bullet$ *and* $t_T^\bullet$ *be naive bootstrap test statistics computed from the i.i.d. bootstrap resample with the locations of missing observations fixed and identical to the missing locations of the real data. Let* $b \in (0,1]$ *be fixed and suppose* $M = bT$*. Then, under Assumption* NR' *with Assumption* NR'*3-5 strengthened to Assumption* NR''*3-5, as* $T \to \infty$,

$$W_T^\bullet \overset{p^\bullet}{\Rightarrow} \overline{\mathcal{W}}_q' P(b, \breve{B}_q(r, \{\lambda_i\}))^{-1} \overline{\mathcal{W}}_q$$

and

$$t_T^\bullet \overset{p^\bullet}{\Rightarrow} \frac{\overline{\mathcal{W}}_1}{\sqrt{P(b, \breve{B}_1(r, \{\lambda_i\}))}}.$$

2.4 FINITE SAMPLE PERFORMANCE

In this section we use Motel Carlo simulations to evaluate the finite sample performance of the fixed-b asymptotic approximation of the HAC robust Wald test defined in Section 2.3.

2.4.1 Data Generating Process

We consider a simple location model for the latent process given by,

$$\begin{aligned}
y_t^* &= \beta + u_t^*, \\
u_t^* &= \rho u_{t-1}^* + \sqrt{1-\rho^2}\varepsilon_t^*, \\
\varepsilon_t^* &\sim i.i.d. N(0,1), \\
u_1^* &= 0,
\end{aligned}$$

with $t = 1, 2, ..., T$ so that T is the time span T. We set $\beta = 0$ and $\rho \in \{0, 0.3, 0.6, 0.9\}$. The time series with missing observations is characterized as an AM series

$$y_t = x_t \beta + u_t,$$

where $y_t = a_t y_t^*$, $x_t = a_t$, $u_t = a_t u_t^*$. We use several specifications of the missing process, $\{a_t\}$, as follows.

1. For the random missing process we model $\{a_t\}$ as a Bernoulli(p) process, i.e. $P(a_t = 1) = p$, with $p \in \{0.3, 0.5, 0.7\}$. We provide results for the time span $T \in \{50, 100, 200\}$.

2. We consider three types of non-random missing processes.

   (a) First, we consider what we call *missing in clusters*. There are cases where observations are missing in large clusters with a small number of clusters. Specifically, we consider cases where data are missing in two clusters ($C = 2$) due to World War I (from 1914 to 1918) and World War II (from 1939 to 1945). We generate data both yearly and quarterly where time spans from 1911 to Y, $Y \in \{1946, 1958, 1970\}$. For yearly data, this means that 12 observations are missing out of T observations, $T \in \{36, 48, 60\}$, and the missing process is $a_{[rT]} = 0$ when $r \in (\lambda_1, \lambda_2] \cup (\lambda_3, \lambda_4]$ and $a_{[rT]} = 1$ otherwise, with $\lambda_1 = {}^3/_T$, $\lambda_2 = {}^8/_T$, $\lambda_3 = {}^{28}/_T$, and $\lambda_4 = {}^{35}/_T$ (See Figure 2.3). For quarterly data, this implies that 48 observations are missing out of T time periods, $T \in \{144, 192, 240\}$, and $a_{[rT]} = 0$ when $r \in (\lambda_1, \lambda_2] \cup (\lambda_3, \lambda_4]$

and $a_{[rT]} = 1$ otherwise, with $\lambda_1 = {}^{12}/T$, $\lambda_2 = {}^{32}/T$, $\lambda_3 = {}^{112}/T$, and $\lambda_4 = {}^{140}/T$. Missing locations are fixed across iterations in the simulations.

(b) Second, we consider *initially scarce data* following the simulation setup in Neave (1970) where the sampling point is shortened at some point during the period of observations. Specifically, we think about a case where at first only the quarterly data (N_Q observations) were available but later monthly data (N_M observations) became available. Hence the latent process is monthly data, and during the periods when only the quarterly data are available, every two observations out of three are missing. (See Figure 2.4.) We set the number of observations available monthly to be$N_M = \{12, 24, 48\}$, and the number of observations available quarterly to be $N_Q \in \{12, 24\}$. Under this setting, the number of missing clusters is $C \in \{11, 23\}$, the total time span $T \in \{46, 58, 82, 94, 118\}$, and $\{22, 46\}$ observations are missing.[4] The missing process is $a_{[rT]} = 0$ when $r \in \cup_{n=1}^{N_Q - 1} (\lambda_{2n-1}, \lambda_{2n}]$ and $a_{[rT]} = 1$ otherwise with $\lambda_{2n-1} = {}^{(3n-2)}/T$ and $\lambda_{2n} = {}^{3n}/T$. The missing locations are fixed across iterations.

(c) Third, we consider a *conditional Bernoulli(p)* missing process to compare to the random Bernoulli(p) missing process. The conditional Bernoulli(p) missing process differs from random Bernoulli(p) missing process in the way in which it is simulated. Once the missing process, $\{a_t\}$, is generated from the corresponding Bernoulli(p) process for the first iteration, the missing locations are then fixed for subsequent iterations. Hence, all the iterations have the same missing locations in contrast to the random Bernoulli(p) process where missing locations change for each iteration. As for the random Bernoulli process, we consider $p \in \{0.3, 0.5, 0.7\}$ and the total time span $T \in \{50, 100, 200\}$.

[4]The number of missing clusters is $C = N_Q - 1$. The total time span is $T = (N_Q - 1) \times 3 + 1 + N_M$. The number of missing observations is $2 \times (N_Q - 1)$.

2.4.2 Test statistics and Critical Values

With the data generating process defined in Section 2.4.1, we consider testing the null hypothesis that $\beta = 0$ against the alternative $\beta \neq 0$ at a nominal level of 5%. When computing the HAC estimator, we use $b \in \{0.1, 0.15, \ldots, 0.95, 1\}$ throughout. The HAC robust t-statistic for β is

$$t_T = \frac{\hat{\beta}}{\sqrt{T\left(\sum_{t=1}^{T} x_t^2\right)^{-1} \hat{\Omega} \left(\sum_{t=1}^{T} x_t^2\right)^{-1}}}$$
$$= \frac{\hat{\beta}}{\sqrt{T\left(\sum_{t=1}^{T} a_t^2\right)^{-1} \hat{\Omega} \left(\sum_{t=1}^{T} a_t^2\right)^{-1}}}$$
$$= \frac{\hat{\beta}}{\sqrt{T\left(\sum_{t=1}^{T} a_t\right)^{-2} \hat{\Omega}}},$$

where

$$\hat{\beta} = \left(\sum_{t=1}^{T} x_t^2\right)^{-1} \sum_{t=1}^{T} x_t y_t = \left(\sum_{t=1}^{T} a_t\right)^{-1} \sum_{t=1}^{T} a_t y_t = \left(\sum_{t=1}^{T} a_t\right)^{-1} \sum_{t=1}^{T} y_t,$$

and $\hat{\Omega} = T^{-1} \sum_{i=1}^{T} \sum_{j=1}^{T} k\left(|i-j|/[bT]\right) \hat{v}_i \hat{v}_j$ with $\hat{v}_t = a_t\left(y_t - \hat{\beta}\right)$. We reject the null hypothesis whenever $|t_T| > t_c$ (or reject the null whenever $t_T < t_c^l$ or $t_T > t_c^r$ if $-t_c^l \neq t_c^r$) where t_c is a critical value. Using 10,000 replications, we compute empirical rejection probabilities. As shown from Theorems 2.1(c) and 2.3(c), the test statistics have different asymptotic distributions depending on whether the missing process is random or non-random. Hence critical values are calculated differently for the two cases.

When the missing process is random, t_c is the 97.5% percentile of the standard fixed-b asymptotic distribution derived by Kiefer and Vogelsang (2005) (See Theorem 2.1 (c)). From Section 2.3.1 we know that we can compute the asymptotic critical values by either simulating the distribution itself or by the naive moving block bootstrap which we denote as $(\{t_c^{R-boot,l}, t_c^{R-boot,r}\})$. To evaluate the finite sample performance we use both of the methods to get the critical values. For the naive moving block bootstrap, we use block length $l = 1$ (the *i.i.d.* bootstrap). From the original random sample of T observations, $y_1, \ldots, y_T$, we get 999 bootstrap resamples, $(y_1^{\bullet B}, a_1^{\bullet B}) \ldots, (y_T^{\bullet B}, a_T^{\bullet B}), B = 1, \ldots, 999$. For each bootstrap resample we compute the bootstrap t-statistic as

$$t_T^B = \frac{(\hat{\beta}^B - \hat{\beta})}{\sqrt{T(\sum_{t=1}^T a_t^{\bullet B})^{-2} \hat{\Omega}^B}},$$

where $\hat{\beta}^B = \left(\sum_{t=1}^T a_t^{\bullet B}\right)^{-1} \sum_{t=1}^T y_t^{\bullet B}$, and $\hat{\Omega}^B = T^{-1} \sum_{t=1}^T \sum_{s=1}^T k(|t-s|/[bT]) \hat{v}_t^{\bullet B} \hat{v}_s^{\bullet B}$, where $\hat{v}_t^{\bullet B} = a_t^{\bullet B}(y_t^{\bullet B} - \hat{\beta}^B)$. Then $t_c^{R-boot,l}$ is the 0.025 quantile and $t_c^{R-boot,r}$ is the 0.975 quantile of t_T^B, $B = 1, \ldots, 999$.

When the missing process is non-random, t_c is the 97.5% percentile of the distribution derived in Theorem 2.3 (c). From Section 2.3.2 we know that critical values can be computed either by simulating the limiting distribution or by naive *i.i.d.* bootstrap (see Theorem 2.4) which we denote as $(\{t_c^{NR-boot,l}, t_c^{NR-boot,r}\})$. Because the limiting distribution depends on missing locations, the naive *i.i.d.* bootstrap is more convenient in practice. However, to illustrate the relative finite performance, we compute the critical values using both methods. From the original sample of T observations, $y_1, \ldots, y_T$, we pull out the data from the observed time periods, $\tilde{y}_1, \ldots, \tilde{y}_{\tilde{T}}$, $\tilde{T} = \sum_{t=1}^T a_t$. From these $\tilde{T}$ observations, we resample $\tilde{T}$ observations with replacement. Repeating this procedure 999 times we obtain resamples which we denote $\tilde{y}_1^{\bullet B}, \ldots, \tilde{y}_{\tilde{T}}^{\bullet B}$, $B = 1, \ldots, 999$. By filling observed locations with resampled data, $\tilde{y}_t^{\bullet}$, and filling missing locations with zeros, we obtain the *i.i.d.* bootstrap resamples, which we denote $y_1^{\bullet B}, \ldots, y_T^{\bullet B}$, for $B = 1, \ldots, 999$. We compute the naive bootstrap t-statistic as

$$t_T^B = \frac{(\hat{\beta}^B - \hat{\beta})}{\sqrt{T\left(\sum_{t=1}^T a_t\right)^{-2} \hat{\Omega}^B}}$$

where $\hat{\beta}^B = \left(\sum_{t=1}^T a_t^B\right)^{-1} \sum_{t=1}^T y_t^{\bullet B}$ and $\hat{\Omega}^B = T^{-1} \sum_{t=1}^T \sum_{s=1}^T k(|t-s|/[bT]) \hat{v}_t^{\bullet B} \hat{v}_s^{\bullet B}$, where $\hat{v}_t^{\bullet B} = a_t(y_t^{\bullet B} - \hat{\beta}^B)$. Then $t_c^{NR-boot,l}$ is the 0.025 and $t_c^{NR-boot,r}$ is the 0.975 quantile of t_T^B, $B = 1, \ldots, 999$. Note that we are using a_t rather than $a_t^{\bullet}$ in this case because we are conditioning on the locations of the missing data when resampling.

2.4.3 Finite Sample Performance

We illustrate the non-random missing process case first. Figures 2.5-2.34 show empirical rejection probabilities computed from $10,000$ replications using AM series for the four missing processes defined in Section 2.4.1. Since the missing process is non-random, by Theorem 2.3 (c) the HAC robust test statistics have a fixed-b asymptotic distribution that depends on the missing locations. Critical values are obtained by the naive *i.i.d.* bootstrap with locations of missing observations fixed (labeled *L-bootstrap*) or by directly simulating the limiting fixed-b distributions (labeled *L-fixed-b*). In addition to these two critical values we consider critical values obtained by the naive *i.i.d.* bootstrap that does not condition on missing locations (labeled *bootstrap*) and by simulating the standard fixed-b limit in Kiefer and Vogelsang (2005) (labeled *fixed-b*) for comparison although these two critical values are not theoretically valid.

The first thing we can notice is that the fixed-b critical values that treat missing locations fixed has less size distortions than the standard fixed-b critical values when the sample size is small and/or serial correlation is high. This difference tends to increase as the number of missing observations increase. For the World War missing process, when $T = 36$ and $\rho \in \{0.6, 0.9\}$ (Figure 2.5), empirical rejection probabilities by the fixed-b limit that depends on locations of missing observations has less size distortion than the usual fixed-b limit, and this size difference is bigger when $\rho = 0.9$ than when $\rho = 0.6$. Comparing the World War missing process with $T = 36$ to that of $T = 144$ (Figures 2.5 and 2.8), we see that while for both cases one third of the data are missing, the difference in rejection probabilities between the usual fixed-b and the fixed-b conditional on the locations is bigger when $T = 36$, the smaller sample size. Similar tendency can be found in conditional Bernoulli missing process. Consider the simulation with $T = 50$, $\rho = 0.9$, and $p = 0.3\,(70\%$ missing$)$ as a base case (Figure 2.17). If we compare this base case to the simulations where (i) $T = 100$, $\rho = 0.9$, and $p = 0.3\,(70\%$ missing$)$ (Figure 2.20), (ii) $T = 50$, $\rho = 0.6$, and $p = 0.3\,(70\%$ missing$)$(Figure 2.17), and (iii) $T = 50$, $\rho = 0.9$, and

$p = 0.5$ (50% missing) (Figure 2.18), it is always the base case that has a bigger difference in rejection probabilities between the usual fixed-b and the fixed-b that depends on the locations of missing observations. Relative to the base case, these three cases have one of three features: a bigger sample size (i), less serial correlation (ii), and a smaller missing proportion (iii). Even though rejection rates based on the standard fixed-b critical values and the conditional fixed-b critical values are sometimes similar, the simulations show the more prudent approach is to use the conditional fixed-b critical values as was predicted by the theoretical results.

The simulation results also suggest that one may gain by bootstrapping rather than simulating the fixed-b distribution especially when the serial correlation is high, the sample size is small, or the missing proportion is large. This tendency holds regardless of missing locations being treated as fixed or not in the bootstrap resampling scheme. The empirical rejection probabilities from naive *i.i.d.* bootstrap with missing locations fixed have less size distortion than the empirical rejection probabilities obtained by simulating the fixed-b distribution in Theorem 2.3 (c). The same thing holds between the standard fixed-b limit and the naive *i.i.d* bootstrap that does not condition on missing locations. For example, if we look at the $T = 50$ Bernoulli (Figures 2.17-2.19), $(N_Q = 12, N_M = 12)$ initially scarce (Figure 2.11), and $T = 36$ World War missing (Figure 2.5) cases, all of which have a small time span, we see differences between the rejection rates from bootstrapping and by simulating the fixed-b distribution especially when $\rho = 0.9$. Comparing the World War missing process with $T = 48$ (Figure 2.6) and $T = 144$ (Figure 2.8), we see that even though $T = 144$ has a bigger time span, there is still a bigger bootstrap gain with $T = 144$ because when $T = 48$ around one fourth of the data are missing whereas for $T = 144$ around one third of the observations are missing. Overall, the simulations indicate that the naive *i.i.d.* bootstrap with locations fixed seems most robust to the stationary asymptotic breakdowns.

In addition it appears that the pattern of missing locations matters as well. For the conditional Bernoulli missing process with $p = 0.5$ and $T = 50$ (Figure 2.18) we see a difference between the fixed-b critical value conditional on missing locations and the usual fixed-b critical value. On the other hand, for initially scarce data with $N_Q = 12$ and $N_M = 12$ (Figure 2.11) or $N_Q = 12$ and $N_M = 24$ (Figure 2.12), where both have time span of around 50 and around half of the observations are missing ($^{24}/_{46}$ and $^{24}/_{58}$ respectively), the difference between the usual fixed-b critical value and the conditional fixed-b critical value is quite small. For initially scarce data, the maximum length of the missing cluster is 2 and the observations are missing only at the beginning of the sample.

When the missing process is random, the usual fixed-b limit in Kiefer and Vogelsang (2005) is valid. Hence critical values can be obtained by simulating this distribution or by the naive *i.i.d.* bootstrap. As with the non-random missing process, for comparison, we also consider critical values obtained from the naive *i.i.d.* bootstrap conditional on the missing locations. Figures 2.26-2.34 show empirical rejection probabilities computed from $10,000$ replications using the AM series. Results for the random missing process are similar to the non-random missing case. The conditional fixed-b limit in general performs no worse than the usual fixed-b limit and the conditional fixed-b limit becomes advantageous when the data is not well-behaved. However this tendency is less strong than the non-random Bernoulli missing process. Starting with the $T = 100$ and $p = 0.5$ (50% missing) case and moving to cases where either p or T increase (less missing proportion or increased time span), the difference between the conditional and unconditional fixed-b rejection rates disappears for the random Bernoulli missing process. When $T = 200$, even with $p = 0.3$ (70% missing) and $\rho = 0.9$ (Figure 2.32) there is no difference between the two rejection probabilities. Given that it is better to conditional on locations critical values when the missing process is non-random and given that it appears conditioning on locations causes no harm when the missing process is random, our simulations suggest the use of conditional on locations critical values in practice. The bootstrap is the most convenient way to obtain these critical values given that each application with missing data will have application specific missing locations.

2.5 WHEN MISSING OBSERVATIONS ARE IGNORED

In practice an empirical researcher might be tempted to simply ignore any missing data problems and estimate the time series regression with the data that is observed. From the perspective of estimating β this has no consequences because one obtains the same estimator of β as is obtained when missing observations are replaced with zeros. For the computation of long run variance estimators, ignoring missing data matters because the time distances between observations is skewed for many pairs of time periods. Thus, robust test statistics are computationally different when ignoring missing data verses replacing missing data with zeros. A reasonable conjecture is that ignoring missing data invalidates inference using HAC robust test statistics. Surprisingly, fixed-b asymptotic theory suggests otherwise. As we show in this section, ignoring the missing data leads to HAC robust tests that have *standard* fixed-b limits. This is true whether the missing process is random or non-random. In contrast to the AM series approach, the empirical researcher does not have to worry about robustness to whether missing dates are best viewed as random or non-random.

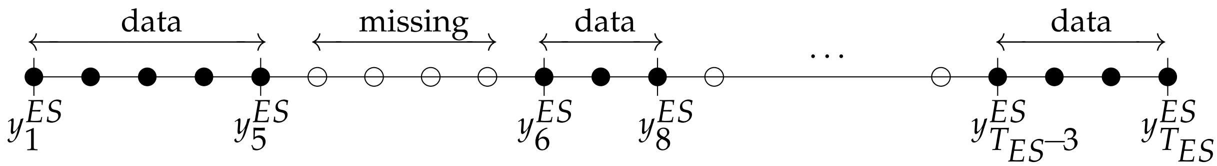
2.5.1 Models and Test statistics

In terms of the regression model, ignoring missing observations amounts to stacking only the observed observations as if they are equally spaced in time. (See Figure 2.2.) Taking out the missing observations from the latent process and relabeling observed observations, the regression model becomes

$$y_t^{ES} = x_t^{ES'}\beta + u_t^{ES} \qquad \left(t = 1, 2, \ldots, T_{ES}\right), \tag{2.5}$$

where $T_{ES} = \sum_{t=1}^{T} a_t$ is the number of non-missing observations. Following Datta and Du (2012) we call this model the ***equal space (ES) regression model***.

Figure 2.2: Equal Space Regression Model



As with the AM series, the ES regression model uses only the observed data. No attempt is made to forecast or proxy missing observations. However unlike the AM series, the original time distances between observations are not preserved in the ES regression model. The distance between the t^{th} and s^{th} observations (in terms of the latent process) is not necessarily $|t-s|$ but is instead equal to $|\sum_{i=1}^{t} a_i - \sum_{i=1}^{s} a_i|$ which is the number of observed data points between time periods t and s. Only when there are no missing observations between time periods t and s will the time distance remain $|t-s|$ in (2.5).

The OLS estimator of the ES regression model is defined as

$$\hat{\beta}^{ES} = \left(\sum_{t=1}^{T_{ES}} x_t^{ES} x_t^{ES\prime} \right)^{-1} \sum_{t=1}^{T_{ES}} x_t^{ES} y_t^{ES}.$$

Recall that missing observations are replaced with zeros in the AM series and missing observations are deleted in the ES regression model. Because the only difference between $\sum_{t=1}^{T} x_t x_t'$ and $\sum_{t=1}^{T_{ES}} x_t^{ES} x_t^{ES\prime}$ comes from the missing observations which are set to zeros, it follows that $\sum_{t=1}^{T} x_t x_t' = \sum_{t=1}^{T_{ES}} x_t^{ES} x_t^{ES\prime}$. By the same reasoning, $\sum_{t=1}^{T} x_t y_t = \sum_{t=1}^{T_{ES}} x_t^{ES} y_t^{ES}$. Therefore, it follows that

$$\hat{\beta} = \hat{\beta}^{ES}.$$

Hence, in terms of the OLS estimator, the ES regression model provides the same estimate of β as the AM series.

Let $\Omega^{ES} = \lim_{T\to\infty} \mathrm{Var}\left(T^{-1/2} \sum_{t=1}^{T_{ES}} v_t^{ES} \right)$, where $v_t^{ES} = x_t^{ES} u_t^{ES}$. Then, the usual kernel based HAC estimator for Ω^{ES} is defined as

$$\hat{\Omega}^{ES} = \hat{\Gamma}_0^{ES} + \sum_{j=1}^{T_{ES}-1} k\left(\frac{j}{M_{ES}} \right) \left(\hat{\Gamma}_j^{ES} + \hat{\Gamma}_j^{ES\prime} \right),$$

where $\hat{\Gamma}_j^{ES} = T_{ES}^{-1} \sum_{t=j+1}^{T_{ES}} \hat{v}_t^{ES} \hat{v}_{t-j}^{ES\prime}$ are the sample autocovariances of $\hat{v}_t^{ES} = x_t^{ES} \hat{u}_t^{ES}$ and $\hat{u}_t^{ES} = y_t^{ES} - x_t^{ES\prime} \hat{\beta}^{ES}$ are OLS residuals from the ES regression model. As before, $k(x)$ is a kernel function such that $k(x) = k(-x)$, $k(0) = 1$, $|k(x)| \leq 1$, continuous at $x = 0$, and $\int_{-\infty}^{\infty} k^2(x) < \infty$ and M_{ES} is the bandwidth.[5]

By well known algebra we can rewrite $\hat{\Omega}^{ES}$ as

$$\hat{\Omega}^{ES} = T_{ES}^{-1} \sum_{n=1}^{T_{ES}} \sum_{m=1}^{T_{ES}} k\left(\frac{n-m}{M_{ES}}\right) \hat{v}_n^{ES} \hat{v}_m^{ES\prime}.$$

Recall that $\hat{v}_t = x_t \hat{u}_t$ where $\hat{u}_t$ are the OLS residuals from the AM series. Therefore, by construction $\hat{v}_t = x_t \hat{u}_t = a_t x_t (y_t - x_t' \hat{\beta})$ which implies that $\hat{v}_t = 0$ at missing dates. Because the ES regression model and the AM series share the same $\hat{\beta}$, $\hat{v}_t^{ES}$ is obtained by dropping missing observations from $\hat{v}_t$.[6] Using these facts, we can recast $\hat{\Omega}^{ES}$, which is a weighted sum of $\hat{v}_n^{ES} \hat{v}_m^{ES\prime}$, instead as a weighted sum of $\hat{v}_t \hat{v}_s'$ because all the elements of $\hat{v}_t^{ES} \hat{v}_s^{ES\prime}$ are found among those of $\hat{v}_t \hat{v}_s'$ with the remaining elements of $\hat{v}_t \hat{v}_s'$ being zeros. The only complication that arises when rewriting $\hat{\Omega}^{ES}$ in terms of $\hat{v}_t \hat{v}_s'$ lies in matching the kernel weights used on $\hat{v}_t \hat{v}_s$ to those that are used by $\hat{\Omega}^{ES}$ which are different from the weights used by $\hat{\Omega}$.

The time distance between $\hat{v}_t$ and $\hat{v}_s$ in the ES regression model is $|\sum_{i=1}^{t} a_i - \sum_{i=1}^{s} a_i|$ and we can rewrite $\hat{\Omega}^{ES}$ as

$$\hat{\Omega}^{ES} = T_{ES}^{-1} \sum_{t=1}^{T} \sum_{s=1}^{T} k\left(\frac{\sum_{i=1}^{t} a_i - \sum_{i=1}^{s} a_i}{M_{ES}}\right) \hat{v}_t \hat{v}_s'.$$

Recall that the HAC estimator of the AM series is given by (2.3). Both $\hat{\Omega}^{ES}$ and $\hat{\Omega}$ are weighted sums of $\hat{v}_t \hat{v}_s'$, $t, s = 1, \ldots, T$, but with different weights. For the ES regression

[5] We denote the bandwidth of the ES regression model as M with subscript ES because if we fix $b = M_{ES}/T_{ES}$, then M_{ES} depends on the time span of the ES regression model (T_{ES}).

[6] To be more specific, $\hat{v}_t = \hat{v}_g^{ES}$ with $g = \sum_{i=1}^{t} a_i$ whenever $a_t = 1$. $\hat{v}_g^{ES}$ for all $g = 1, \ldots, T_{ES}$ can be defined this way. When $a_i = 0$, there is no term in the ES regression model that matches $\hat{v}_i$ (which is zero) because missing observations are dropped in the ES regression model.

model, by taking out the missing observations, the time distances between observations become shorter than the true time distances, $|t-s|$, unless there are no missing observations between t and s. Therefore, $\hat{\Omega}^{ES}$ gives $\hat{v}_t\hat{v}_s'$ weights at least as big as $\hat{\Omega}$ if the same bandwidth is used: $k\left(\left(\sum_{i=1}^{t} a_i - \sum_{i=1}^{s} a_i\right)/M_{ES}\right) \geq k\left((t-s)/M\right)$ if $M = M_{ES}$.

We now revisit testing the null hypothesis $H_0 : r\left(\beta_0\right) = 0$ against $H_A : r\left(\beta_0\right) \neq 0$. We define the ES HAC robust Wald statistic as

$$W_T^{ES} = T_{ES} r\left(\hat{\beta}\right)' \left[R\left(\hat{\beta}\right) \hat{Q}^{ES-1} \hat{\Omega}^{ES} \hat{Q}^{ES-1} R\left(\hat{\beta}\right)'\right]^{-1} r\left(\hat{\beta}\right),$$

and when $q = 1$, t-statistics of the form

$$t_T^{ES} = \frac{\sqrt{T_{ES}} r\left(\hat{\beta}\right)}{\sqrt{R\left(\hat{\beta}\right) \hat{Q}^{ES-1} \hat{\Omega}^{ES} \hat{Q}^{ES-1} R\left(\hat{\beta}\right)'}},$$

where $\hat{Q}^{ES} = T_{ES}^{-1} \sum_{t=1}^{T_{ES}} x_t^{ES} x_t^{ES\prime}$. Note that $\sum_{t=1}^{T} x_t x_t' = \sum_{t=1}^{T_{ES}} x_t^{ES} x_t^{ES\prime}$ implies $\hat{Q}^{ES} = \left(T/T_{ES}\right) \hat{Q}$. We therefore can write W_T^{ES} as

$$W_T^{ES} = Tr\left(\hat{\beta}\right)' \left[\left(\frac{T_{ES}}{T}\right) R\left(\hat{\beta}\right) \hat{Q}^{-1} \hat{\Omega}^{ES} \hat{Q}^{-1} R\left(\hat{\beta}\right)'\right]^{-1} r\left(\hat{\beta}\right).$$

Other than the scaling factor T_{ES}/T and $\hat{\Omega}^{ES}$, the other terms in W_T^{ES} are identical to W_T. Therefore, in terms of test statistics, choosing between the AM series statistics and the ES statistic boils down to choosing the kernel weights when computing the HAC estimator.

2.5.2 Asymptotic Theory

As with the AM series, we are mainly interested in the fixed-b asymptotic limits of W_T^{ES} and t_T^{ES} under the null hypothesis H_0 defined in Section 2.5.1.

2.5.2.1 Non-random missing process

We first consider the non-random missing process case. Because the fixed-b asymptotic distributions depend on the kernels used to compute the HAC estimators, we need to define some random matrices that appear in the asymptotic results. The random matrices in Definition 1 no longer work here because the kernel weights in $\hat{\Omega}^{ES}$ are different from those of $\hat{\Omega}$. In fact because the kernel weight for $\hat{v}_t\hat{v}_s'$ depends on the number of missing observations between t and s, unlike the random matrices defined in Definition 1, the random matrices that appear in the asymptotic approximation of $\hat{\Omega}^{ES}$ depend on the missing locations $\{\lambda_i\}_{i=0}^{2C+1}$. Note that for two numbers r and s, $r \wedge s$ denotes the minimum of r and s and $r \vee s$ denote the maximum of r and s.

Definition 3. *Let $h > 0$ be an integer. Let $\{\lambda_i\}_{i=0}^{2C+1}$ be given by Assumption NR.1 and let λ be given by* (2.4). *Let $B_h(r, \{\lambda_i\})$ denote a generic $h \times 1$ vector of stochastic process that depends on $\{\lambda_i\}$. Let the random matrix, $P^{ES}(b, B_h(\{\lambda_i\}))$, be defined as follows for $b \in (0, 1]$:*

Case (i) : if $k(x)$ is twice continuously differentiable everywhere,

$$P^{ES}\left(b, B_h(\{\lambda_i\})\right) \equiv -\frac{1}{b^2\lambda^3}\sum_{n=0}^{C}\sum_{l=0}^{C}\int_{\lambda_{2n}}^{\lambda_{2n+1}}\int_{\lambda_{2l}}^{\lambda_{2l+1}}$$
$$\left[k''\left((\lambda b)^{-1}\left[\sum_{j=1}^{2n+1}(-1)^{j+1}(r\wedge\lambda_j) - \sum_{j=1}^{2l+1}(-1)^{j+1}(u\wedge\lambda_j)\right]\right)\right.$$
$$\left.\times B_h(r,\{\lambda_i\})B_h(u,\{\lambda_i\})'\right] dudr$$

Case (ii) : if $k(x)$ is continuous, $k(x) = 0$ for $|x| \geq 1$ and $k(x)$ is twice continuously differentiable everywhere except for $|x| = 1$,

$$P^{ES}\left(b, B_h(\{\lambda_i\})\right) \equiv -\frac{1}{b^2\lambda^3}\sum_{n=0}^{C}\sum_{l=0}^{C}\int_{\lambda_{2n}}^{\lambda_{2n+1}}\int_{\lambda_{2l}}^{\lambda_{2l+1}}$$
$$\left[\mathbb{1}\left\{|r-u| < b\lambda + \sum_{j=2(n\wedge l)+1}^{2(n\vee l)}(-1)^j\lambda_j\right\}\right.$$
$$k''\left((\lambda b)^{-1}\left[\sum_{j=1}^{2n+1}(-1)^{j+1}\left(r\wedge\lambda_j\right) - \sum_{j=1}^{2l+1}(-1)^{j+1}\left(u\wedge\lambda_j\right)\right]\right)$$
$$\left.B_h(r,\{\lambda_i\})B_h(u,\{\lambda_i\})'\right] drdu$$
$$+\frac{k'(1)_-}{b\lambda^2}\sum_{n=0}^{C}\sum_{l=0}^{n}\int_{\lambda_{2l}}^{\lambda_{2l+1}}$$
$$\left[\mathbb{1}\left\{\lambda_{2n} - b\lambda - \sum_{j=2l+1}^{2n}(-1)^j\lambda_j < u \leq \lambda_{2n+1} - b\lambda - \sum_{j=2l+1}^{2n}(-1)^j\lambda_j\right\}\right.$$
$$\times\left\{B_h\left(u + b\lambda + \sum_{j=2l+1}^{2n}(-1)^j\lambda_j, \{\lambda_i\}\right)B_h(u,\{\lambda_i\})'\right.$$
$$\left.\left.+B_h(u,\{\lambda_i\})B_h\left(u + b\lambda + \sum_{j=2l+1}^{2n}(-1)^j\lambda_j, \{\lambda_i\}\right)'\right\}\right] du,$$

where $k_-(1)' = \lim_{h\to 0}\left[(k(1) - k(1-h))/h\right]$, *Case (iii) : if $k(x)$ is the Bartlett kernel,*

$$P^{ES}\left(b, B_h(\{\lambda_i\})\right) \equiv \frac{2}{b\lambda^2}\sum_{n=0}^{C}\int_{\lambda_{2n}}^{\lambda_{2n+1}} B_h\left(r,\{\lambda_i\}\right) B_h\left(r,\{\lambda_i\}\right)' dr$$
$$-\frac{1}{b\lambda^2}\sum_{n=0}^{C}\sum_{l=0}^{n}\int_{\lambda_{2l}}^{\lambda_{2l+1}}$$
$$\left[\mathbb{1}\left\{\lambda_{2n} - b\lambda - \sum_{k=2l+1}^{2n}(-1)^k\lambda_k \le u \le \lambda_{2n+1} - b\lambda - \sum_{k=2l+1}^{2n}(-1)^k\lambda_k\right\}\right.$$
$$\times\left\{B_h\left(u,\{\lambda_i\}\right) B_h\left(u + b\lambda + \sum_{k=2l+1}^{2n}\lambda_k(-1)^k, \{\lambda_i\}\right)'\right.$$
$$\left.\left.+B_h\left(u + b\lambda + \sum_{k=2l+1}^{2n}\lambda_k(-1)^k, \{\lambda_i\}\right) B_h\left(u,\{\lambda_i\}\right)'\right\}\right] du.$$

When the missing process is non-random, the asymptotic theory for the ES regression model is based on Assumption NR$'$ which is the same assumption that the AM series results are based on. Theorem 2.5 below provides the asymptotic limits of $\hat{\Omega}^{ES}$ and W_T^{ES} (t_T^{ES} when $q = 1$) when the missing process is non-random. Because the ES regression model and the AM series model have the same OLS estimator, we do not restate the asymptotic result of the OLS estimator given by Theorem 2.3 (a). The proof for Theorem 2.5 is provided in Appendix D.

Theorem 2.5. *Let $\overline{\mathcal{W}}_k$ be defined as in Theorem 2.3. Let $\breve{B}_k\left(r,\{\lambda_i\}\right)$ be a $k \times 1$ vector of stochastic processes defined as*

$$\breve{B}_k\left(r,\{\lambda_i\}\right) \equiv$$
$$\sum_{n=0}^{C}\mathbb{1}\left\{\lambda_{2n} < r \le \lambda_{2(n+1)}\right\}\sum_{j=1}^{2n+1}(-1)^{j+1}\left(\mathcal{W}_k\left(r\wedge\lambda_j\right) - \left(r\wedge\lambda_j\right)\lambda^{-1}\overline{\mathcal{W}}_k\right),$$

for $r \in (0,1]$. Assume $M_{ES} = bT_{ES}$ where $b \in (0,1]$ is fixed. Then under Assumption NR$'$, as $T \to \infty$,

(a). (Fixed-b asymptotic approximation of $\hat{\Omega}^{ES}$)

$$\hat{\Omega}^{ES} \Rightarrow \Lambda^* P^{ES}\left(b, \breve{B}_k(\{\lambda_i\})\right)\Lambda^{*\prime},$$

(b). (Fixed-b asymptotic distribution of W_T^{ES}) under H_0,

$$W_T^{ES} \Rightarrow \overline{\mathcal{W}}_q'\left[\lambda P^{ES}\left(b, \breve{B}_q(\{\lambda_i\})\right)\right]^{-1}\overline{\mathcal{W}}_q$$

and when $q = 1$,

$$t_T^{ES} \Rightarrow \frac{\overline{\mathcal{W}}_1}{\sqrt{\lambda P^{ES}\left(b, \breve{B}_1(\{\lambda_i\})\right)}}.$$

Although the difference of the limits of $\hat{\Omega}$ and $\hat{\Omega}_{ES}$ is in the form of the functions $P(\cdot)$ and $P^{ES}(\cdot)$ because of the different relative distances between observations, it is proportional to Ω^* and is a function of $\breve{B}_q(r, \{\lambda_i\})$ like $\hat{\Omega}$. Similar to W_T, W_T^{ES} has a limiting distribution that is non-standard and depends on the locations of the missing data but remains pivotal with respect to Ω^* and Q^*.

Surprisingly, it turns out that the asymptotic distribution in Theorem 2.5 (b) is equivalent to the standard fixed-b asymptotic distribution in Kiefer and Vogelsang (2005) with $b = M_{ES}/T_{ES}$. To establish this result we first consider the special case where the latent process is *i.i.d*. When the latent process is *i.i.d.*, the ES regression model is a time series regression with T_{ES} observations and there is no serial correlation in the data. Therefore, W_T^{ES} is the usual HAC statistic computed with T_{ES} observations. For $M_{ES} = bT_{ES}$, W_T^{ES} has the usual fixed-b limit because the results of Kiefer and Vogelsang (2005) directly apply. Intuitively speaking, when the data is *i.i.d.*, the time distances between observations do not matter and missing observations only reduce the sample size. Therefore, the fixed-b theory goes through as usual. This result for the *i.i.d*. case is formally stated in the following Lemma.

Lemma 1. *Let the missing process* $\{a_t\}$ *be non-random. The latent process is given by* (2.1). *Suppose that* $\{(x_t^*, u_t^*)\}$ *is i.i.d. Assume* $M_{ES} = bT_{ES}$ *where* $b \in (0, 1]$ *is fixed. Then under* H_0 *as* $T \to \infty$,

$$W_T^{ES} \Rightarrow \mathcal{W}_q' P\left(b, \tilde{B}_q\right)^{-1} \mathcal{W}_q$$

and when $q = 1$,

$$t_T^{ES} \Rightarrow \frac{\mathcal{W}_1}{\sqrt{P\left(b, \tilde{B}_q\right)}}.$$

Because the *i.i.d.* assumption needed for Lemma 1 is a special case of the conditions needed for Theorem 2.5, the fixed-b limits given by Theorem 2.5 (b) and Lemma 1 are distributionally equivalent. Because the limits in Theorem 2.5(b) continue to hold when the data is not *i.i.d.*, the following Theorem holds as a direct consequence of Lemma 1:

Theorem 2.6. *Assume $M_{ES} = bT_{ES}$ where $b \in (0, 1]$ is fixed. Then under Assumption NR$'$ and H_0, as $T \to \infty$,*

$$W_T^{ES} \Rightarrow \mathcal{W}_q' P(b, \tilde{B}_q)^{-1} \mathcal{W}_q$$

and when $q = 1$,

$$t_T^{ES} \Rightarrow \frac{\mathcal{W}_1}{\sqrt{P(b, \tilde{B}_1)}}.$$

2.5.2.2 Random missing process

Now we consider the case where the missing process is random and explore the asymptotic properties of W_T^{ES} (t_T^{ES} when $q = 1$). From Theorem 2.6 we can easily deduce the asymptotic limit of W_T^{ES} in the missing at random case. Suppose that Assumption R$'$ holds. Suppose we condition on the missing process $\{a_t\}$. Conceptually this is the same as treating the missing process as non-random. Recall that Assumption R$'$ and Assumption NR$'$ are identical in terms of the latent process. Thus the fixed-b limiting distribution of W_T^{ES} *conditional on the missing process* is given by Theorem 2.6. Because the conditional distribution in Theorem 2.6 is the standard fixed-b distribution in Kiefer and Vogelsang (2005) and does not depend on the conditioning process $\{a_t\}$ (does not depend on $\{\lambda_i\}$), it directly follows that the unconditional limiting distribution of W_T^{ES} must also be the distribution in Theorem 2.6. Formally, we have the following result.

Theorem 2.7. *Assume $M_{ES} = bT_{ES}$ where $b \in (0, 1]$ is fixed. Then under Assumption R$'$ and H_0, as $T \to \infty$,*

$$W_T^{ES} \Rightarrow \mathcal{W}_q' P(b, \tilde{B}_q)^{-1} \mathcal{W}_q$$

and when $q = 1$,

$$t_T^{ES} \Rightarrow \frac{\mathcal{W}_1}{\sqrt{P(b, \tilde{B}_1)}}.$$

Remarkably for the ES regression model, regardless of whether the missing process is random or non-random, the HAC robust wald statistic and the t-statistic have usual fixed-b asymptotic distribution as in Kiefer and Vogelsang (2005). As discussed for the AM series case with a random missing process, the standard fixed-b critical values can be obtained using various simulation and numerical methods. It is worth noting that for the equally spaced case, the *i.i.d*. bootstrap can be applied to the equally spaced data as the results of Gonçalves and Vogelsang (2011) directly apply under the assumptions of Theorem 2.4.

2.5.3 Finite Sample Properties

In this section we analyze the finite sample performance of W_T^{ES} using Monte Carlo simulations. We use the same data generating process defined in Section 2.4.1. From the simple location model in Section 2.4.1 that lies on the time span T we construct the ES regression model

$$y_t^{ES} = \beta + u_t^{ES}, t = 1, \ldots, T_{ES},$$

with $T_{ES} = \sum_{t=1}^{T} a_t$. We set $\beta = 0$ and $\rho \in \{0, 0.3, 0.6, 0.9\}$ as for the AM series. The HAC robust t-statistic for β is

$$t_T^{ES} = \frac{\sqrt{T_{ES}}\hat{\beta}}{\sqrt{\hat{\Omega}_{ES}}},$$

where $\hat{\beta} = \sum_{t=1}^{T} y_t / T_{ES}$ and $\hat{\Omega}_{ES} = T_{ES}^{-1} \sum_{i=1}^{T_{ES}} \sum_{j=1}^{T_{ES}} k\left(|i-j| / [bT_{ES}] \right) \hat{v}_i^{ES} \hat{v}_j^{ES}$ with $\hat{v}_t^{ES} = y_t^{ES} - \hat{\beta}$. As with the AM series, we use $b \in \{0.1, 0.15, \ldots, 1\}$. We reject the null hypothesis whenever $\left| t_T^{ES} \right| > t_c$ (or reject the null whenever $t_T^{ES} < t_c^l$ or $t_T^{ES} > t_c^r$ if $-t_c^l \neq t_c^r$) where t_c is a critical value. From Section 2.5.2, we know that t_T^{ES} has the standard fixed-b limiting distribution in Kiefer and Vogelsang (2005) whether the missing process is random or non-random. Therefore t_c is the 97.5% percentile of the fixed-b asymptotic distribution in Kiefer and Vogelsang (2005). Critical values can be computed either by directly simulating the limiting distribution (t_c^{ES}) or by using the naive *i.i.d.* bootstrap ($\{t_c^{ES-boot,l}, t_c^{ES-boot,r}\}$). We compute critical values using both methods. From the original sample of T_{ES} observed observations, $y_1^{ES}, \ldots, y_{T_{ES}}^{ES}$, we resample T_{ES} observations with replacement. Repeating this procedure 999 times we obtain *i.i.d.* bootstrap resamples for ES regression model which we denote $y_1^{ES\bullet B}, \ldots, y_{T_{ES}}^{ES\bullet B}$, $B = 1, \ldots, 999$. We compute the naive bootstrap t-statistic

$$t_T^{ES,B} = \frac{\sqrt{T_{ES}}(\hat{\beta}^B - \hat{\beta})}{\sqrt{\hat{\Omega}_{ES}^B}},$$

where $\hat{\beta}^B = \sum_{t=1}^{T} y_t^{\bullet B}/T_{ES}$ and $\hat{\Omega}_{ES}^B = 1/T_{ES} \sum_{t=1}^{T_{ES}} \sum_{s=1}^{T_{ES}} k(|t-s|/[bT_{ES}]) \hat{v}_t^{ES\bullet B} \hat{v}_s^{ES\bullet B}$ with $\hat{v}_t^{ES\bullet B} = y_t^{ES\bullet B} - \hat{\beta}^B$. Then $t_c^{ES-boot,l}$ is the 0.025 quantile and $t_c^{ES-boot,r}$ is the 0.975 quantile of $t_T^{ES,B}$ for $B = 1, \ldots, 999$.

Figures 2.35-2.64 show the empirical rejection probabilities computed from 10,000 replications using the ES regression model. For all four cases of missing processes defined in Section 2.4.1, the empirical rejection probabilities are computed using critical values obtained by the naive *i.i.d.* bootstrap and by simulating the fixed-b limiting distribution in Kiefer and Vogelsang (2005). We can see that ES regression model works reasonably well regardless of the methods used to compute the critical values even when a large portion of the data are missing. For example consider the World Wars missing process. When $T = 36$ (yearly case, Figure 2.35) and $T = 144$ (monthly case, Figure 2.38), one third of the data are missing. In these two cases for $\rho = 0, 0.3$, there are mild over-rejection problems if any. For $T = 36$ some over-rejection problems appear with $\rho = 0.6$ and become severe with $\rho = 0.9$. For $T = 144$ over-rejection problems are much less severe for $\rho = 0.6, 0.9$. Similar patterns hold for initially scarce and Bernoulli missing processes. This over-rejection tendency when the data is highly correlated is something that is routinely found when no observations are missing.

In addition we see in all Figures 2.35-2.64 that rejection rates from the two critical values are close together nearly all the time. Minor exceptions occur when the sample size is small and the latent process is highly correlated in which case the two rejection rates show some differences. For the World War missing process with $T = 36$ (Figure 2.35), one third of data are missing and thus we only have $T_{ES} = 24$. For initially scarce data with $N_Q = 12$ and $N_M = 12$ (Figure 2.41), we have 24 observations missing out of 46 ($T_{ES} = 22$). For the random and non-random Bernoulli(0.3) missing process with $T = 50$ (Figure 2.47) we have $T_{ES} \approx 15$. It is when T_{ES} is very small that we can see some difference between the two empirical rejection rates but only when $\rho = 0.9$. This is not surprising because with small T_{ES} and ρ close to 1 the asymptotic approximation provided fixed-b is likely to be inaccurate. When there is a difference between the two rejection probabilities, it is always the case that the naive *i.i.d*. bootstrap has the better size properties. See Figures 2.35,2.41,2.47 and 2.56. Ultimately, our simulation results suggest that in the presence of missing observations, one does well by ignoring missing observations and computing critical values using the naive *i.i.d*. bootstrap.

2.5.4 Comparison of AM and ES Statistics

Figures 2.65-2.94 compare empirical rejection probabilities of the AM series approach to those of the ES regression approach. In Section 2.4.3 we found that for the AM series approach the naive *i.i.d*. bootstrap conditional on the locations of missing observations always performs no worse than directly simulating the asymptotic fixed-b critical values. This is true whether the missing process is random or non-random. Similarly, we found in Section 2.5.3 that the naive *i.i.d*. bootstrap always performs no worse than simulating the standard fixed-b critical values for the ES regression approach as well. Hence, to make comparisons between the AM series approach and the ES regression approach, critical values are computed by the naive *i.i.d*. bootstrap conditional on the locations of missing observations. For the most part the two approaches give similar rejections. However, the

AM series approach has a tendency to outperform the ES regression approach when the latent process is highly serially correlated and this tendency is stronger when the sample size is small. In other words, when the stationarity asymptotic theory is more likely to break down, it is more likely that the AM series approach outperforms the ES regression approach.

2.6 CONCLUSION

In this chapter we discussed the properties of HAC robust test statistics in time series regression setting when there is missing data. We considered two regression models, AM series and ES regression model, both for the random and non-random missing processes. Depending on the regression model used and the missing process being random or non-random, HAC robust tests have different asymptotic limits. From simulation studies we find that the naive *i.i.d*. bootstrap is the most effective and practical way to obtain fixed-b critical values in the presence of missing observations especially when the bootstrap conditions on the locations of the missing data.

Figure 2.3: Missing due to World War I and World War II : Yearly data

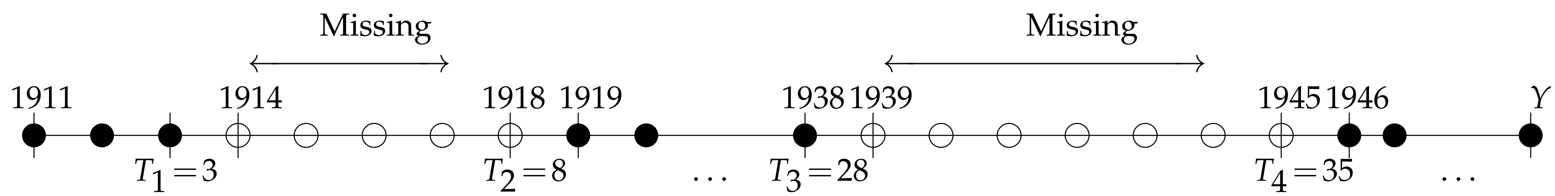


Figure 2.4: Initially Scarce Data

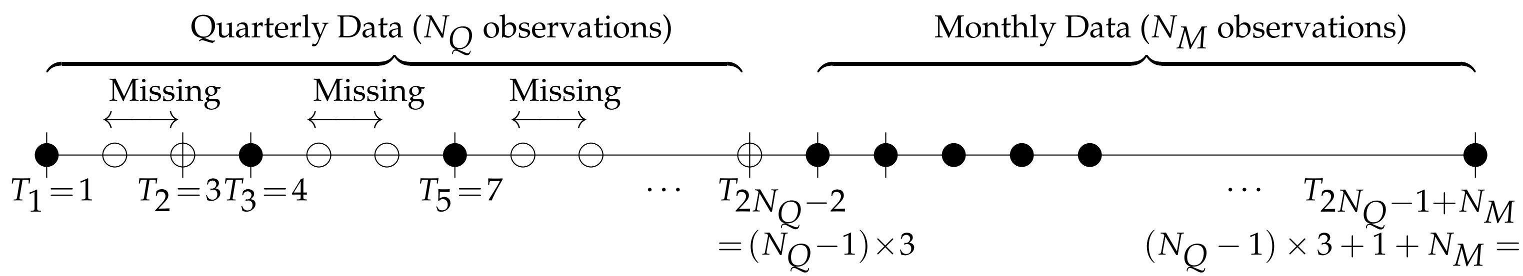


Figure 2.5: AM Series - World War (yearly), Bartlett, $T = 36$

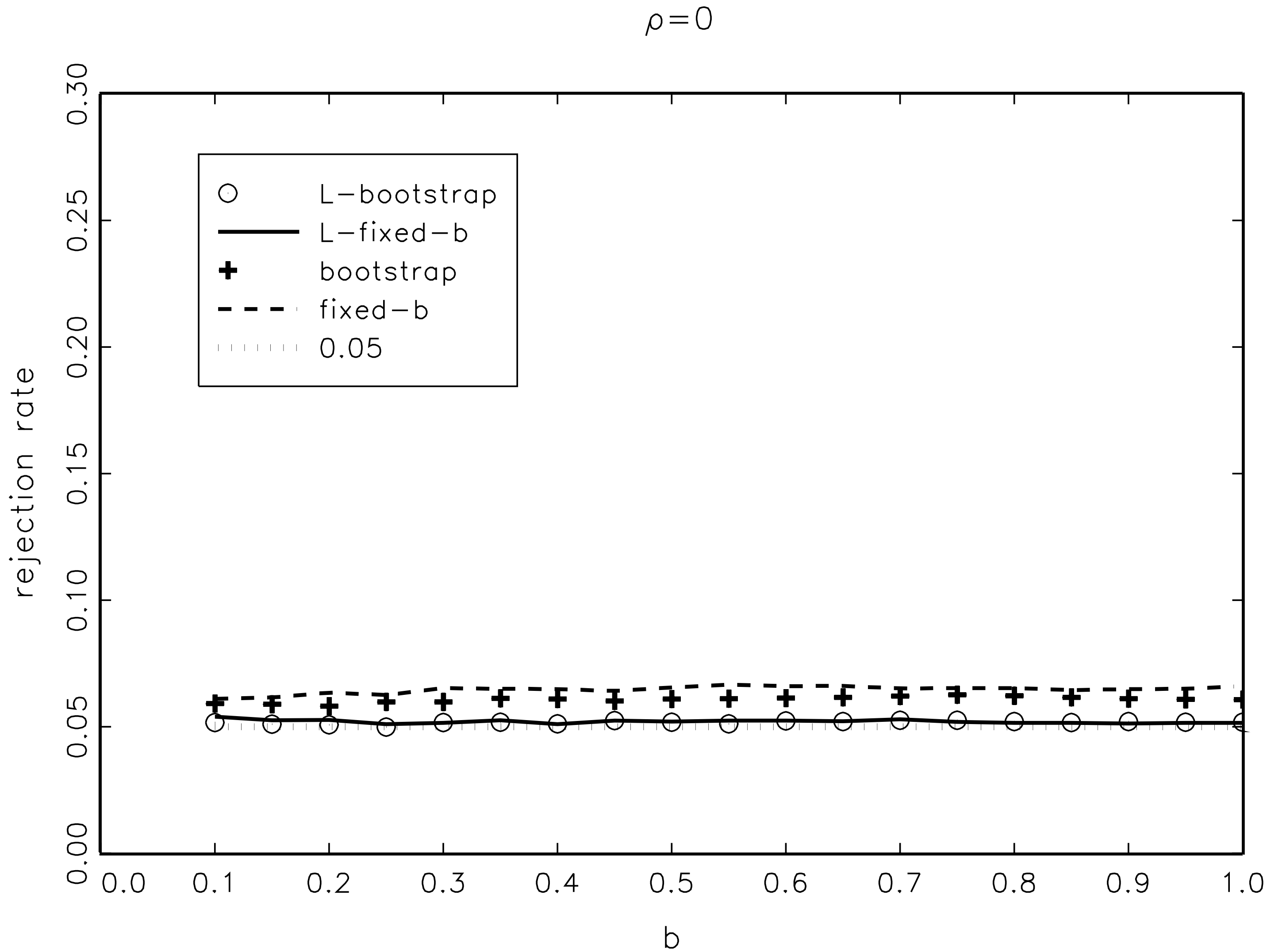


Figure 2.5: (cont'd)

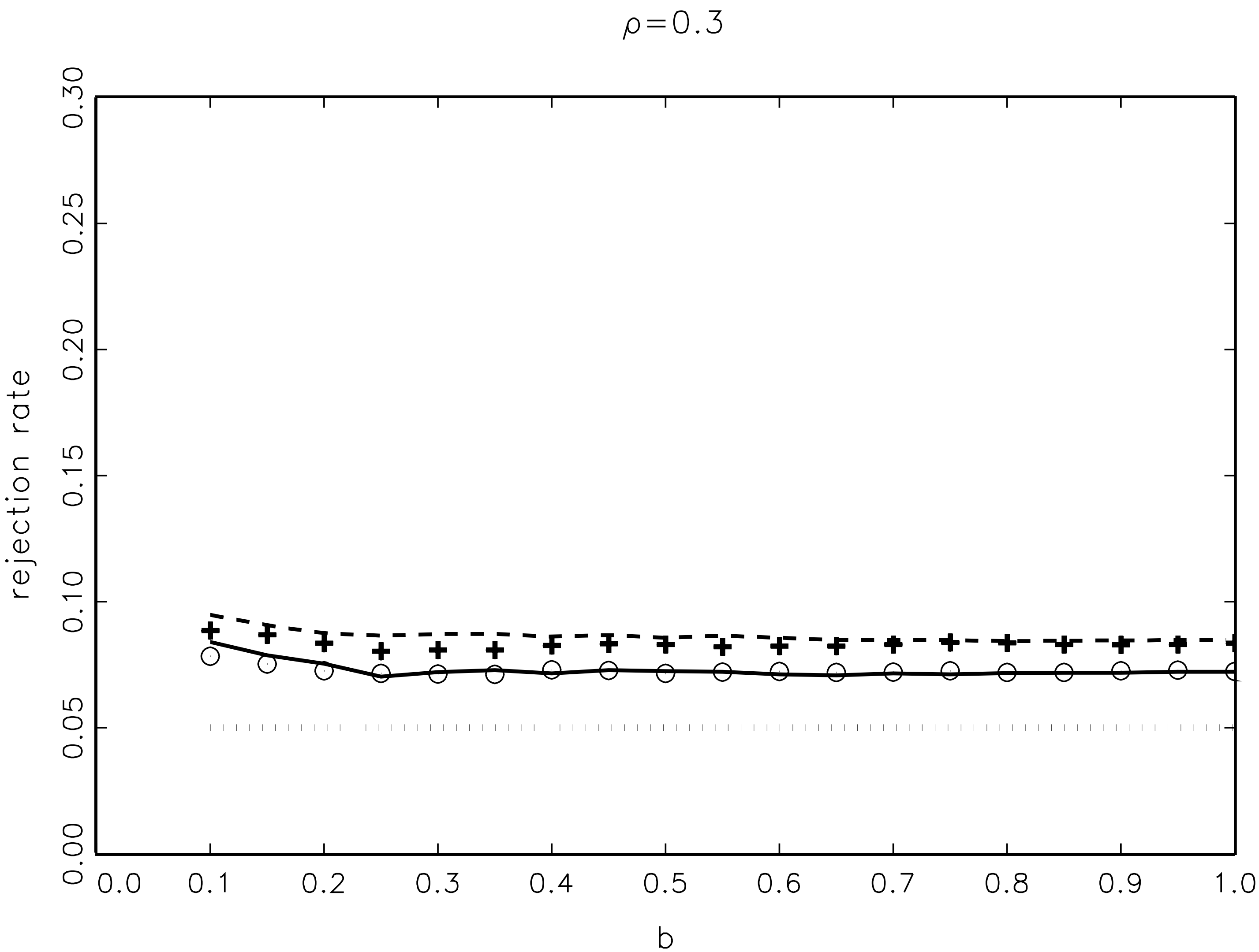


Figure 2.5: (cont'd)

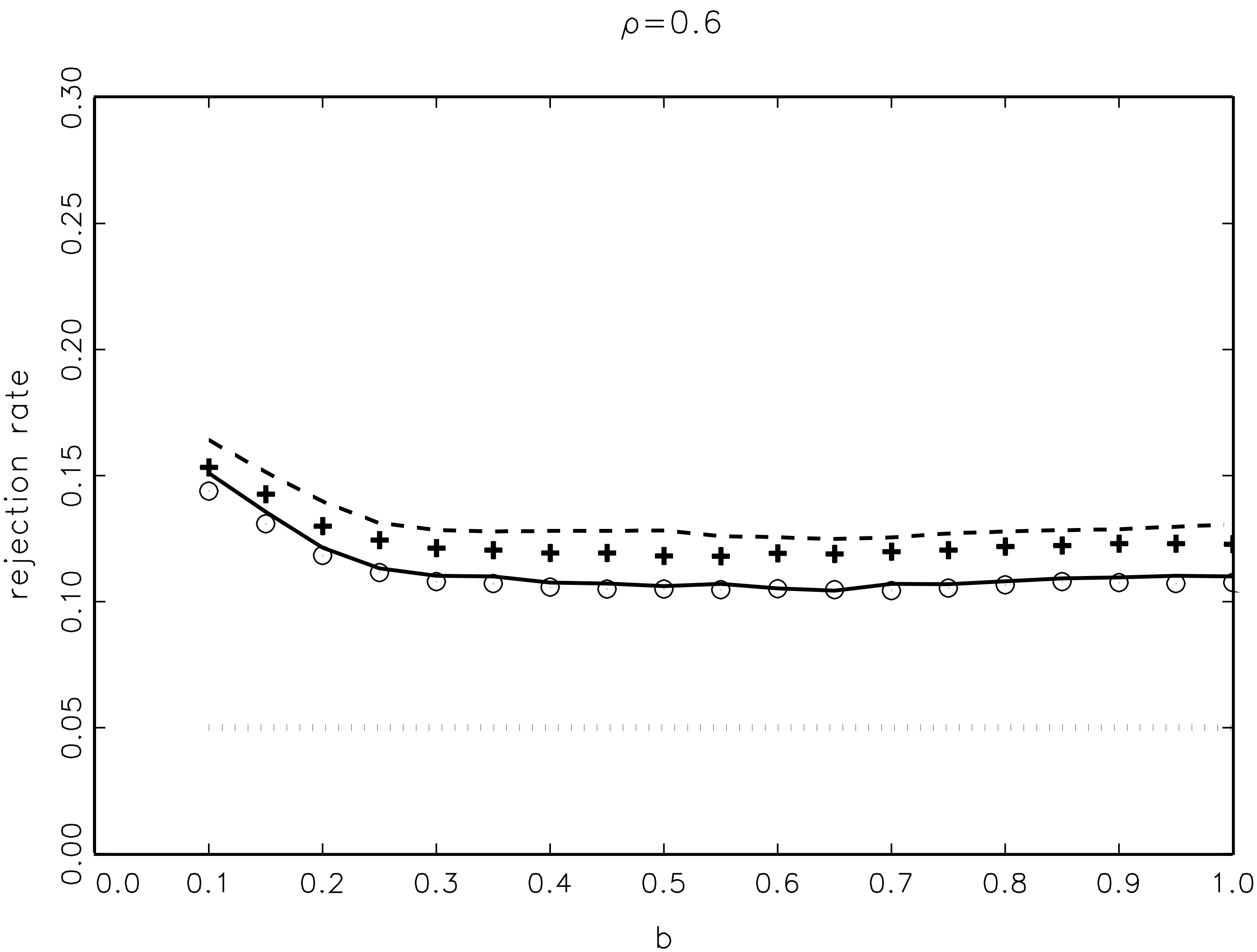


Figure 2.5: (cont'd)

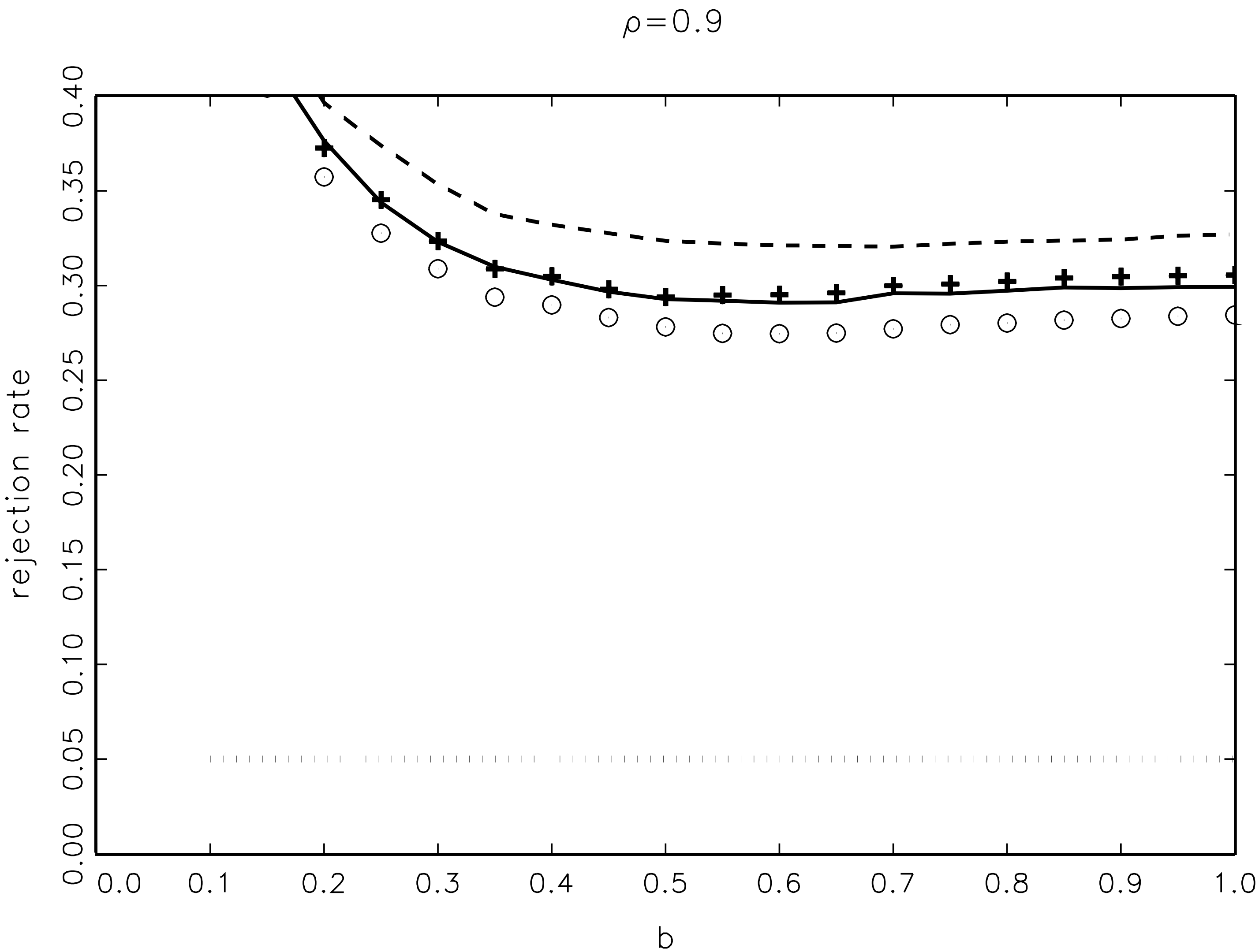


Figure 2.6: AM Series - World War (yearly), Bartlett, $T = 48$

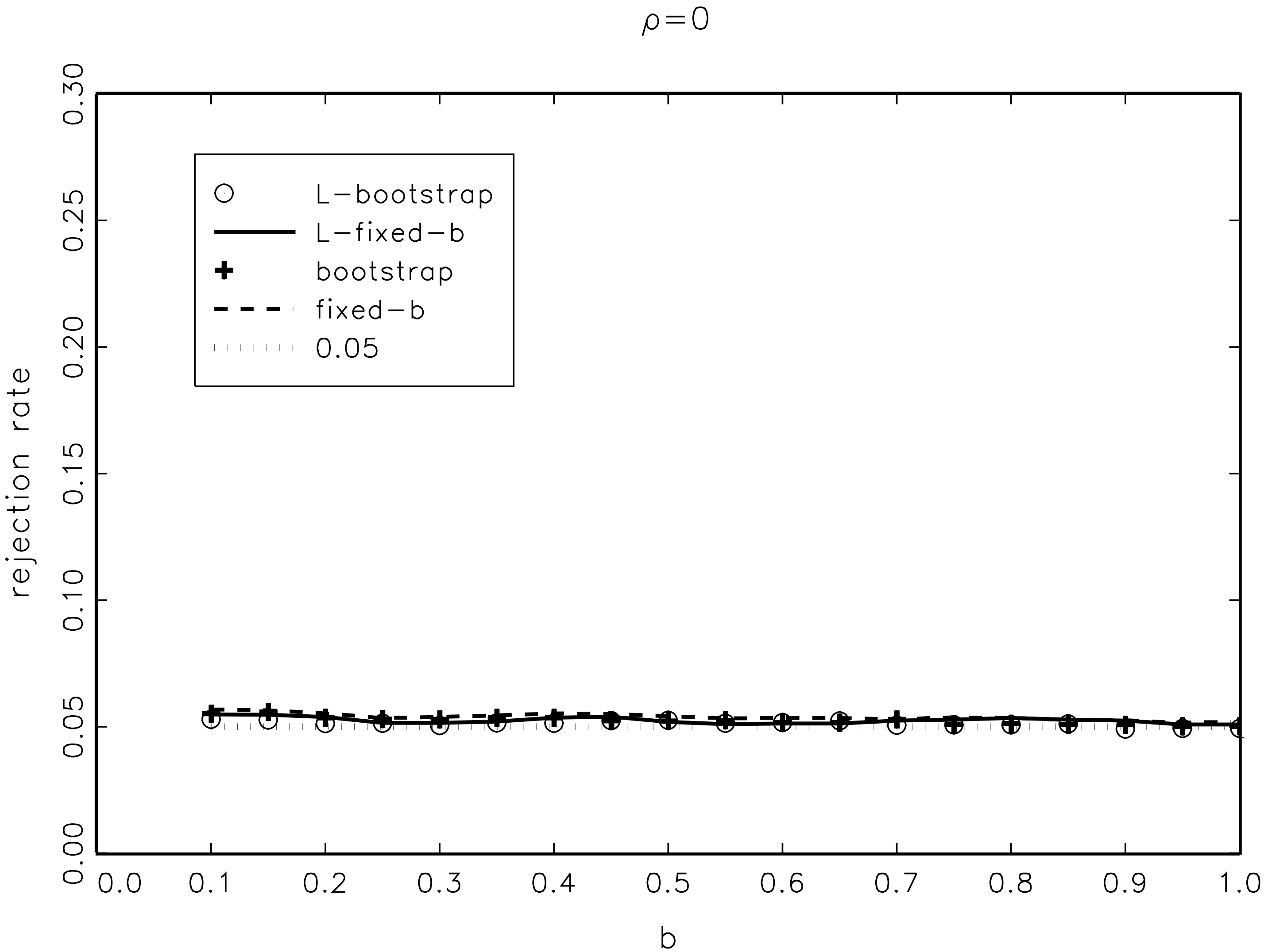


Figure 2.6: (cont'd)

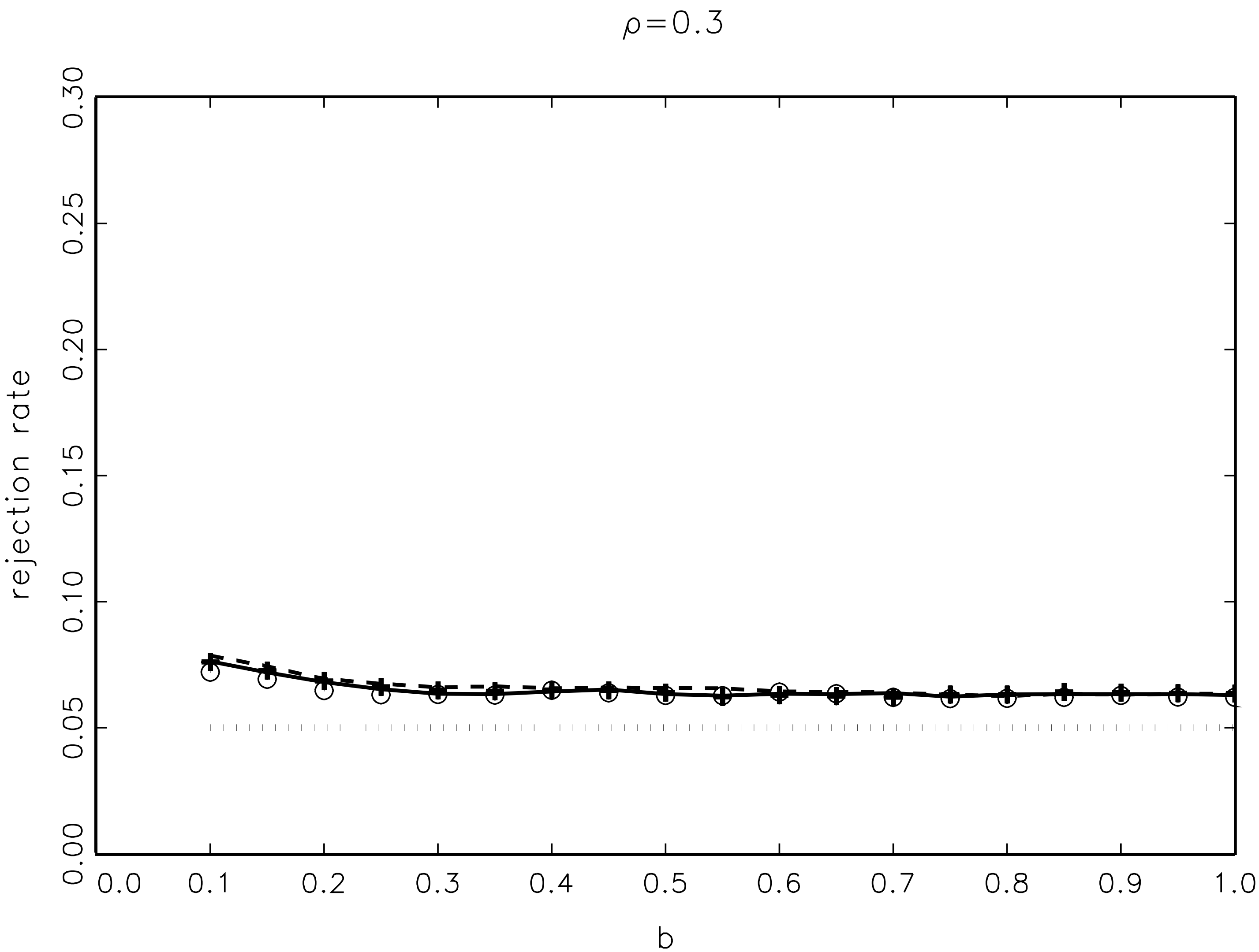


Figure 2.6: (cont'd)

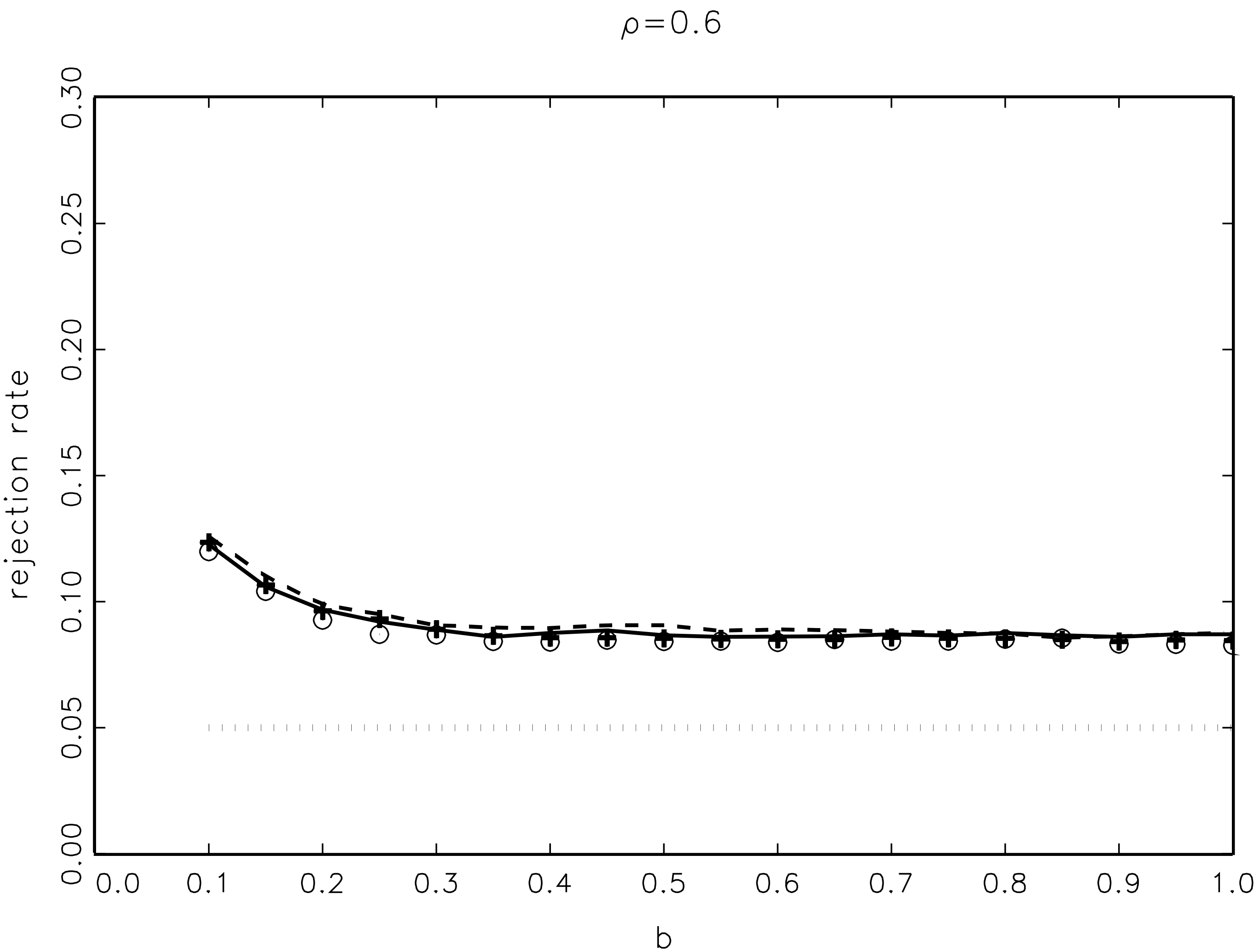


Figure 2.6: (cont'd)

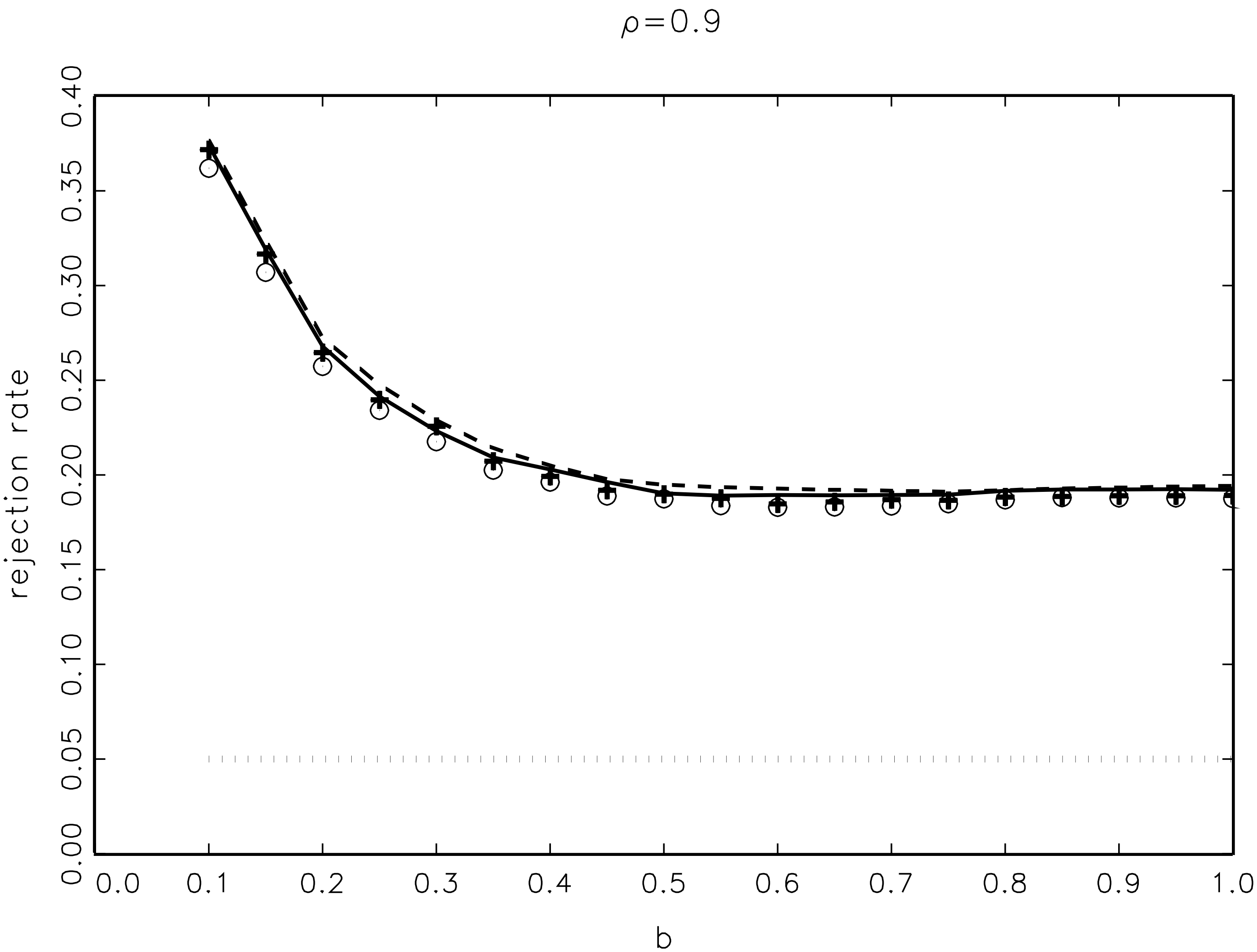


Figure 2.7: AM Series - World War (yearly), Bartlett, $T = 60$

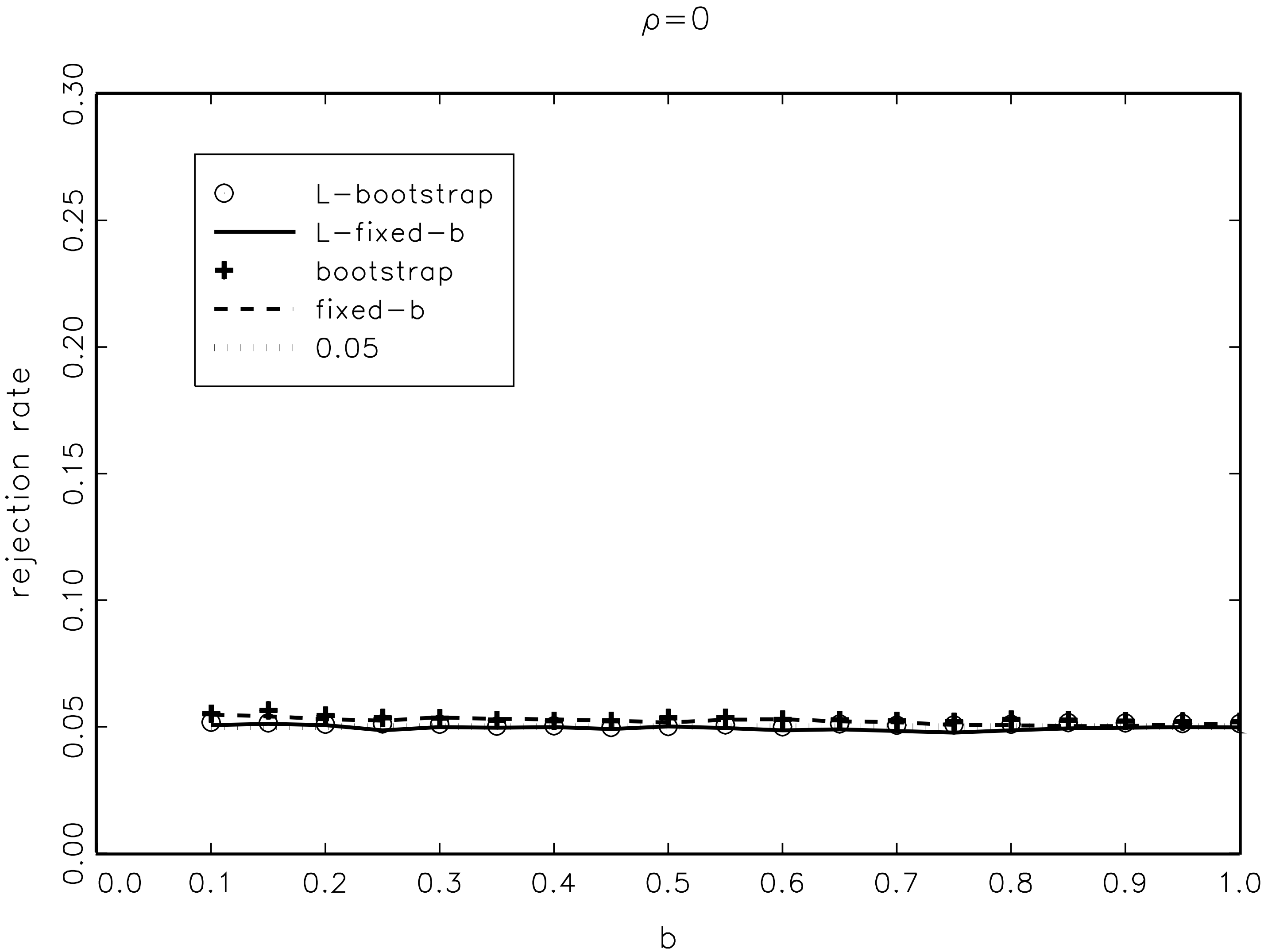


Figure 2.7: (cont’d)

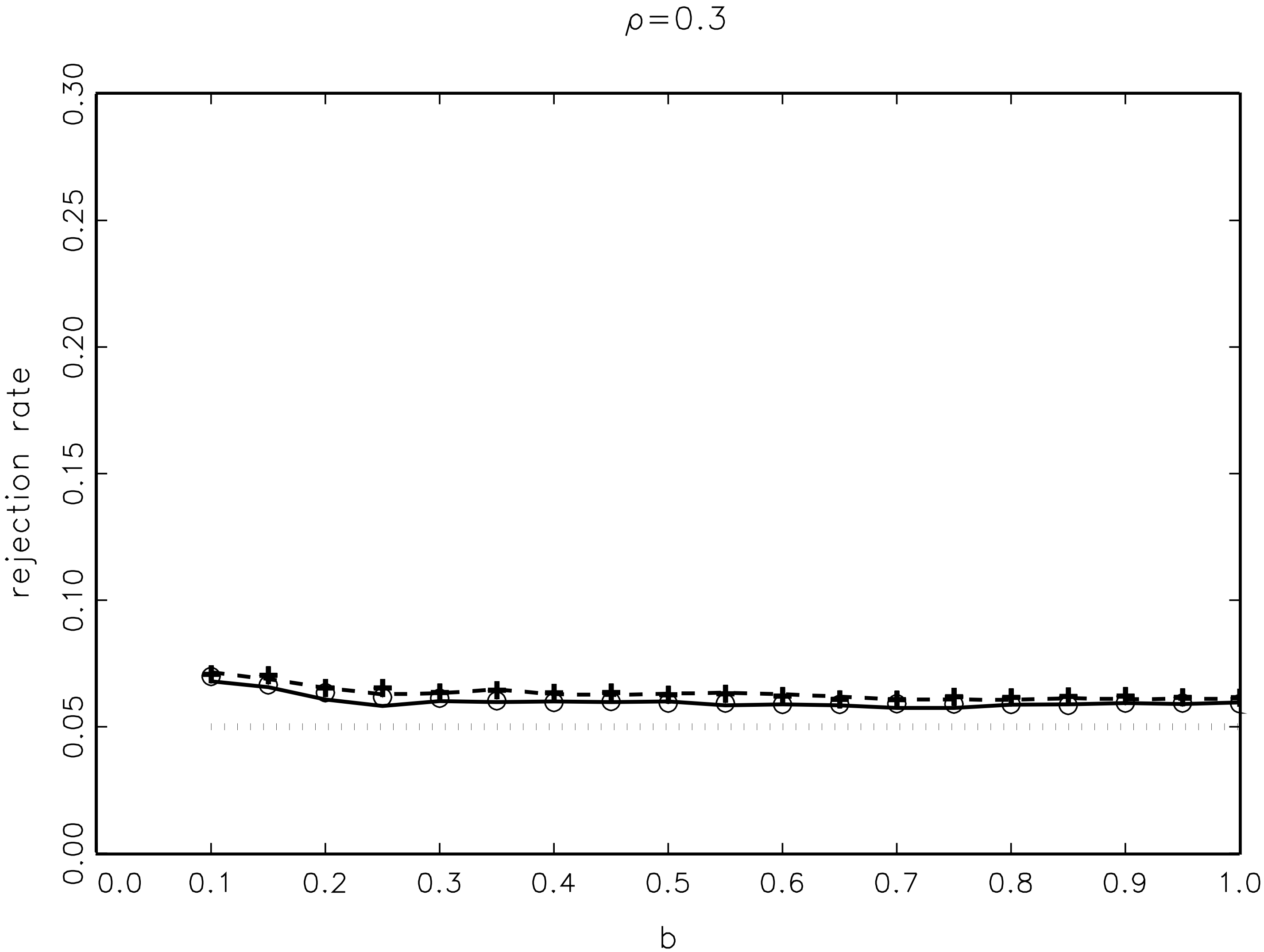


Figure 2.7: (cont'd)

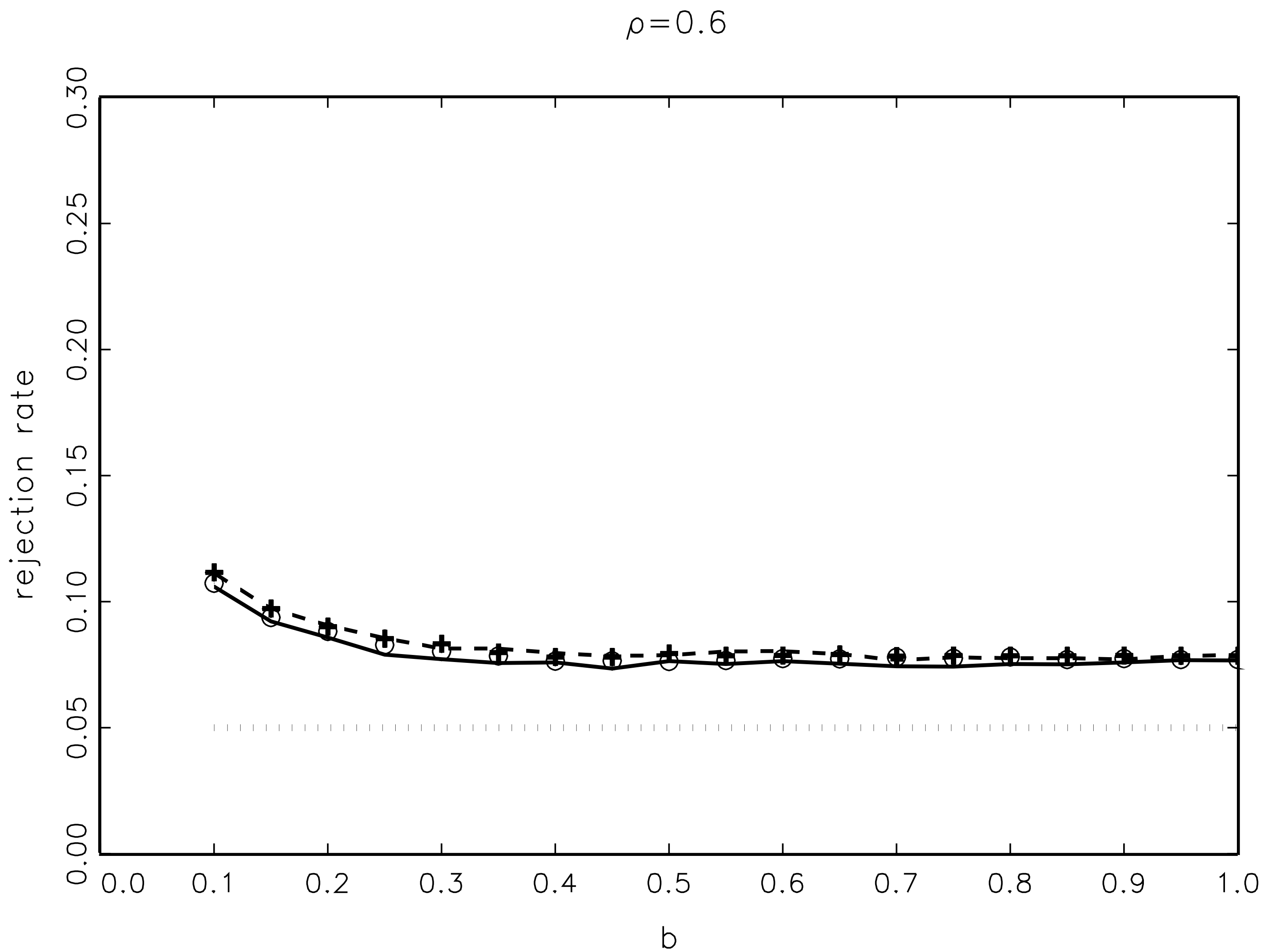


Figure 2.7: (cont'd)

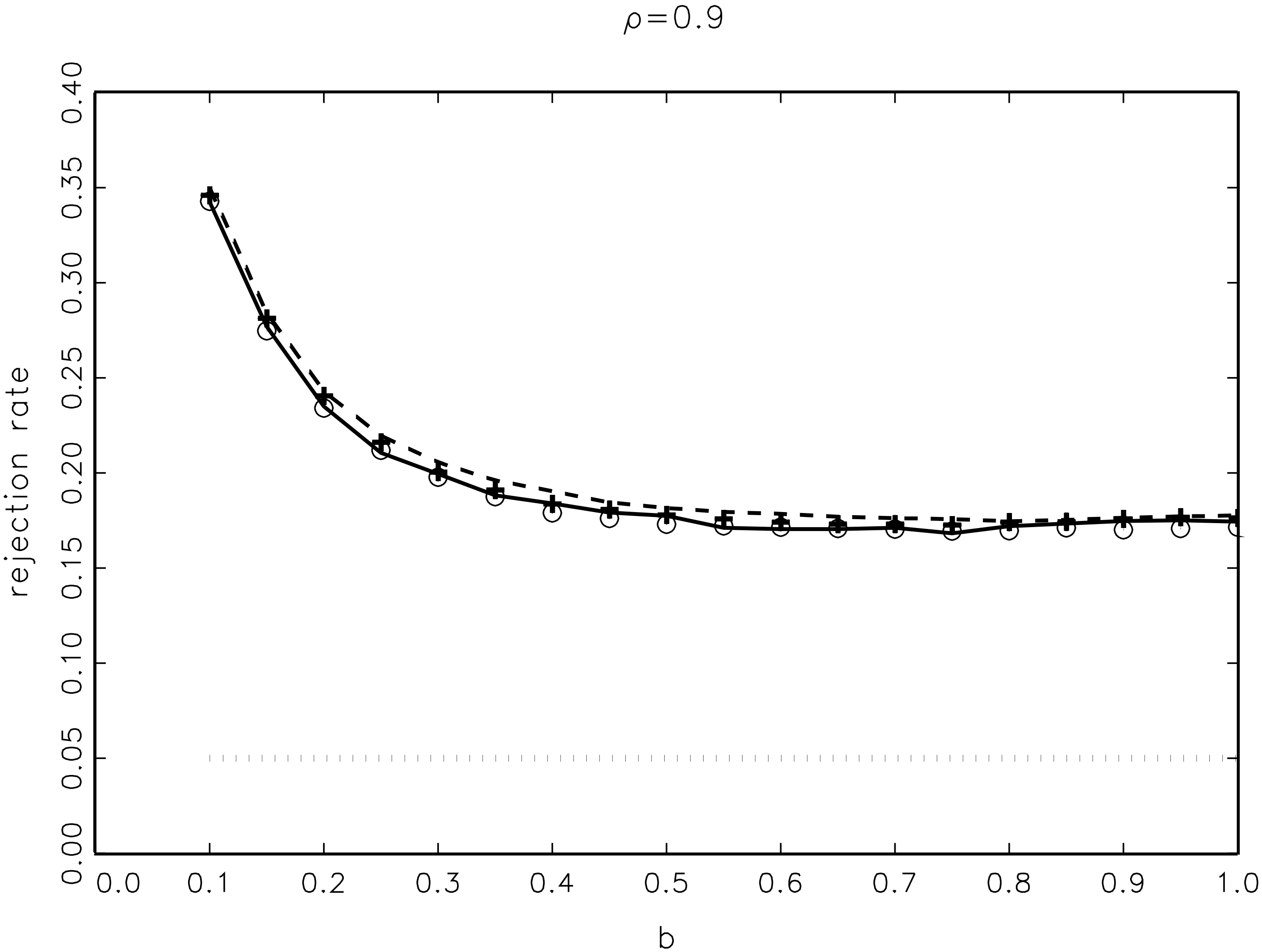


Figure 2.8: AM Series - World War (quarterly), Bartlett, $T = 144$

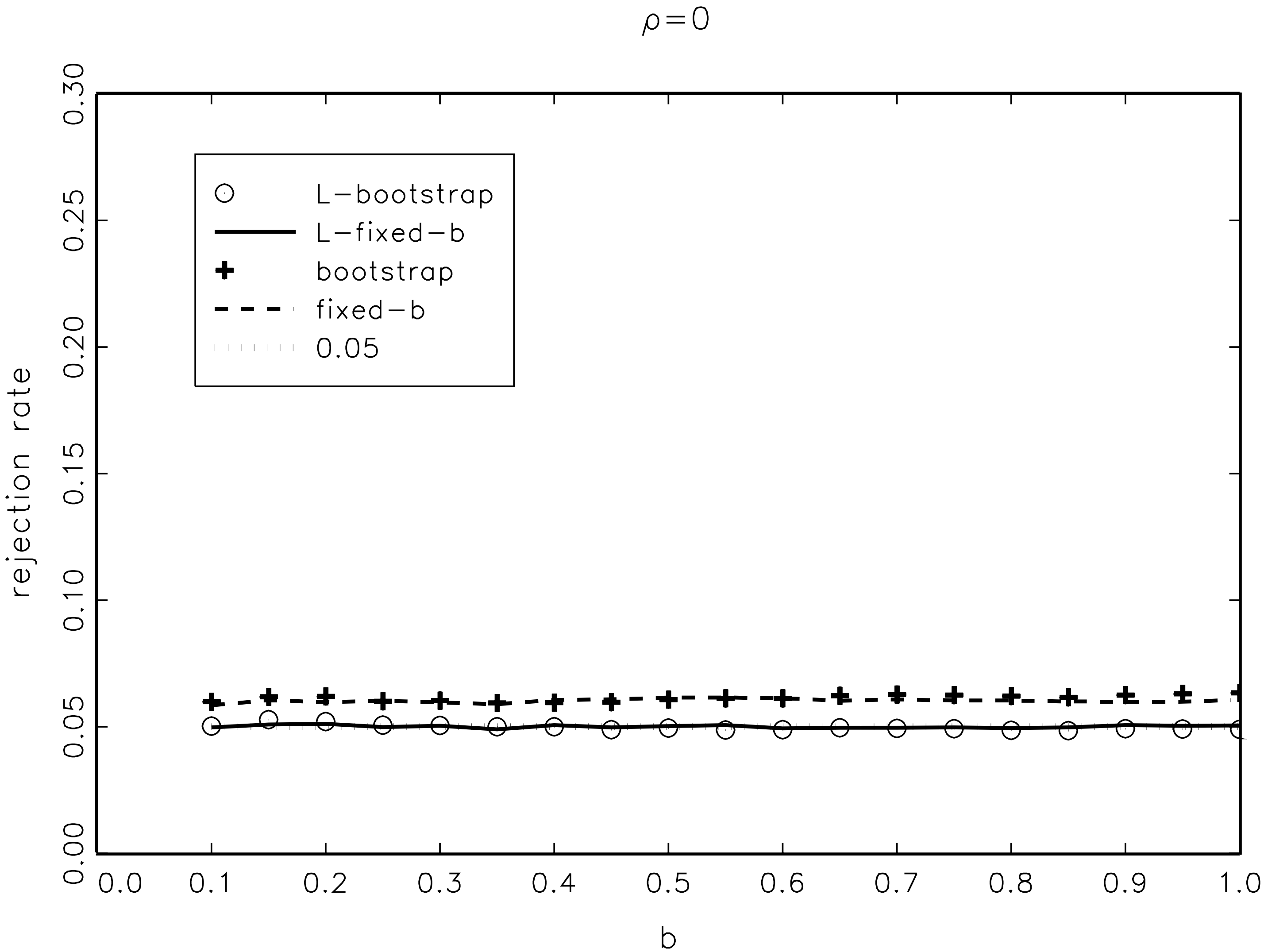


Figure 2.8: (cont'd)

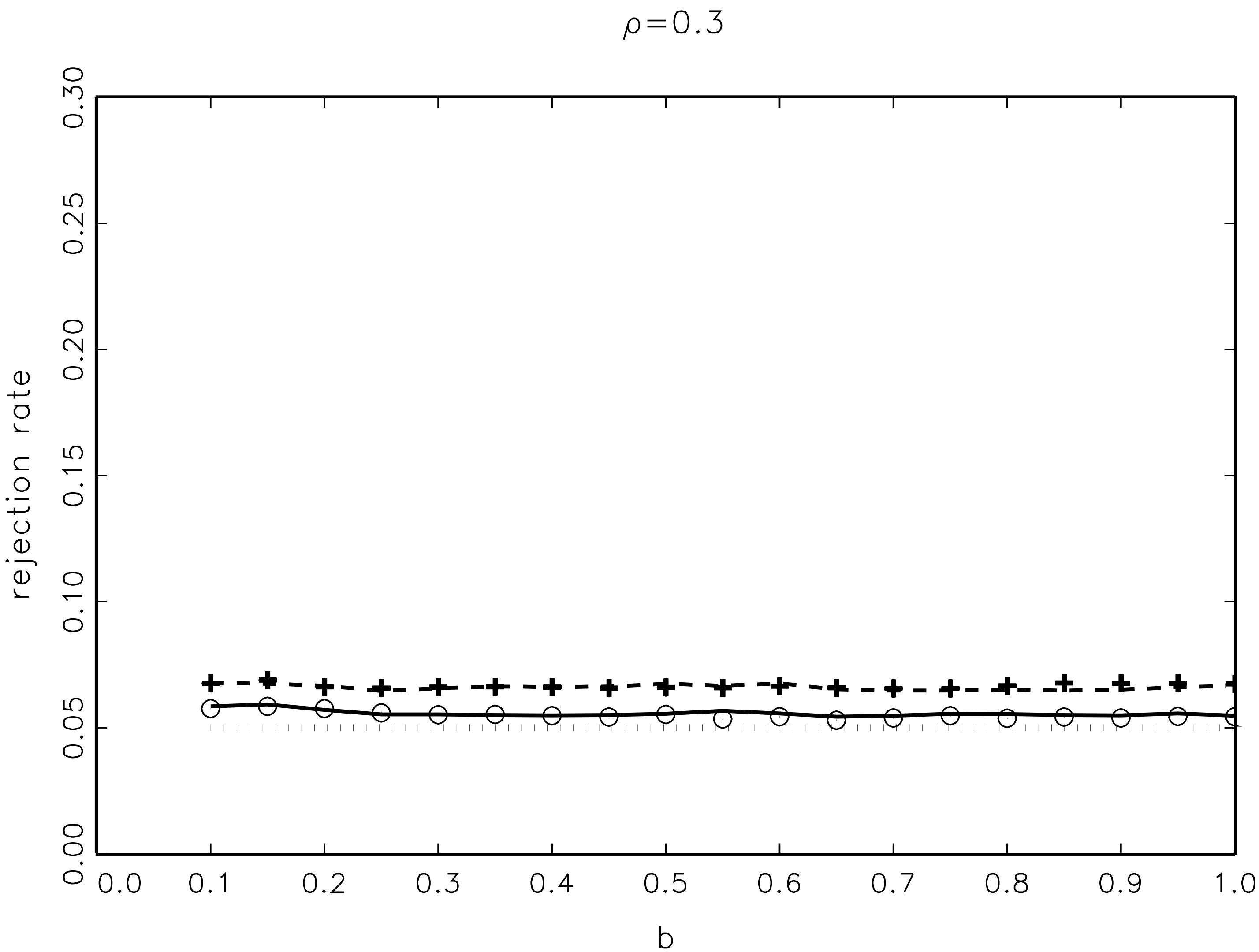


Figure 2.8: (cont'd)

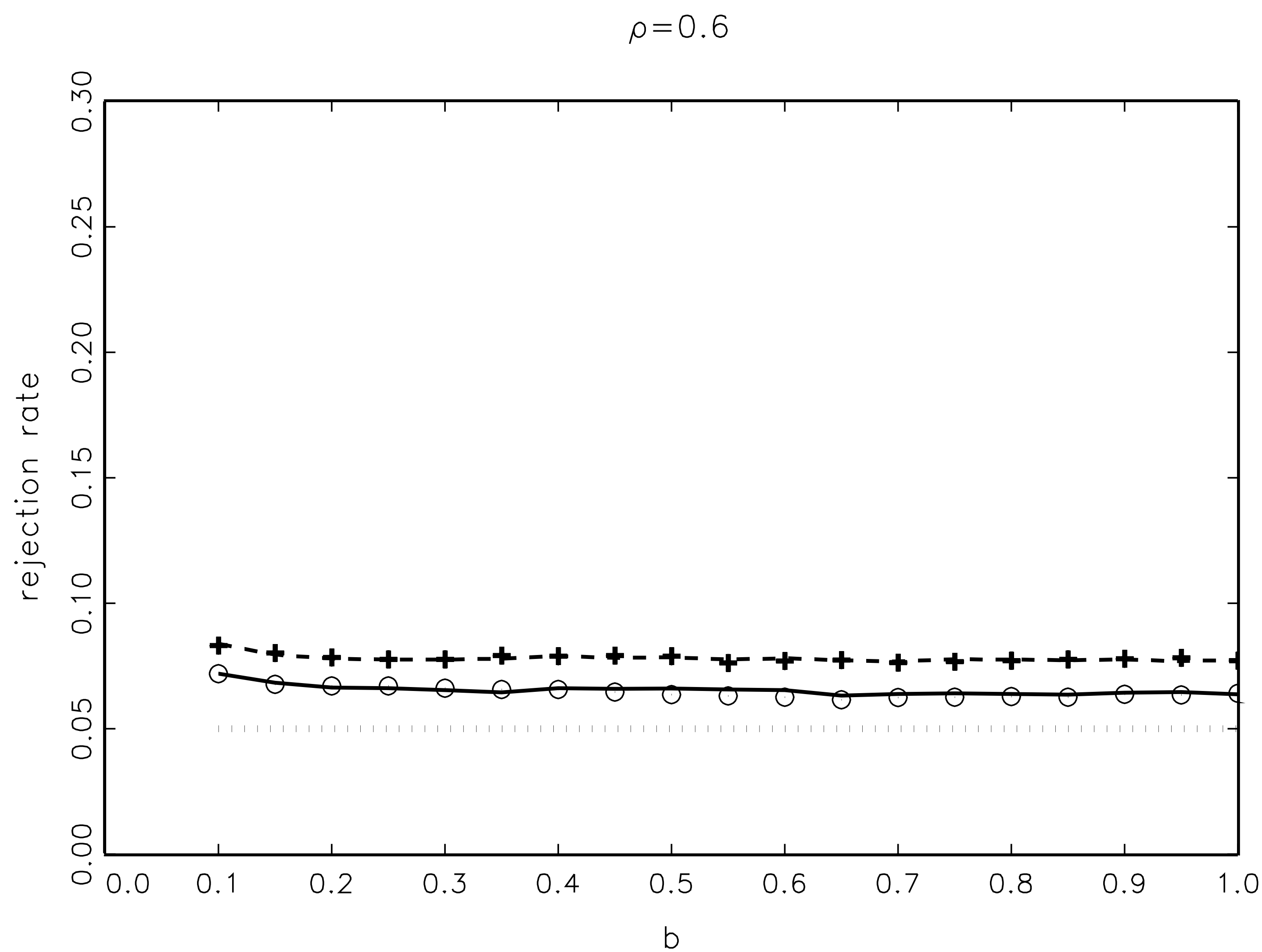


Figure 2.8: (cont'd)

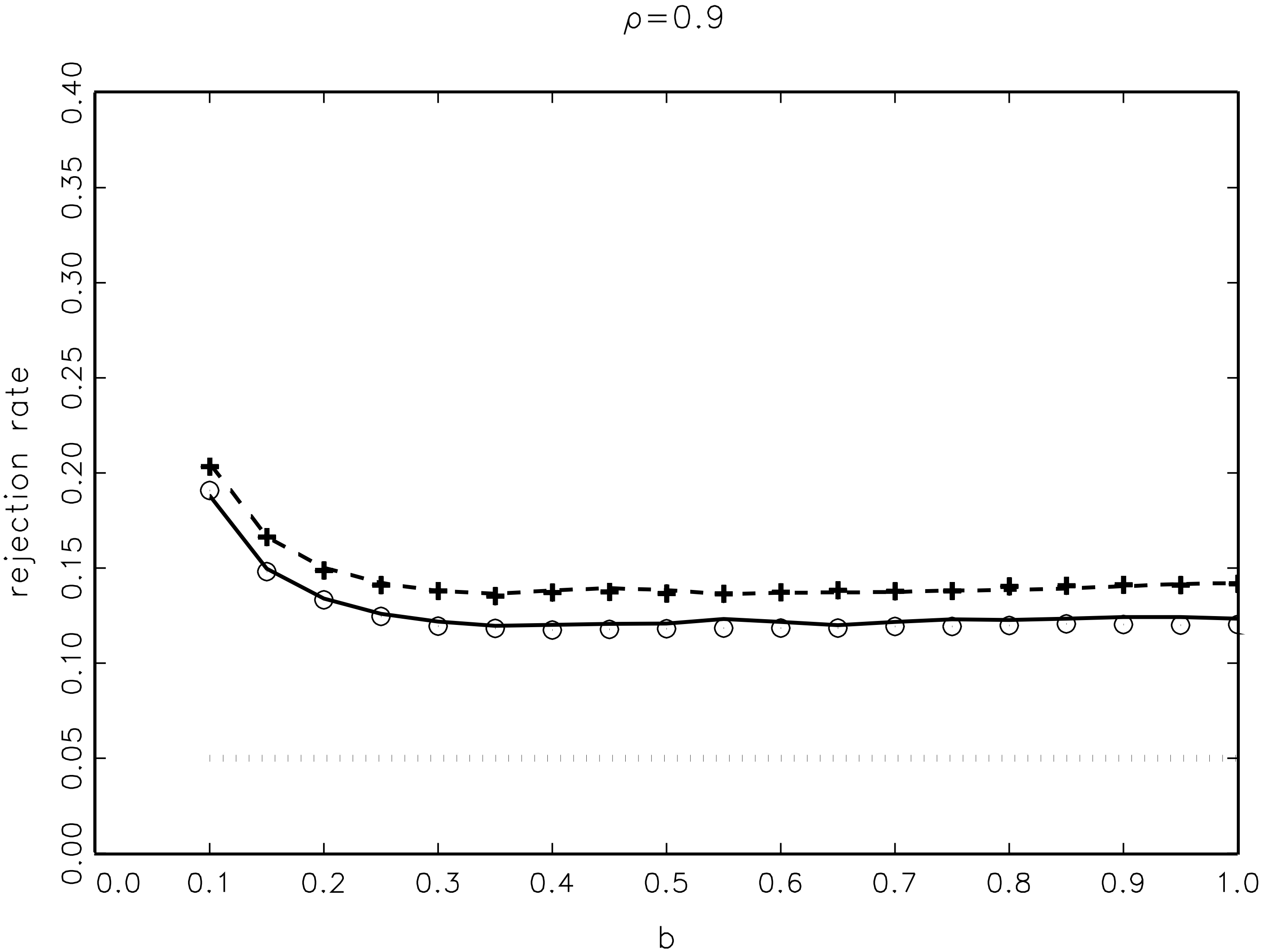


Figure 2.9: AM Series - World War (quarterly), Bartlett, $T = 192$

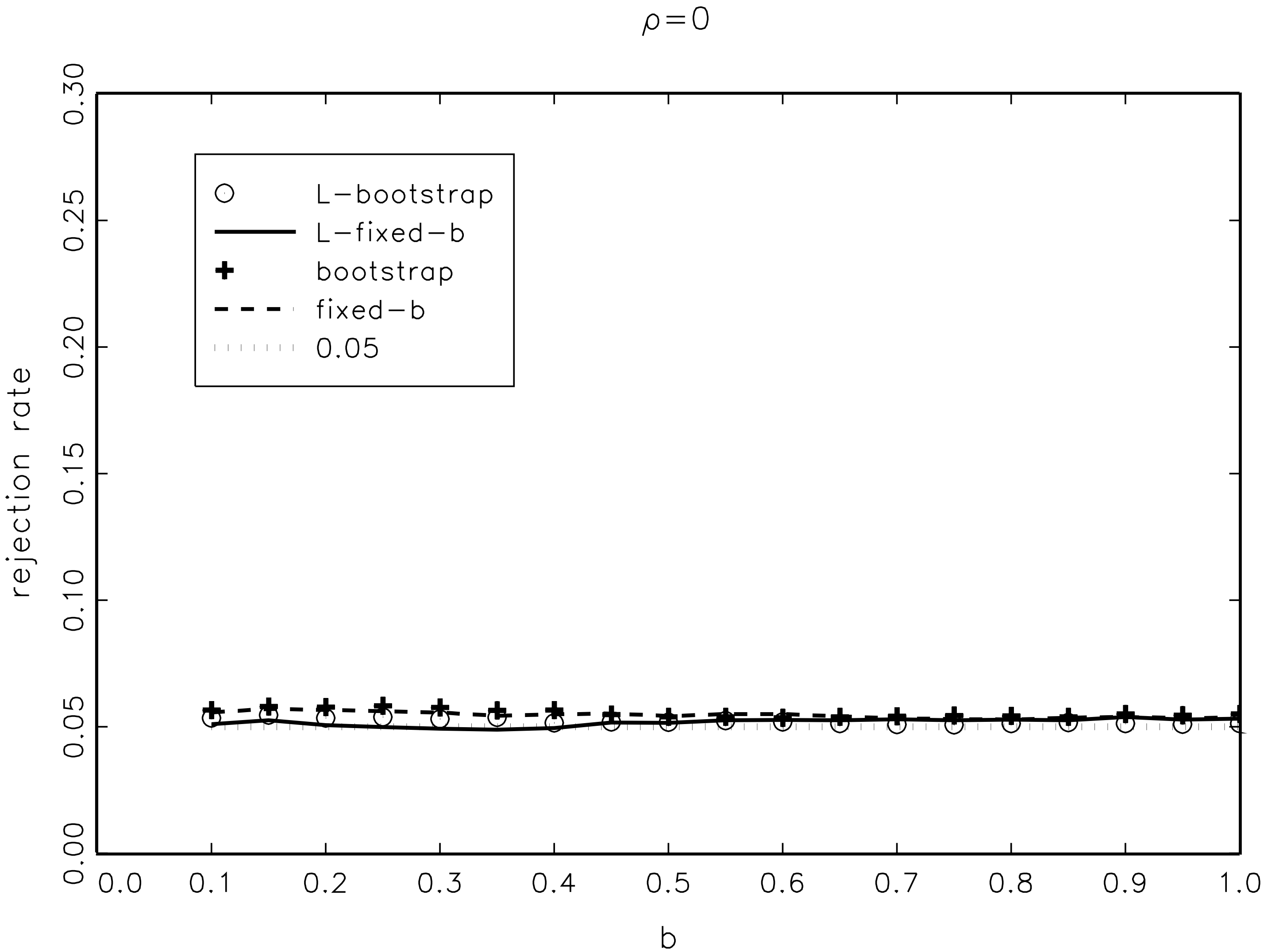


Figure 2.9: (cont'd)

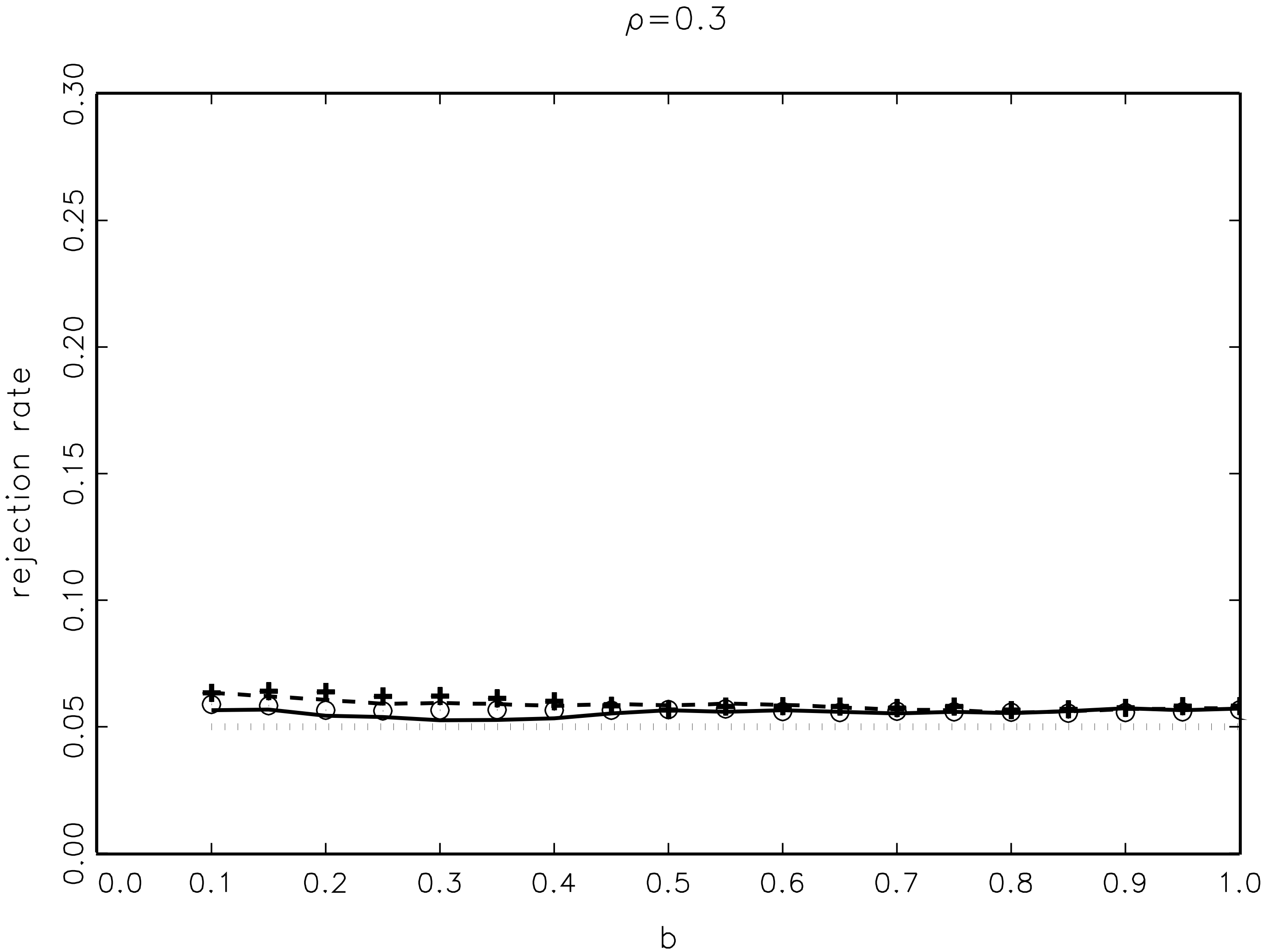


Figure 2.9: (cont'd)

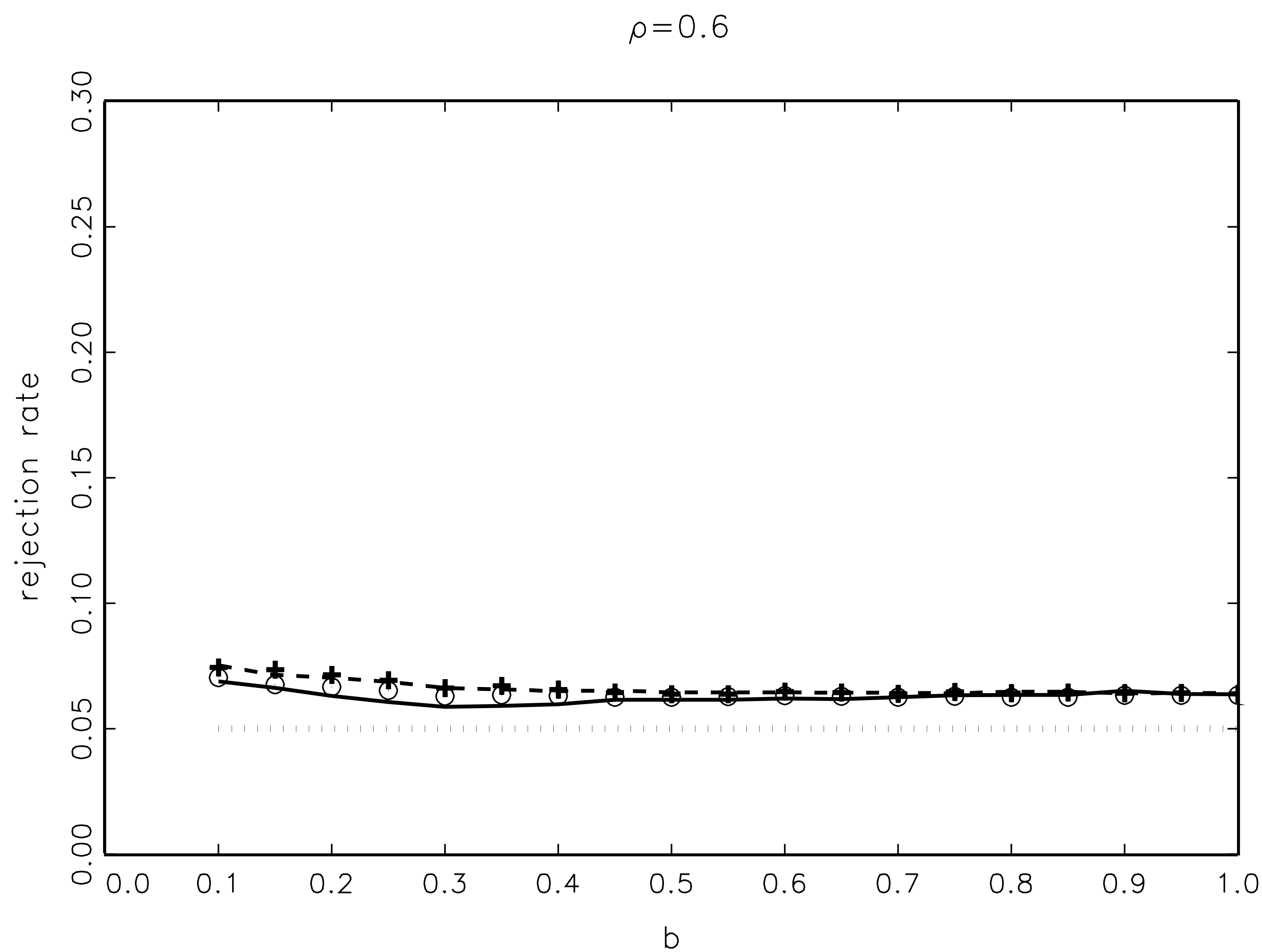


Figure 2.9: (cont'd)

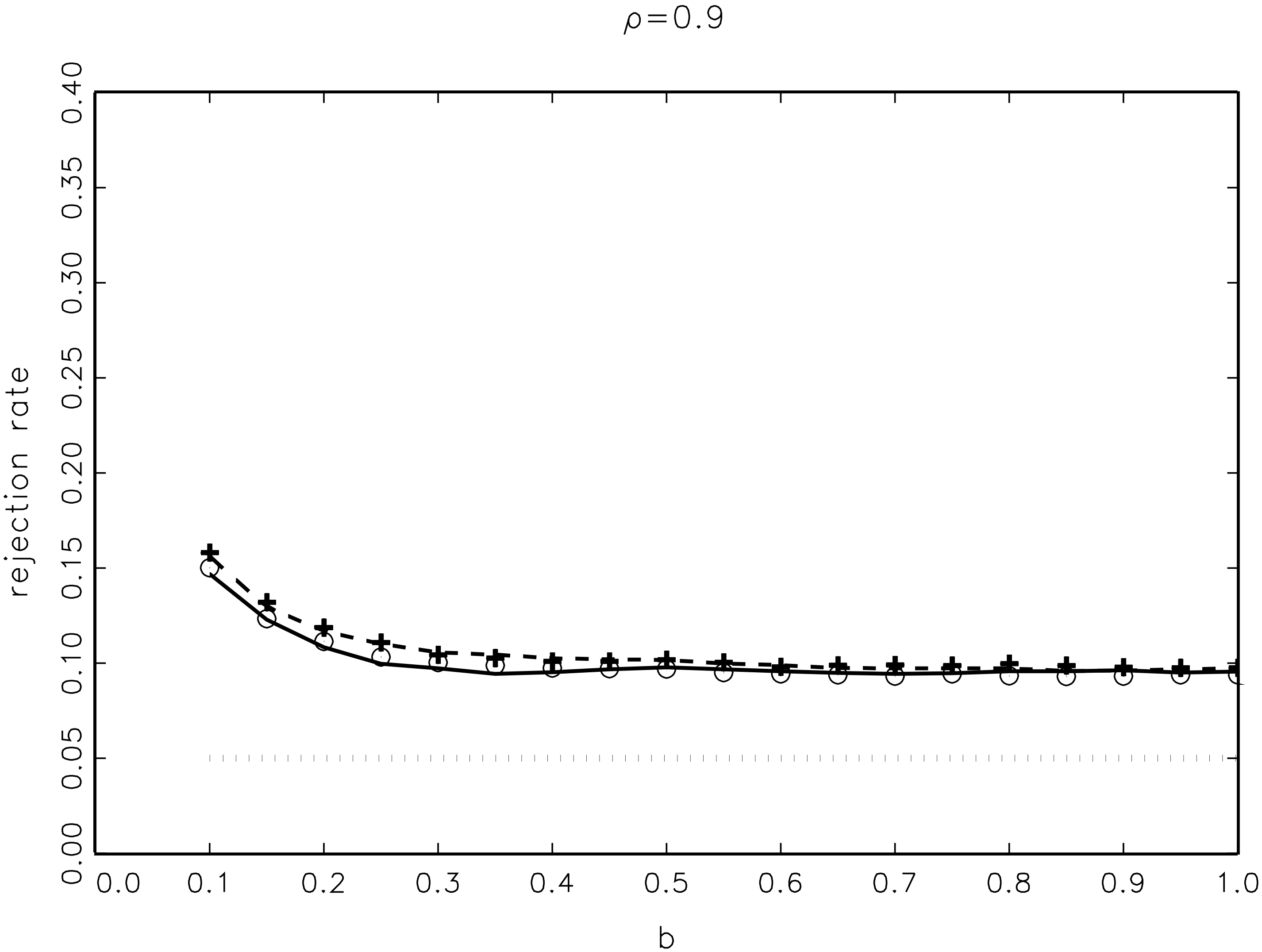


Figure 2.10: AM Series - World War (quarterly), Bartlett, $T = 240$

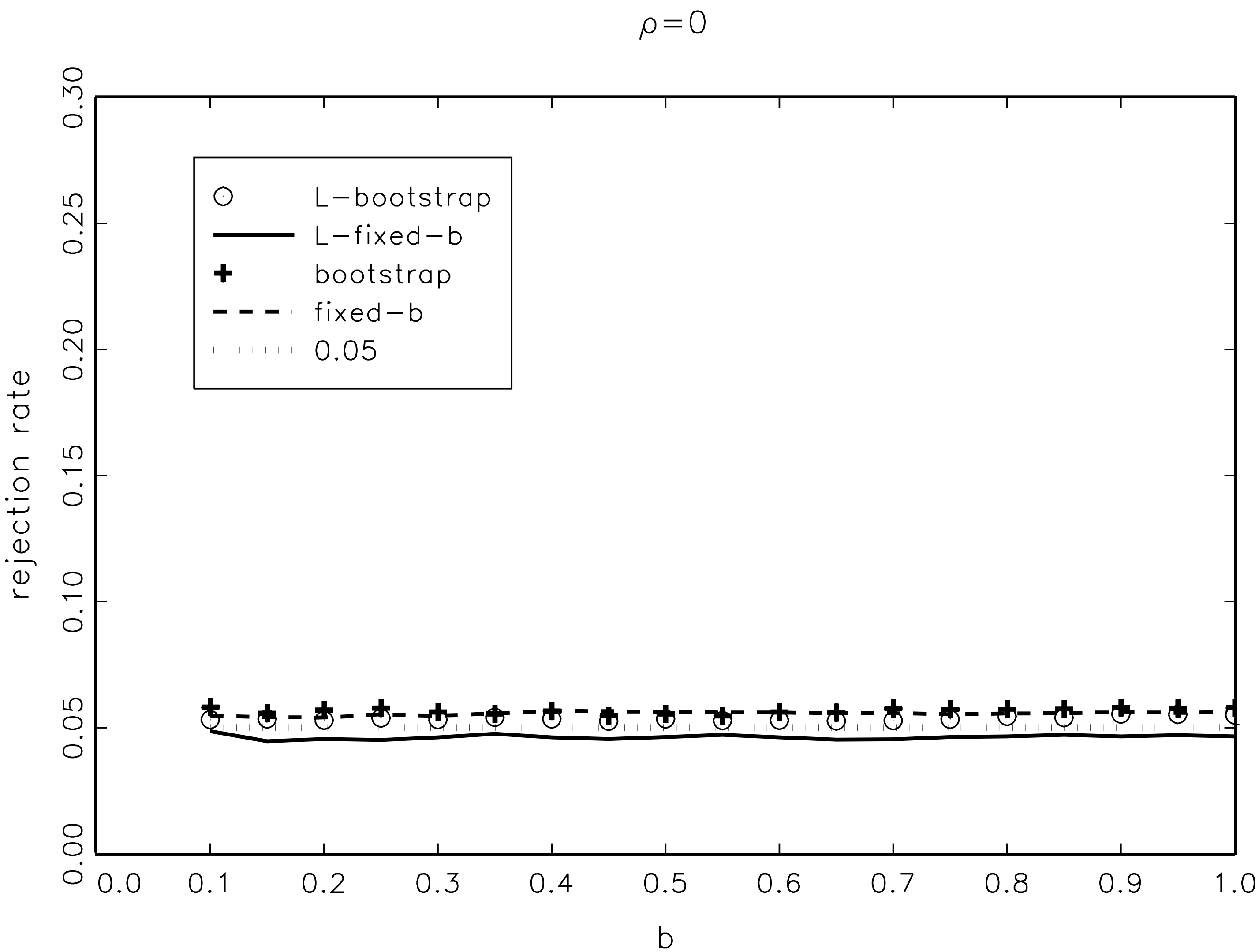


Figure 2.10: (cont'd)

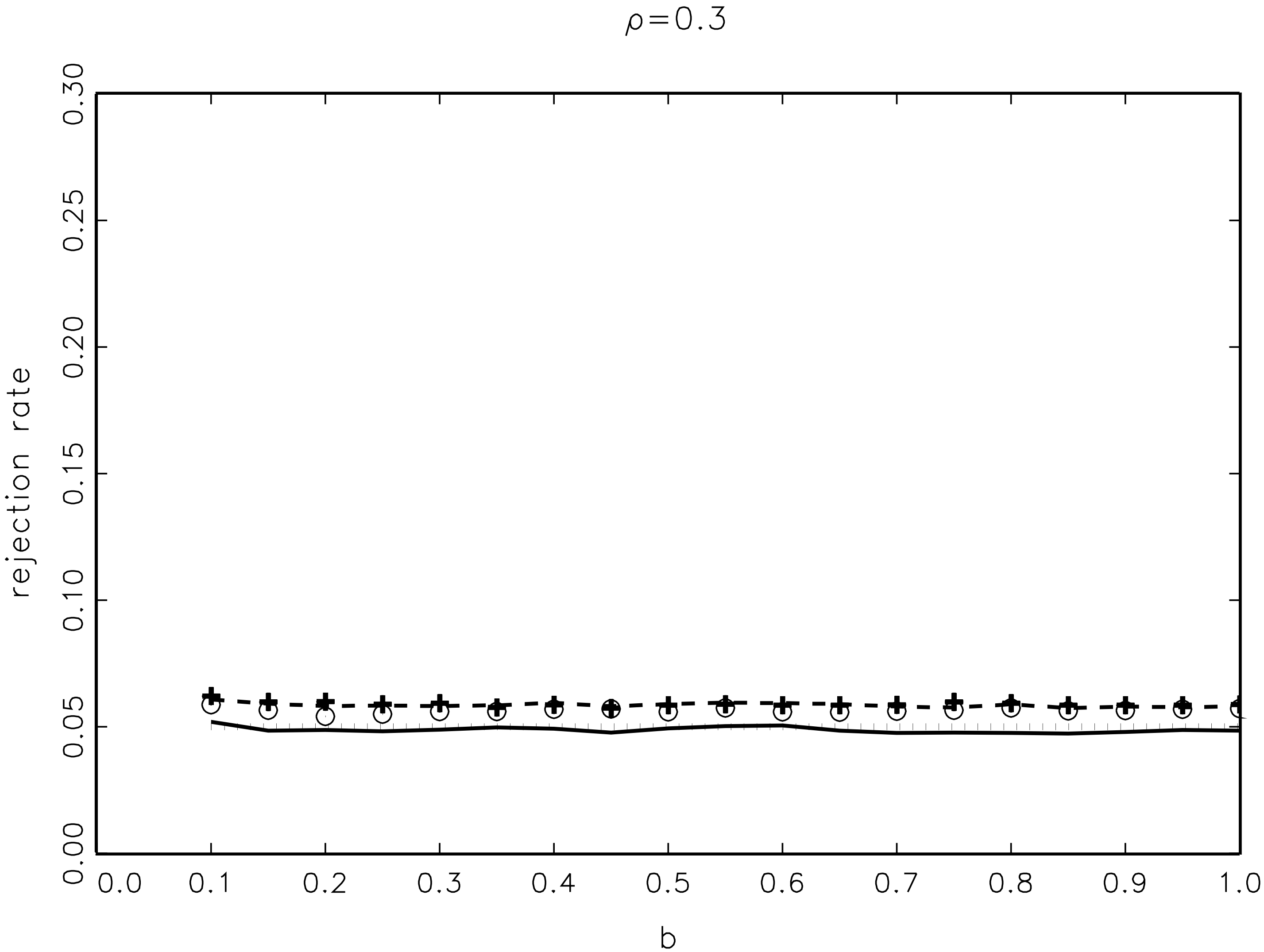


Figure 2.10: (cont'd)

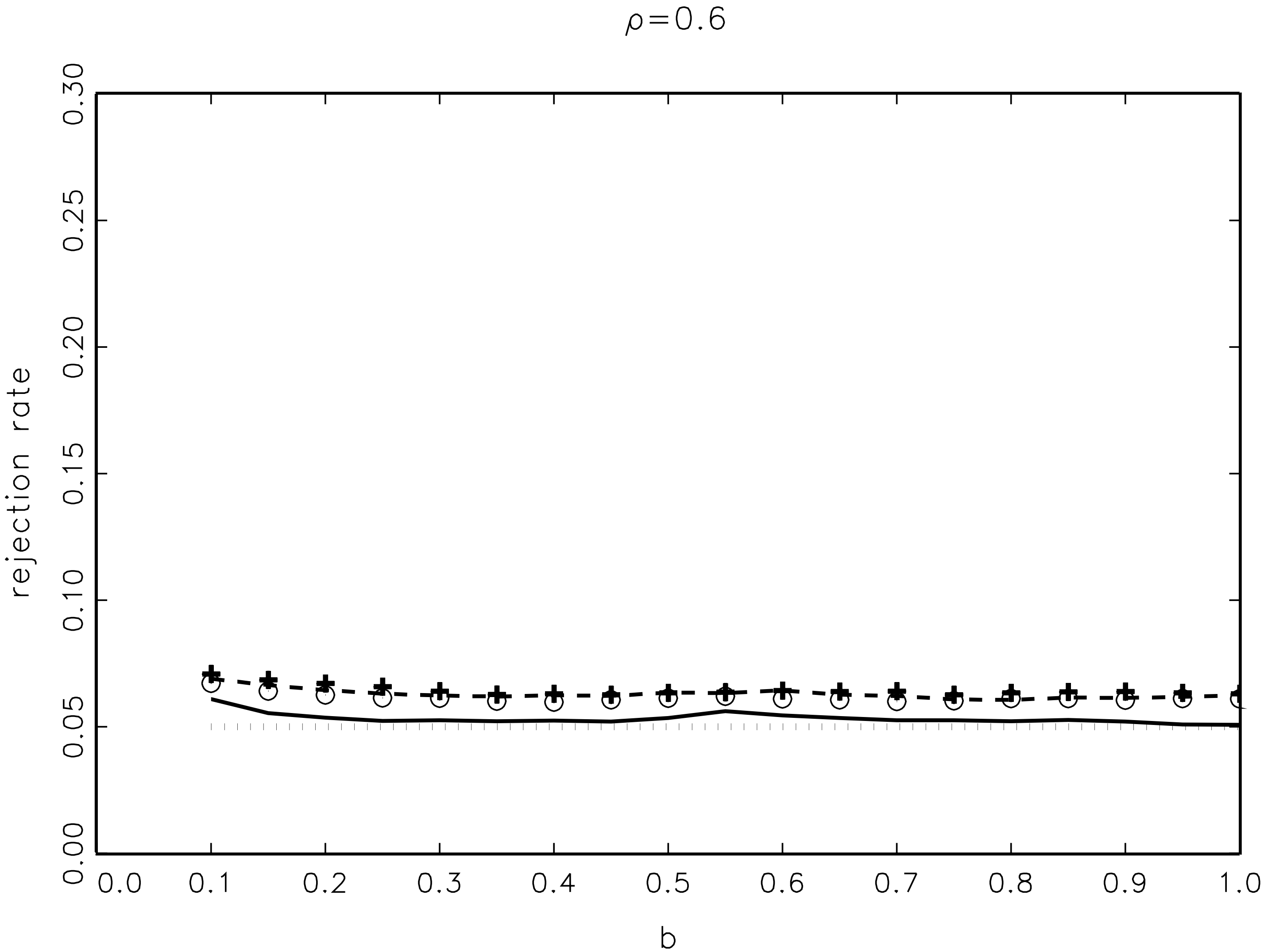


Figure 2.10: (cont'd)

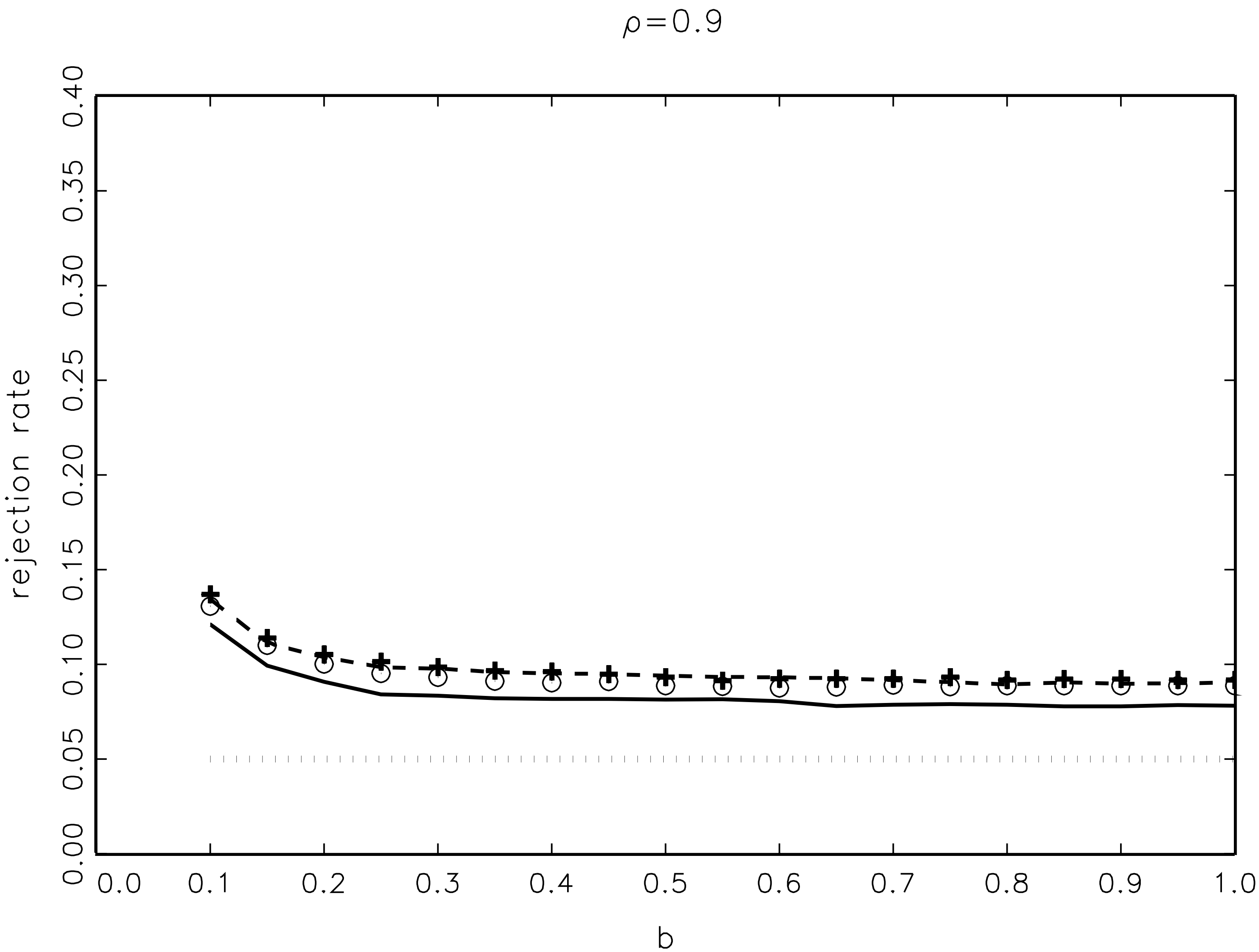


Figure 2.11: AM Series - Initially Scarce Data, Bartlett, $N_Q = 12$ $N_M = 12$

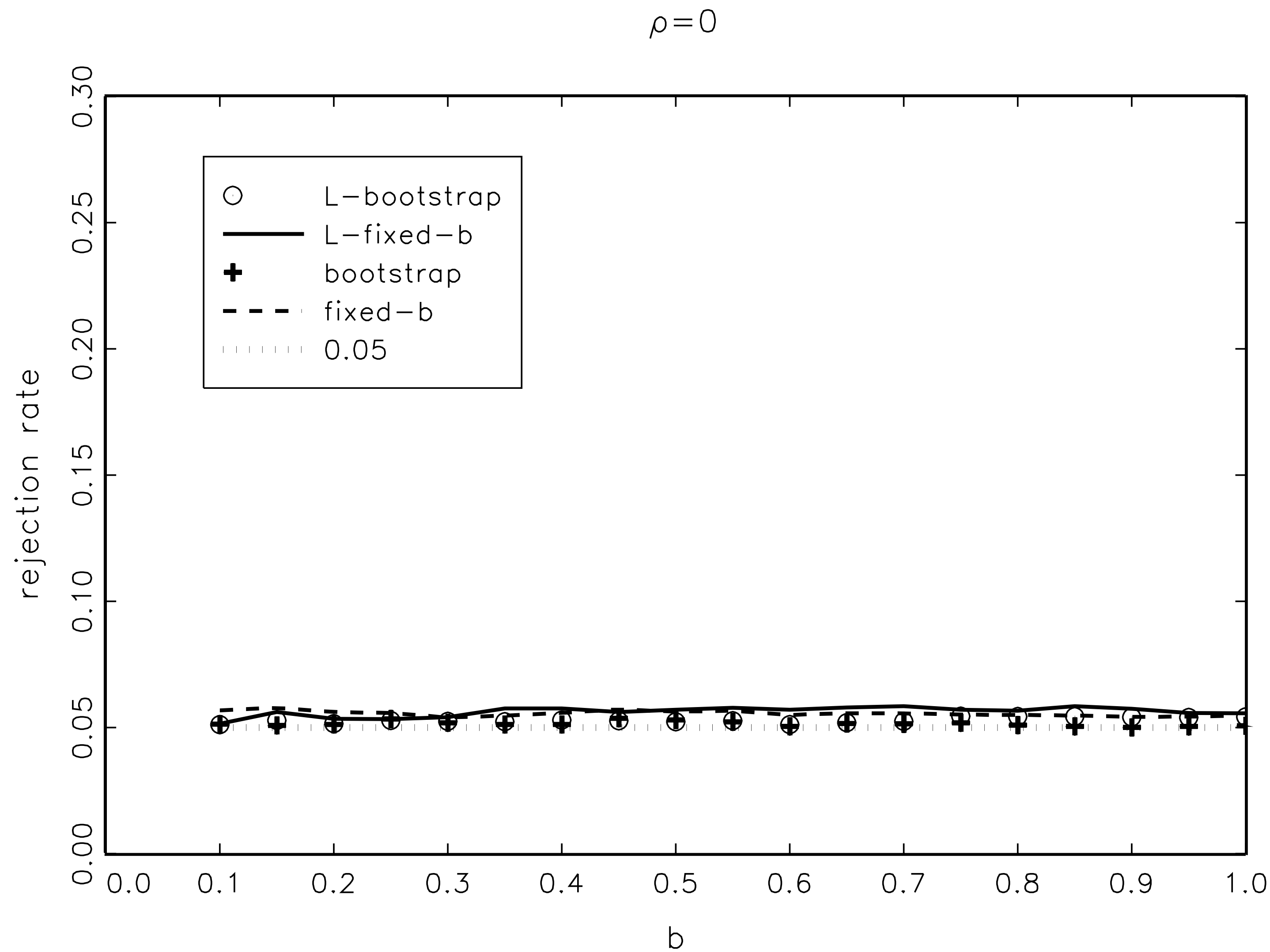


Figure 2.11: (cont'd)

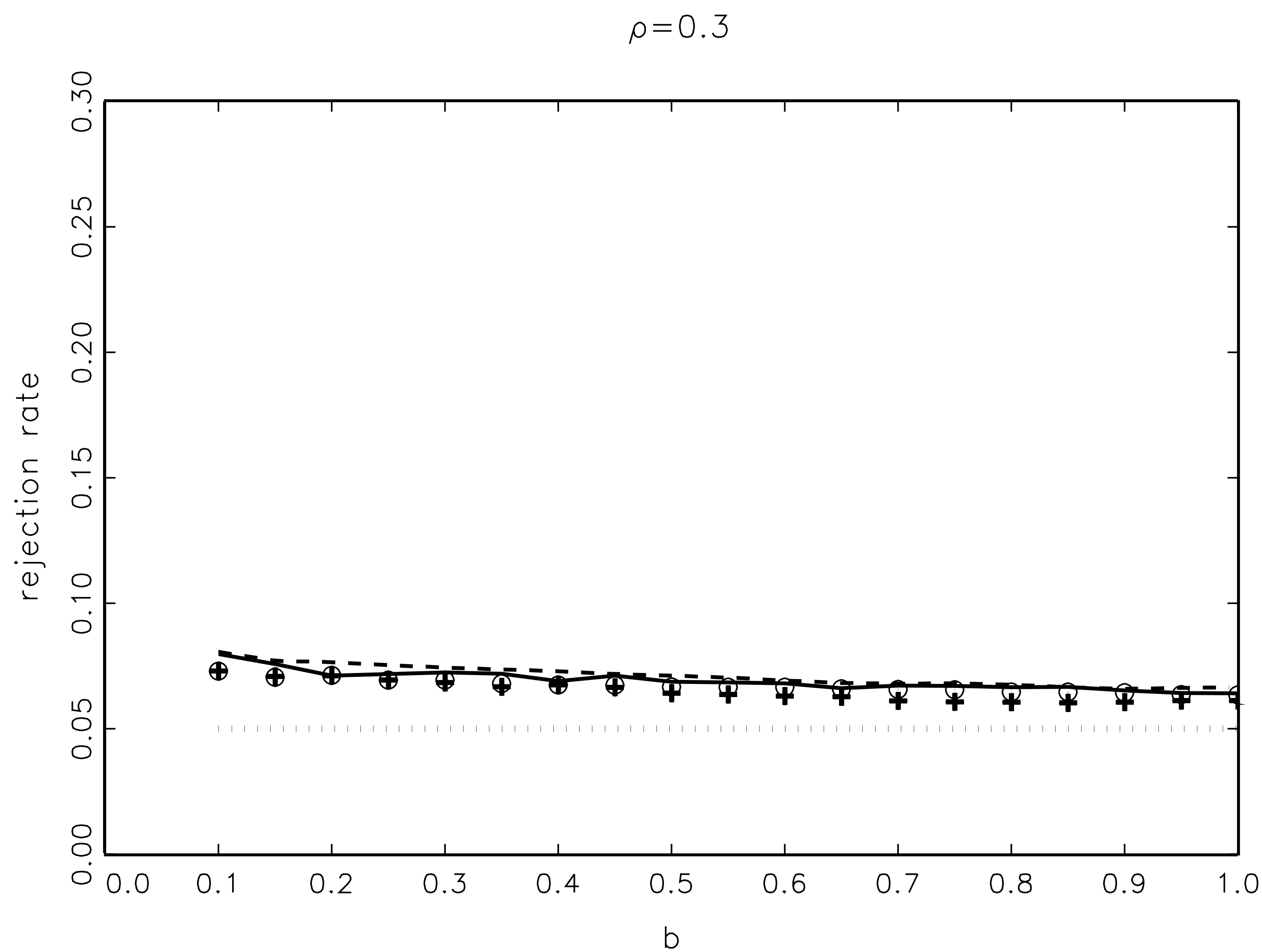


Figure 2.11: (cont'd)

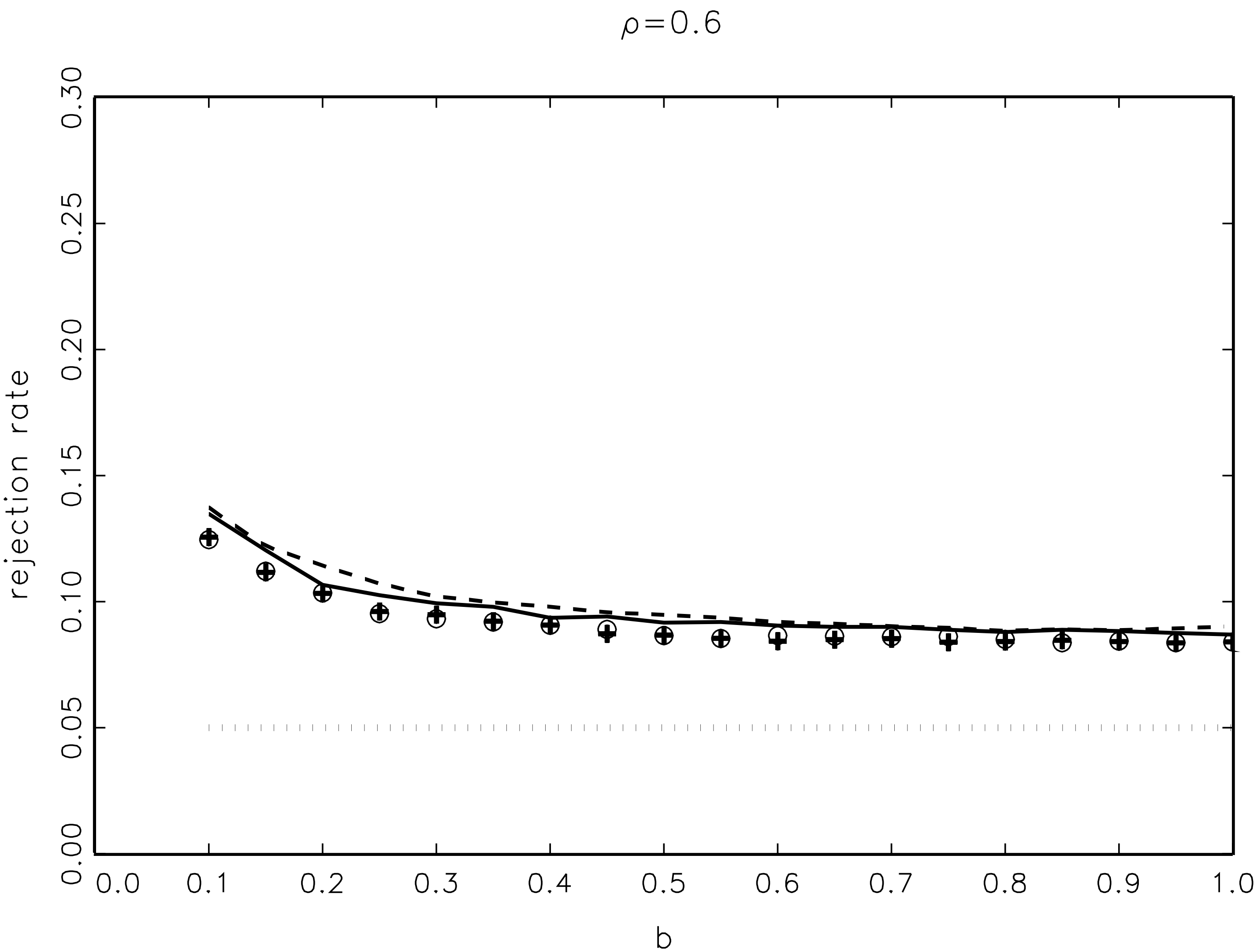


Figure 2.11: (cont'd)

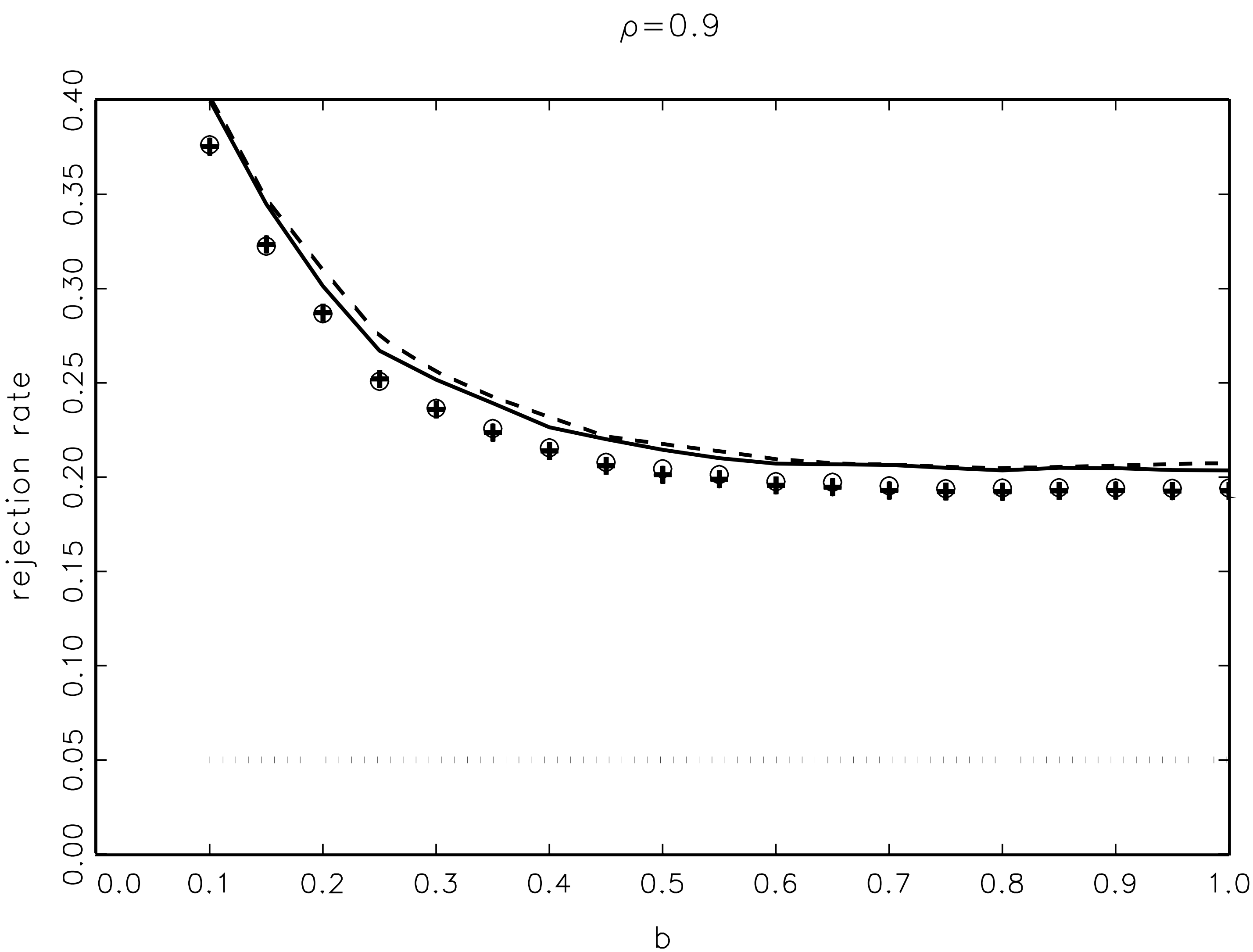


Figure 2.12: AM Series - Initially Scarce Data, Bartlett, $N_Q = 12$ $N_M = 24$

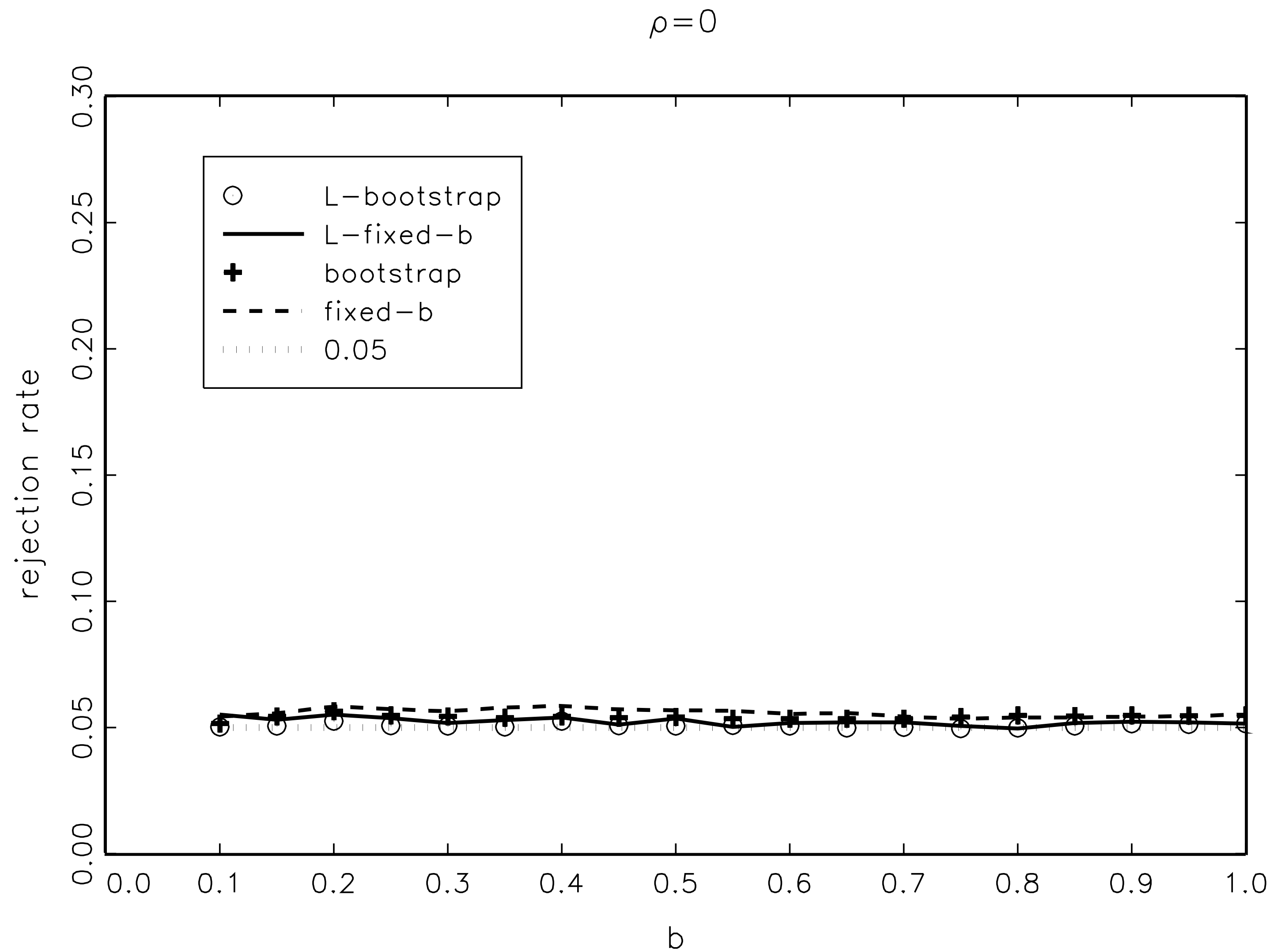


Figure 2.12: (cont'd)

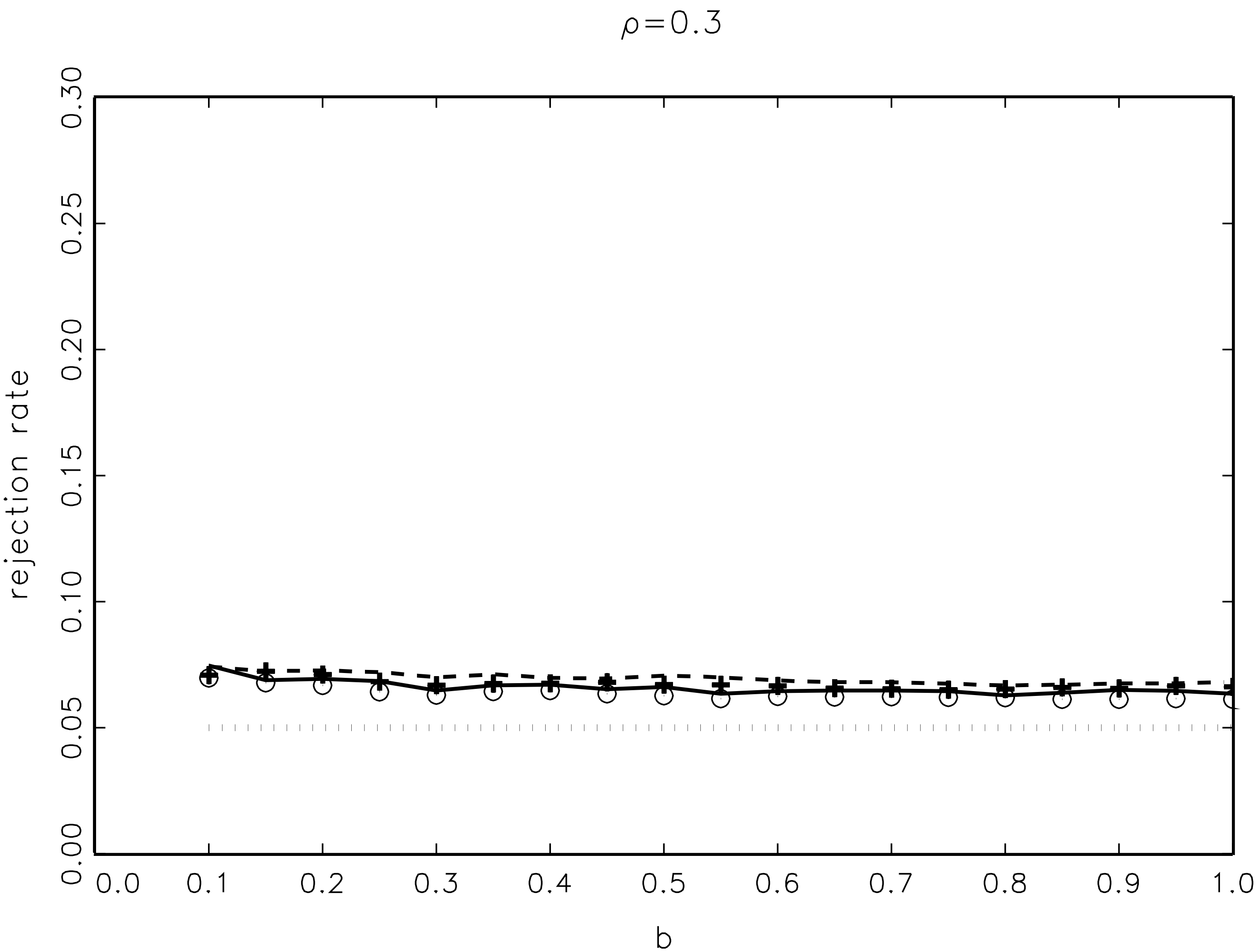


Figure 2.12: (cont'd)

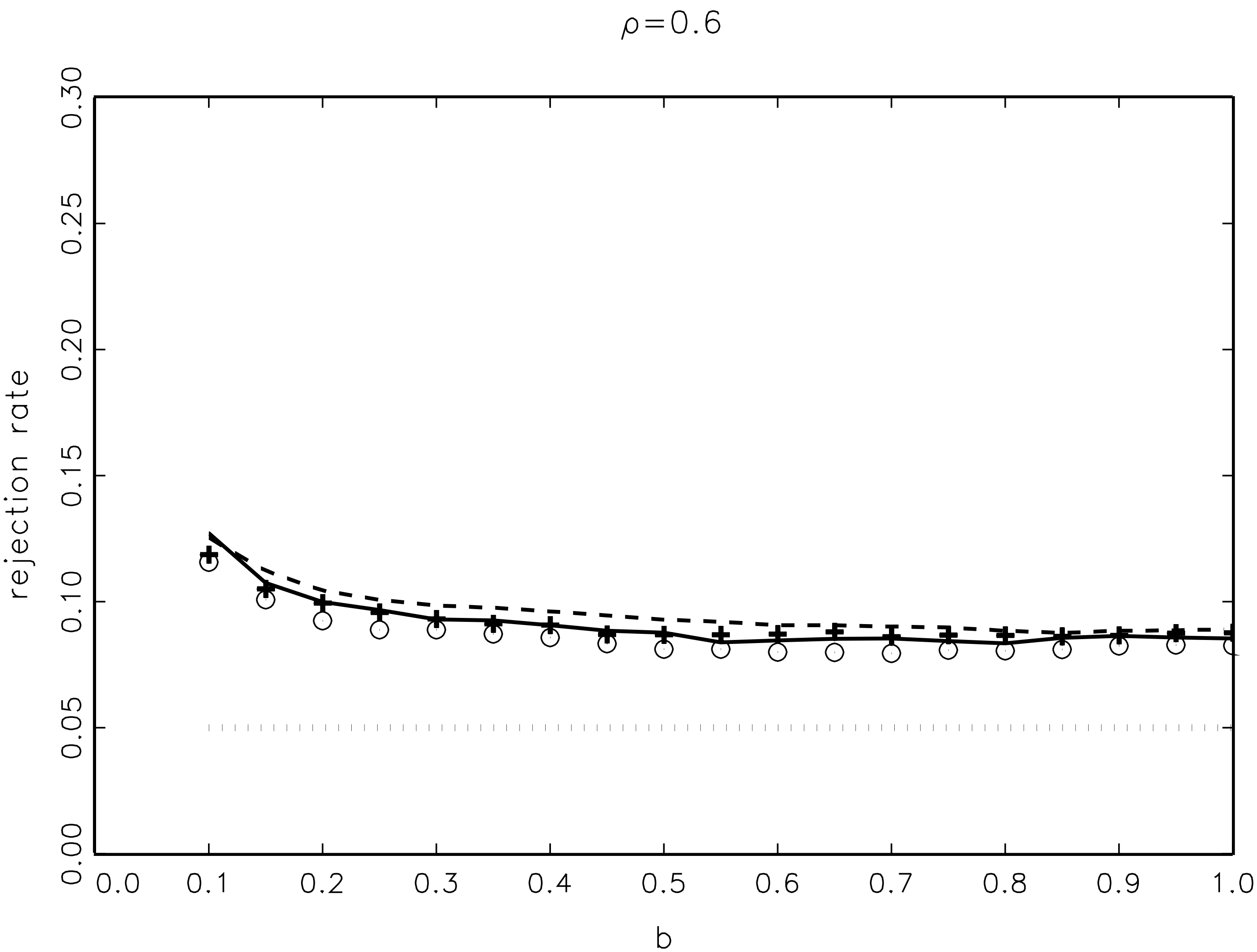


Figure 2.12: (cont'd)

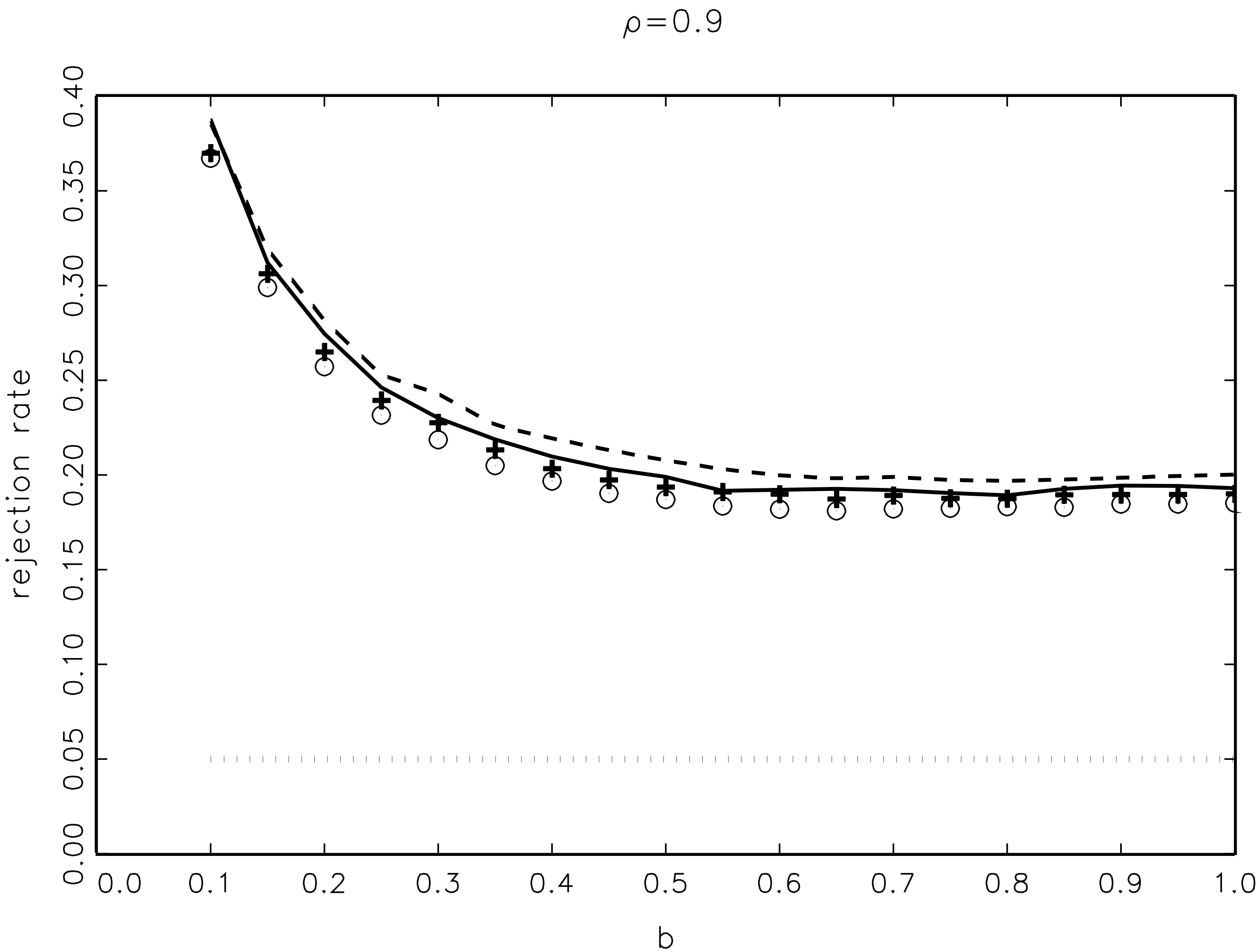


Figure 2.13: AM Series - Initially Scarce Data, Bartlett, $N_Q = 12$ $N_M = 48$

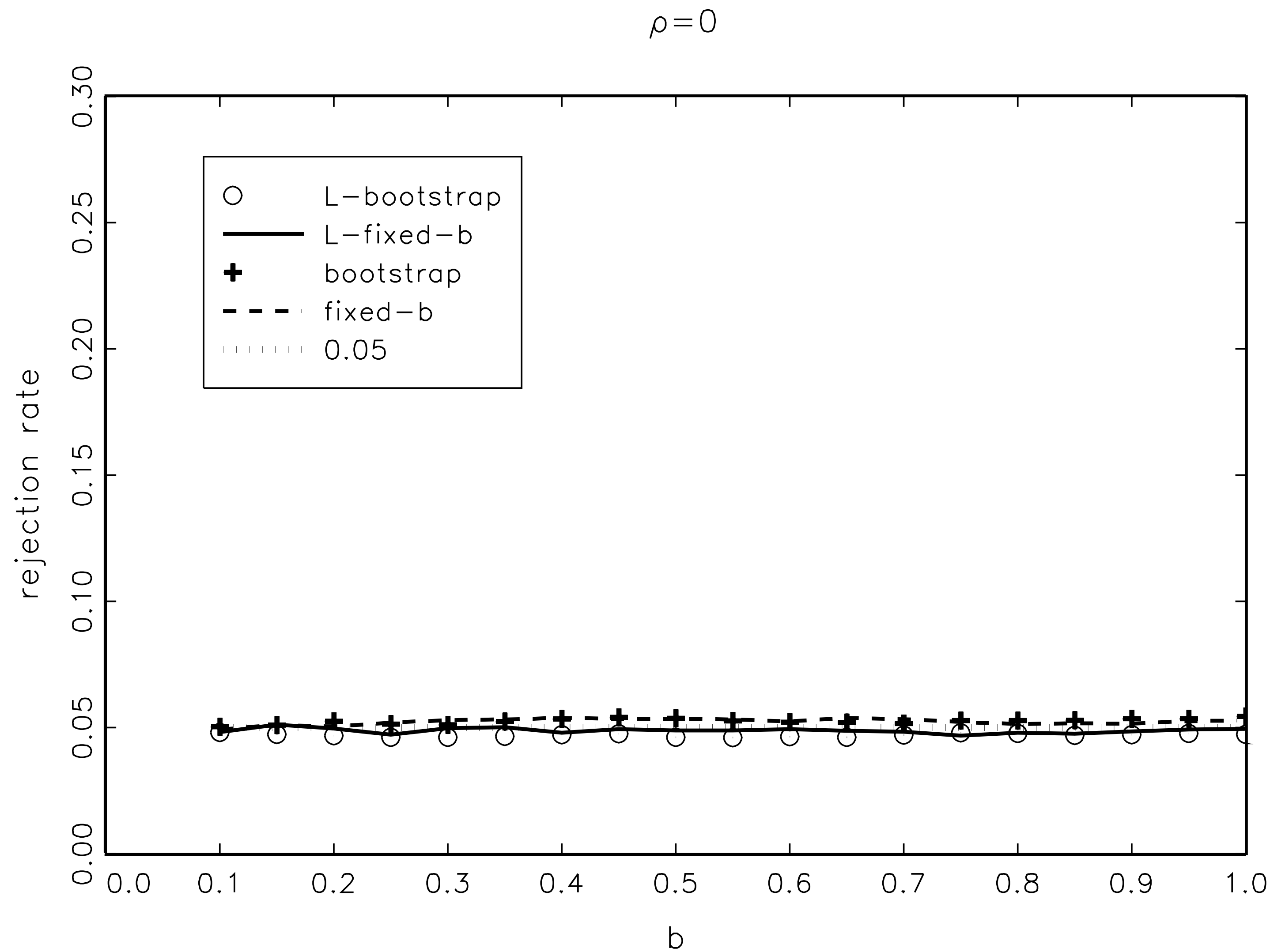


Figure 2.13: (cont'd)

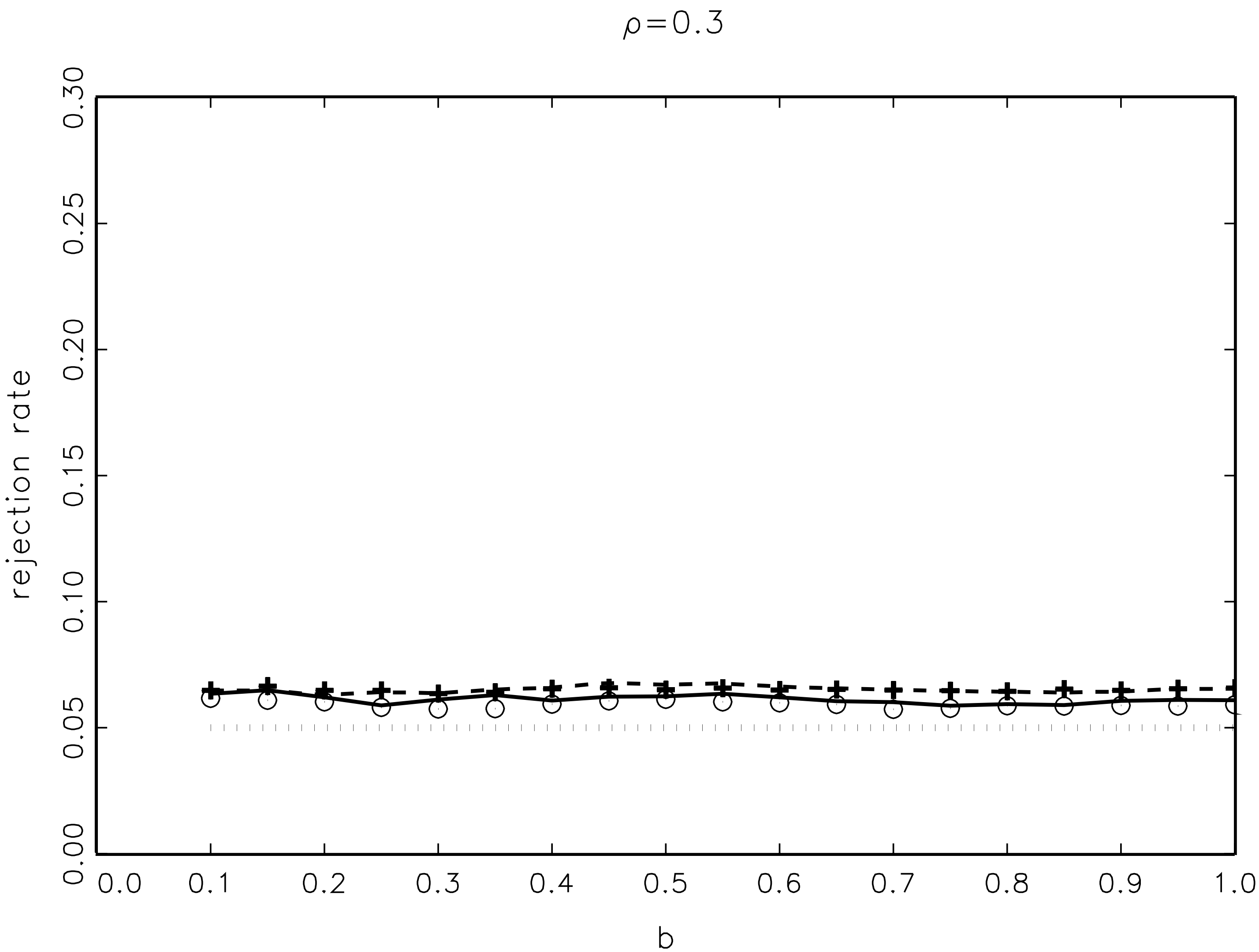


Figure 2.13: (cont'd)

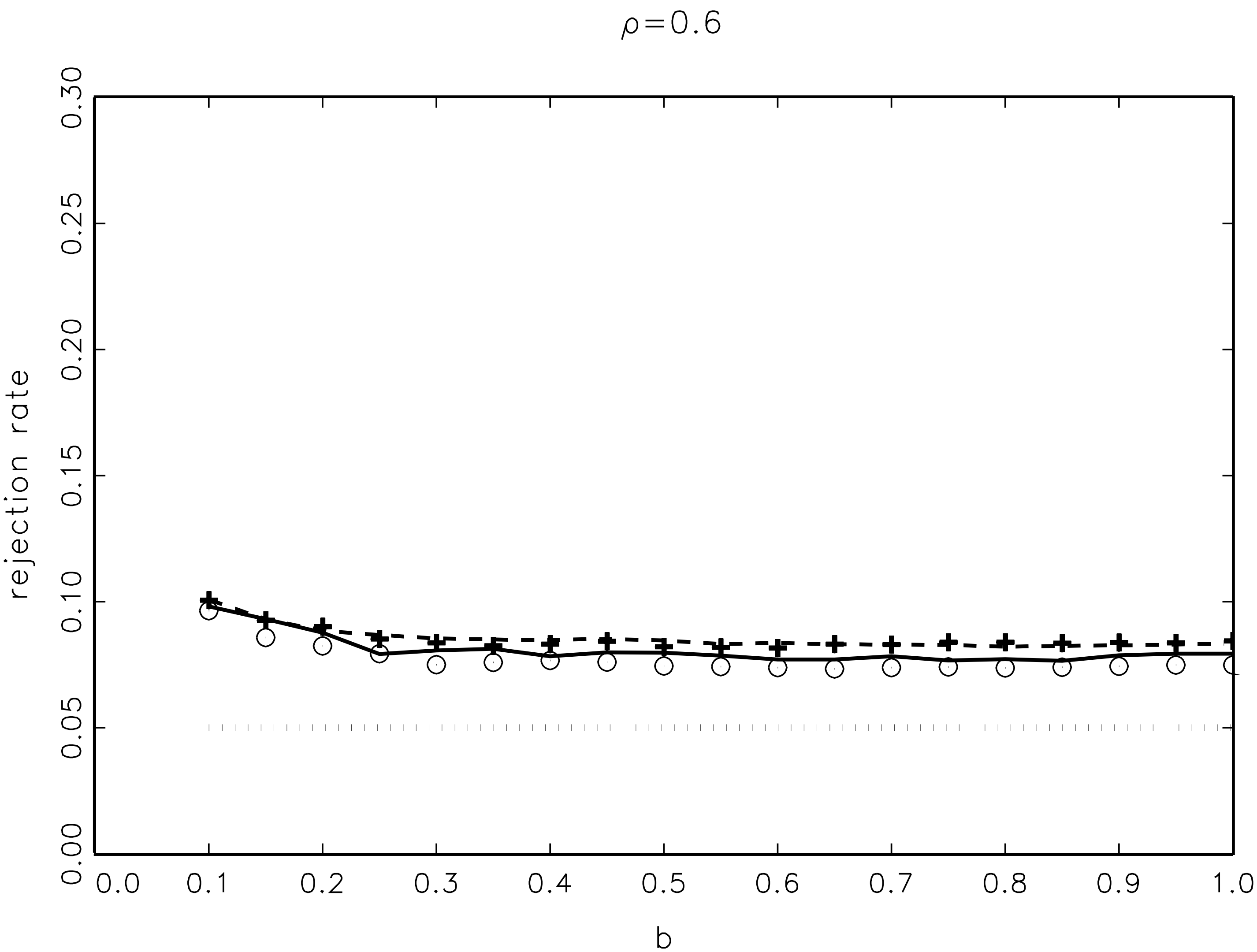


Figure 2.13: (cont'd)

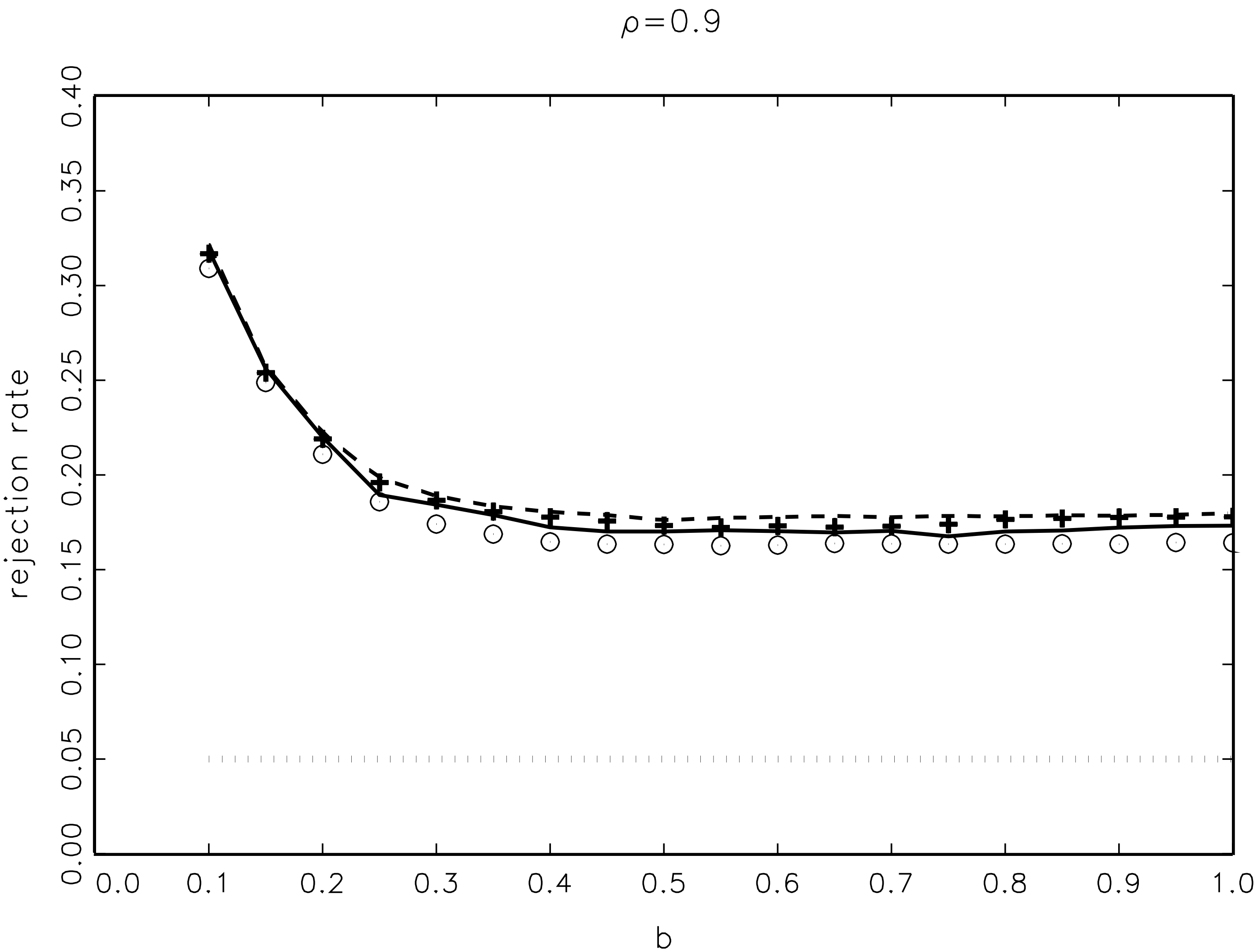


Figure 2.14: AM Series - Initially Scarce Data, Bartlett, $N_Q = 24$ $N_M = 12$

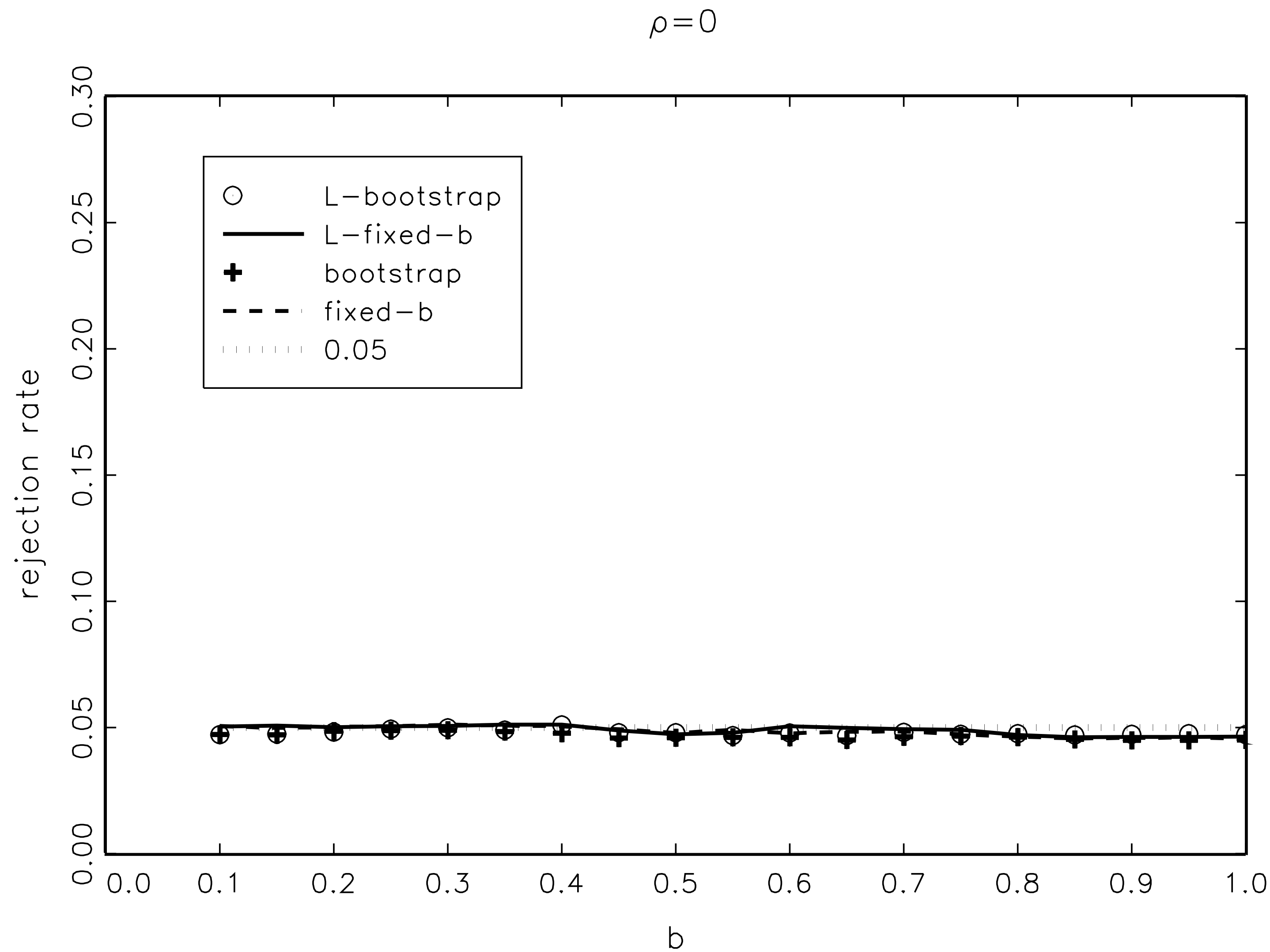


Figure 2.14: (cont'd)

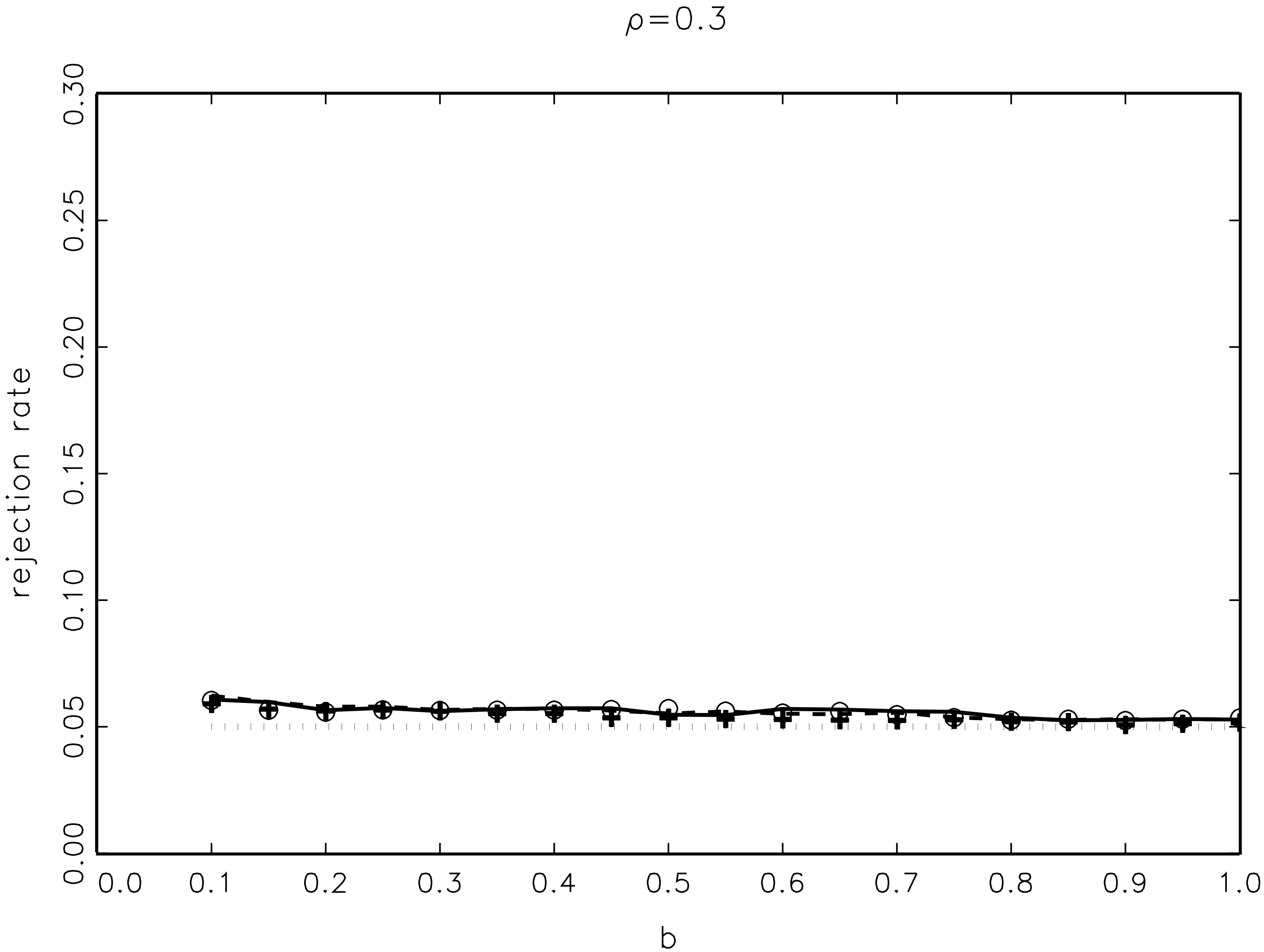


Figure 2.14: (cont'd)

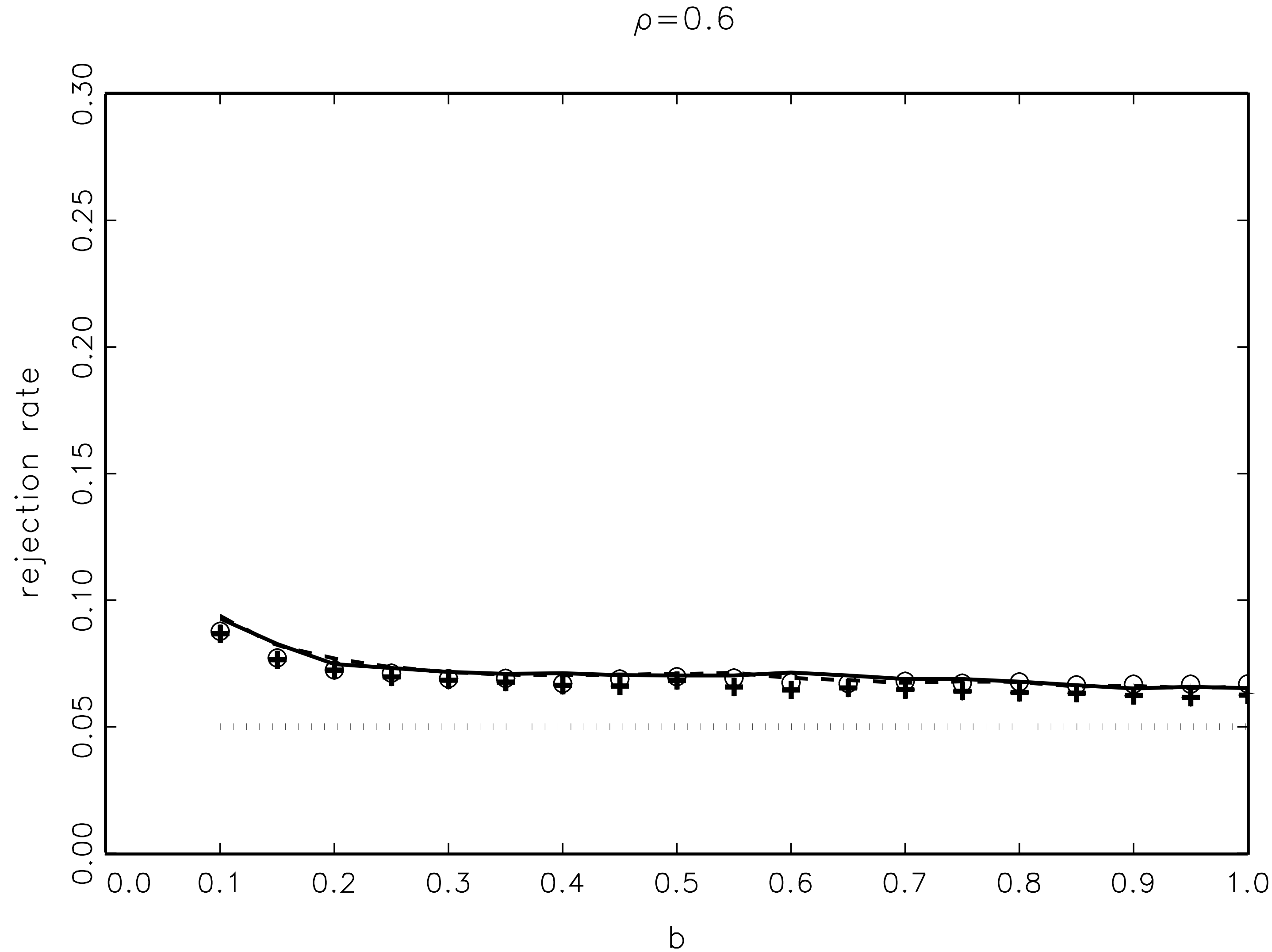


Figure 2.14: (cont'd)

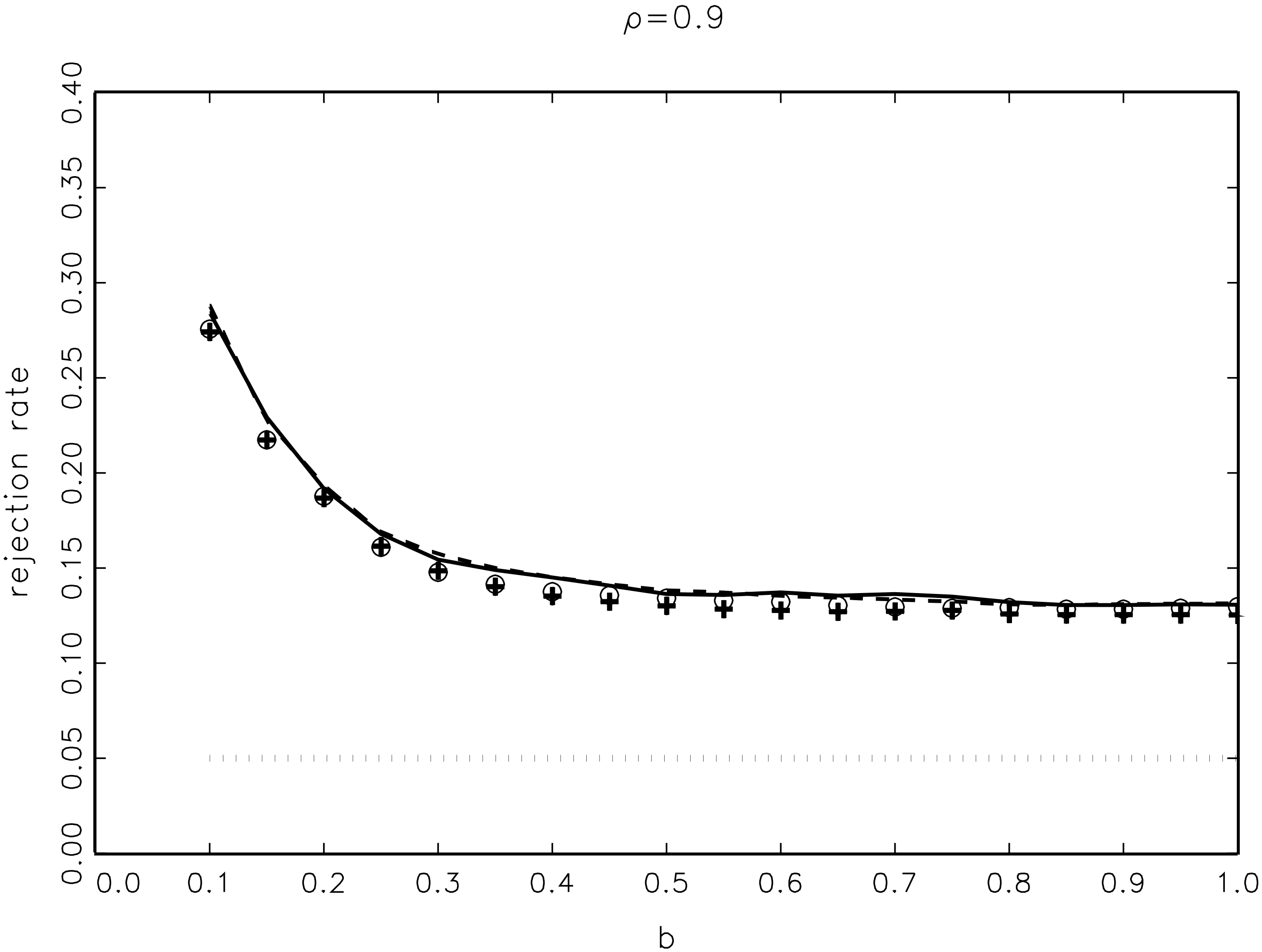


Figure 2.15: AM Series - Initially Scarce Data, Bartlett, $N_Q = 24$ $N_M = 24$

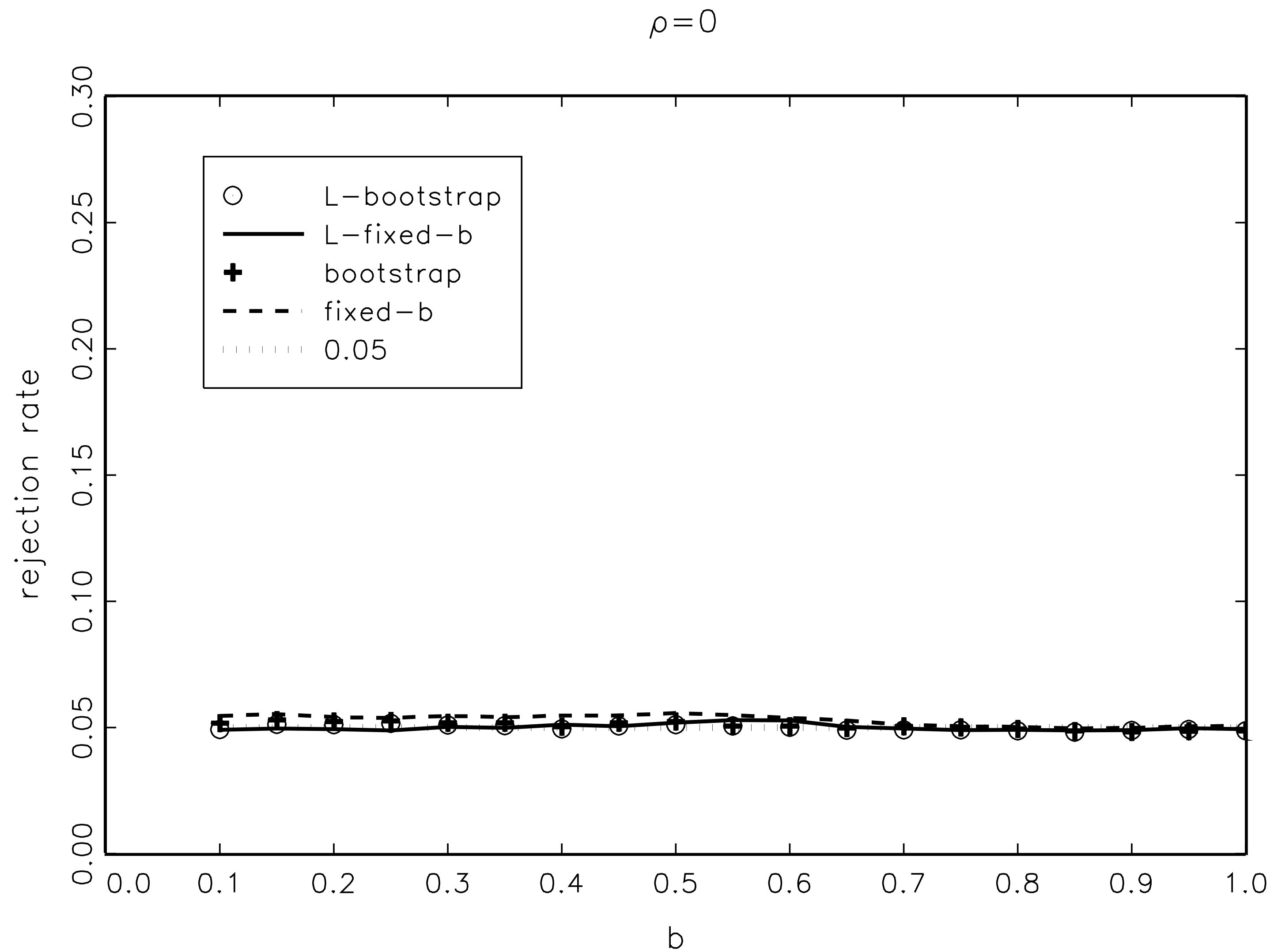


Figure 2.15: (cont'd)

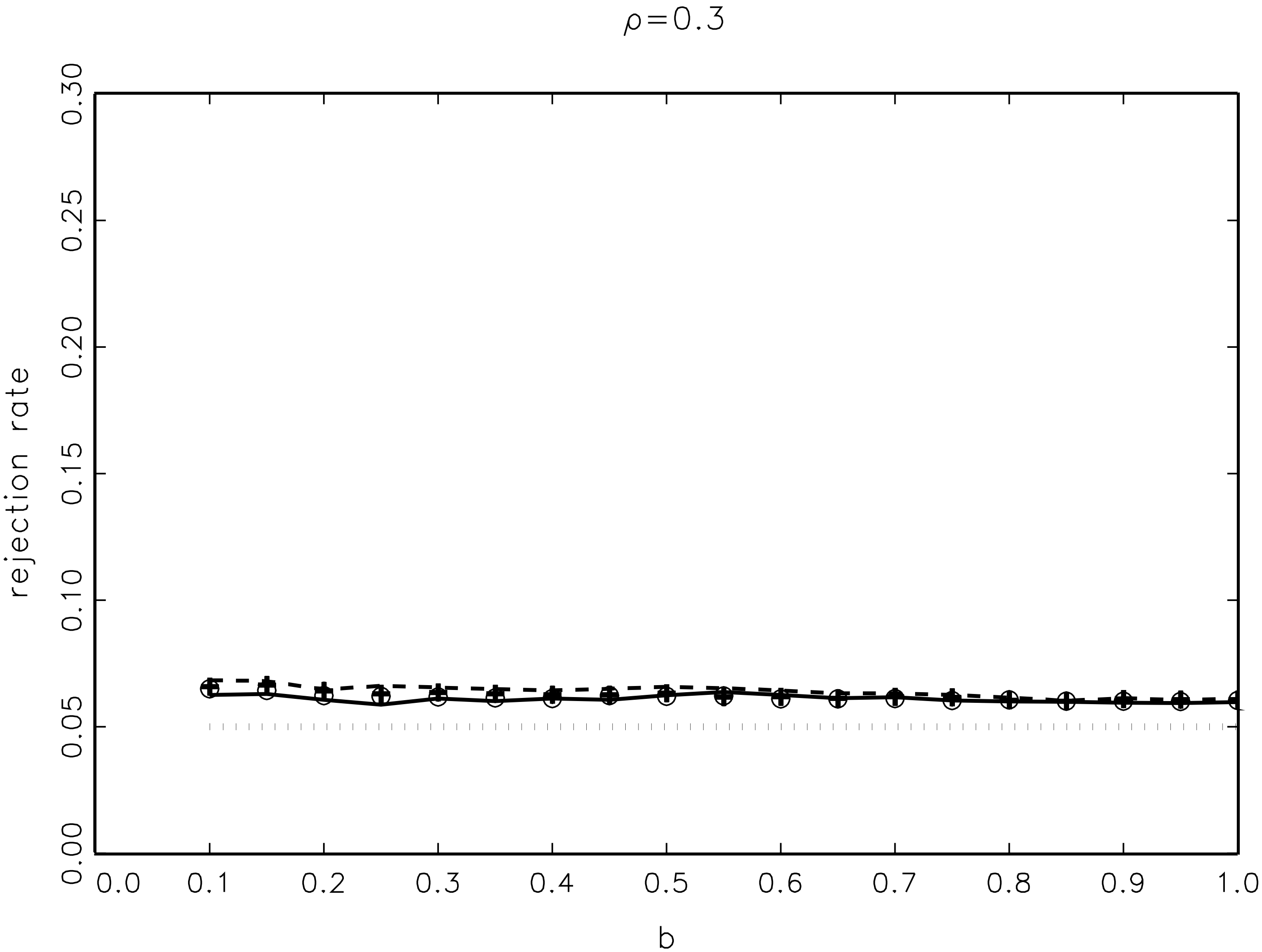


Figure 2.15: (cont'd)

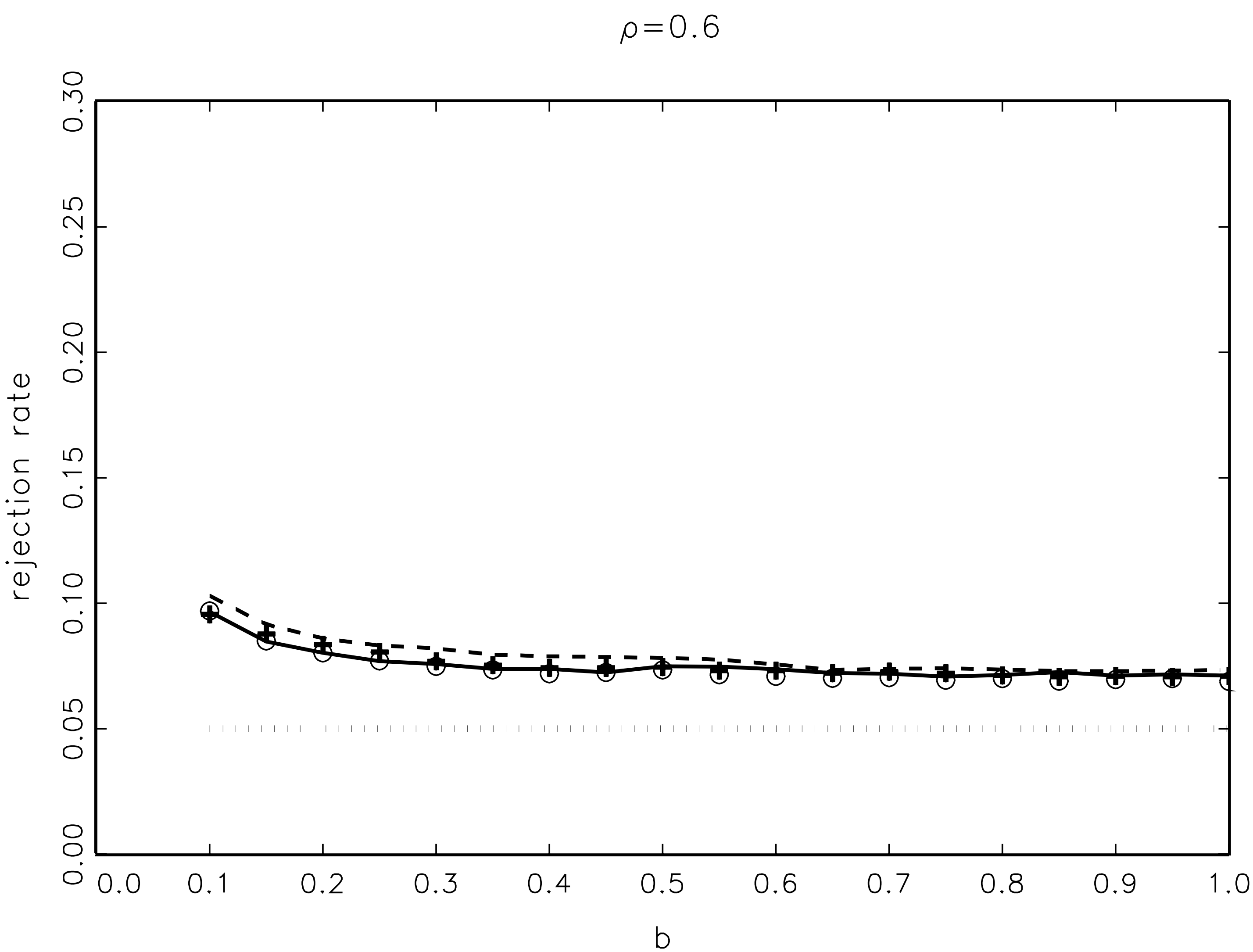


Figure 2.15: (cont'd)

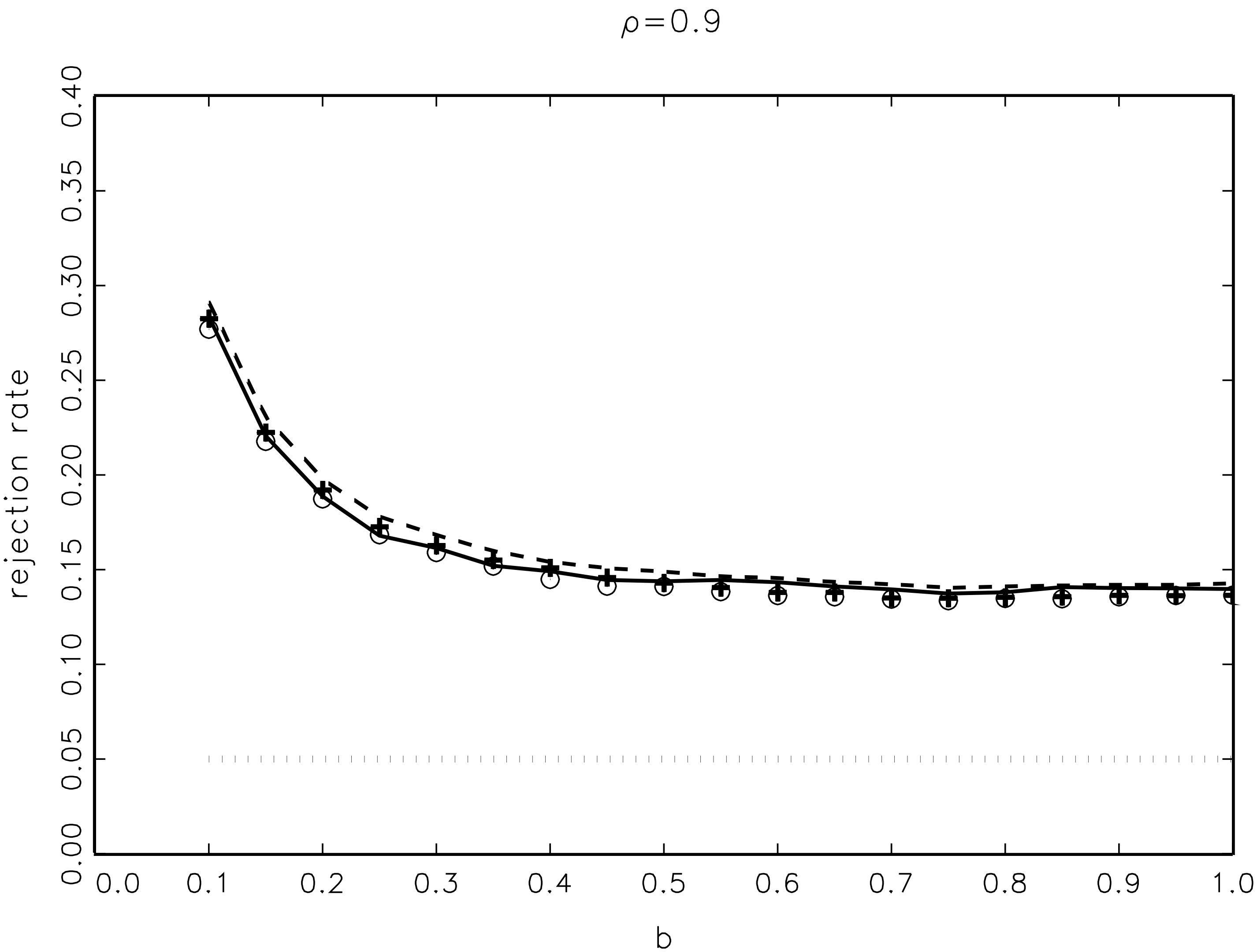


Figure 2.16: AM Series - Initially Scarce Data, Bartlett, $N_Q = 24$ $N_M = 48$

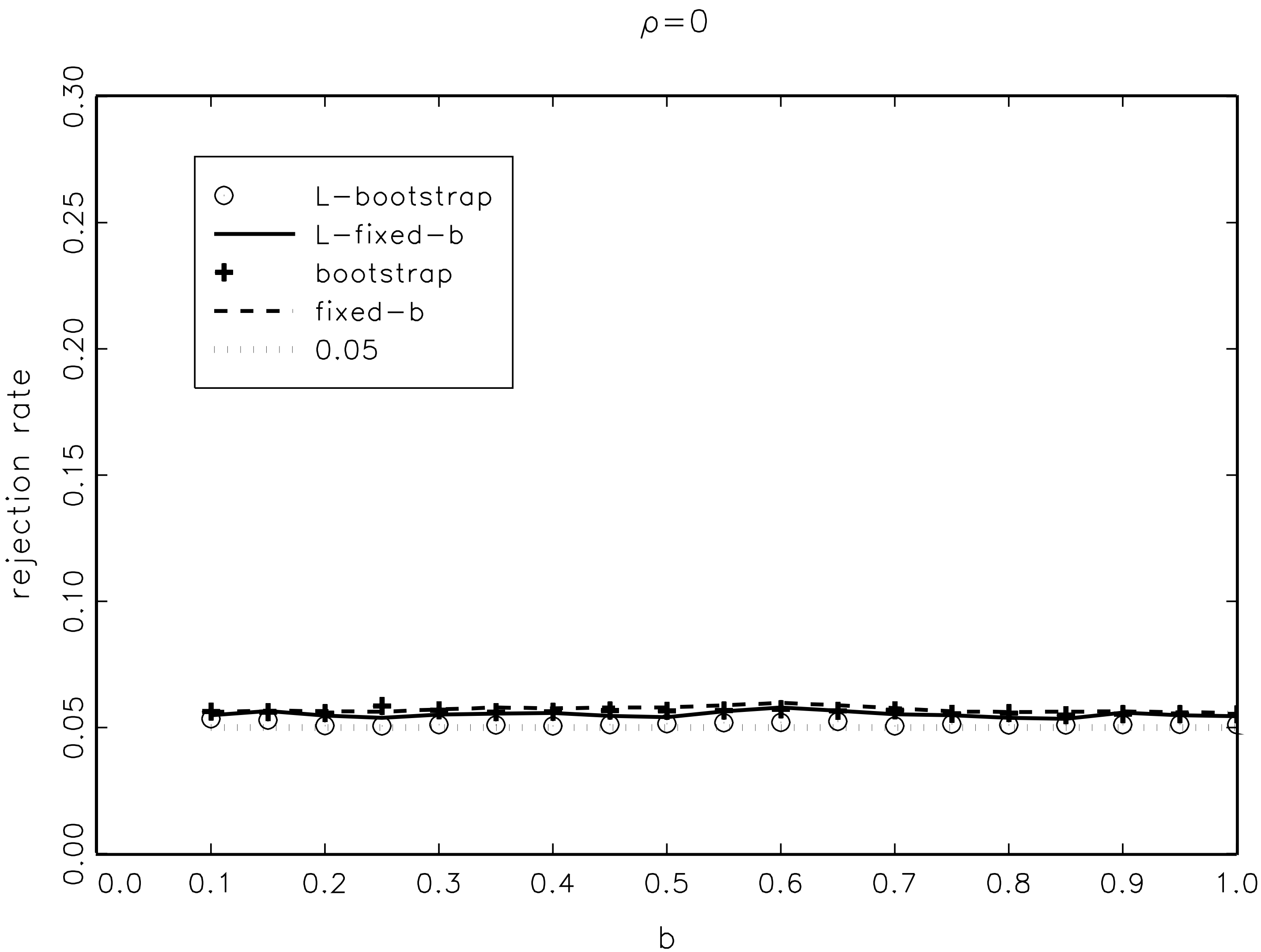


Figure 2.16: (cont'd)

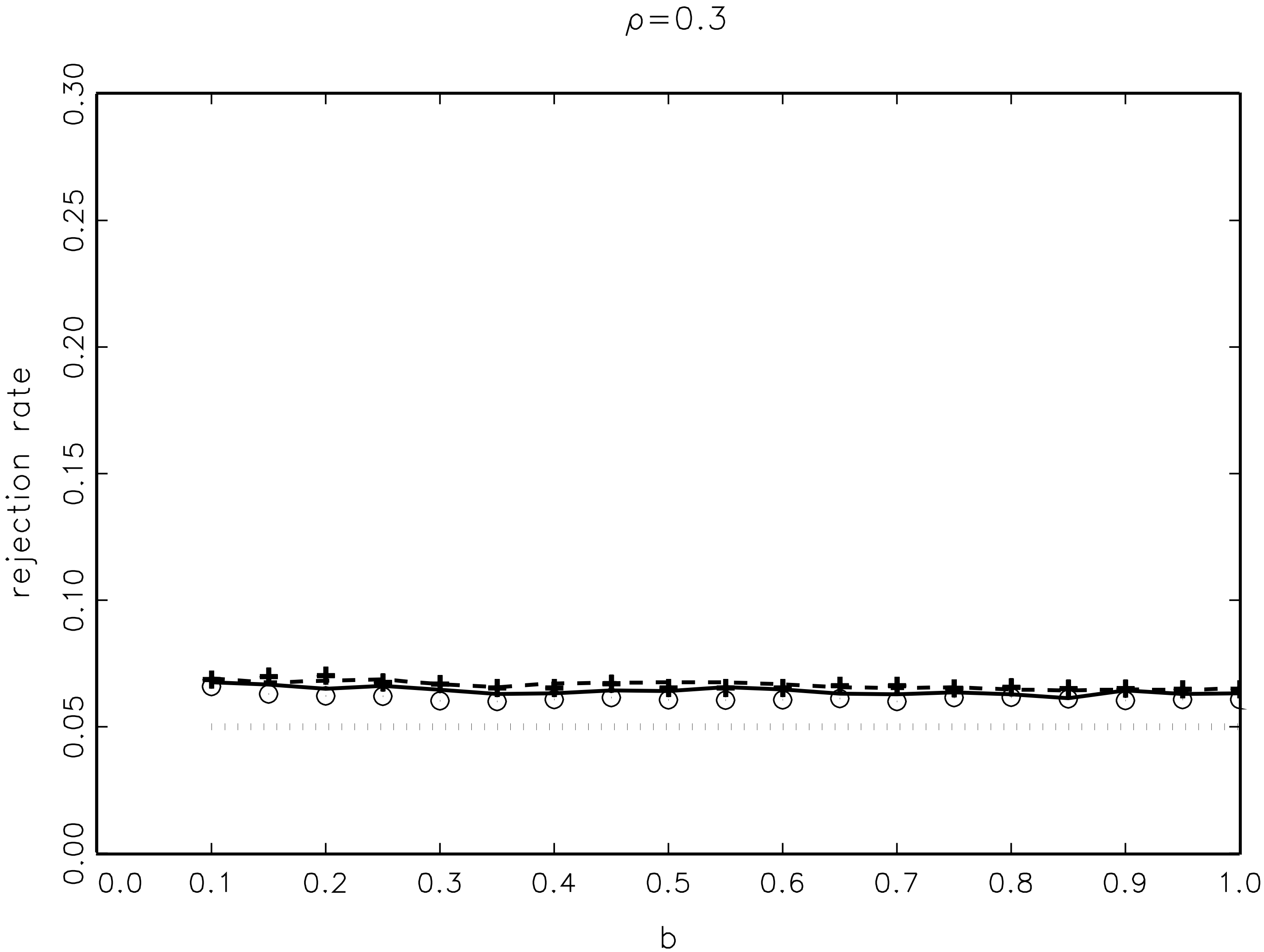


Figure 2.16: (cont'd)

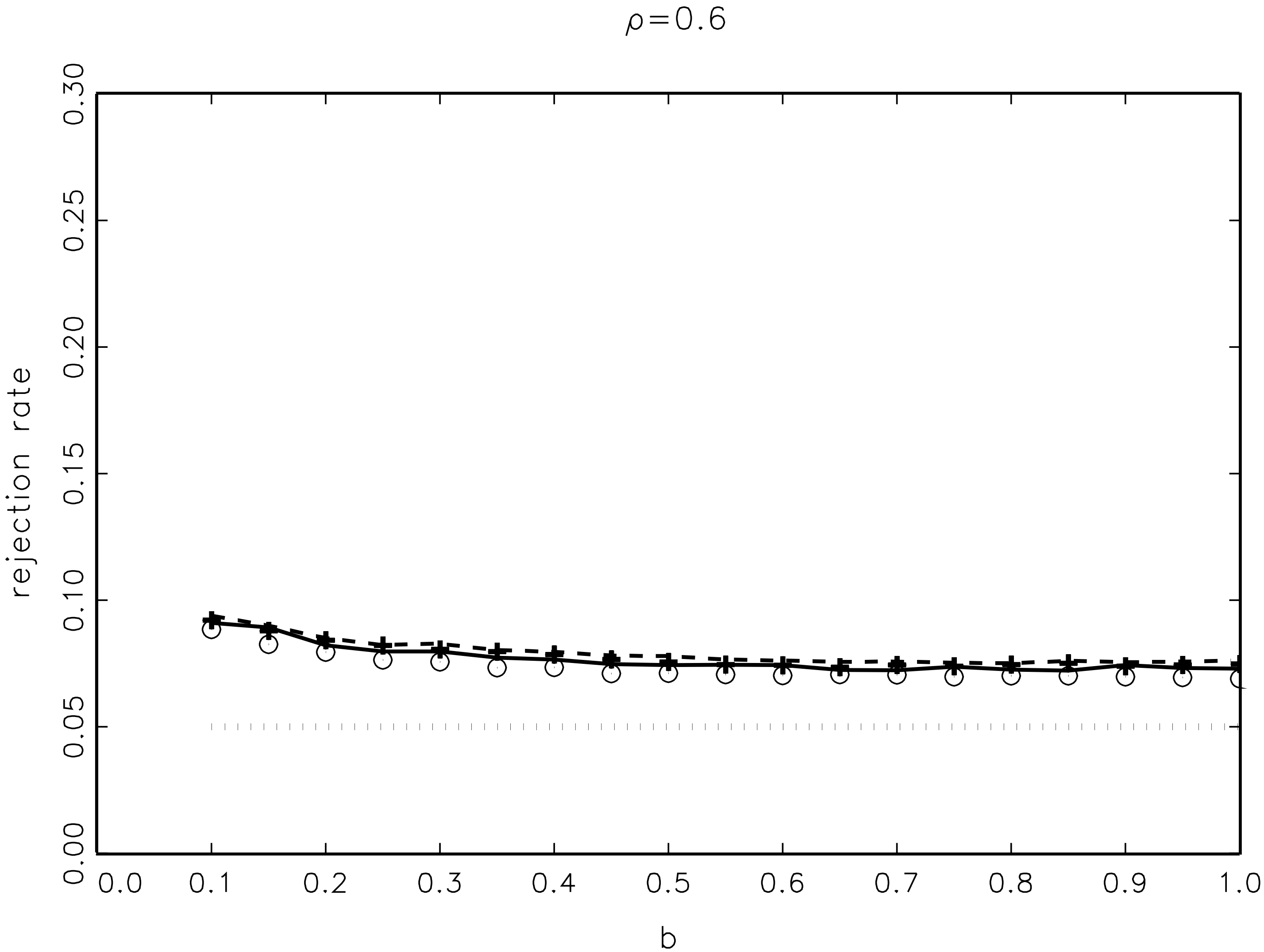


Figure 2.16: (cont'd)

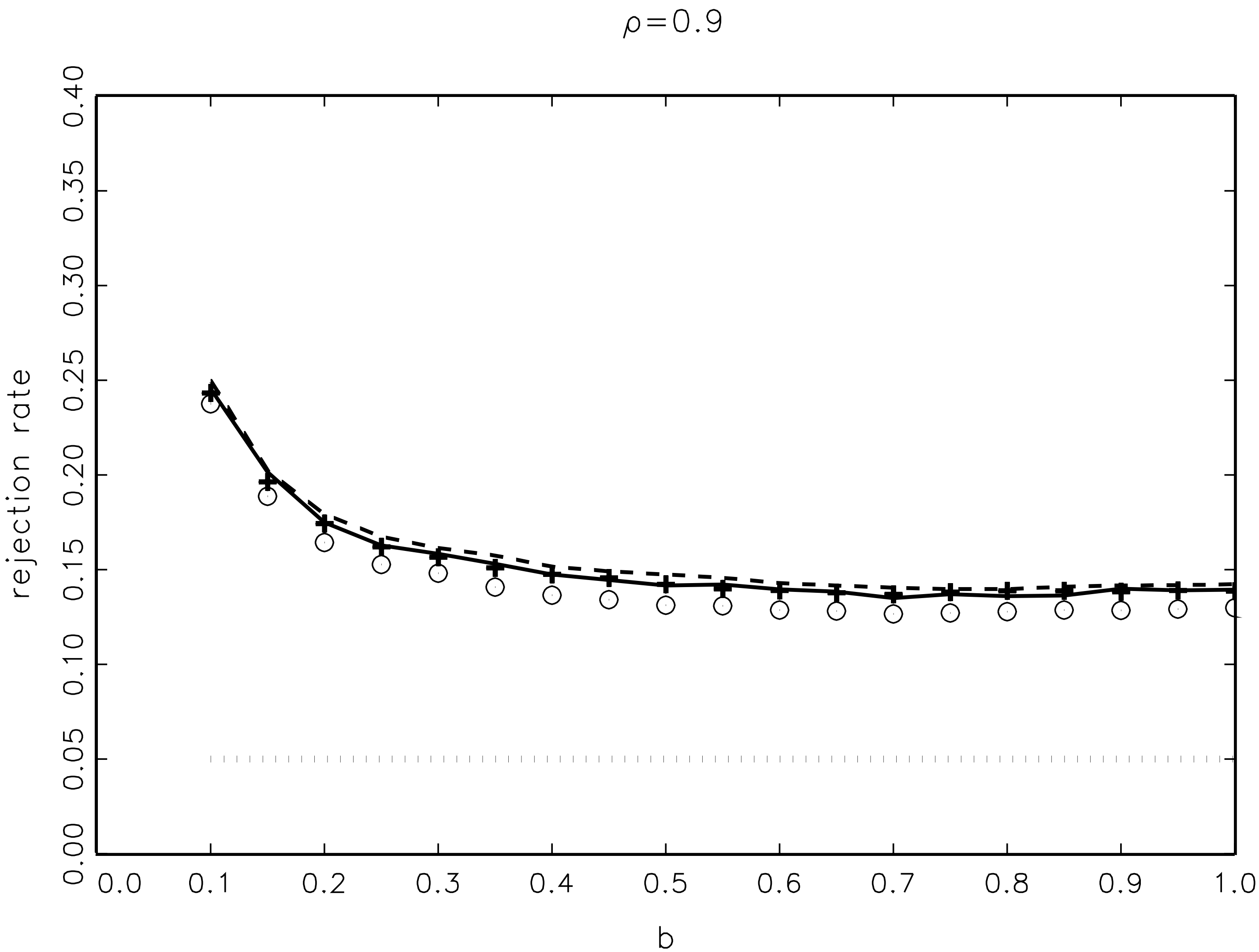


Figure 2.17: AM Series - Conditional Bernoulli ($p = 0.3$), Bartlett, $T = 50$

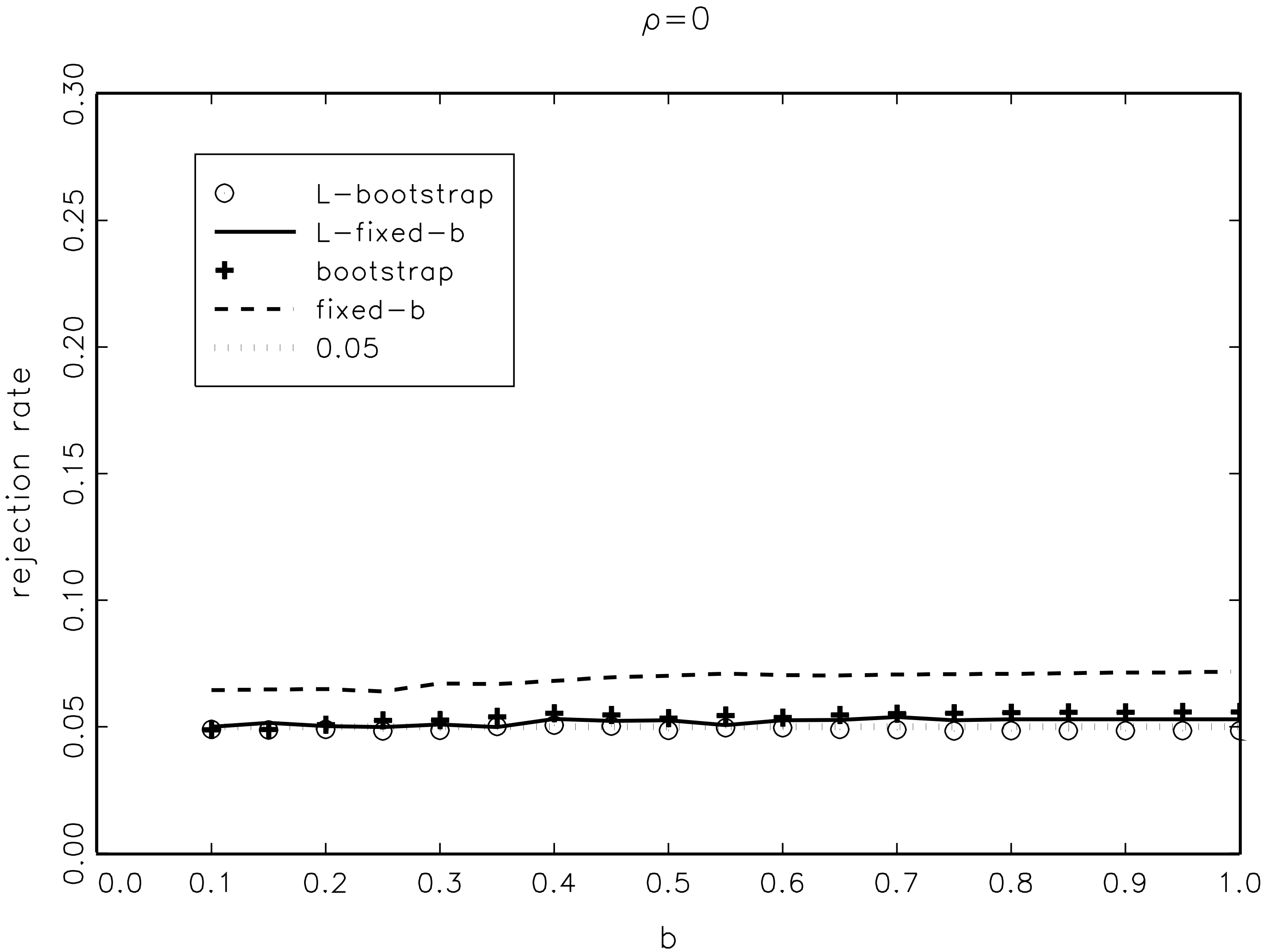


Figure 2.17: (cont'd)

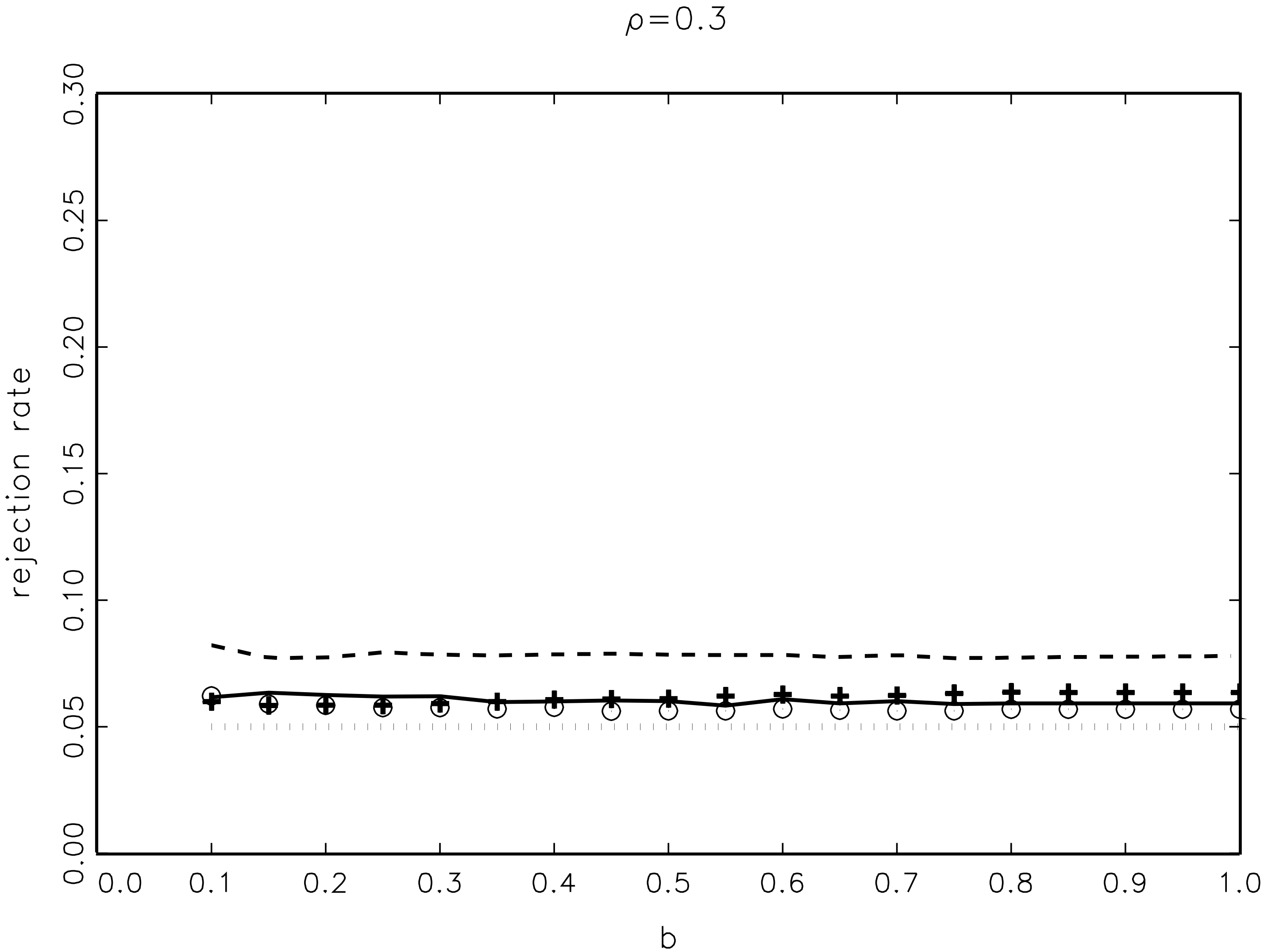


Figure 2.17: (cont'd)

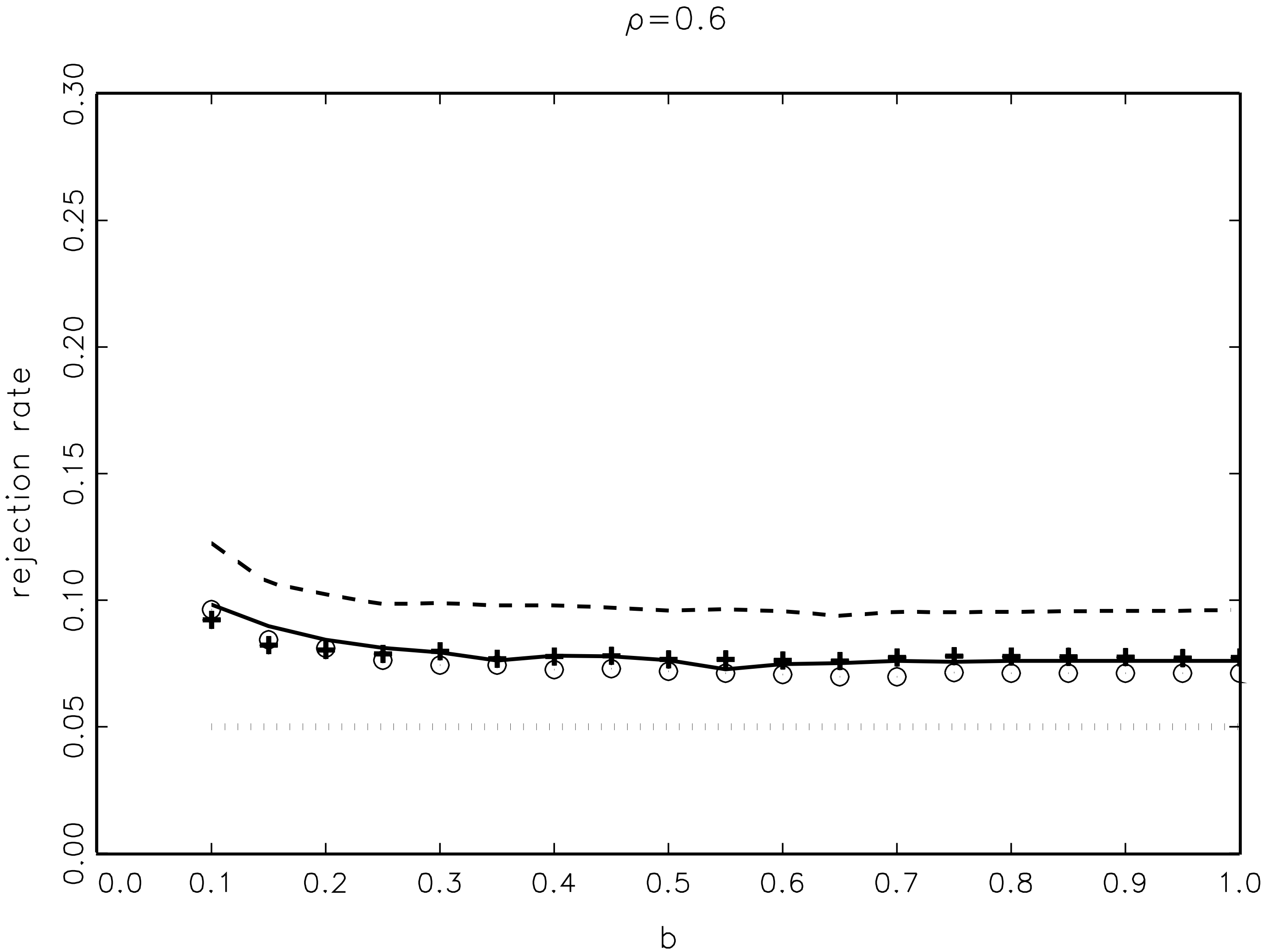


Figure 2.17: (cont'd)

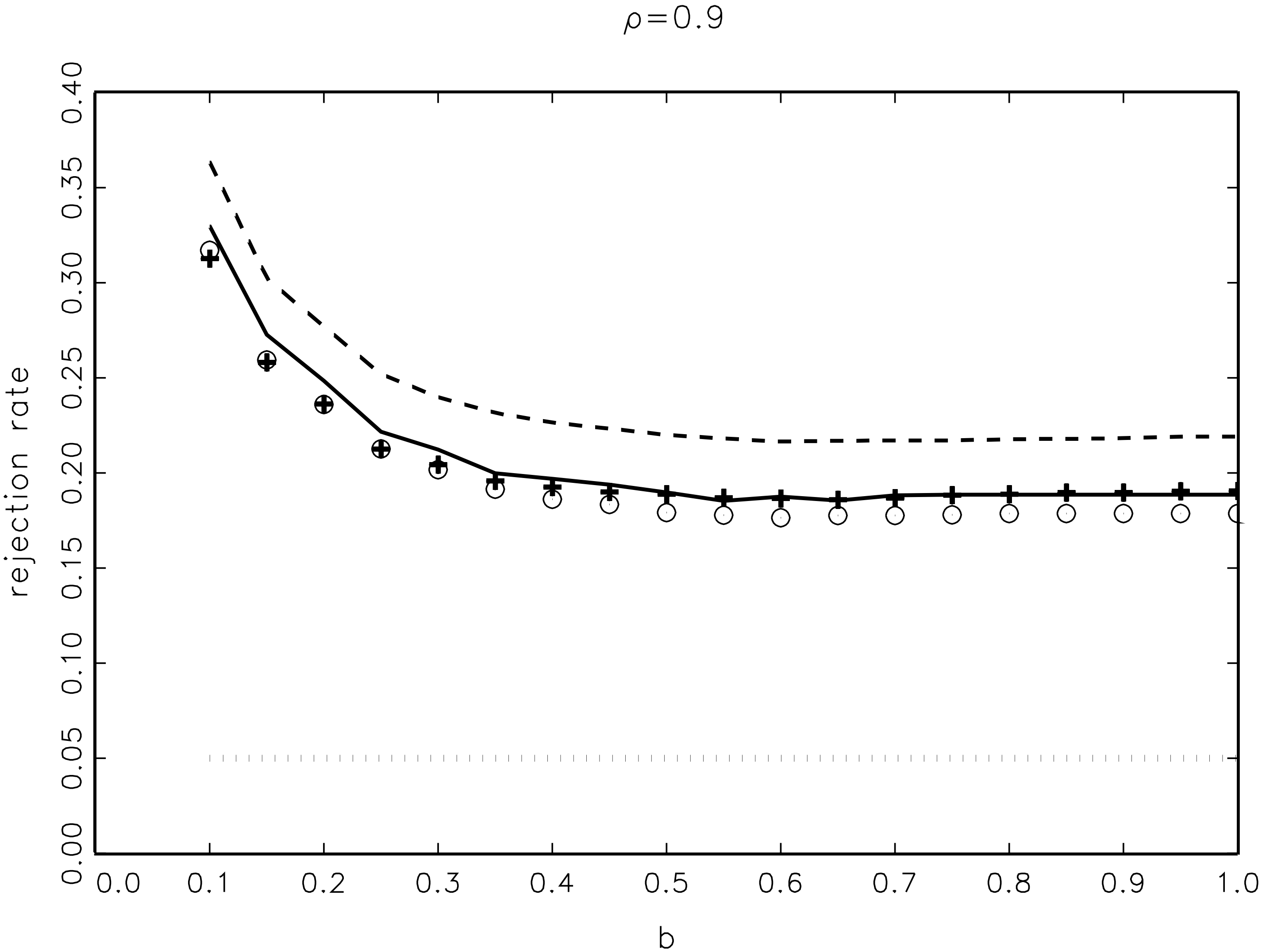


Figure 2.18: AM Series - Conditional Bernoulli ($p = 0.5$), Bartlett, $T = 50$

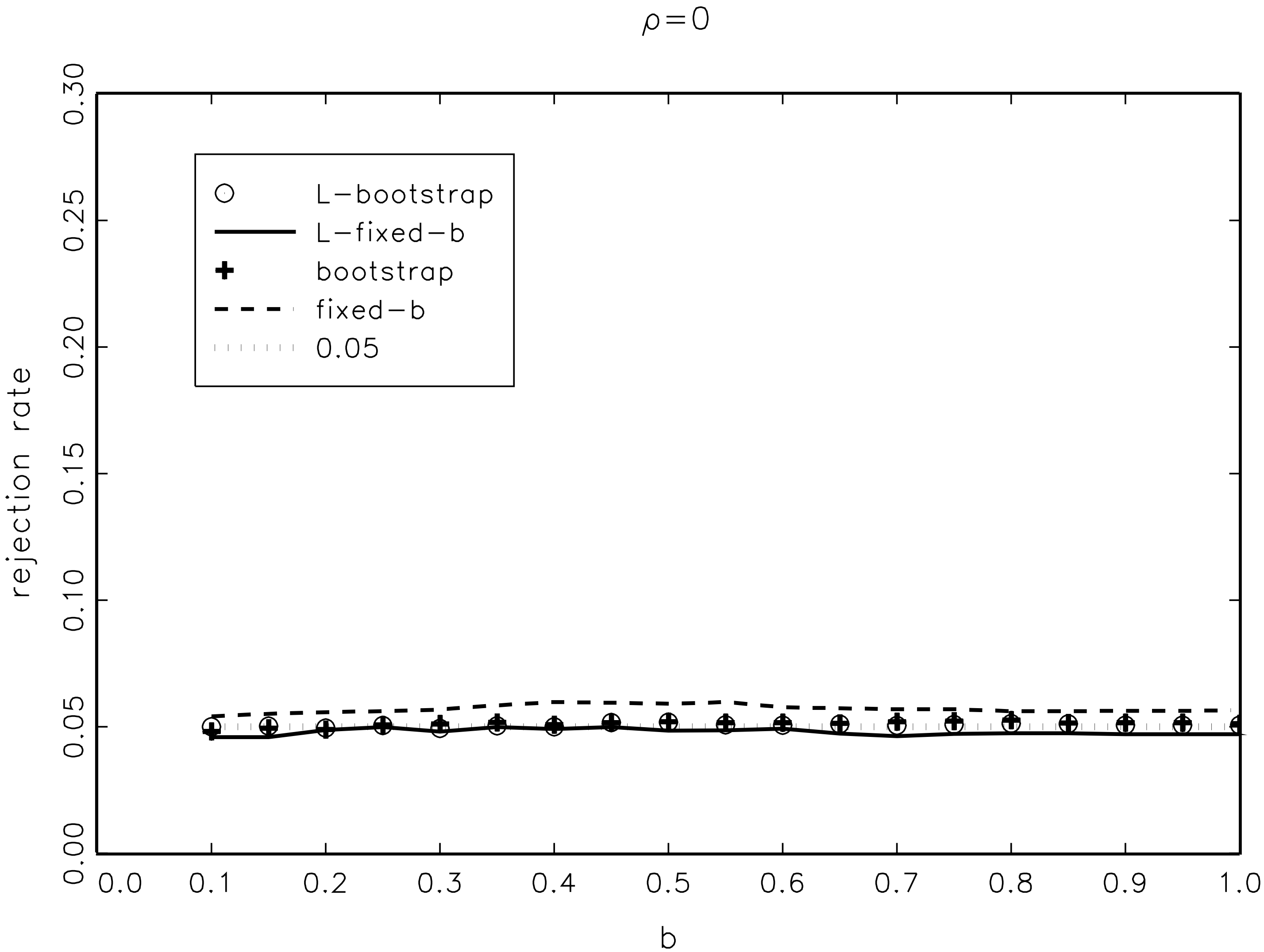


Figure 2.18: (cont'd)

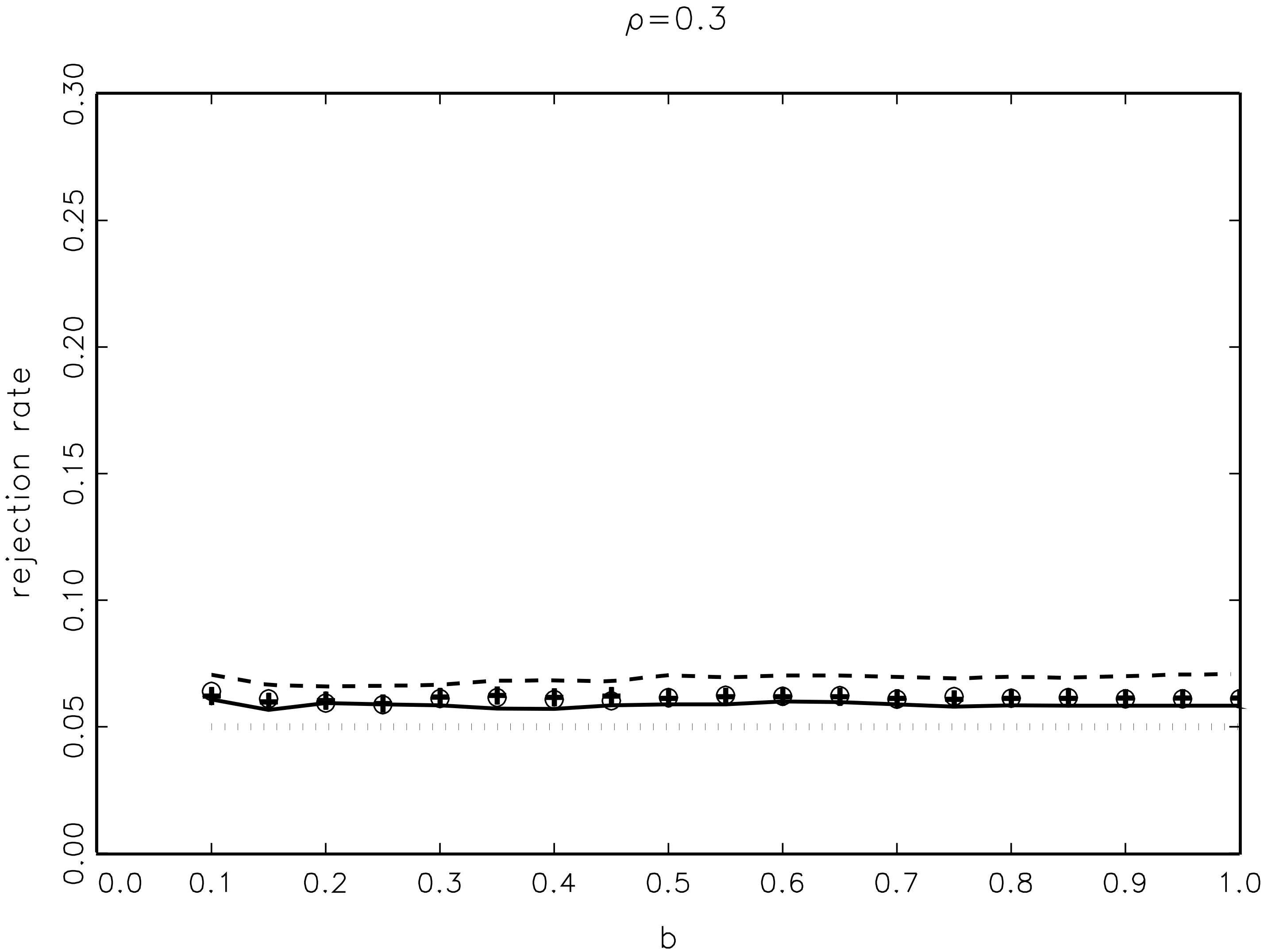


Figure 2.18: (cont'd)

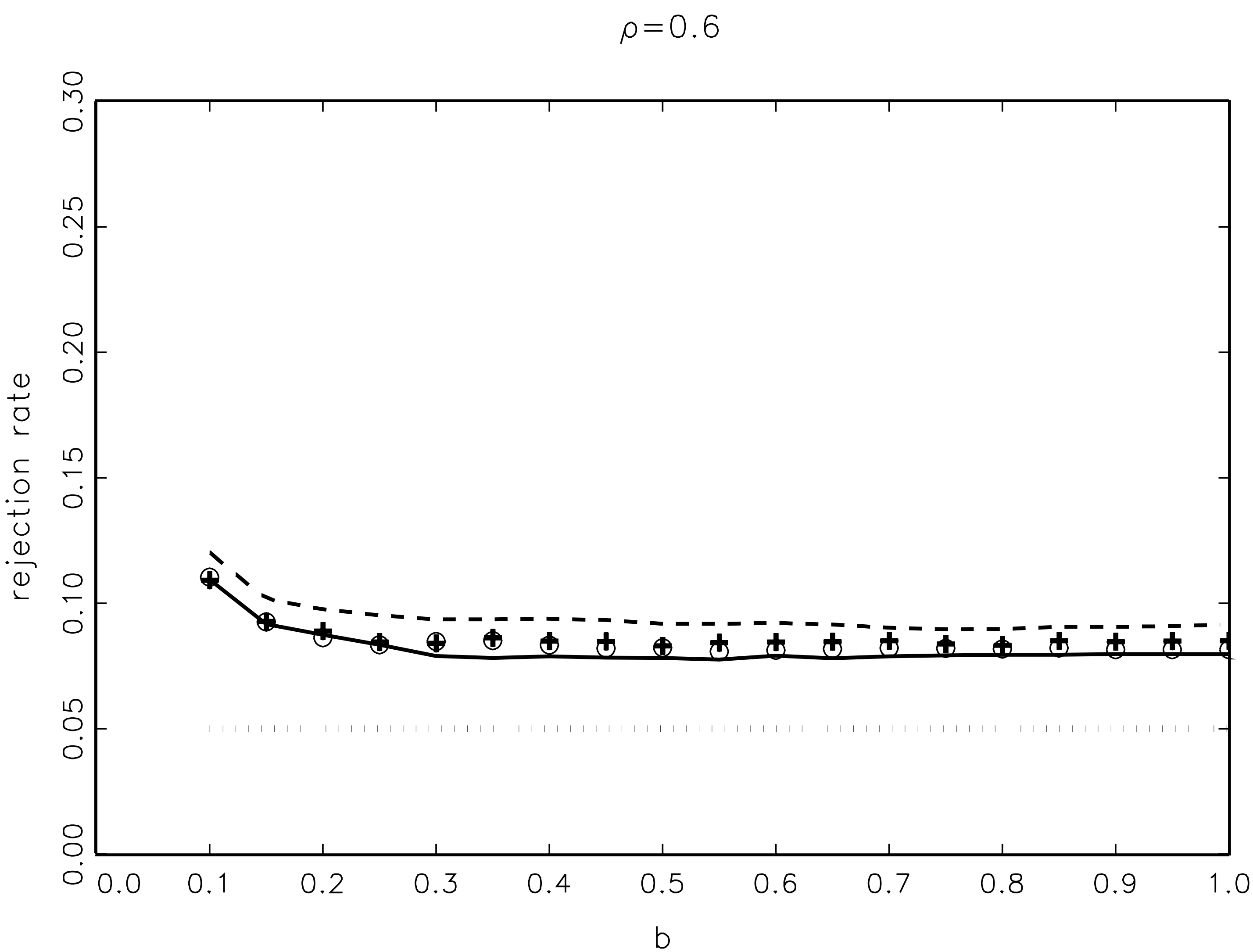


Figure 2.18: (cont'd)

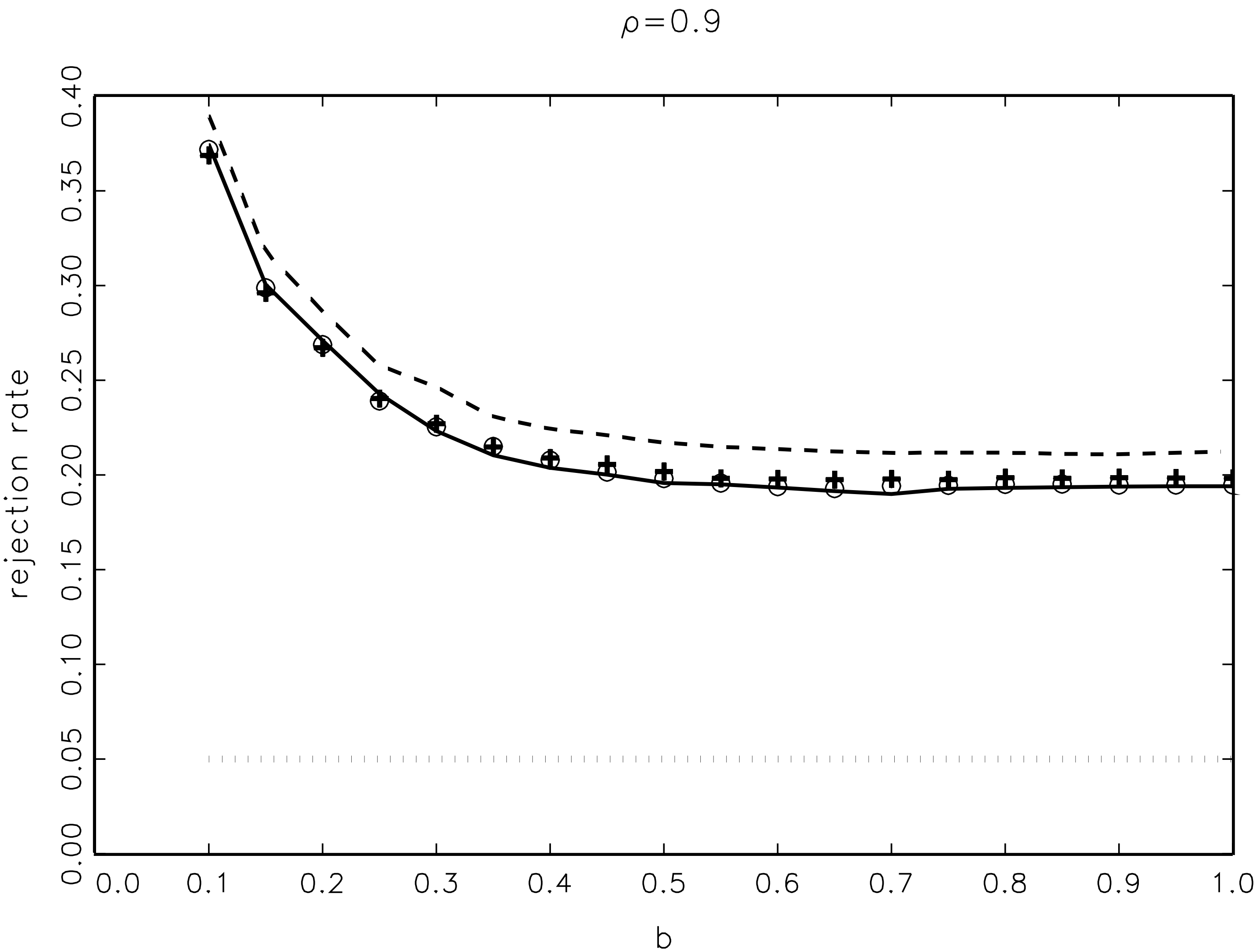


Figure 2.19: AM Series - Conditional Bernoulli ($p = 0.7$), Bartlett, $T = 50$

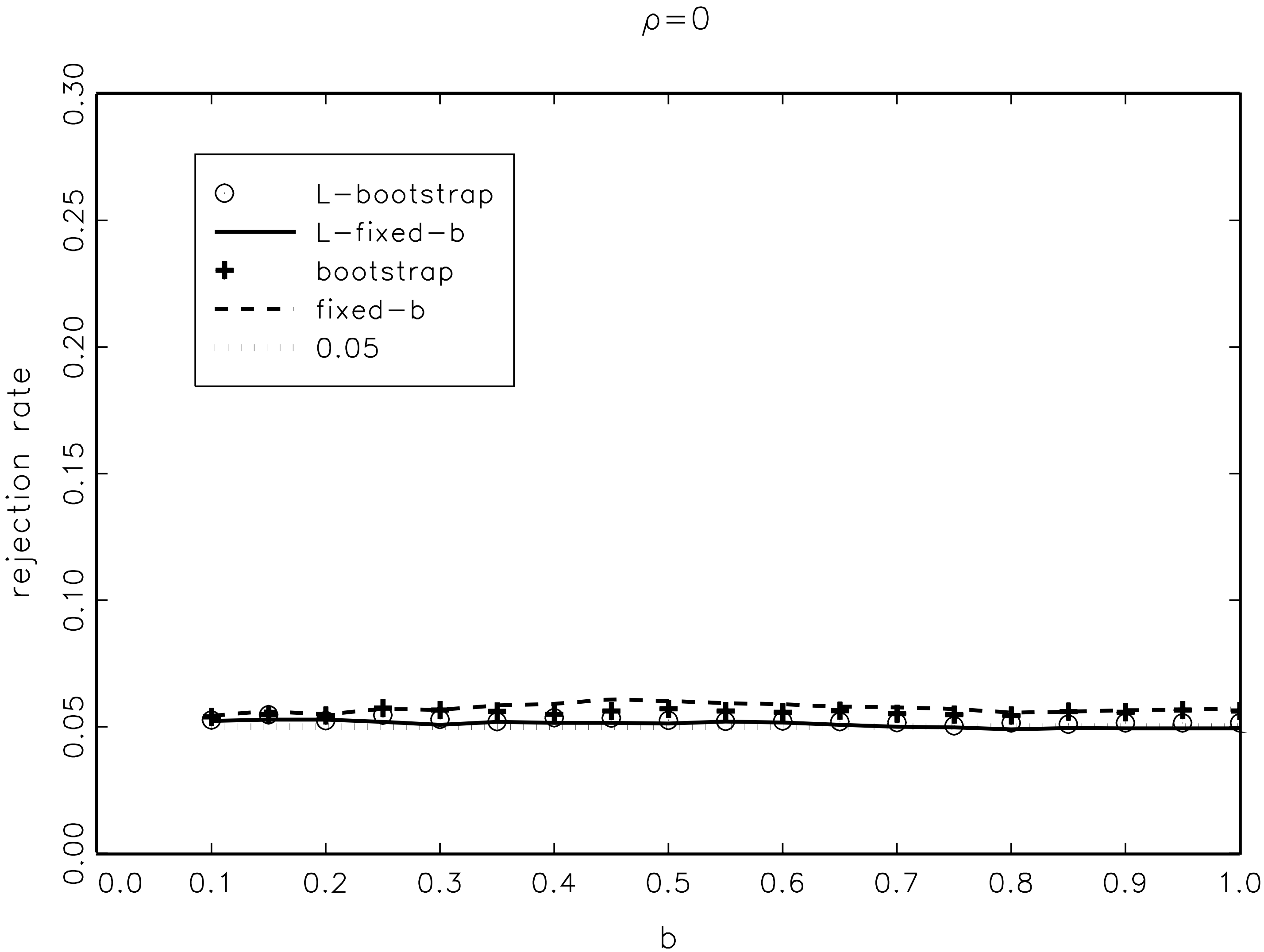


Figure 2.19: (cont'd)

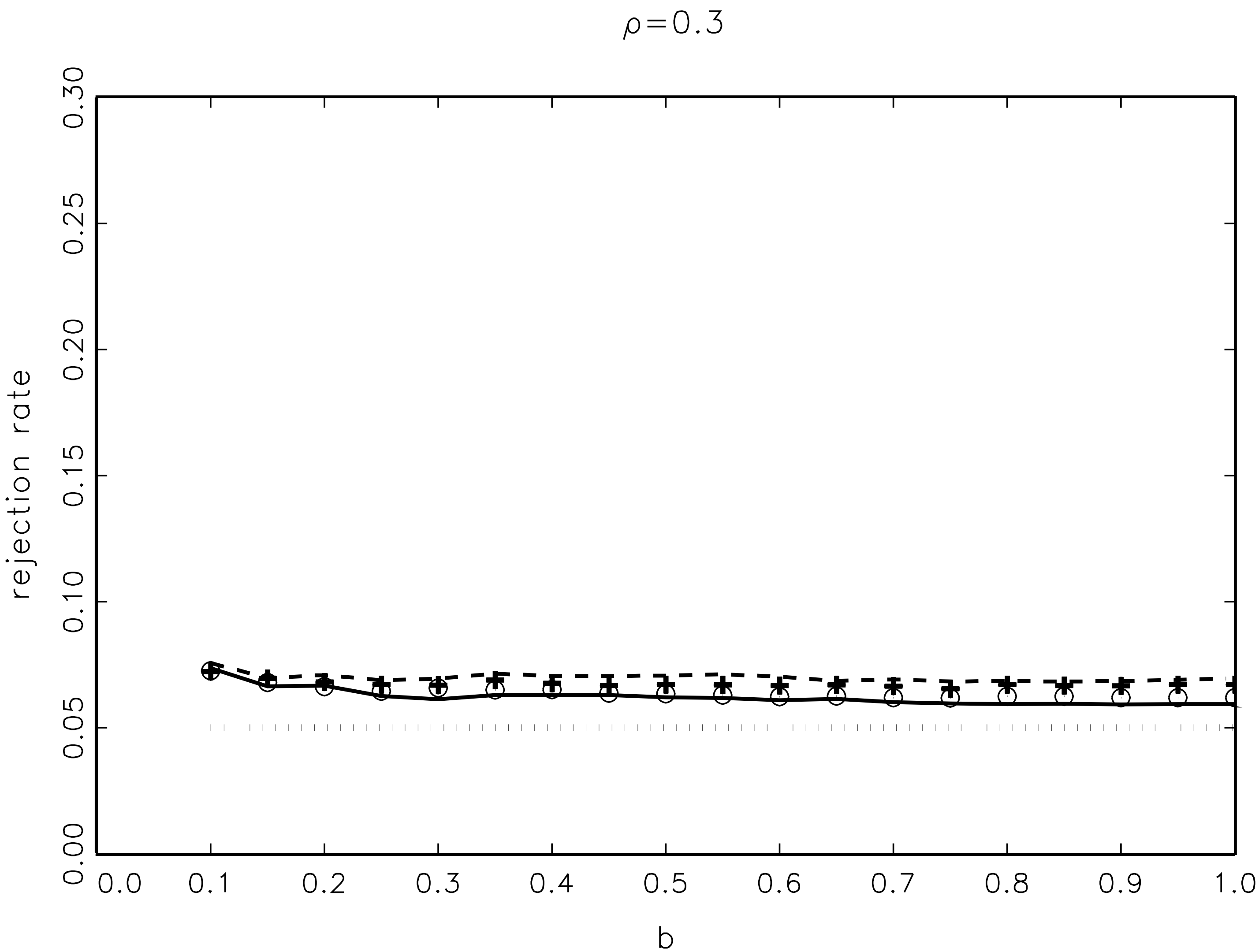


Figure 2.19: (cont'd)

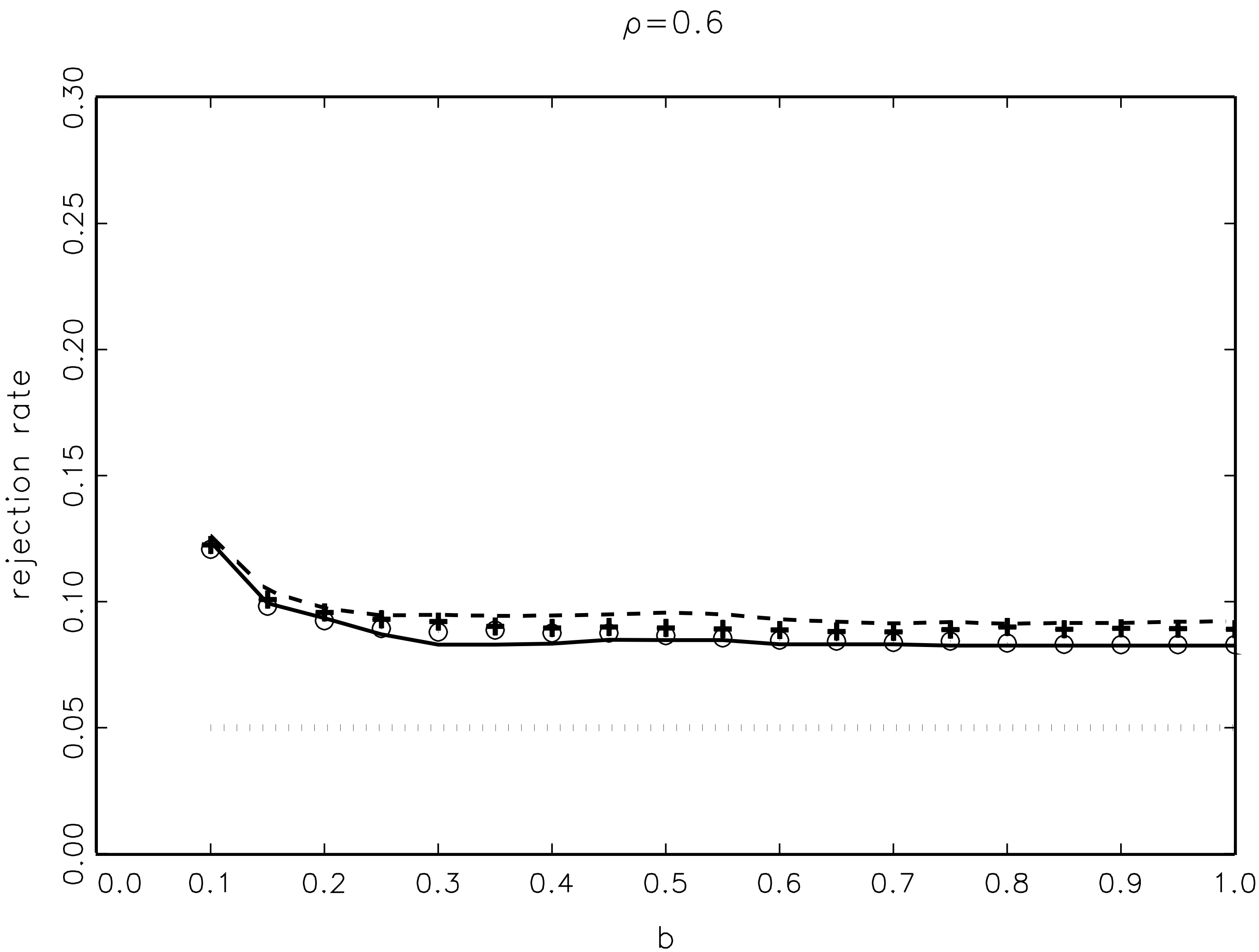


Figure 2.19: (cont'd)

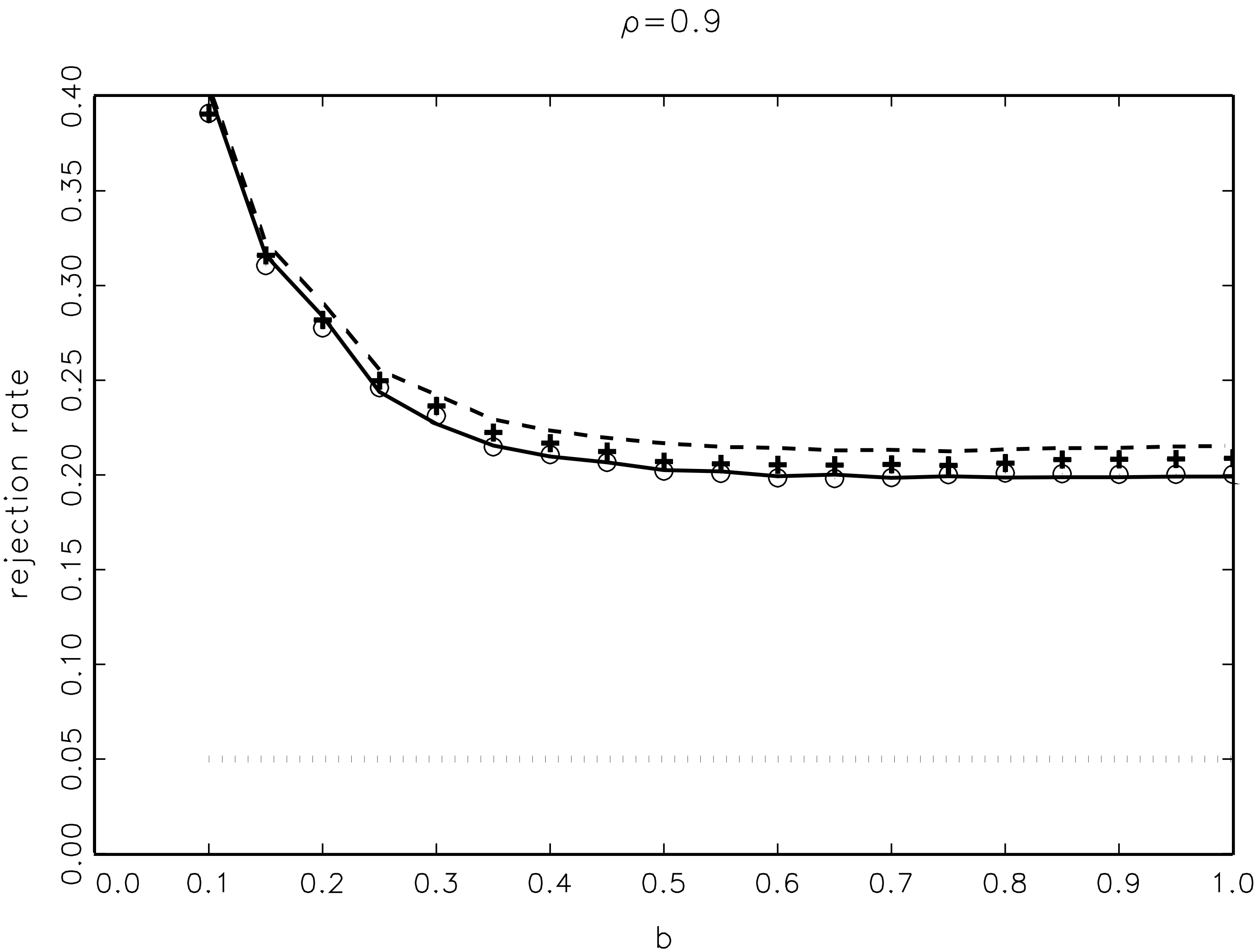


Figure 2.20: AM Series - Conditional Bernoulli ($p = 0.3$), Bartlett, $T = 100$

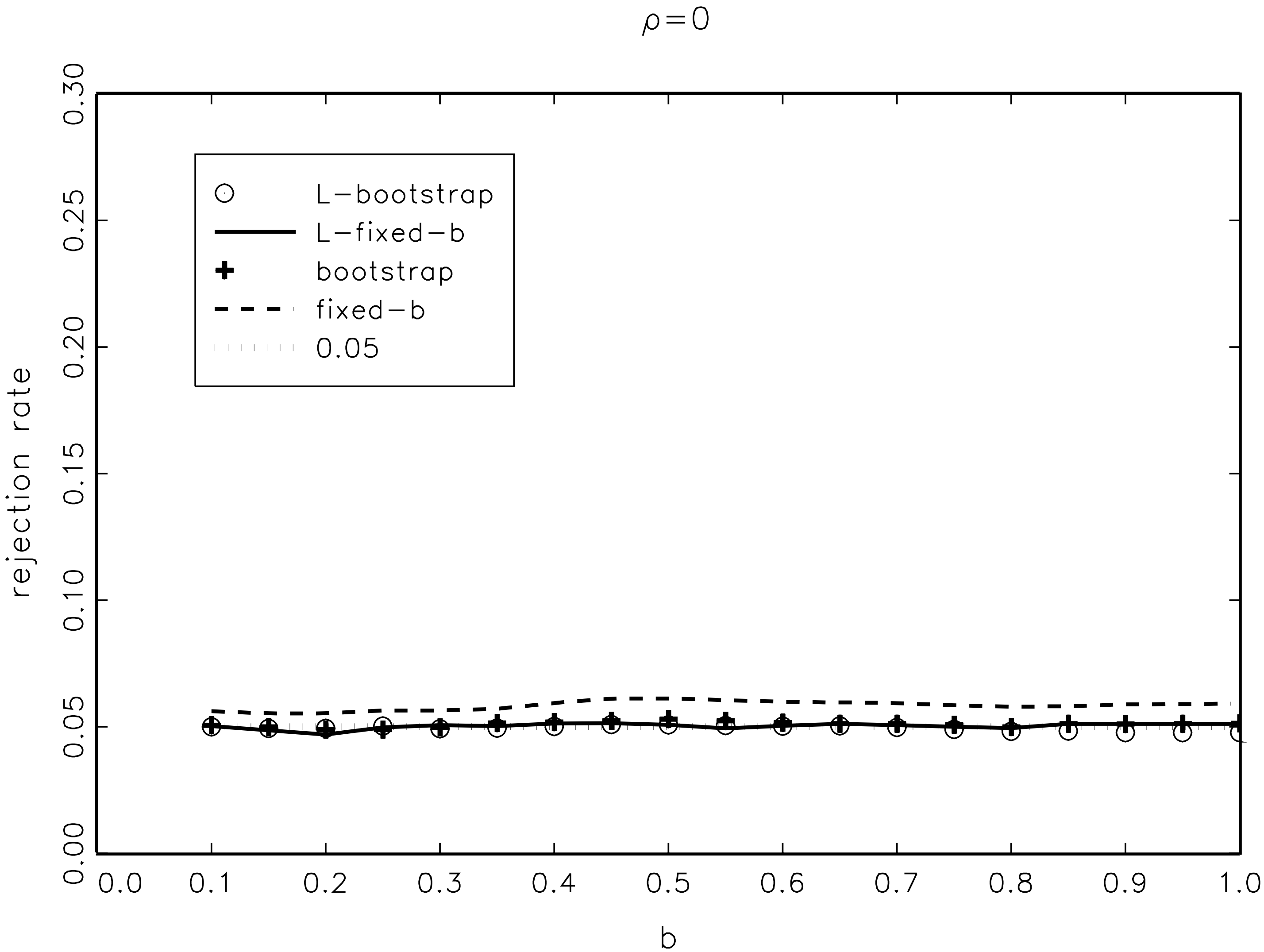


Figure 2.20: (cont'd)

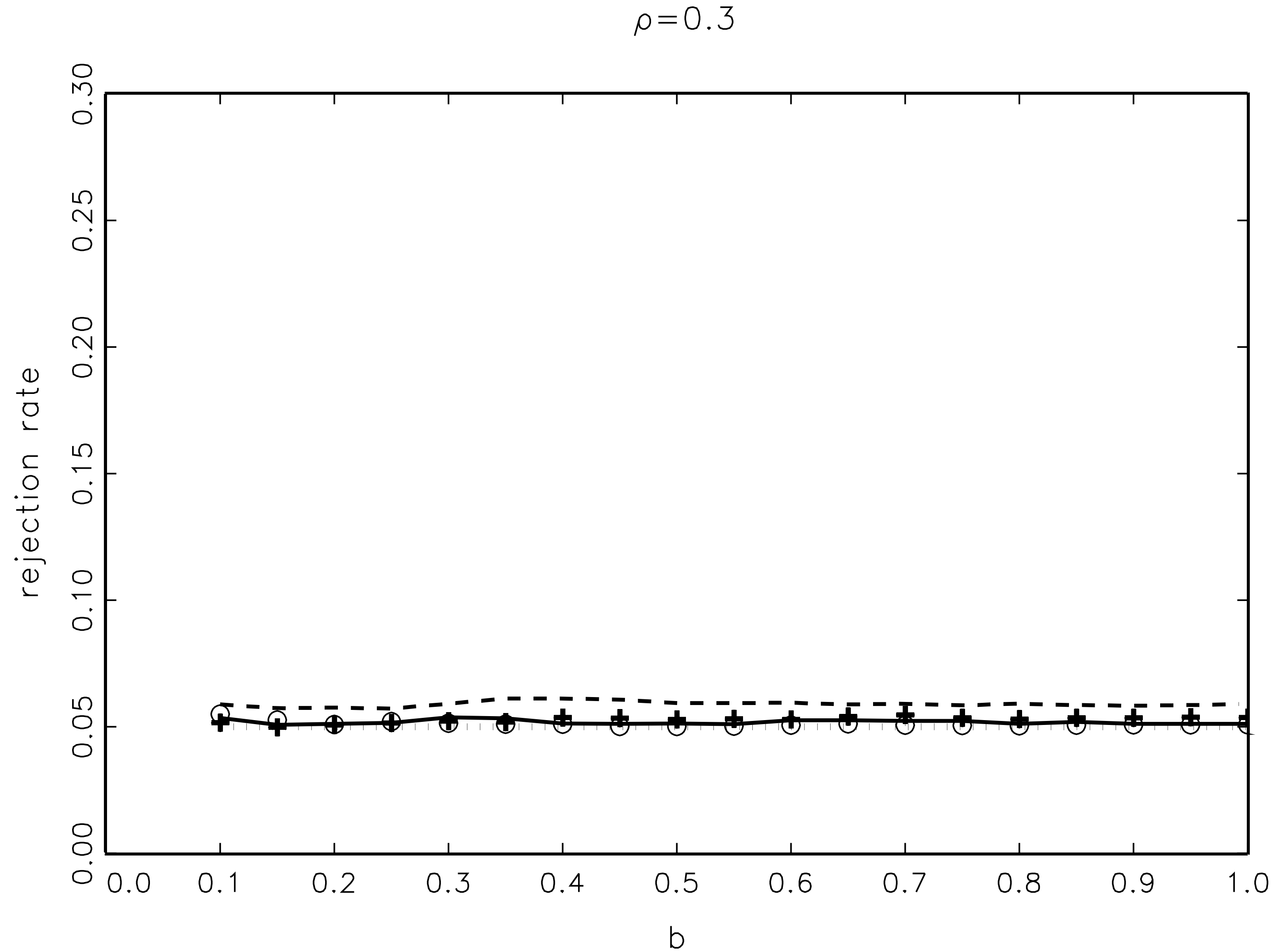


Figure 2.20: (cont'd)

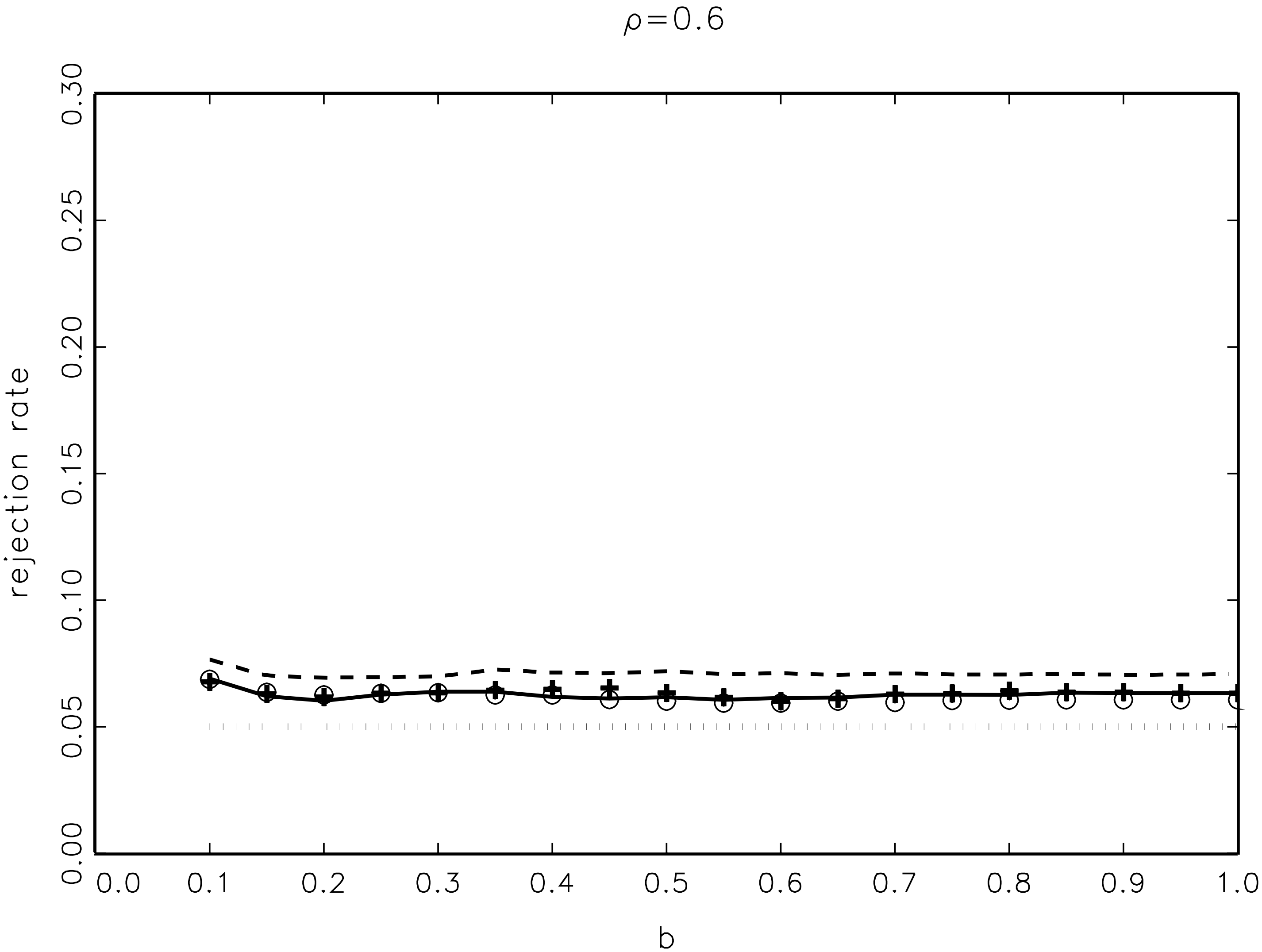


Figure 2.20: (cont'd)

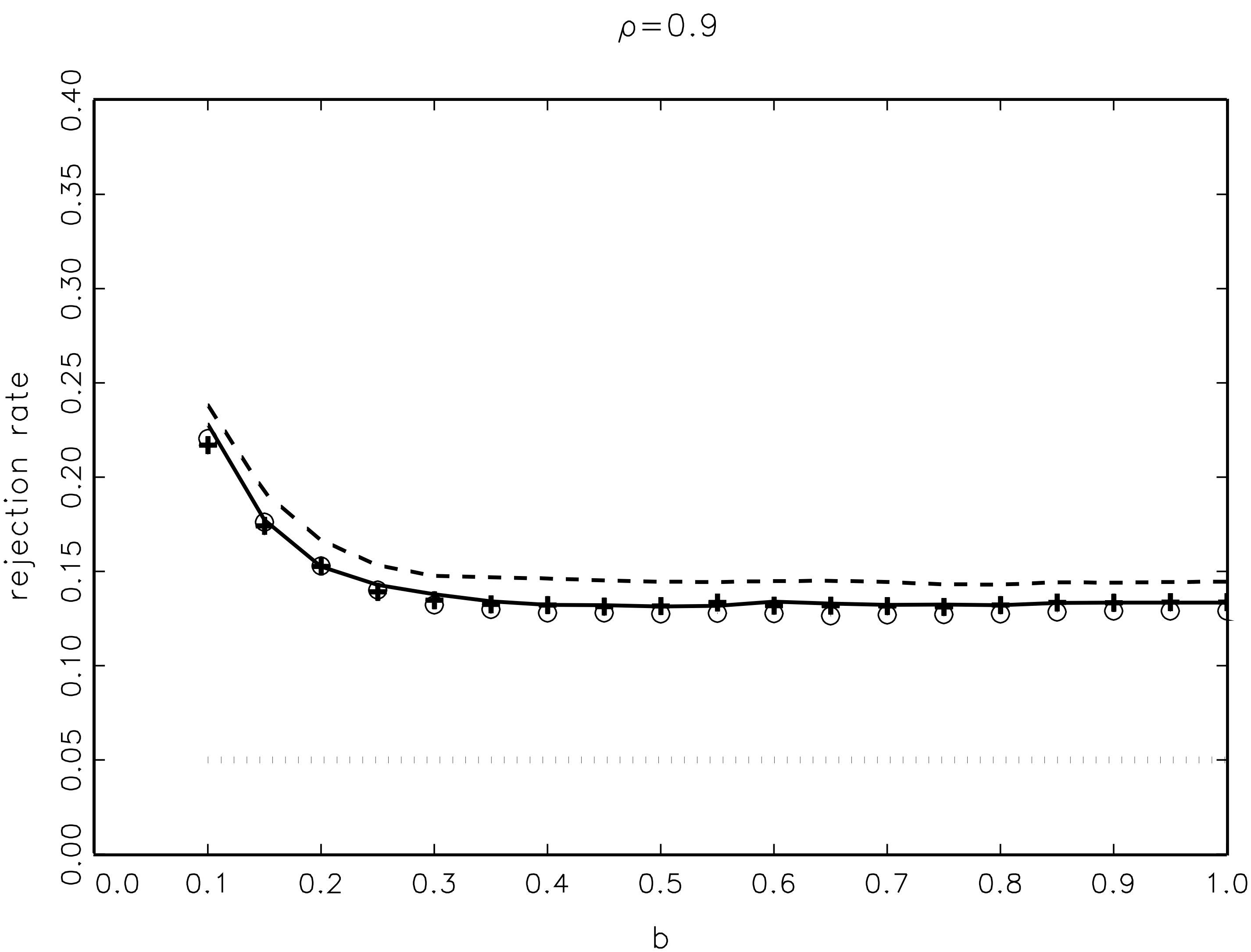


Figure 2.21: AM Series - Conditional Bernoulli ($p = 0.5$), Bartlett, $T = 100$

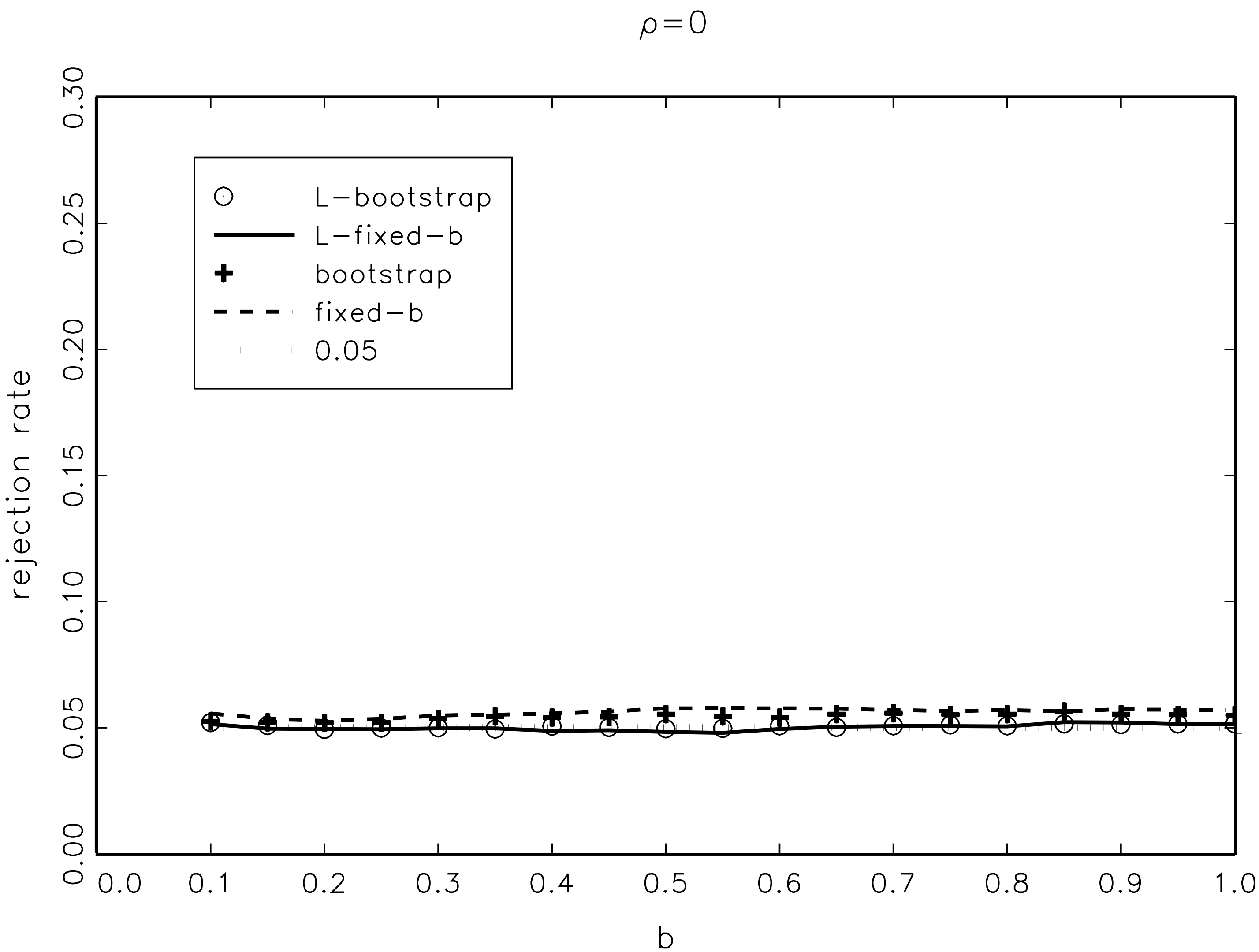


Figure 2.21: (cont'd)

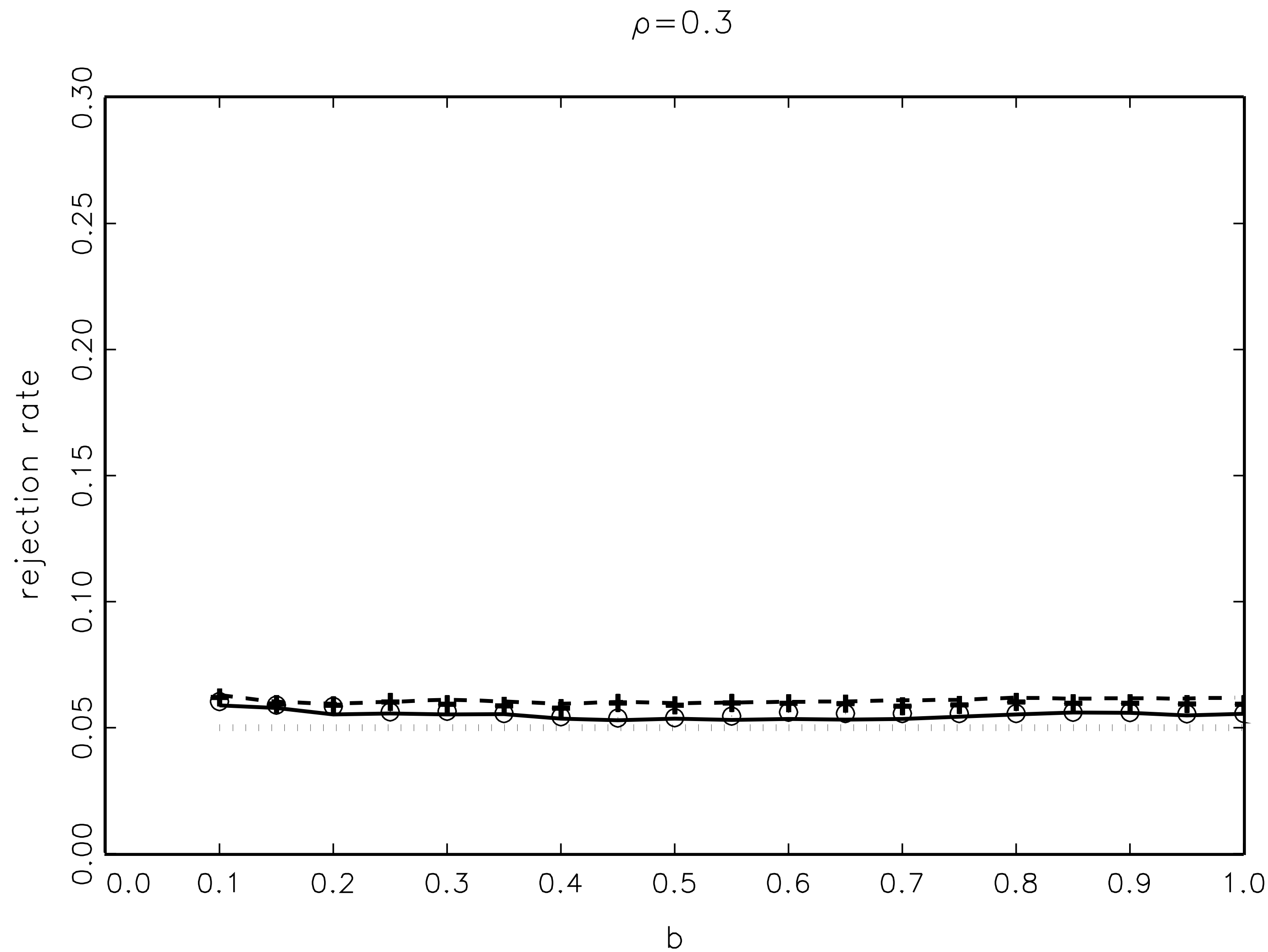


Figure 2.21: (cont'd)

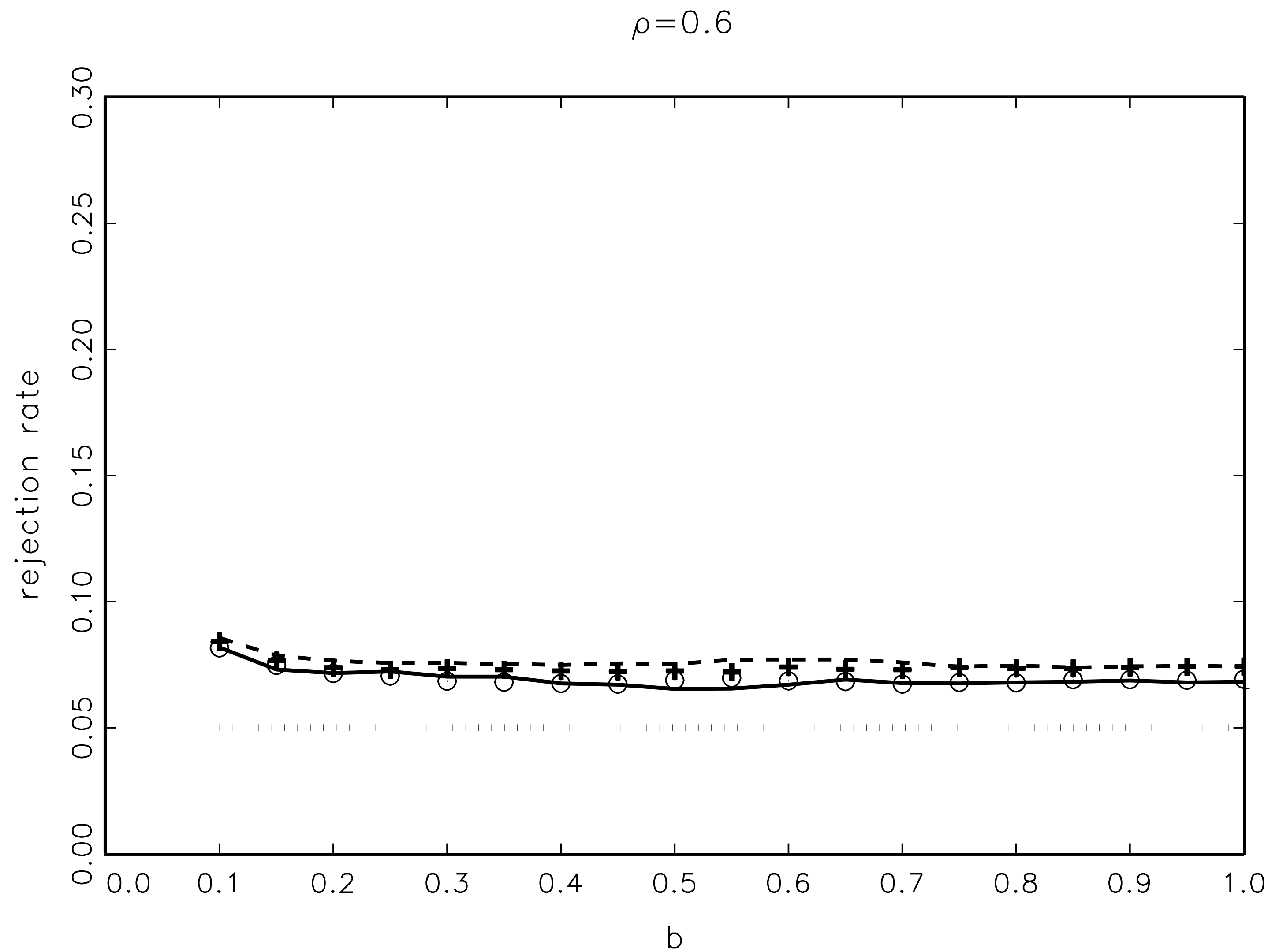


Figure 2.21: (cont'd)

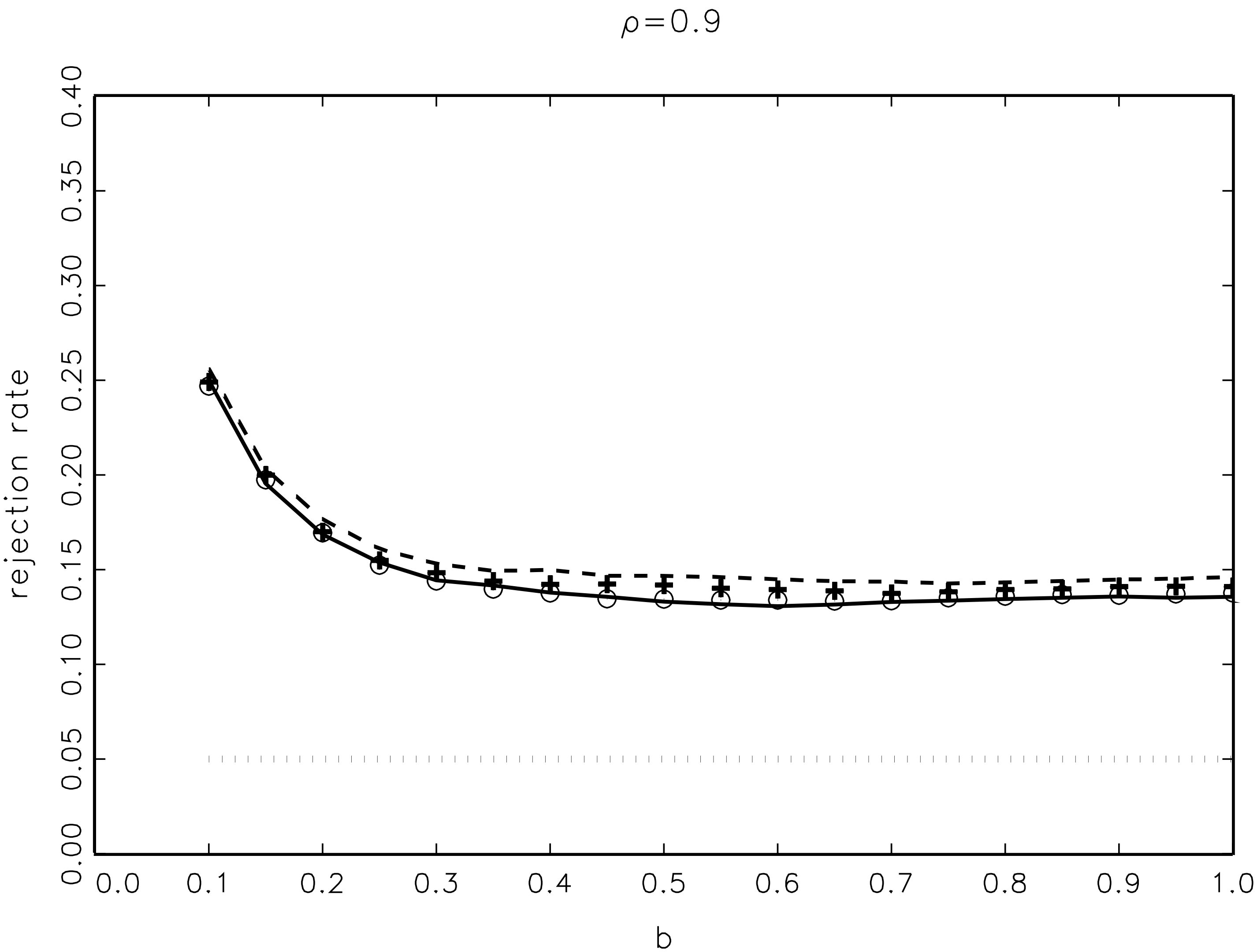


Figure 2.22: AM Series - Conditional Bernoulli ($p = 0.7$), Bartlett, $T = 100$

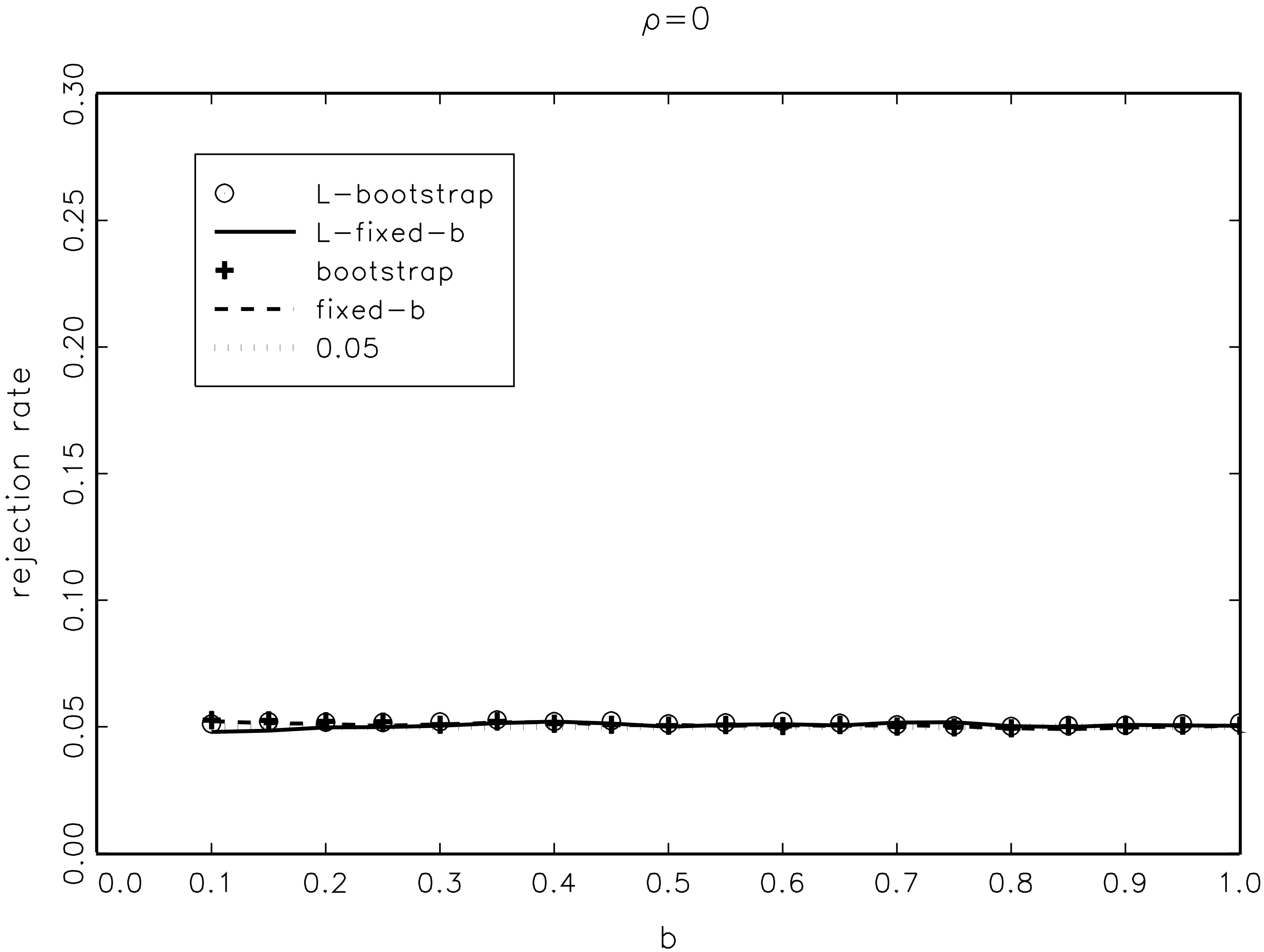


Figure 2.22: (cont'd)

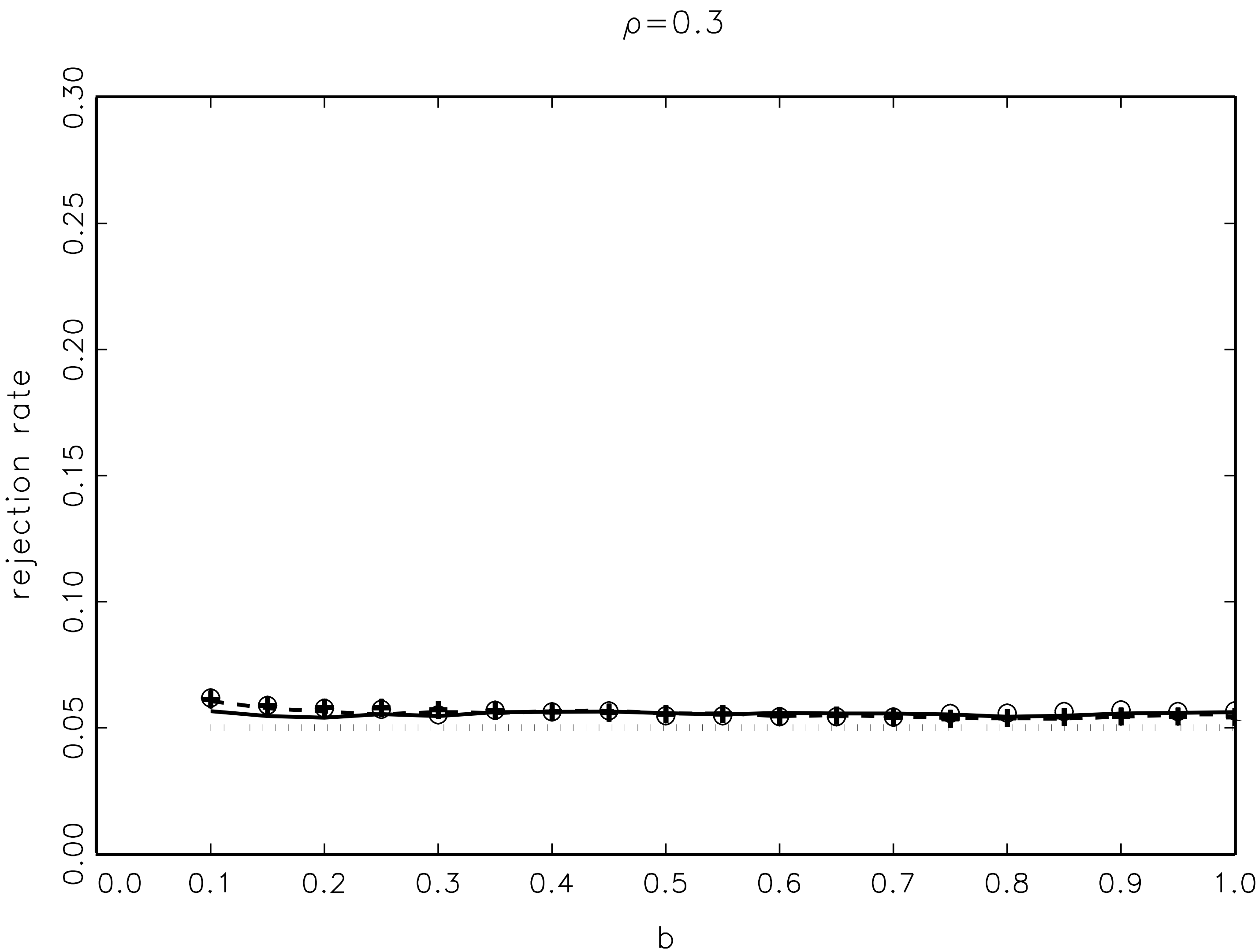


Figure 2.22: (cont'd)

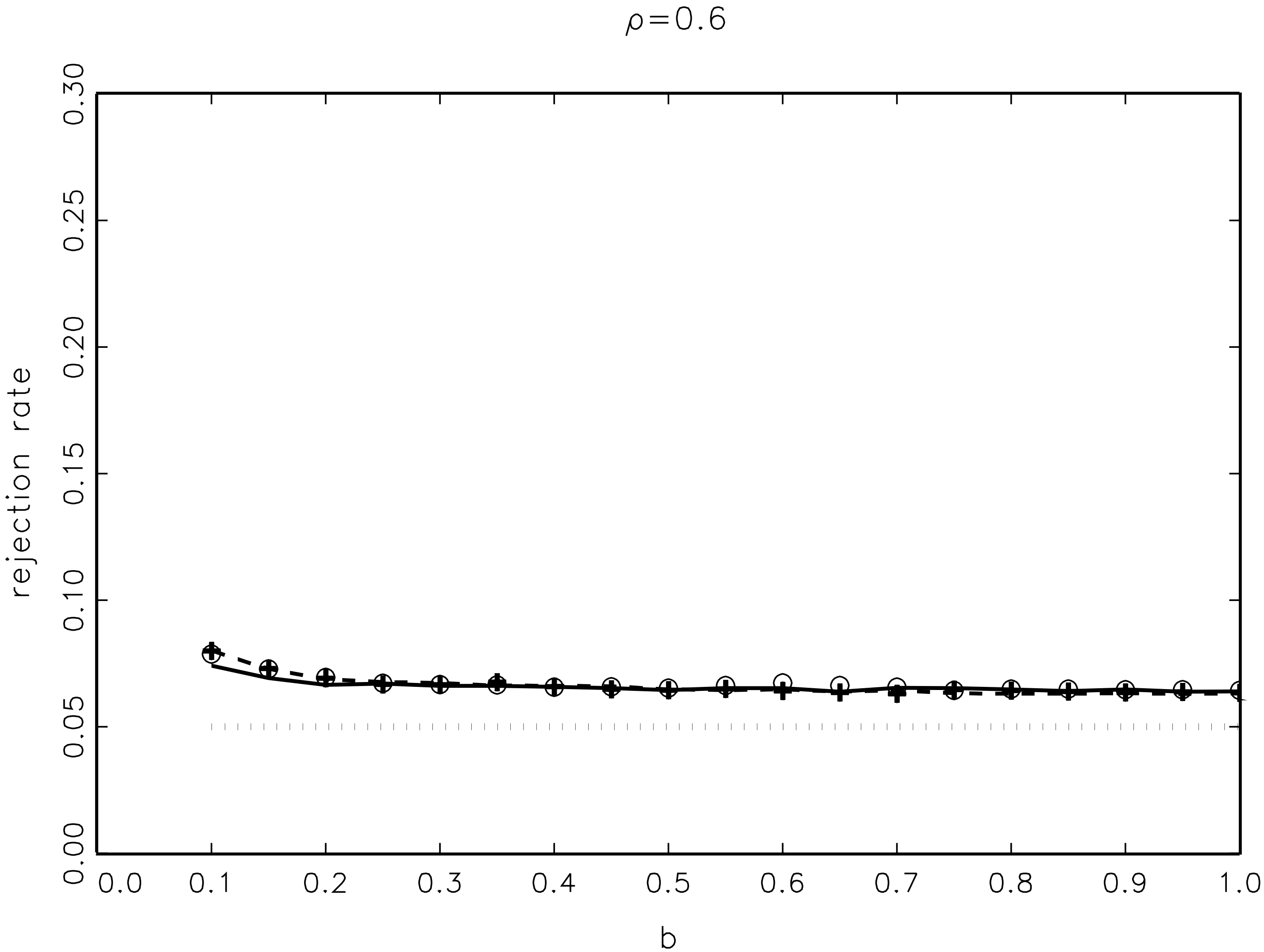


Figure 2.22: (cont'd)

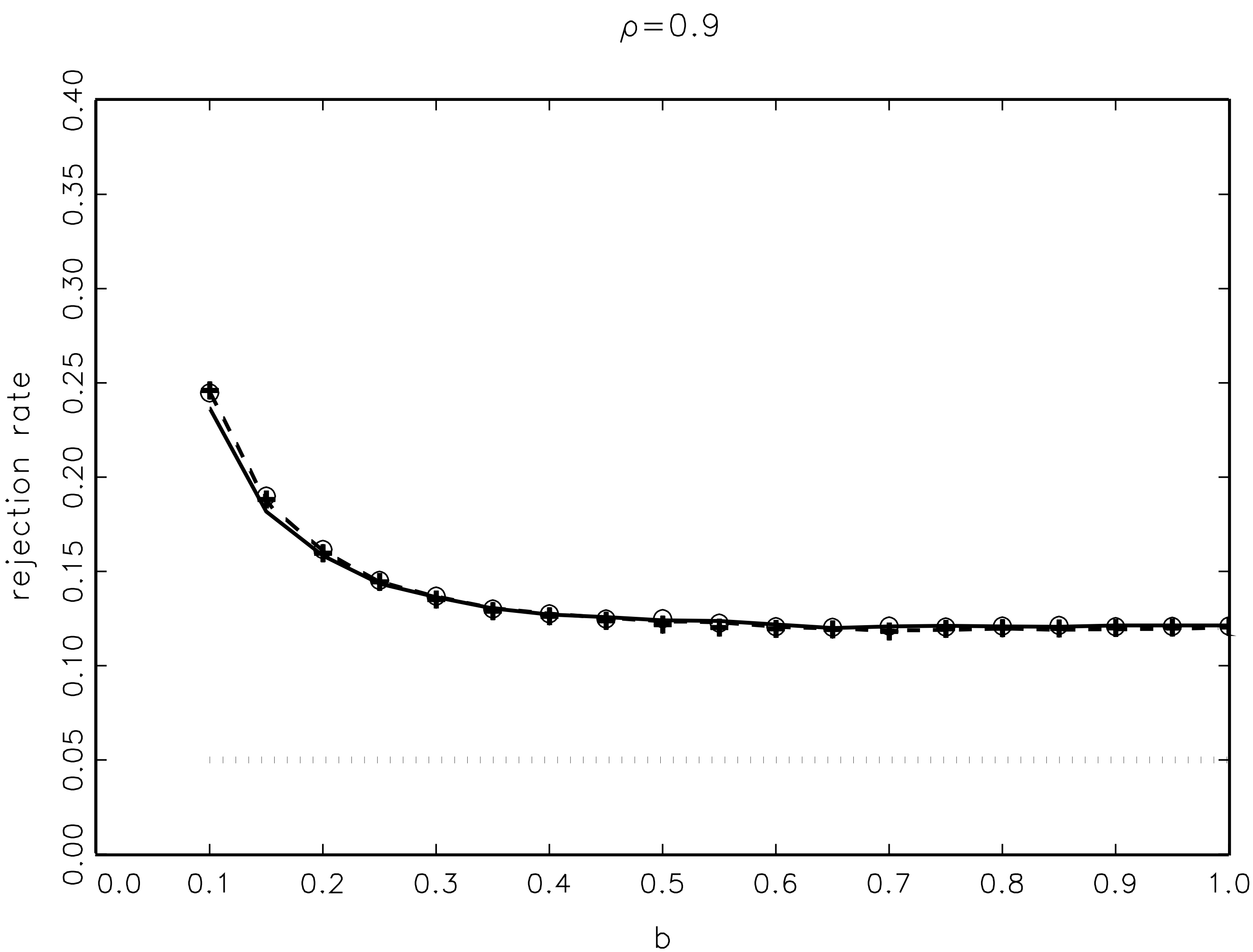


Figure 2.23: AM Series - Conditional Bernoulli ($p = 0.3$), Bartlett, $T = 200$

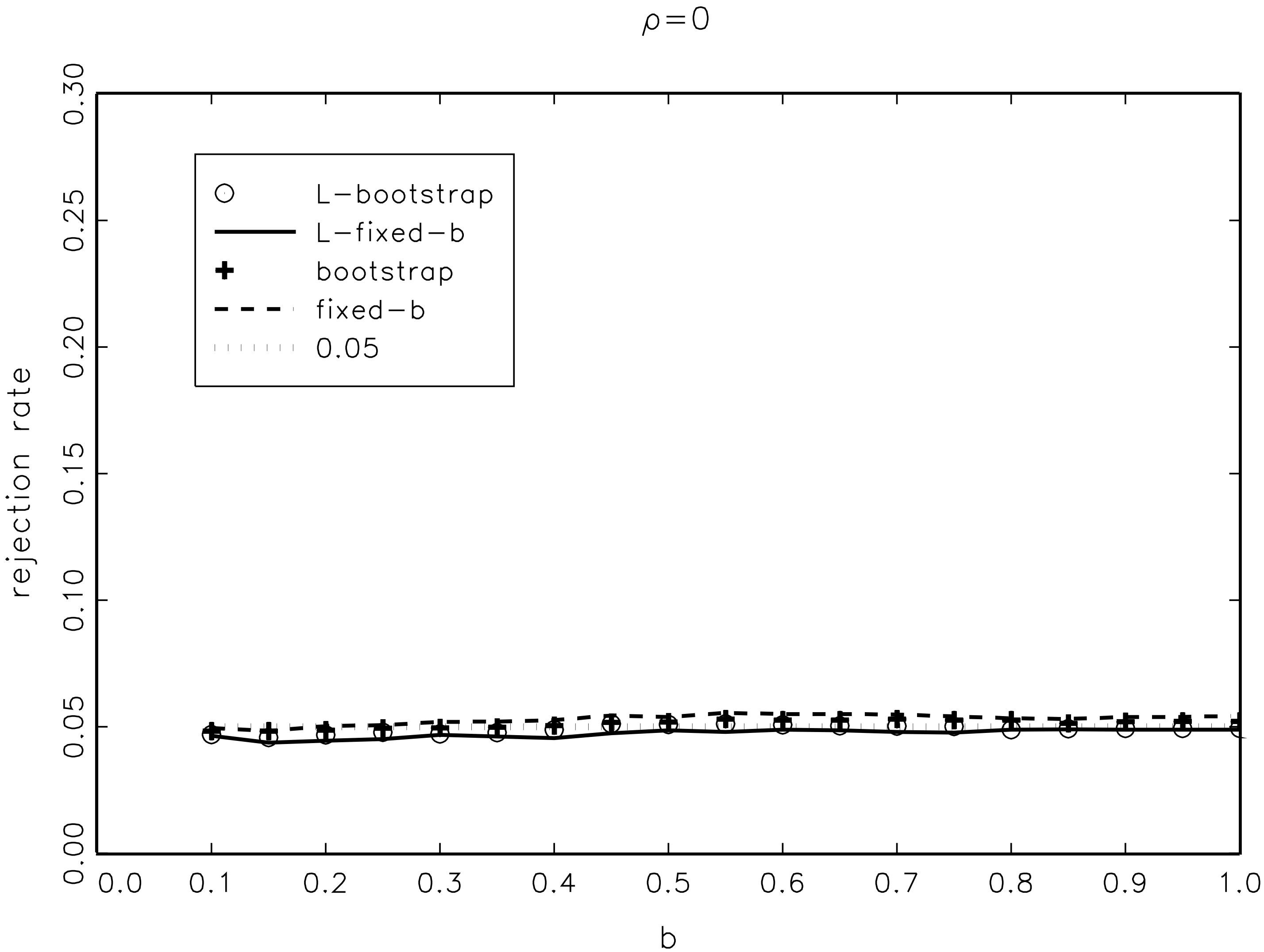


Figure 2.23: (cont'd)

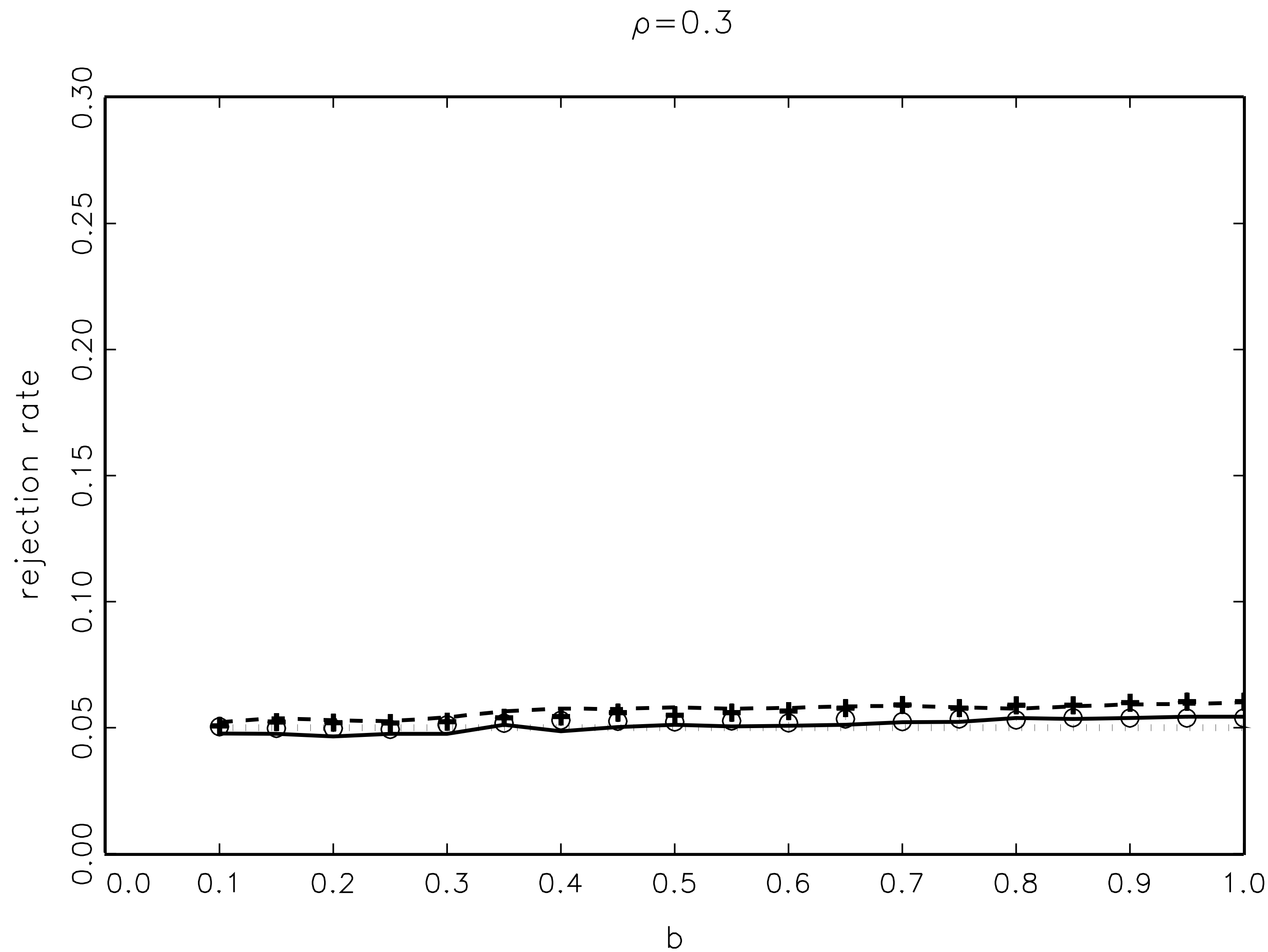


Figure 2.23: (cont'd)

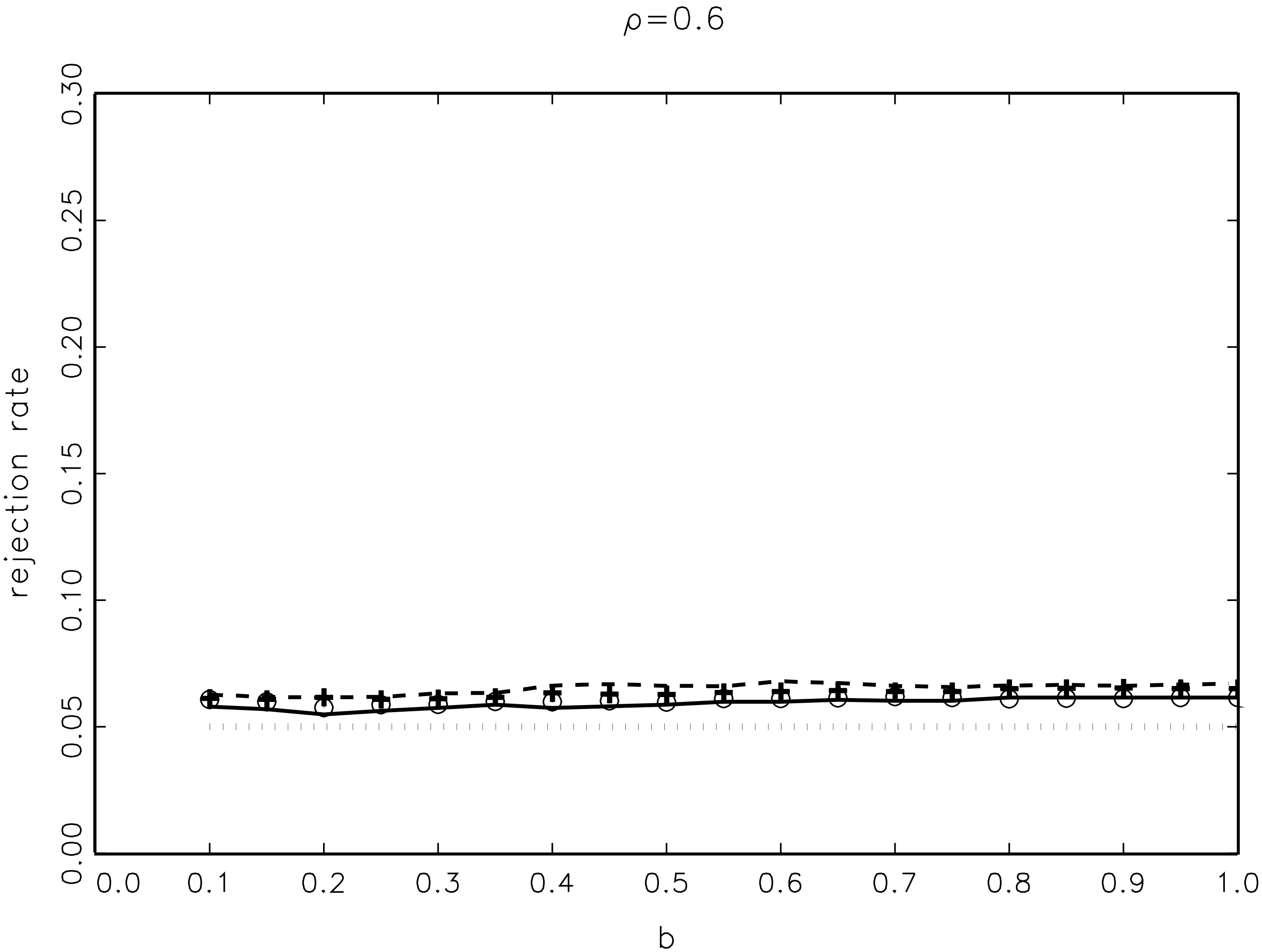


Figure 2.23: (cont'd)

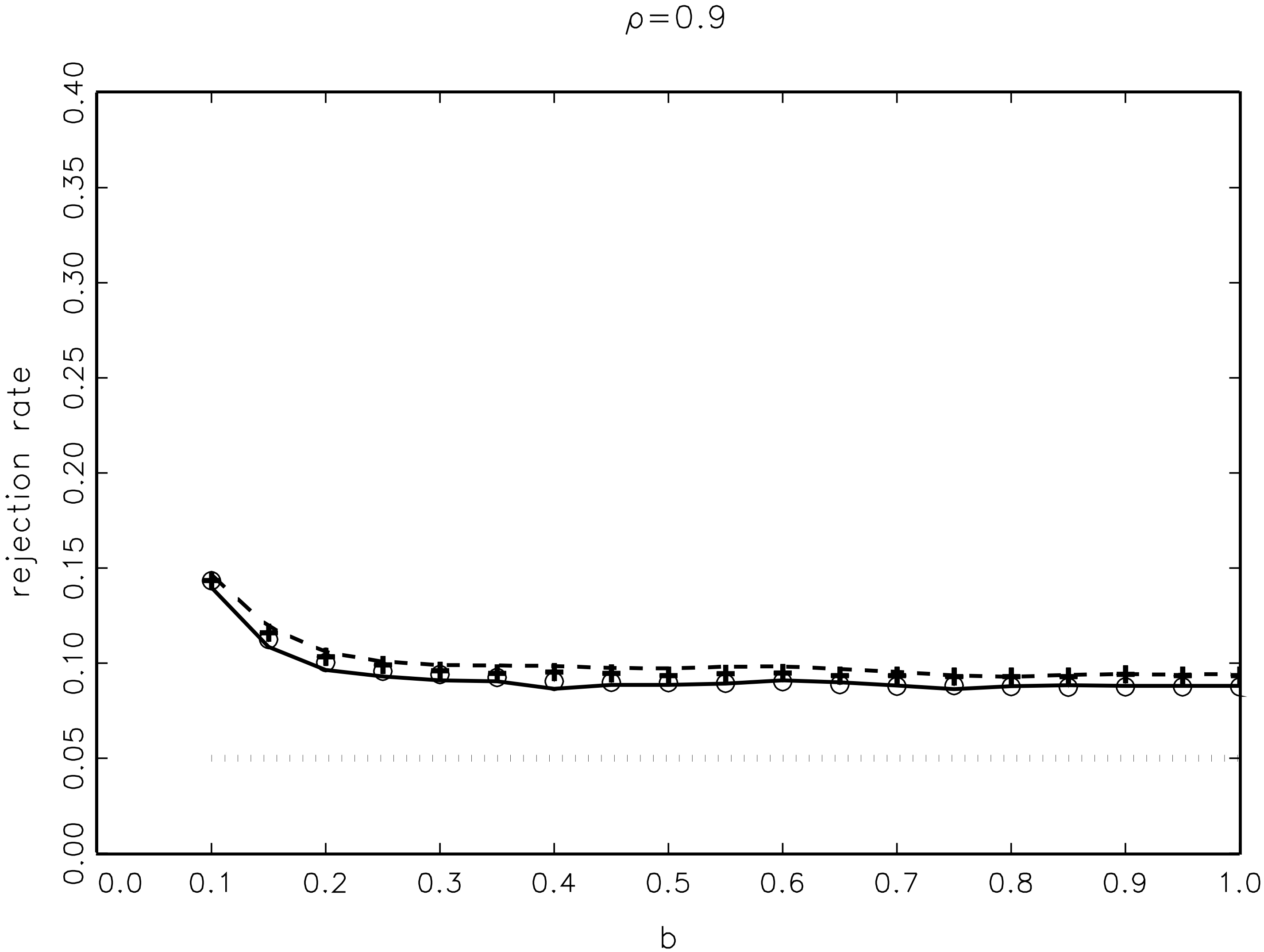


Figure 2.24: AM Series - Conditional Bernoulli ($p = 0.5$), Bartlett, $T = 200$

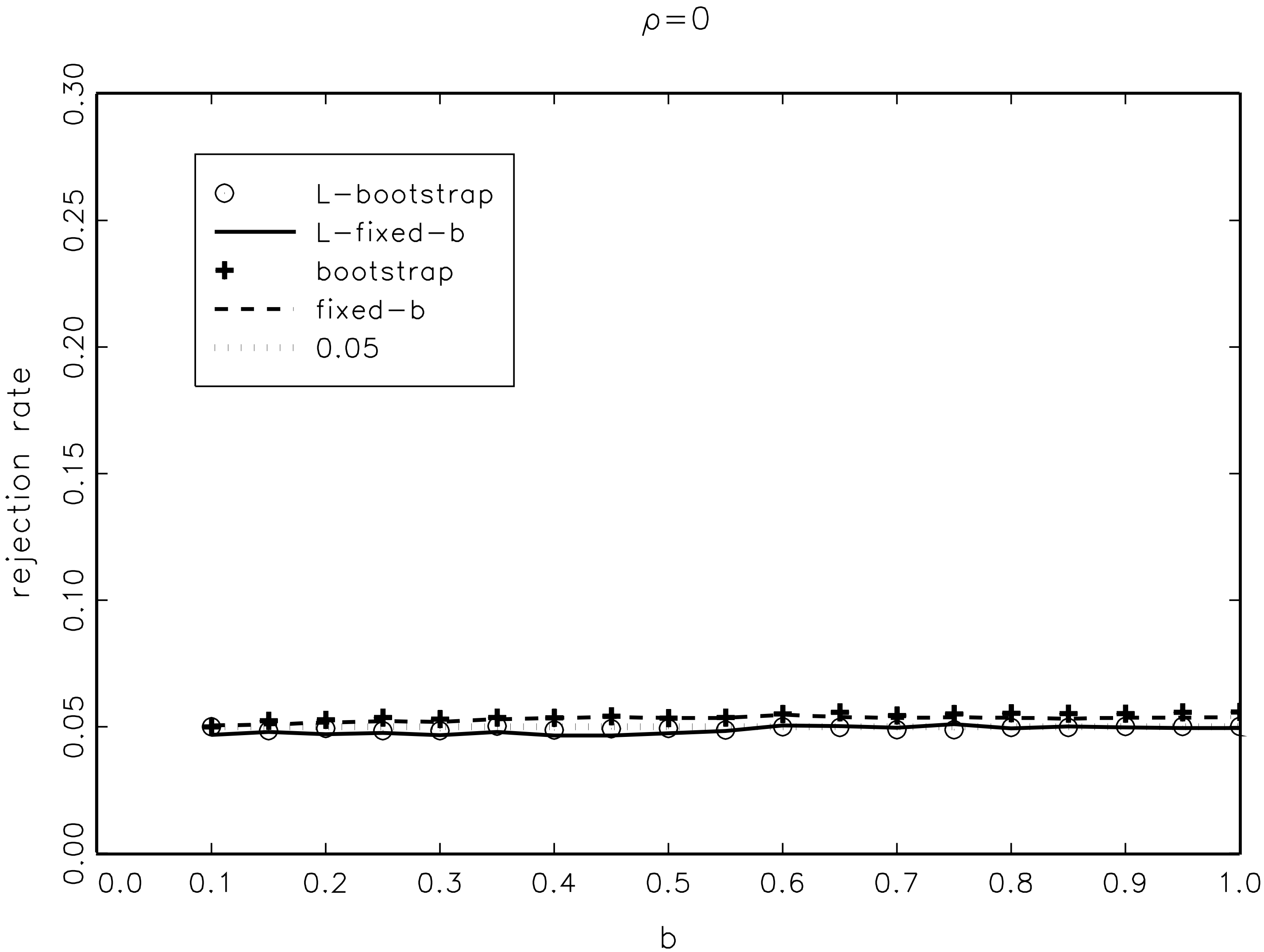


Figure 2.24: (cont'd)

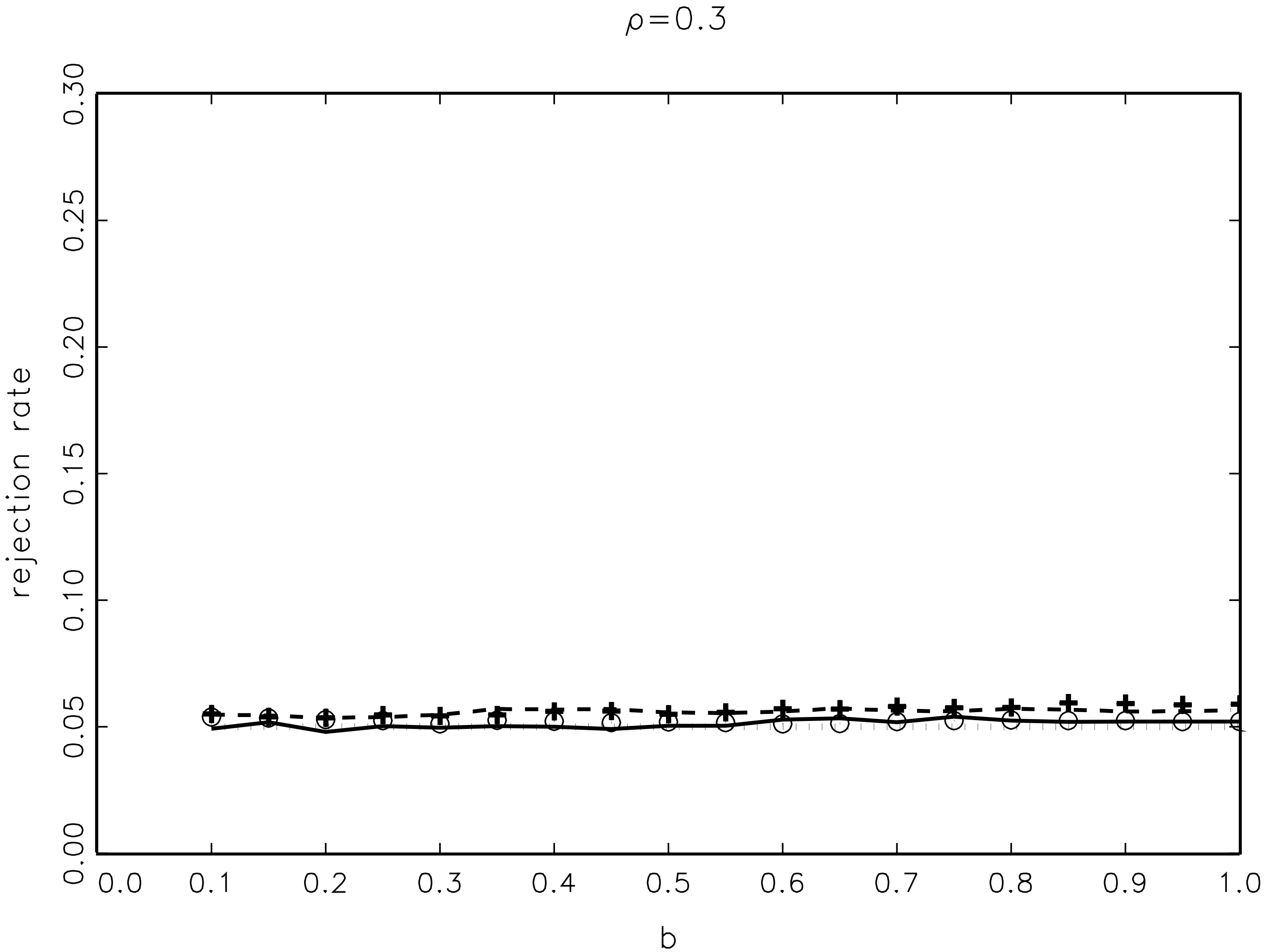


Figure 2.24: (cont'd)

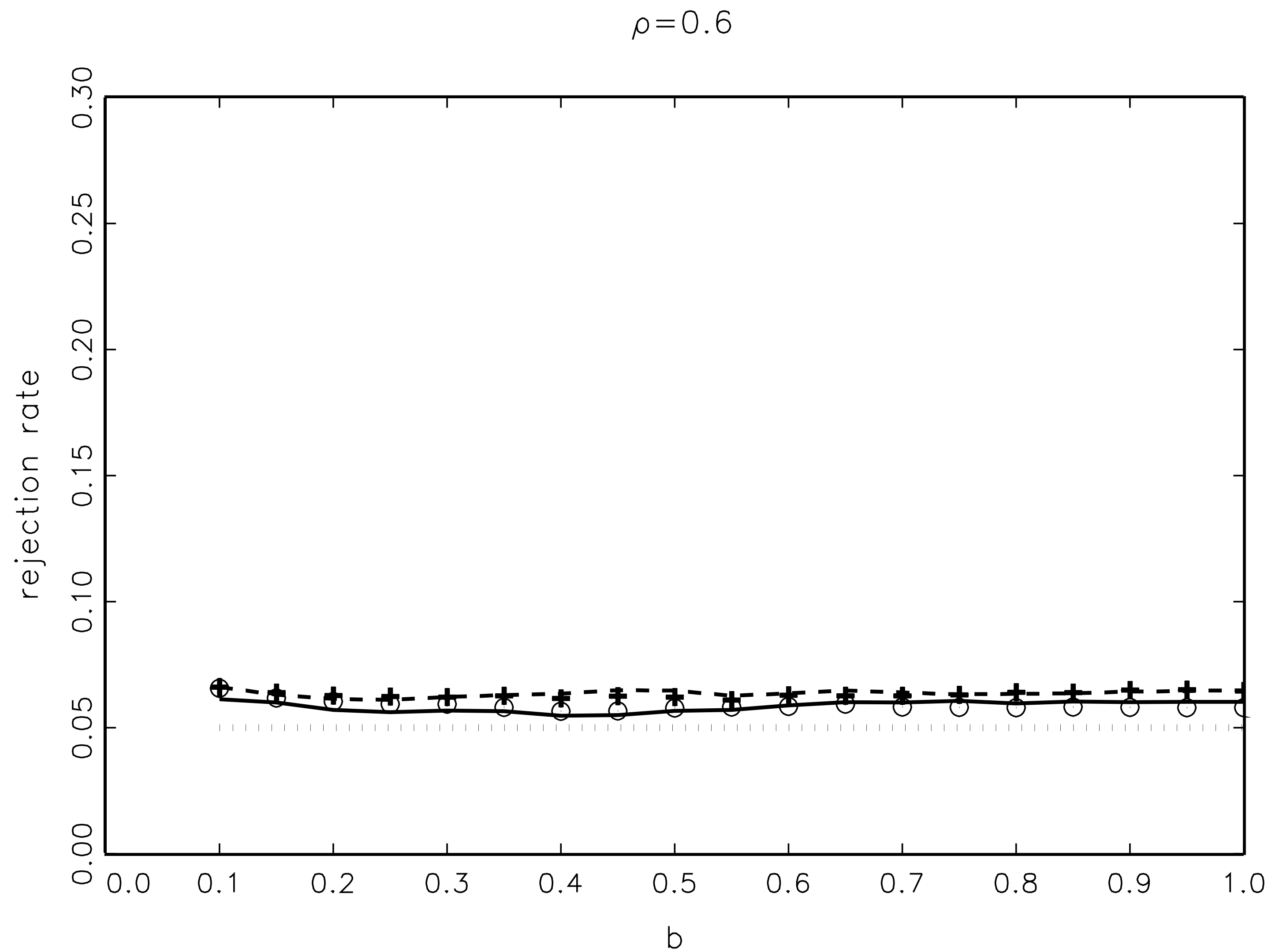


Figure 2.24: (cont'd)

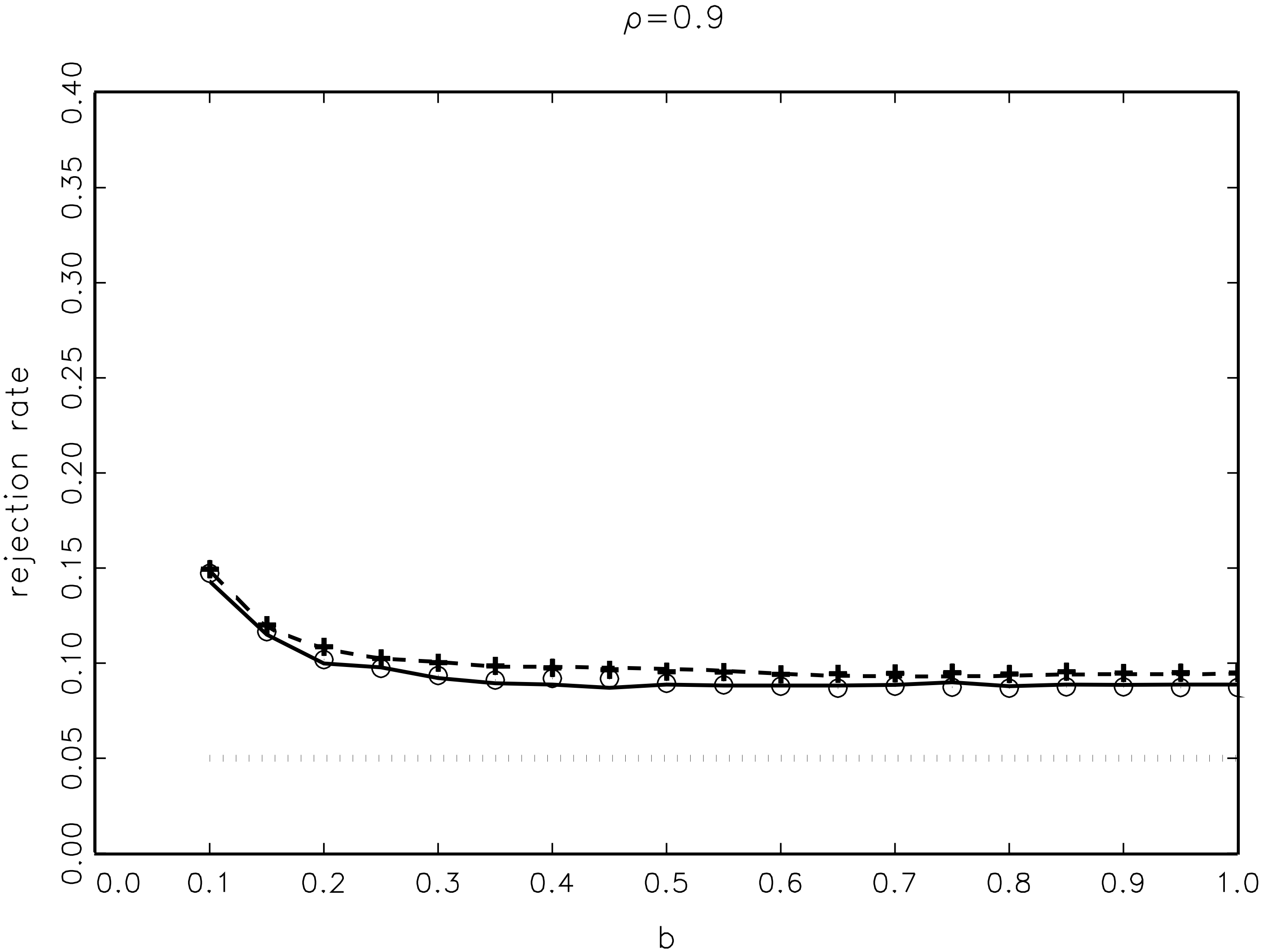


Figure 2.25: AM Series - Conditional Bernoulli ($p = 0.7$), Bartlett, $T = 200$

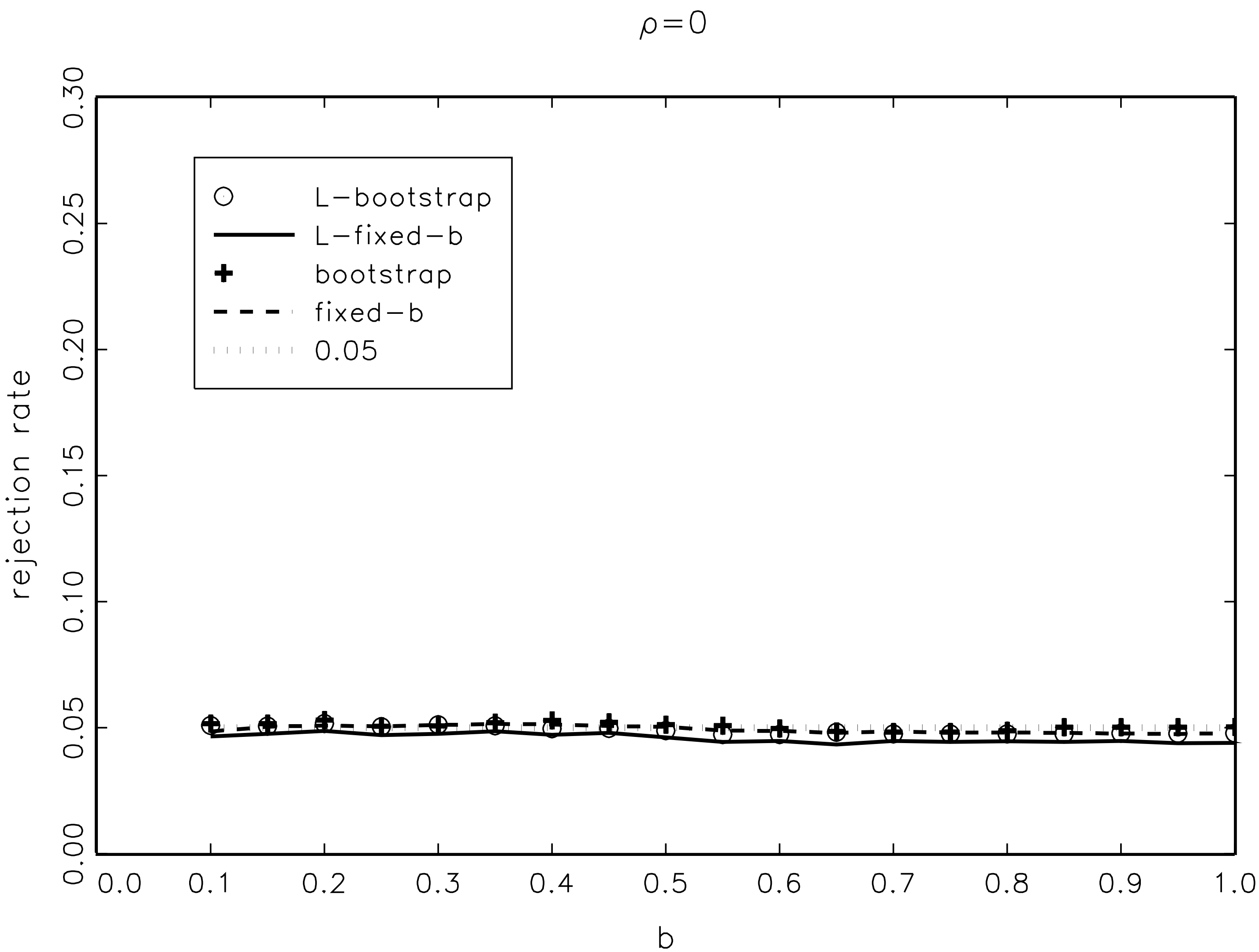


Figure 2.25: (cont'd)

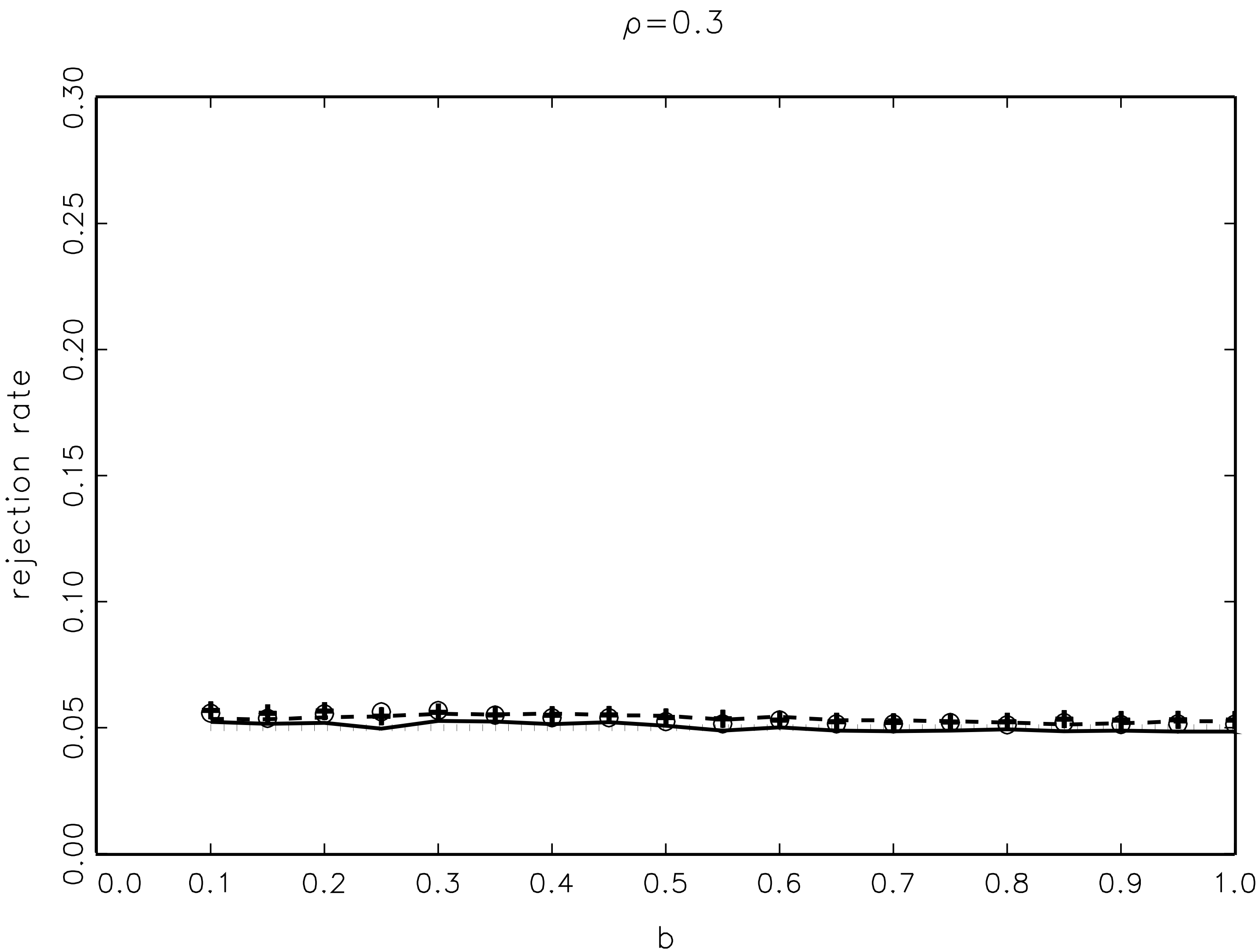


Figure 2.25: (cont'd)

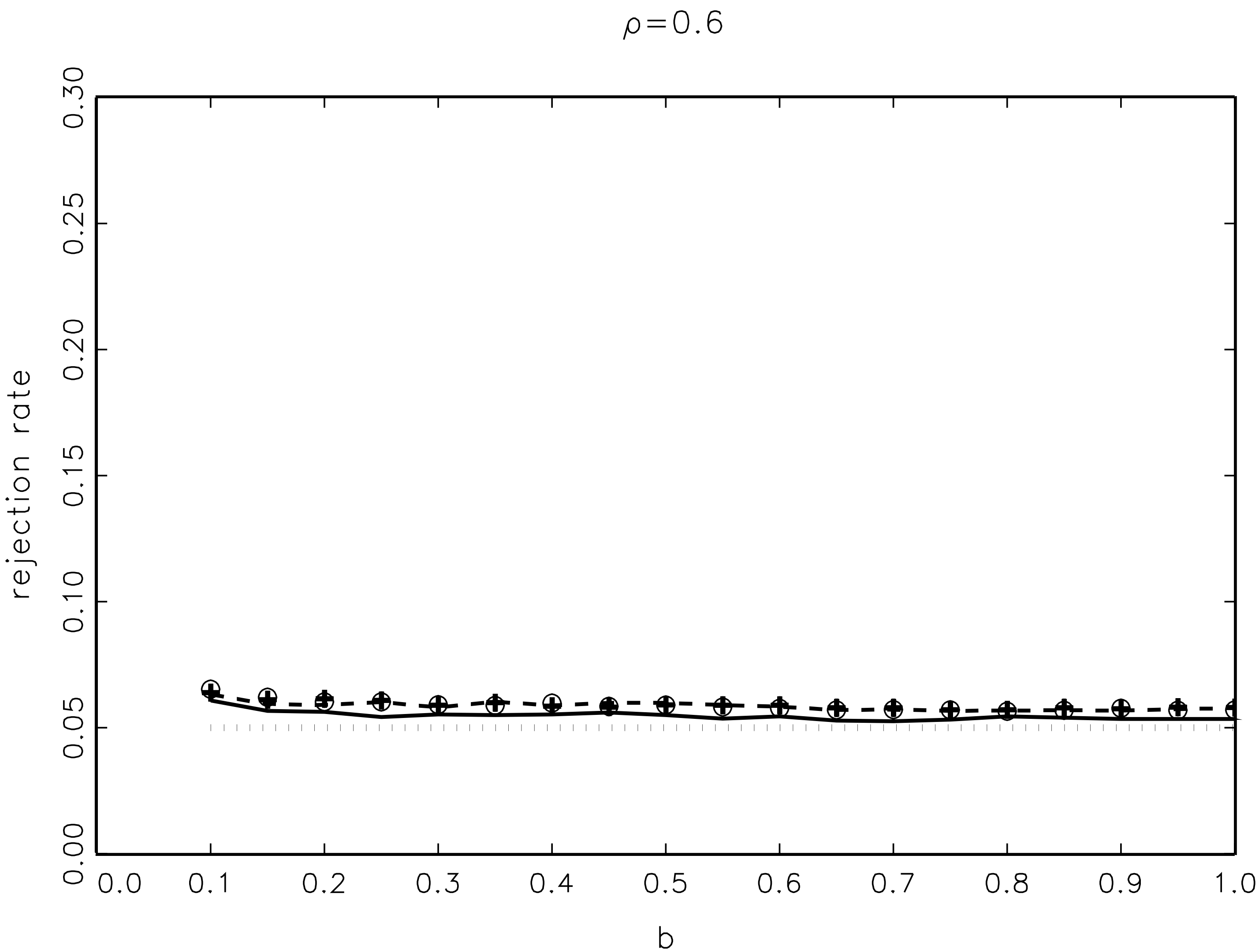


Figure 2.25: (cont'd)

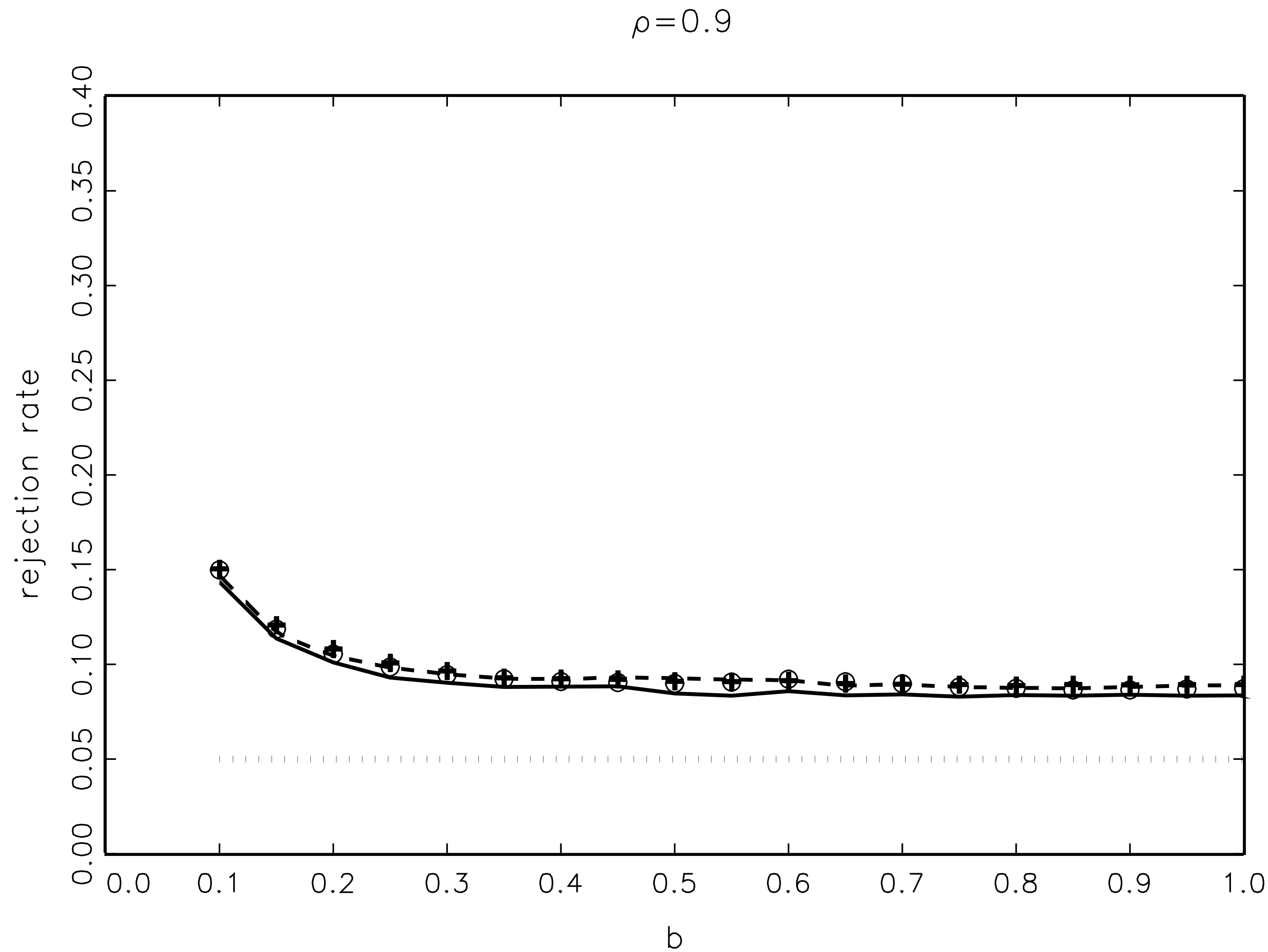


Figure 2.26: AM Series - Random Bernoulli ($p = 0.3$), Bartlett, $T = 50$

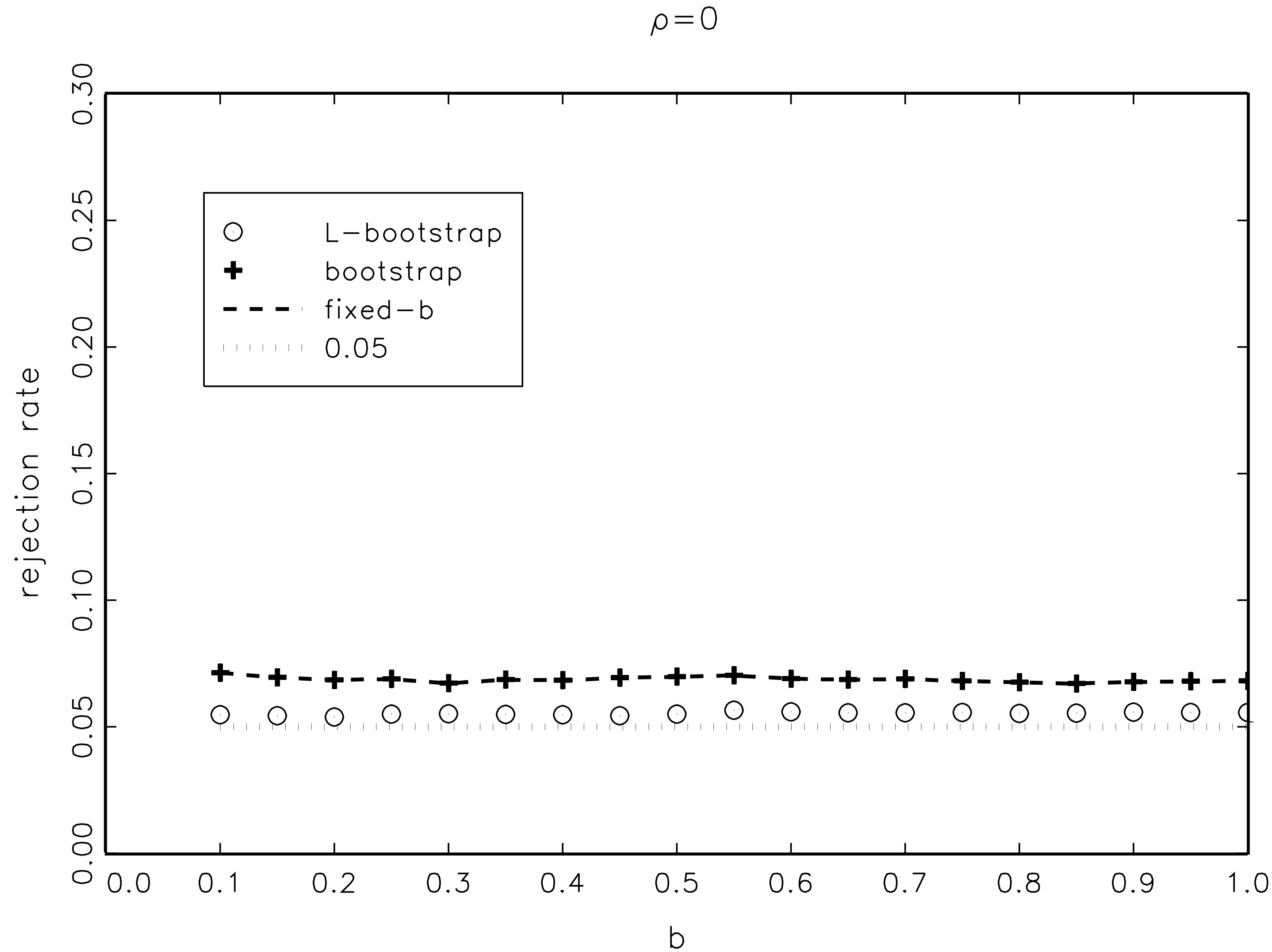


Figure 2.26: (cont'd)

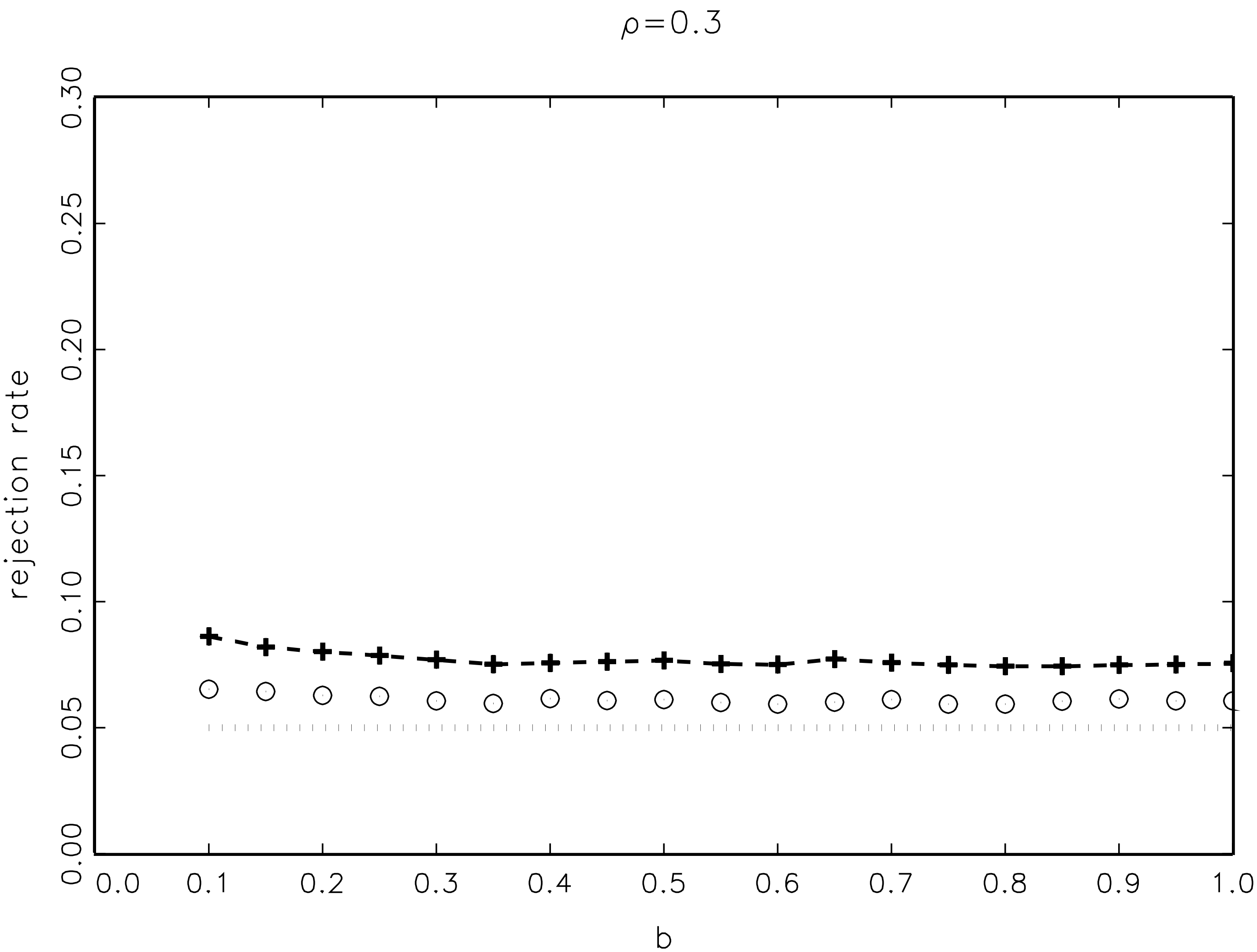


Figure 2.26: (cont'd)

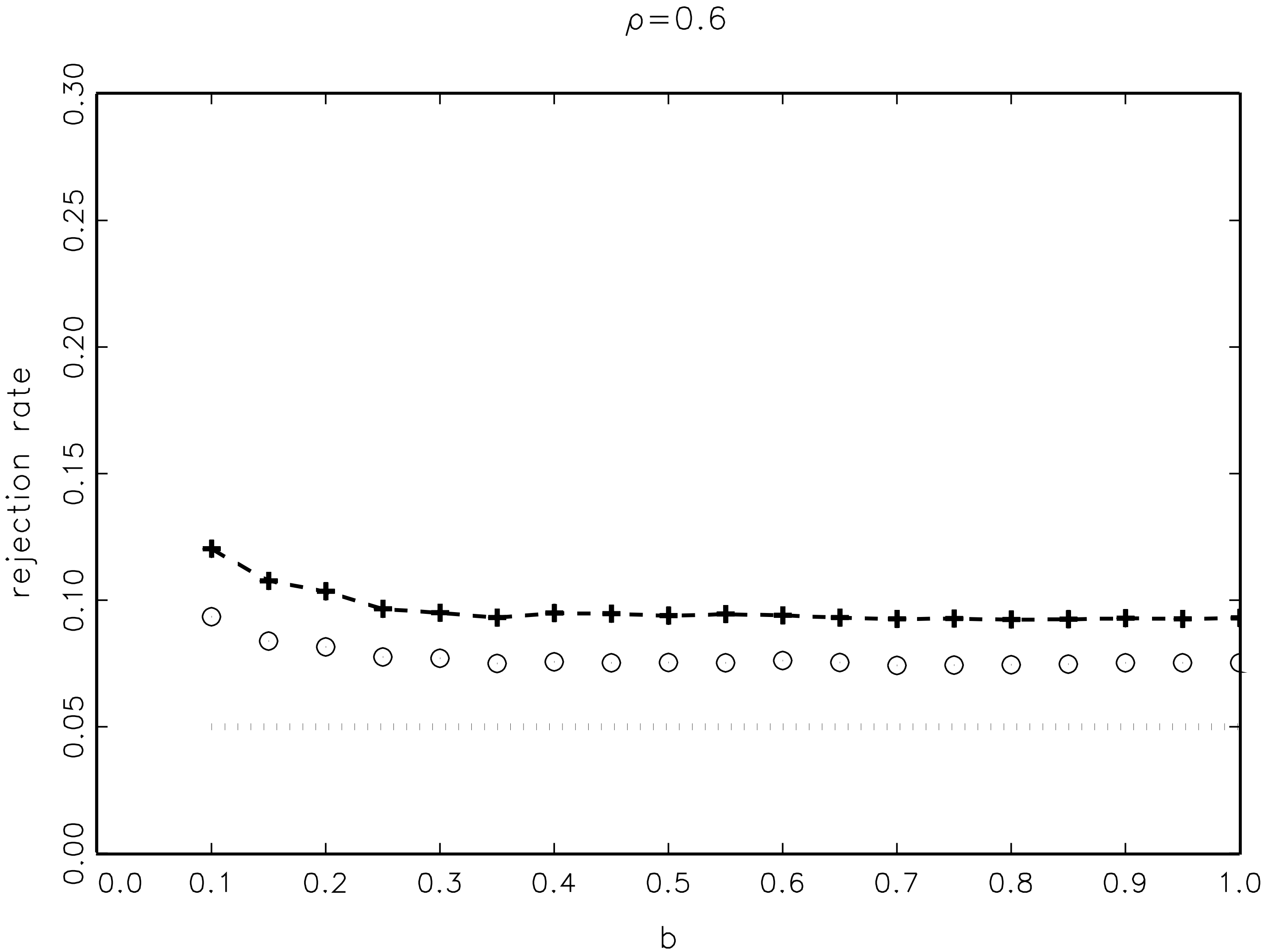


Figure 2.26: (cont'd)

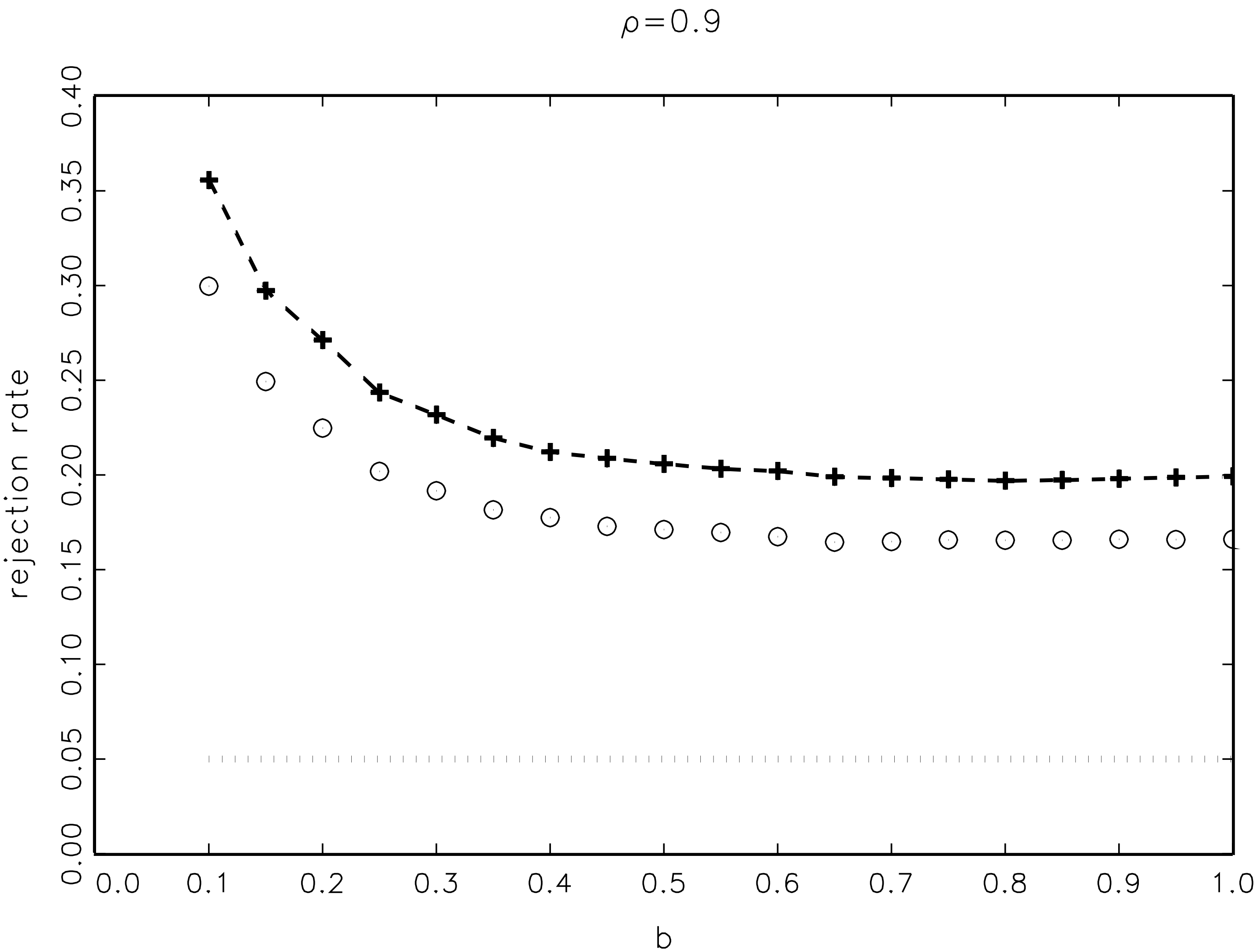


Figure 2.27: AM Series - Random Bernoulli ($p = 0.5$), Bartlett, $T = 50$

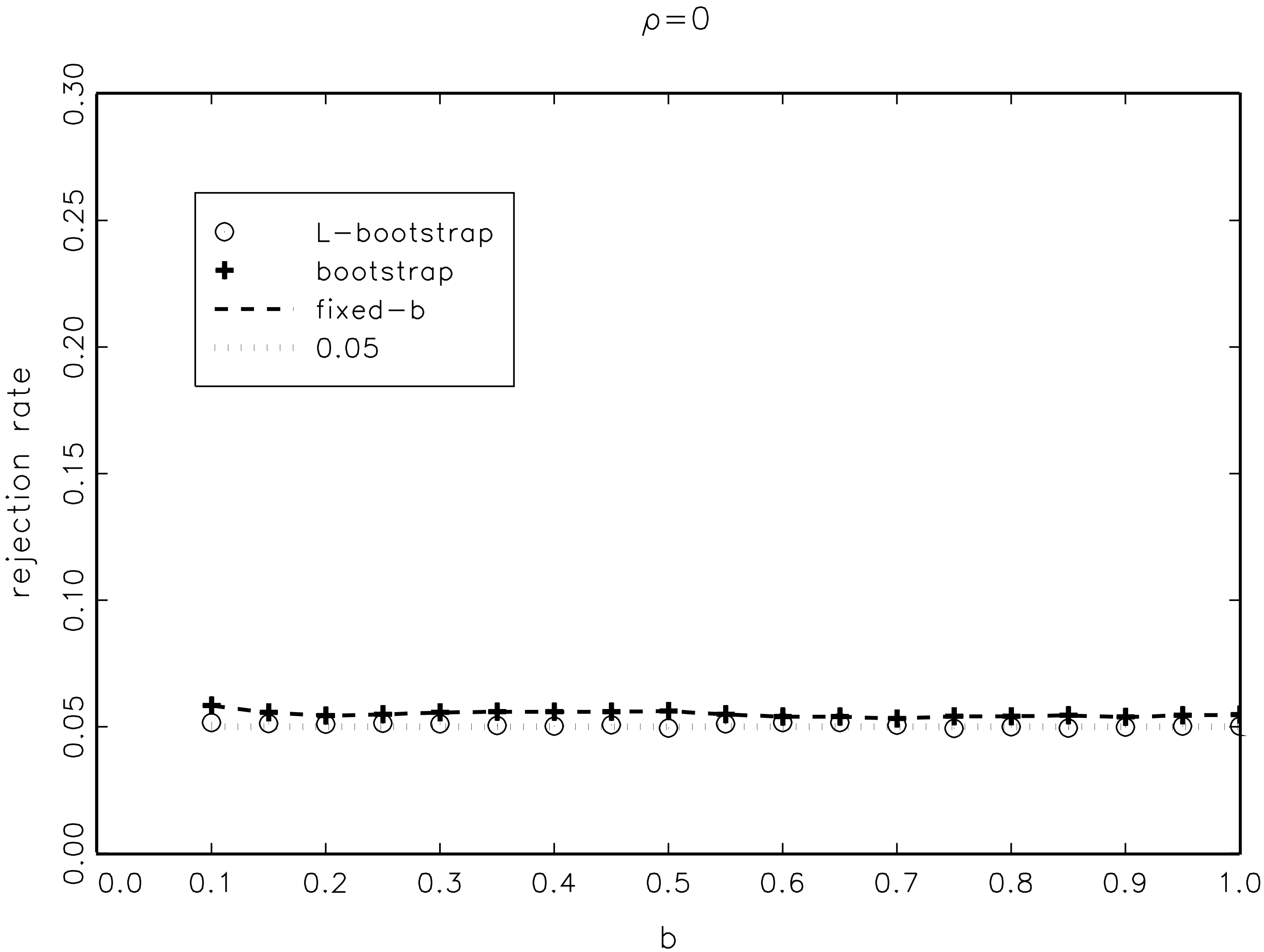


Figure 2.27: (cont'd)

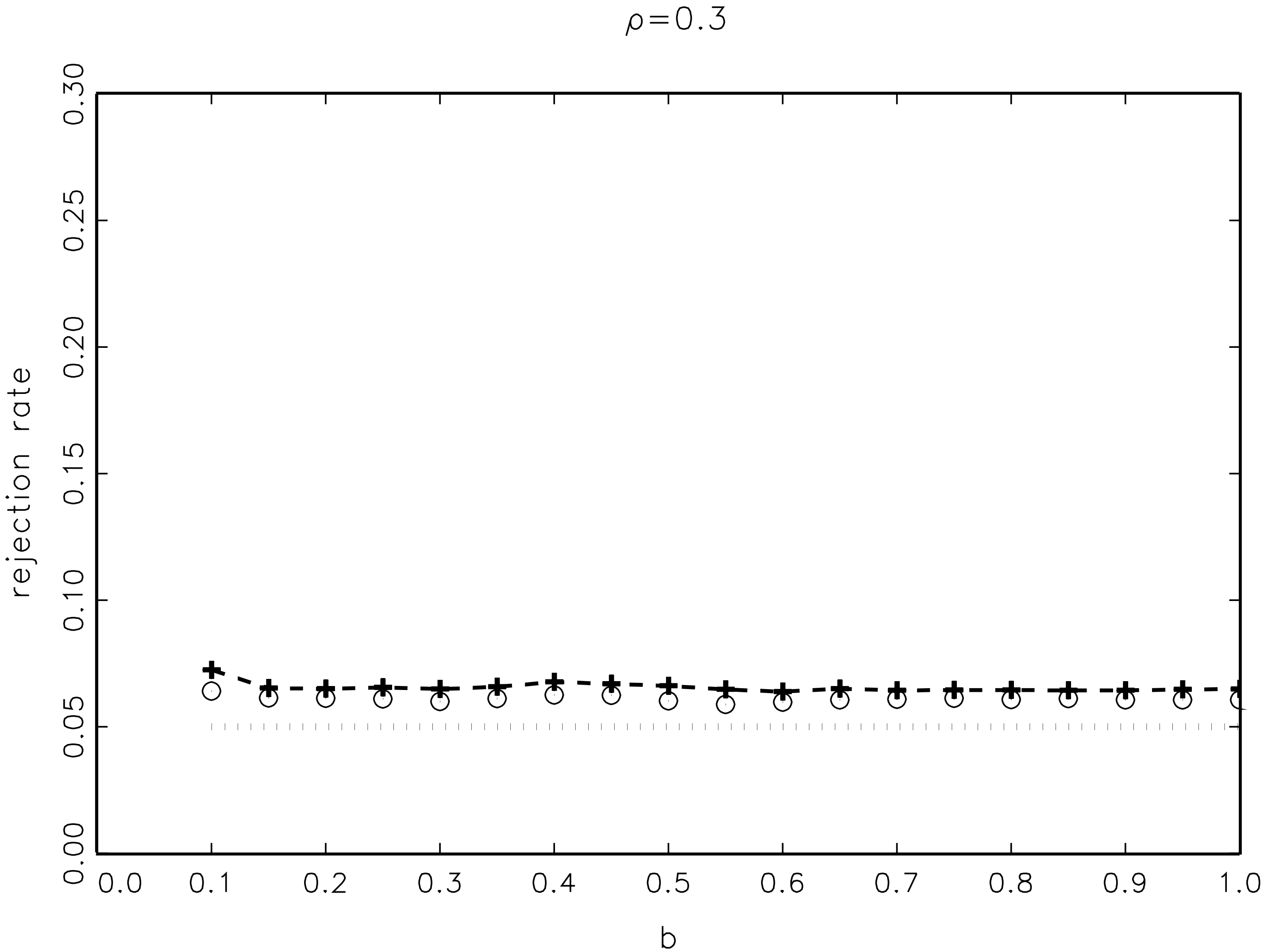


Figure 2.27: (cont'd)

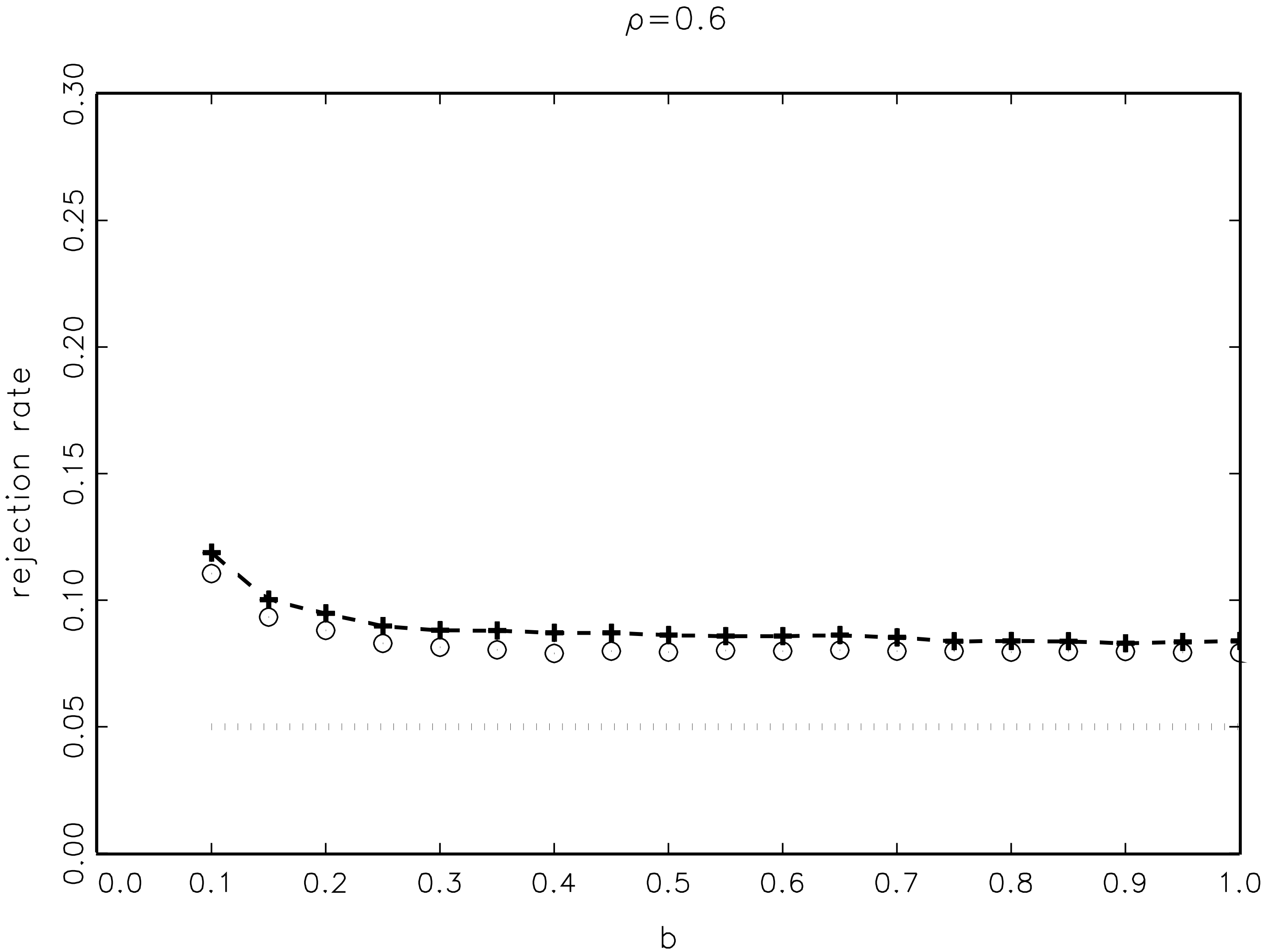


Figure 2.27: (cont'd)

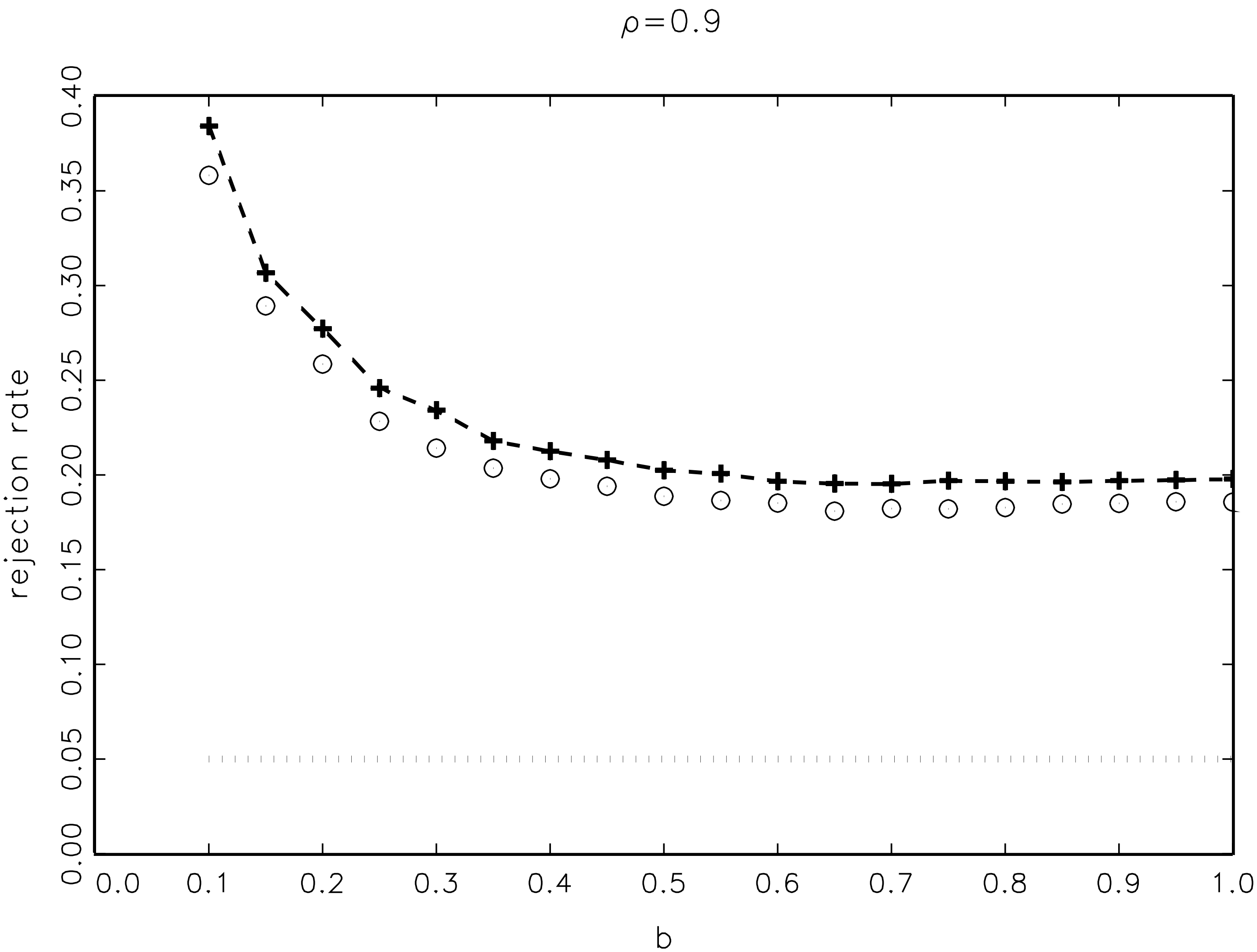


Figure 2.28: AM Series - Random Bernoulli ($p = 0.7$), Bartlett, $T = 50$

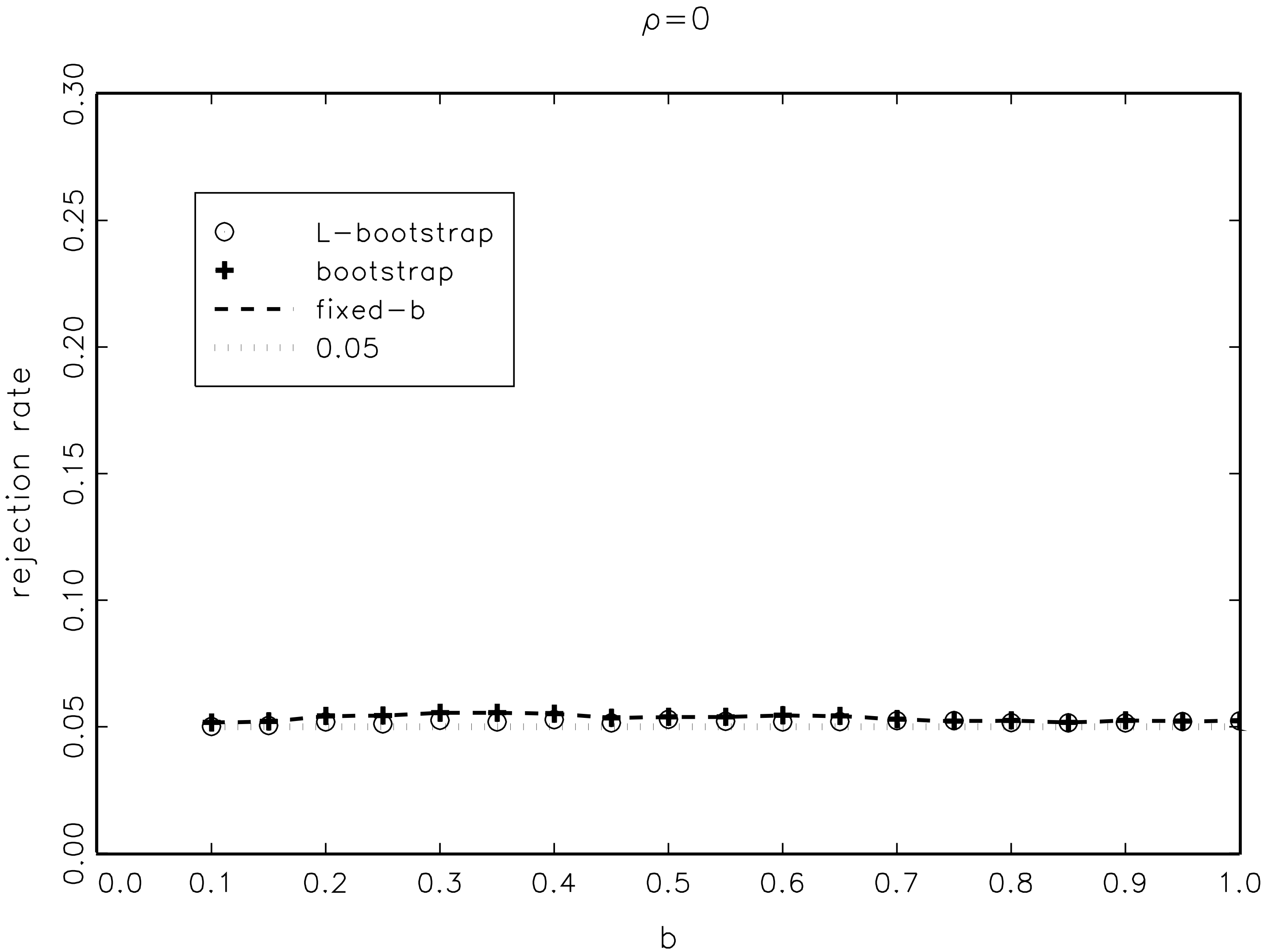


Figure 2.28: (cont'd)

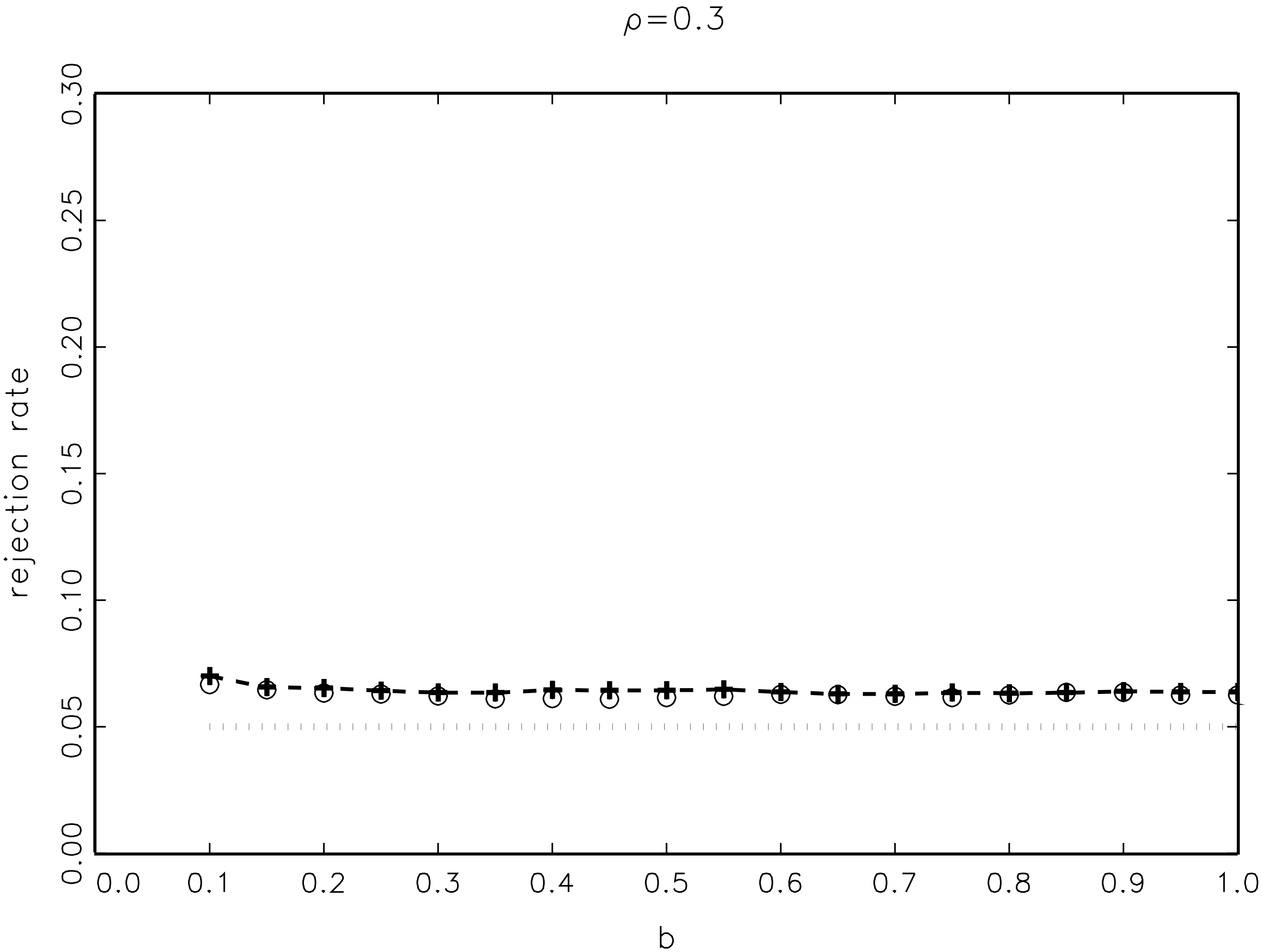


Figure 2.28: (cont'd)

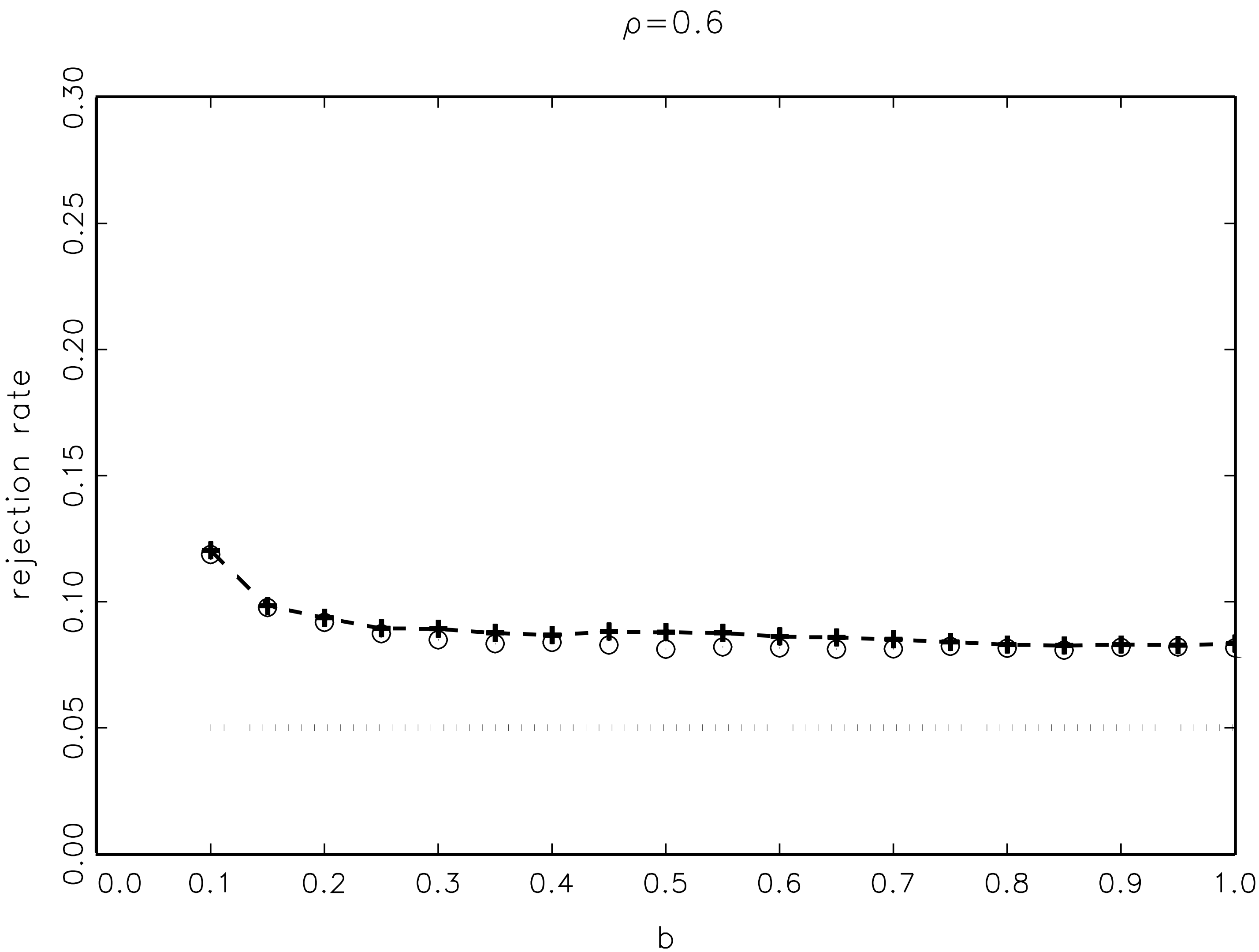


Figure 2.28: (cont'd)

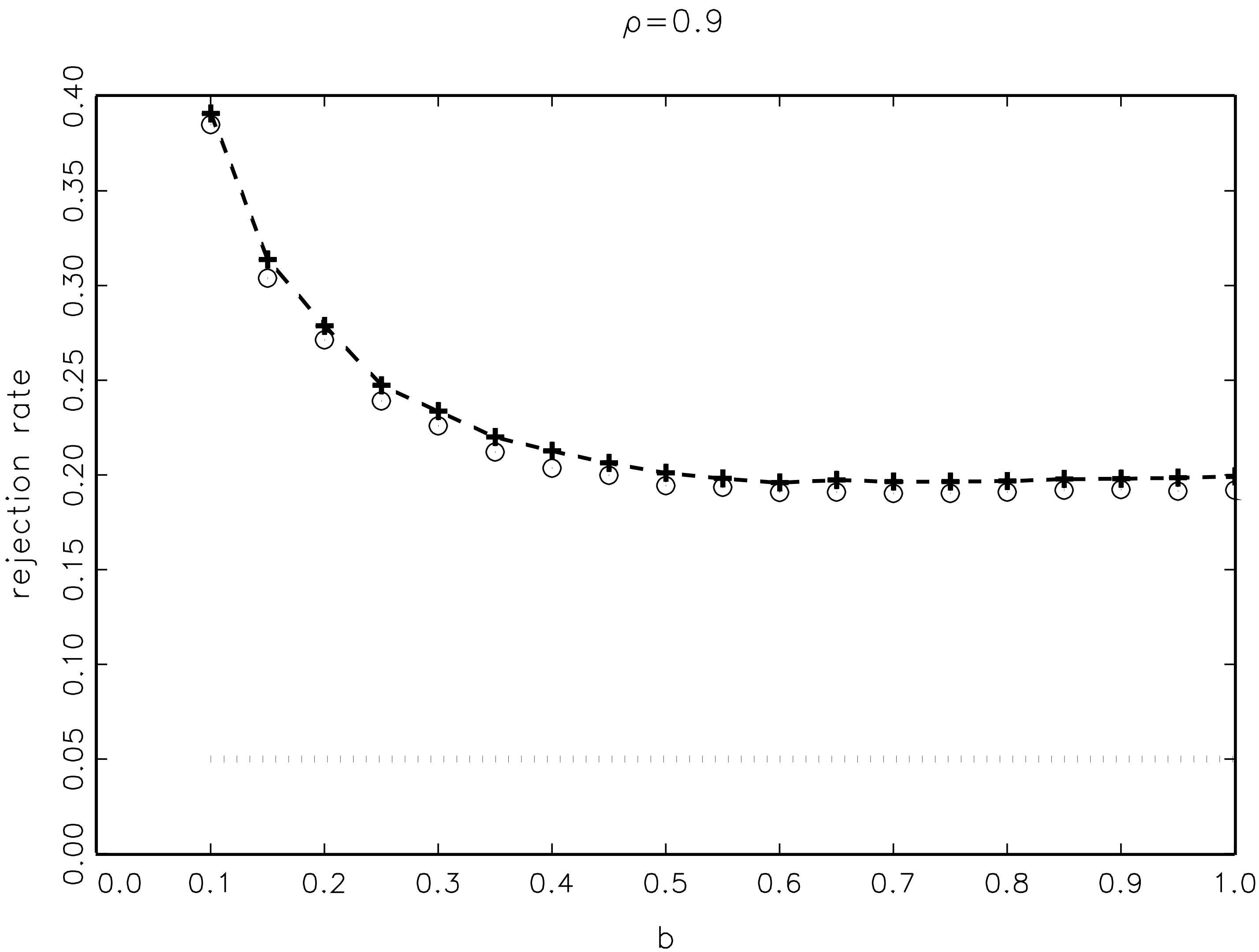


Figure 2.29: AM Series - Random Bernoulli ($p = 0.3$), Bartlett, $T = 100$

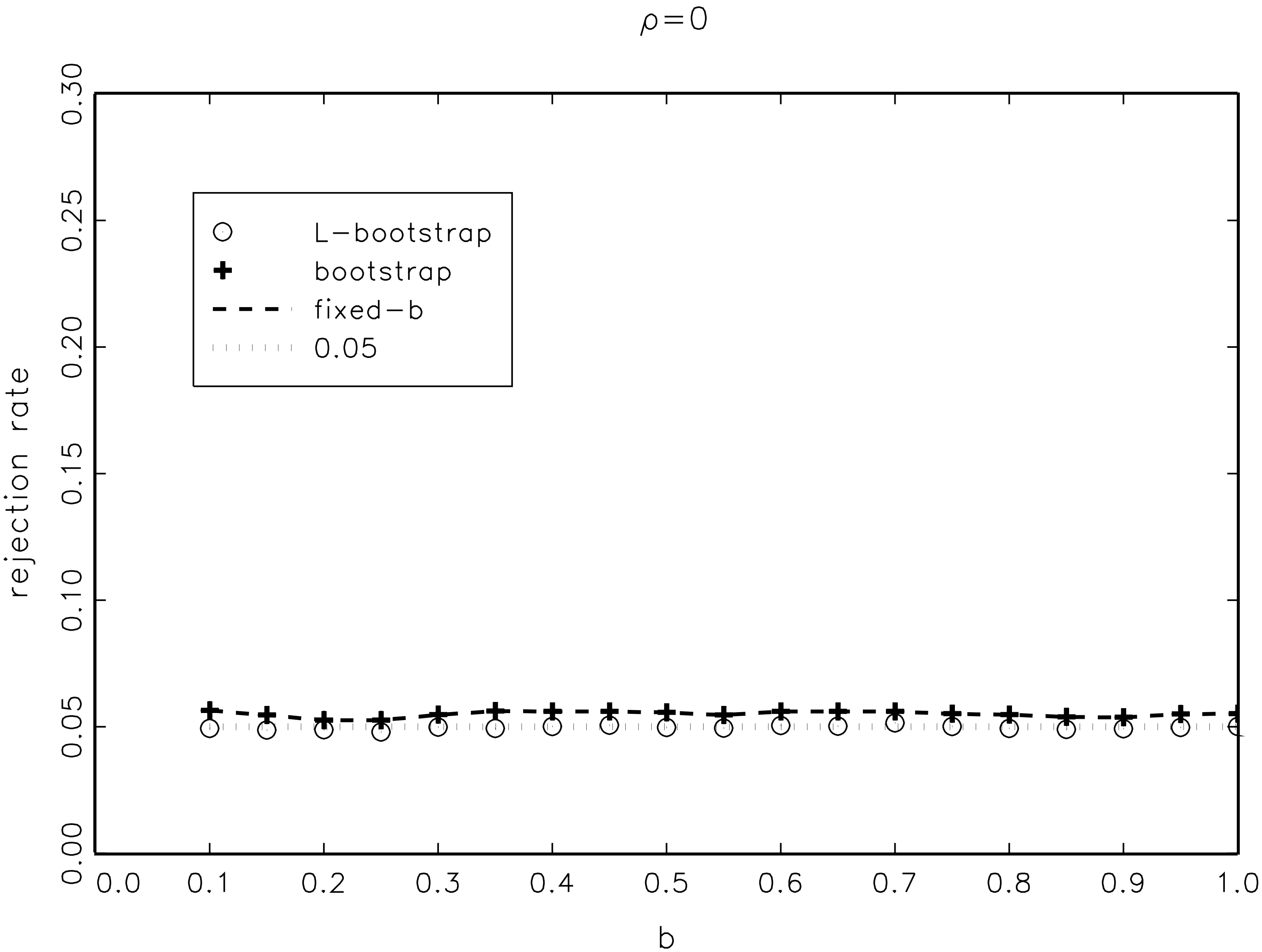


Figure 2.29: (cont'd)

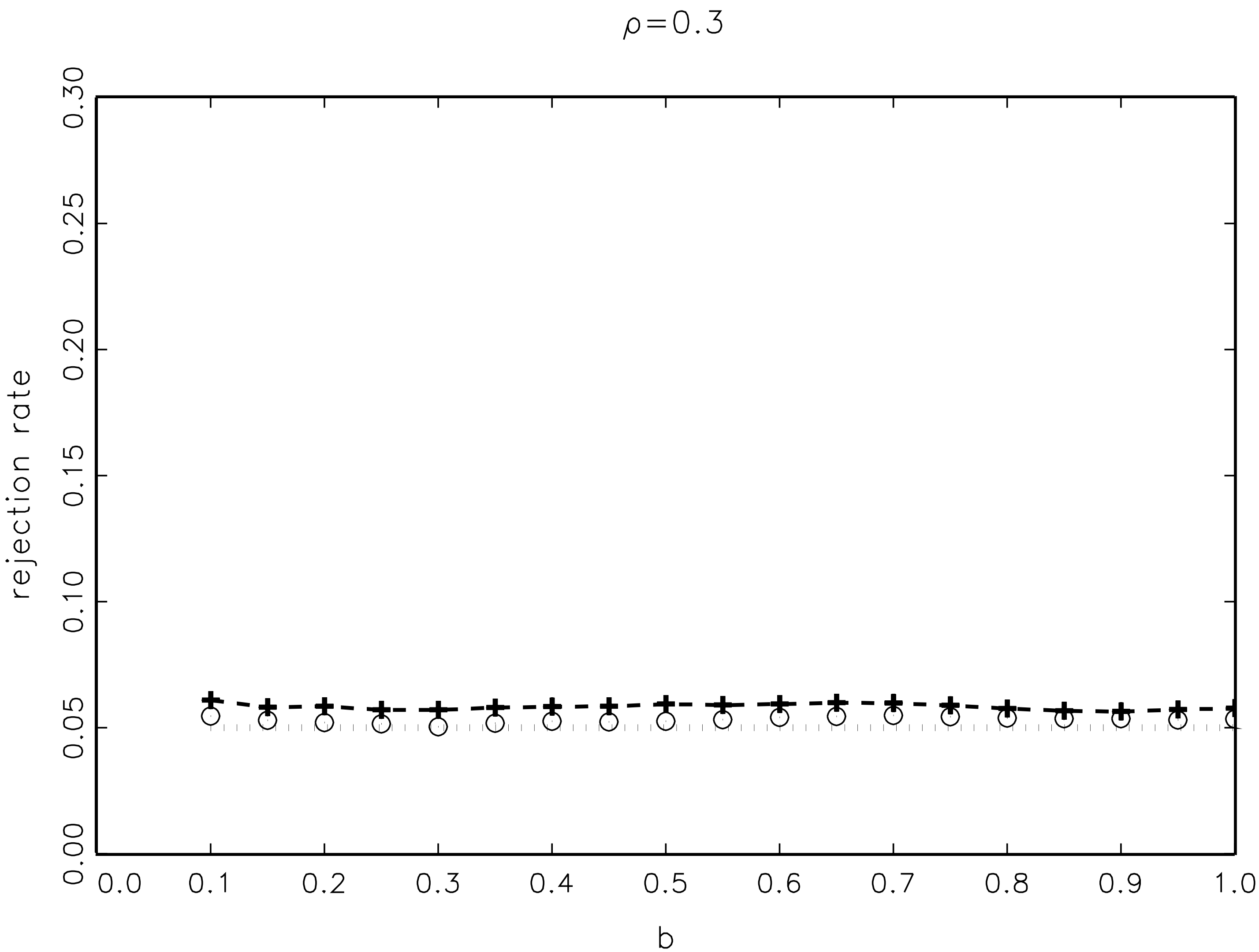


Figure 2.29: (cont'd)

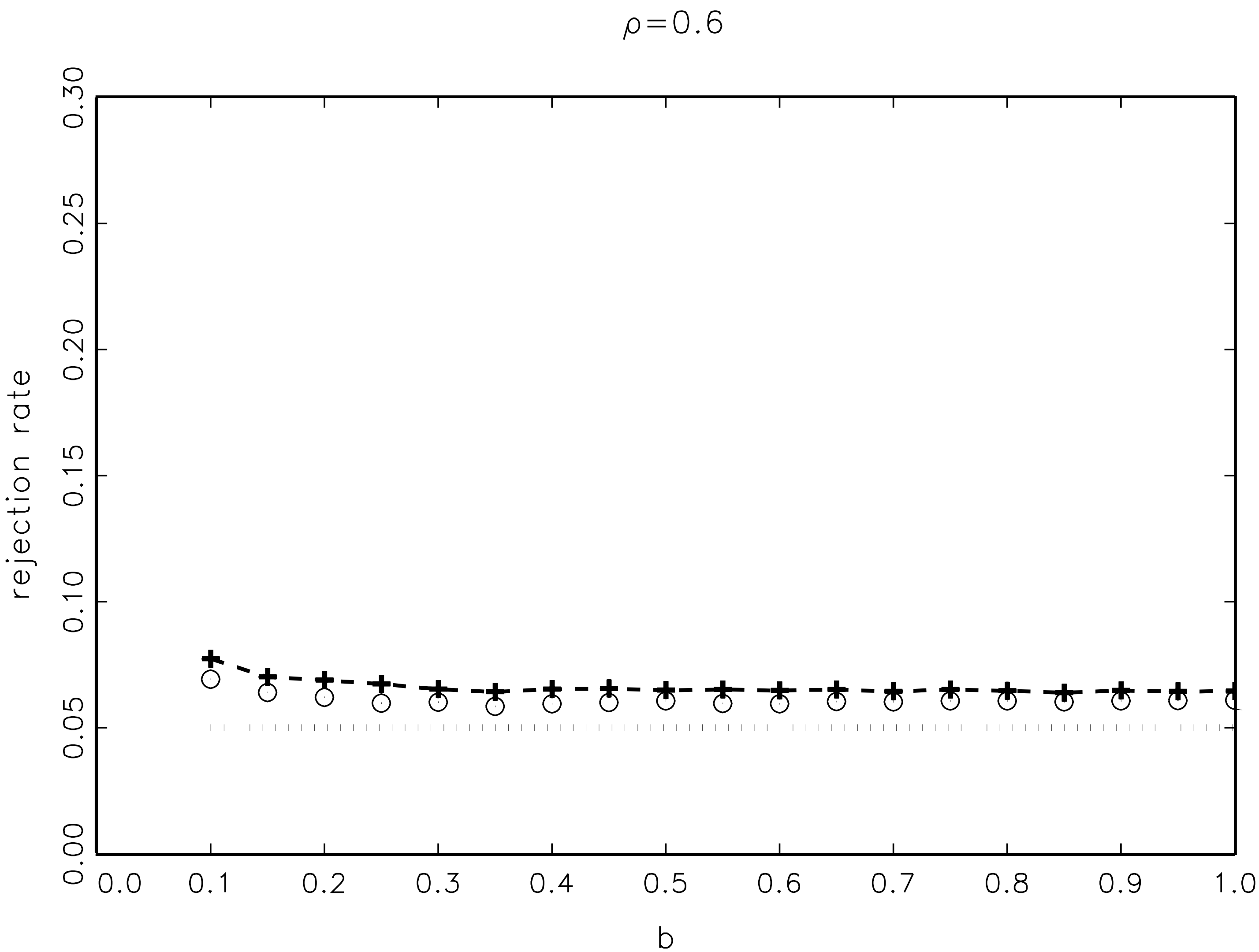


Figure 2.29: (cont'd)

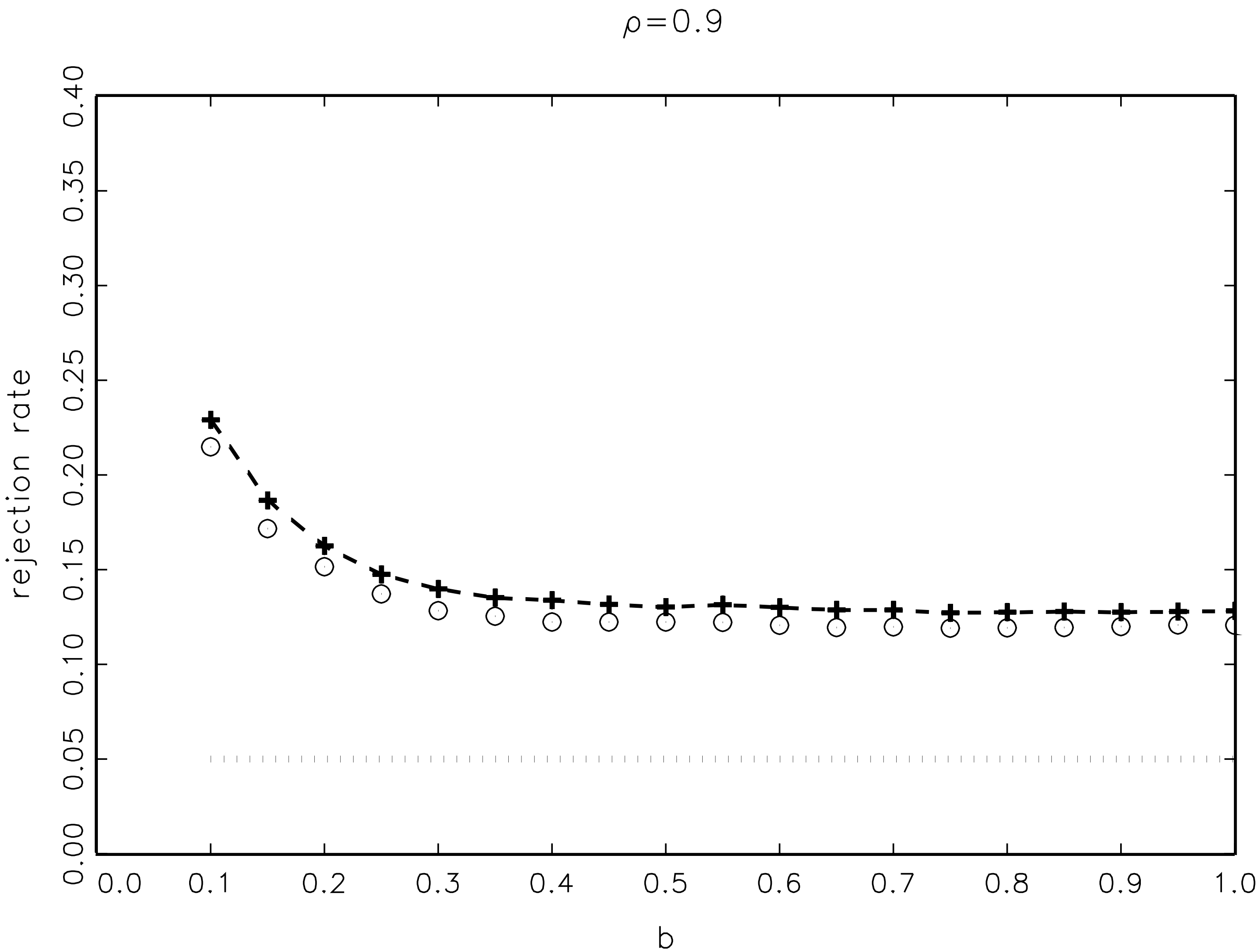


Figure 2.30: AM Series - Random Bernoulli ($p = 0.5$), Bartlett, $T = 100$

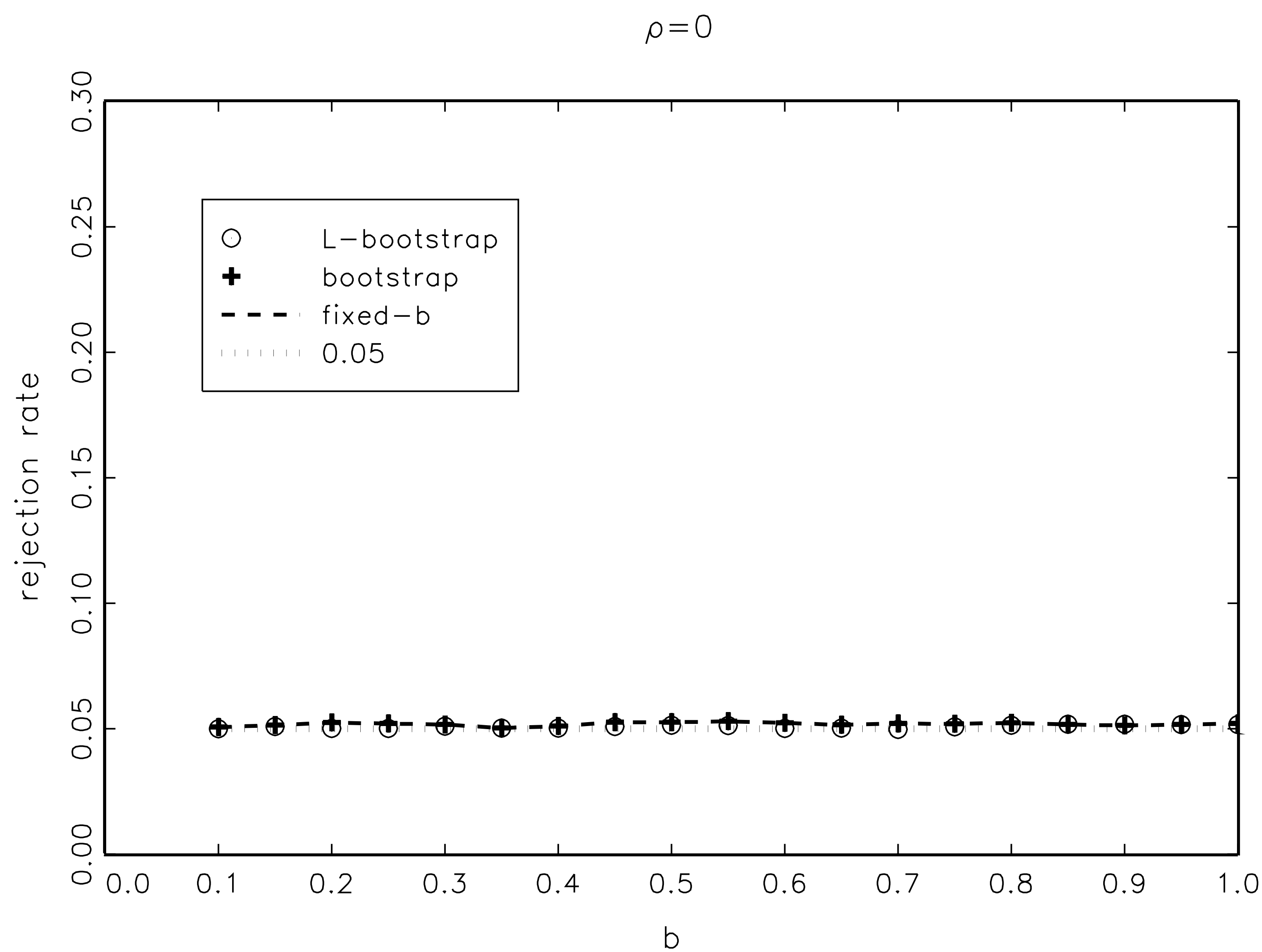


Figure 2.30: (cont'd)

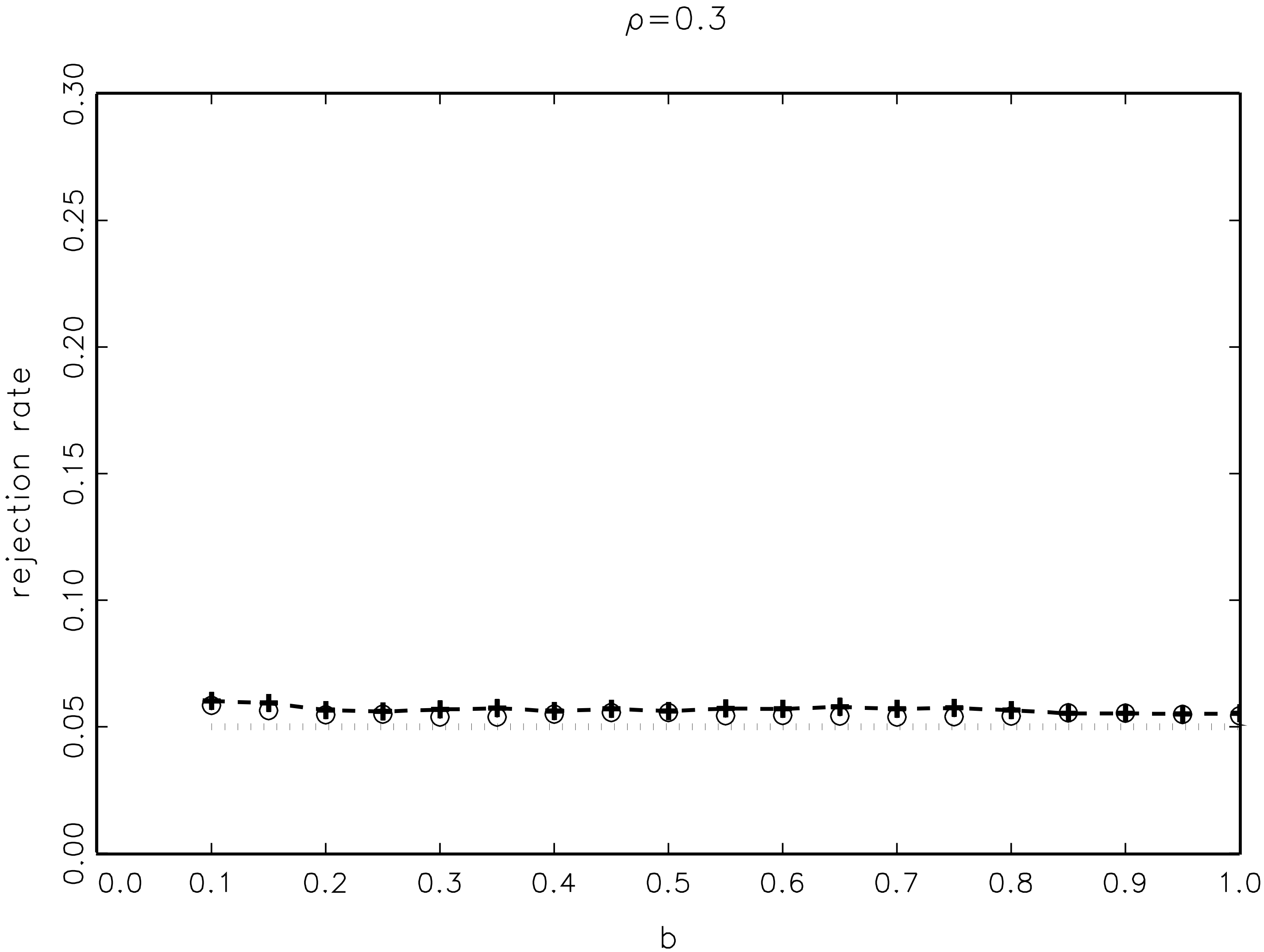


Figure 2.30: (cont'd)

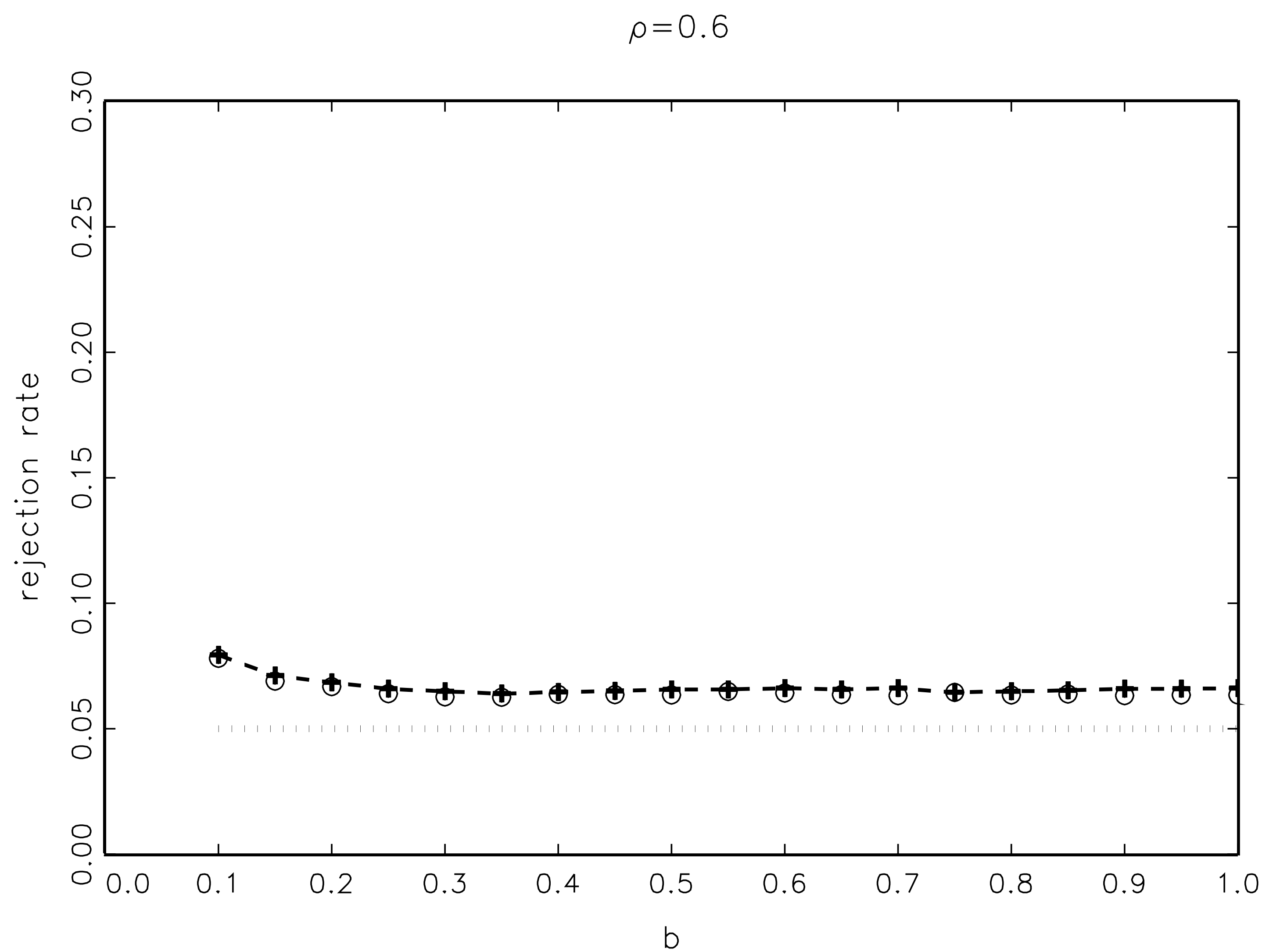


Figure 2.30: (cont'd)

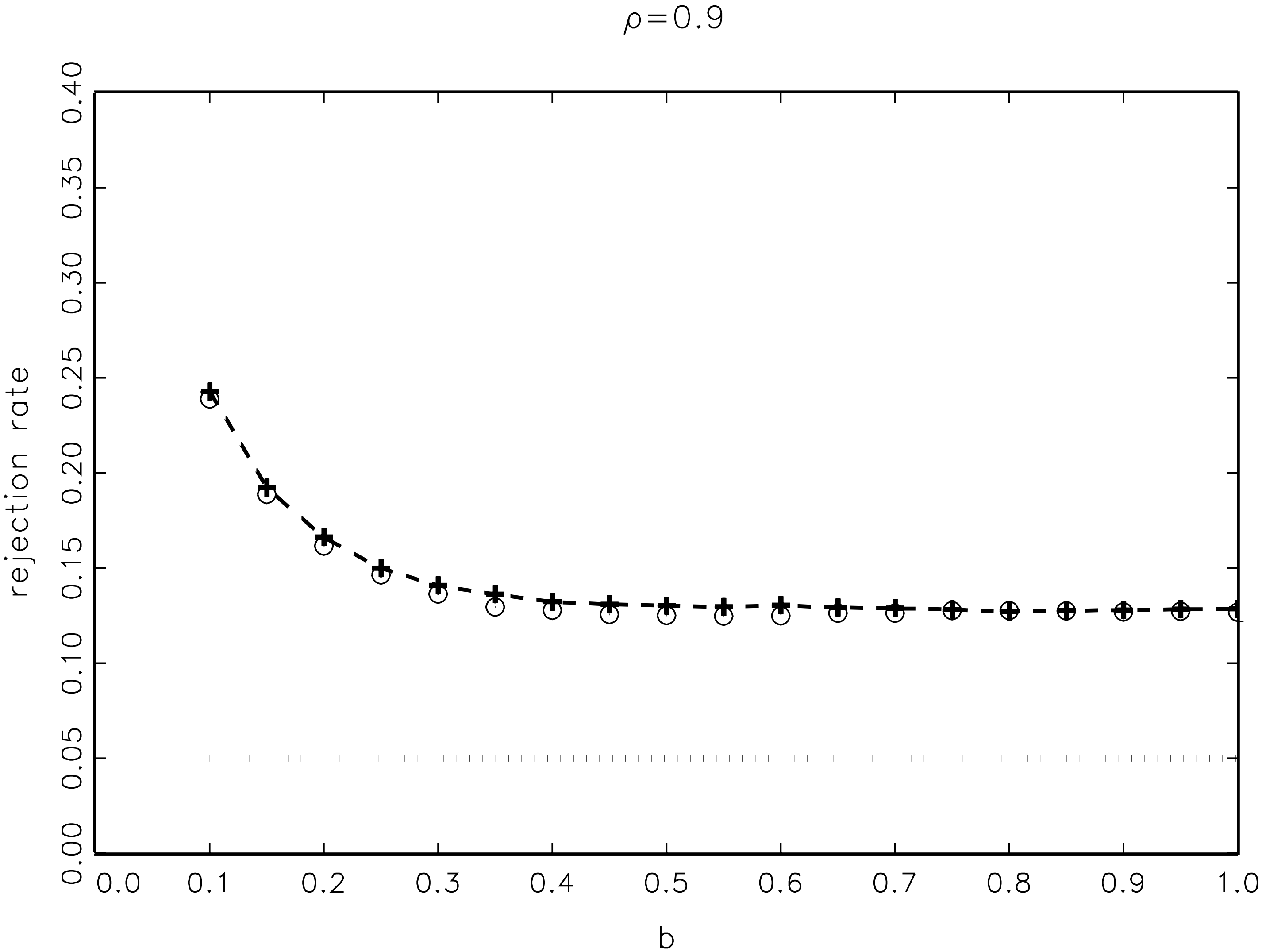


Figure 2.31: AM Series - Random Bernoulli ($p = 0.7$), Bartlett, $T = 100$

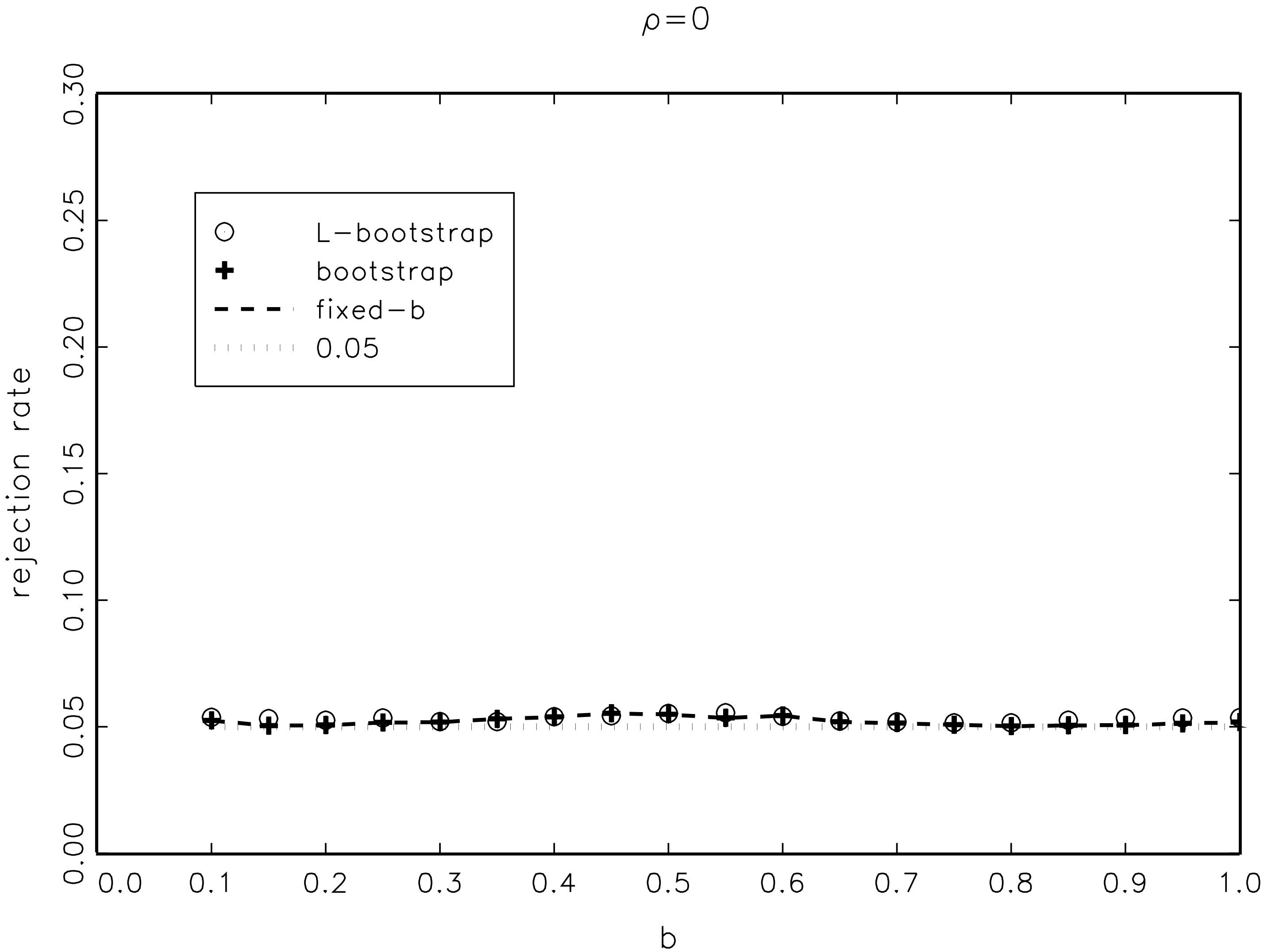


Figure 2.31: (cont'd)

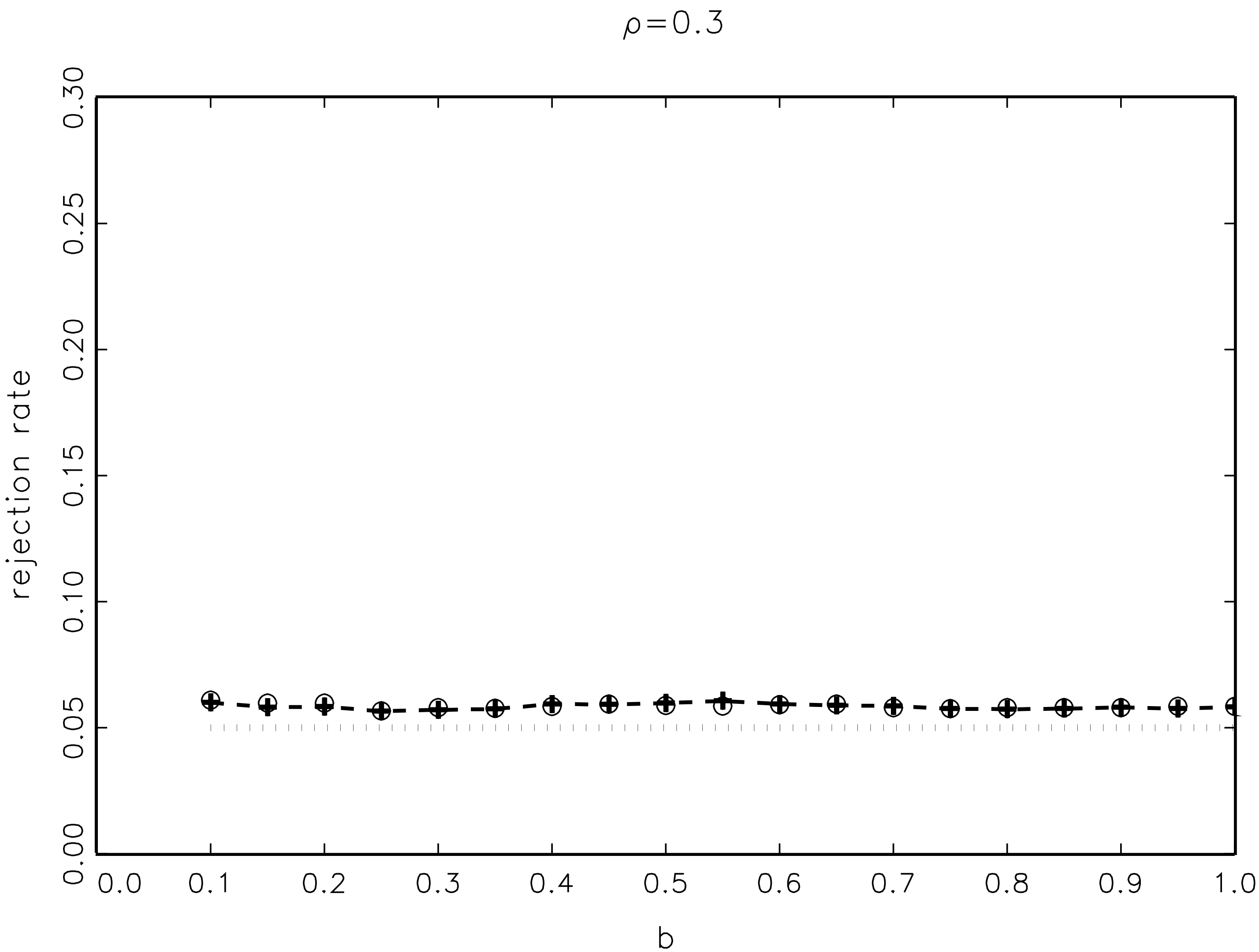


Figure 2.31: (cont'd)

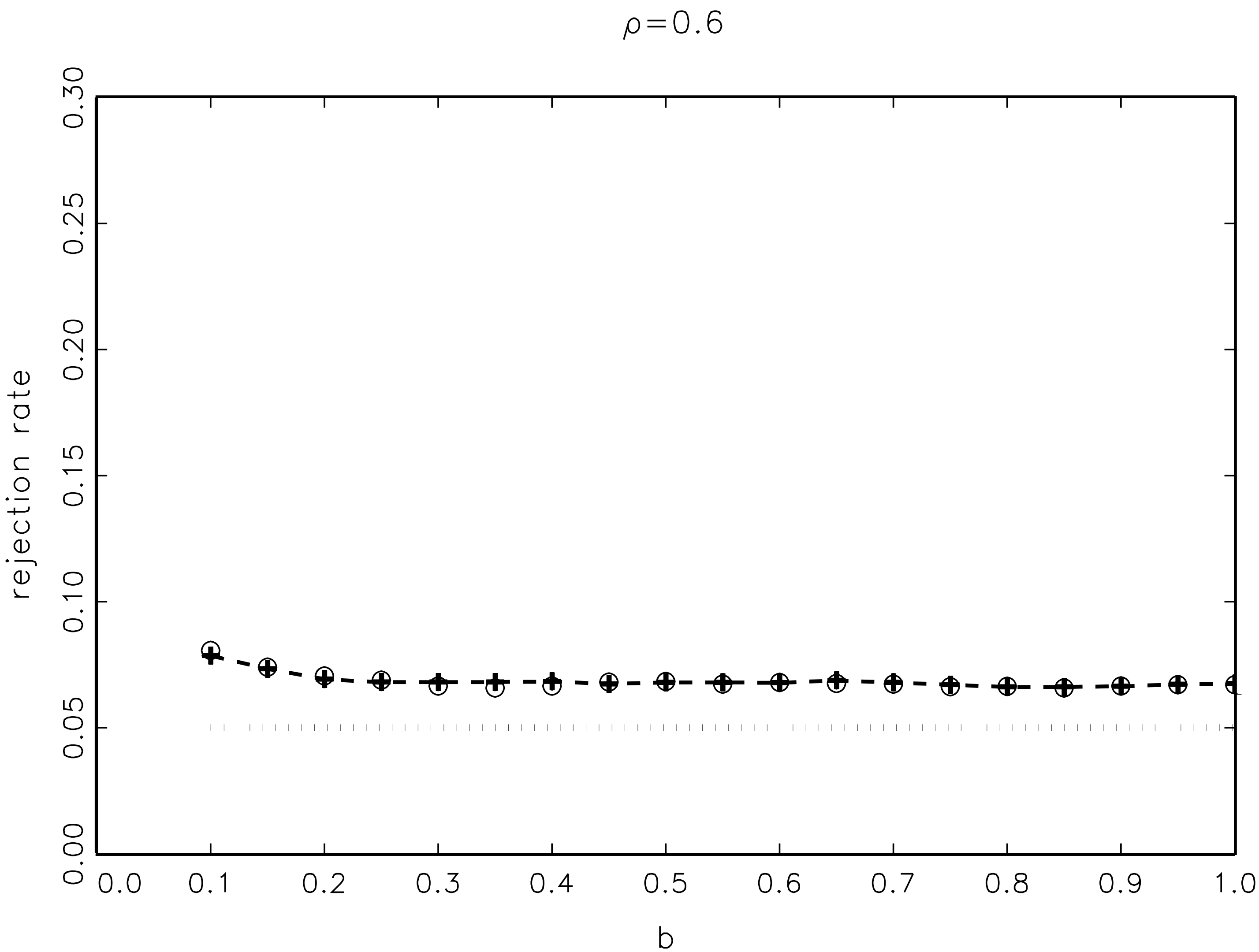


Figure 2.31: (cont'd)

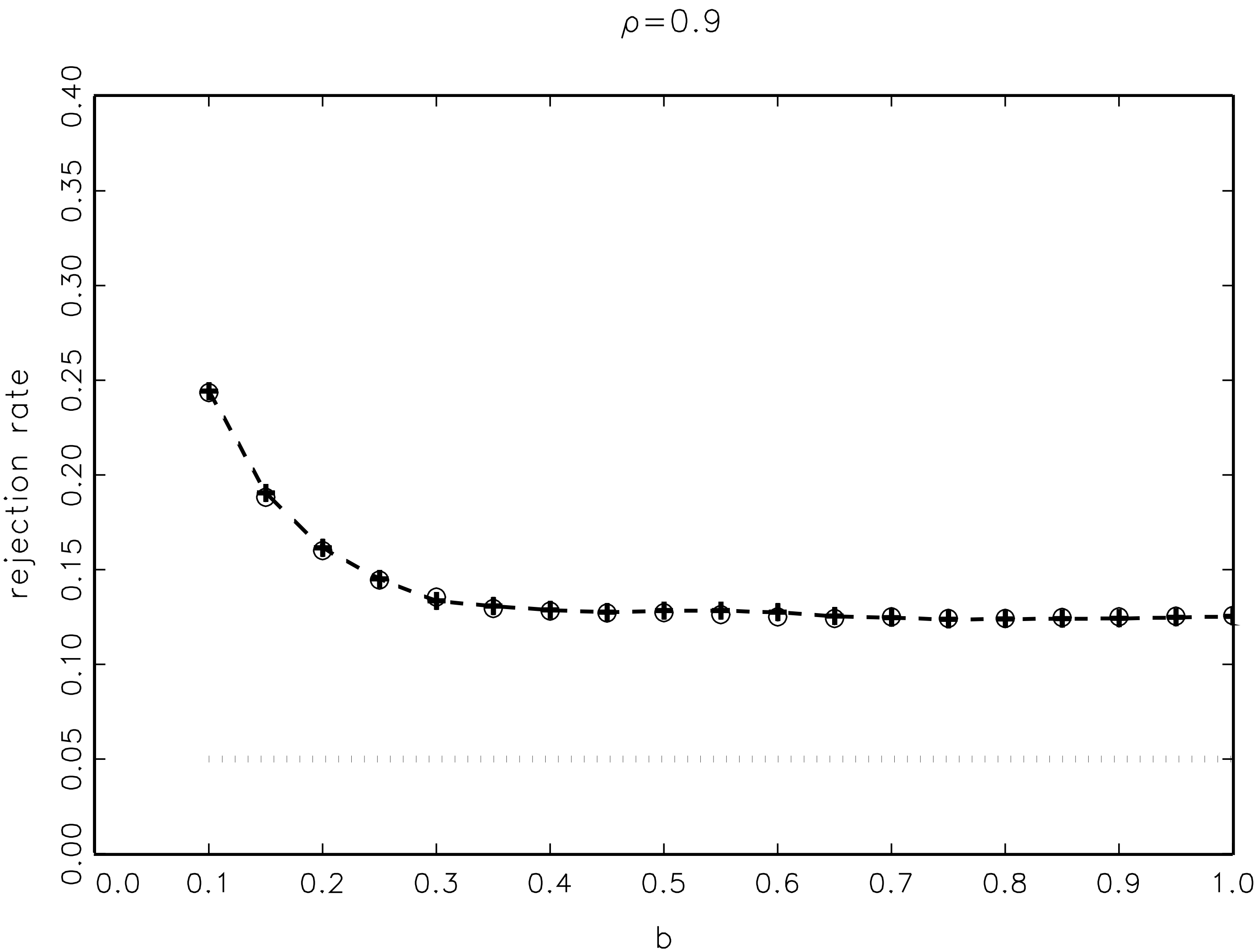


Figure 2.32: AM Series - Random Bernoulli ($p = 0.3$), Bartlett, $T = 200$

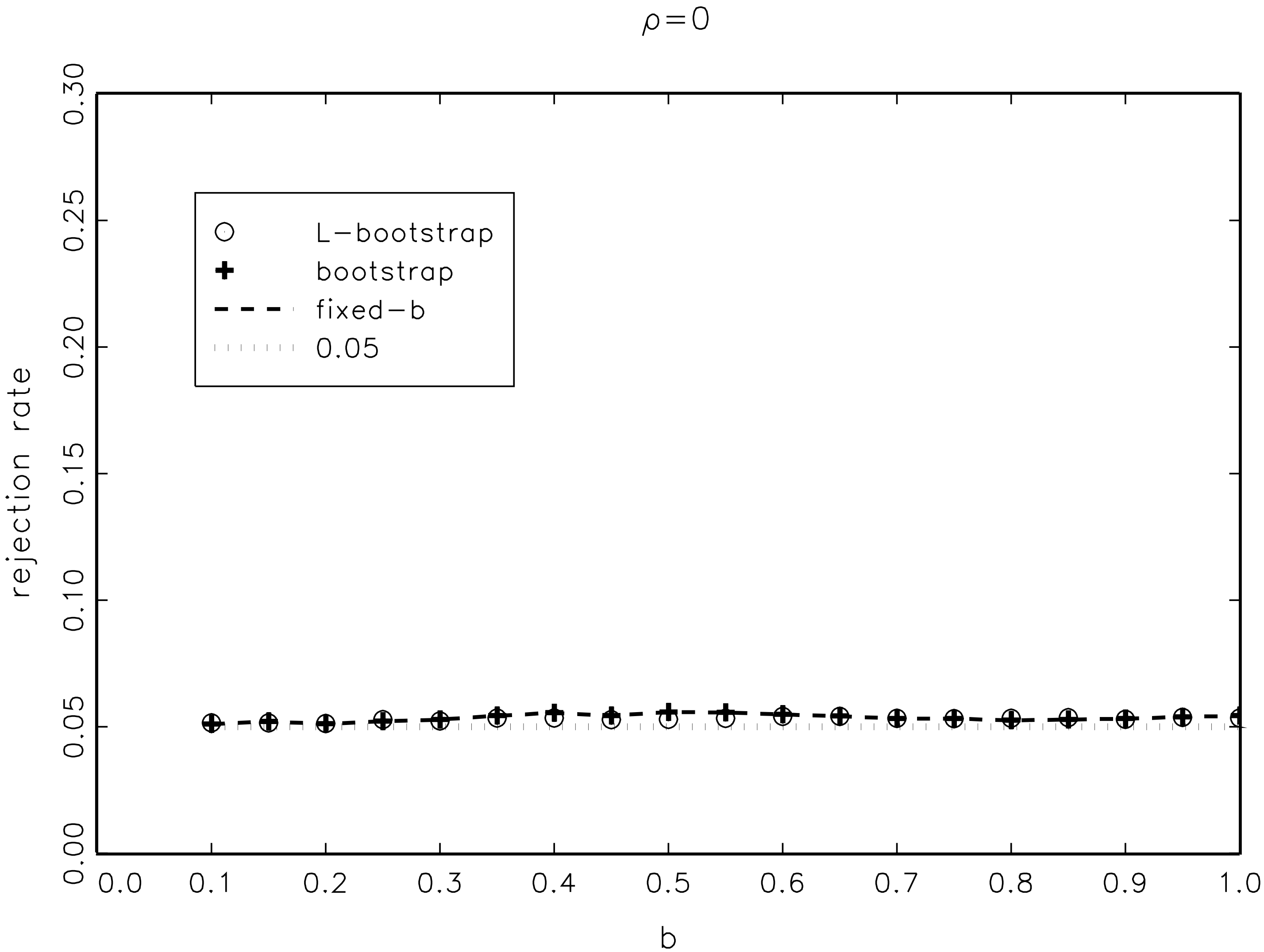


Figure 2.32: (cont'd)

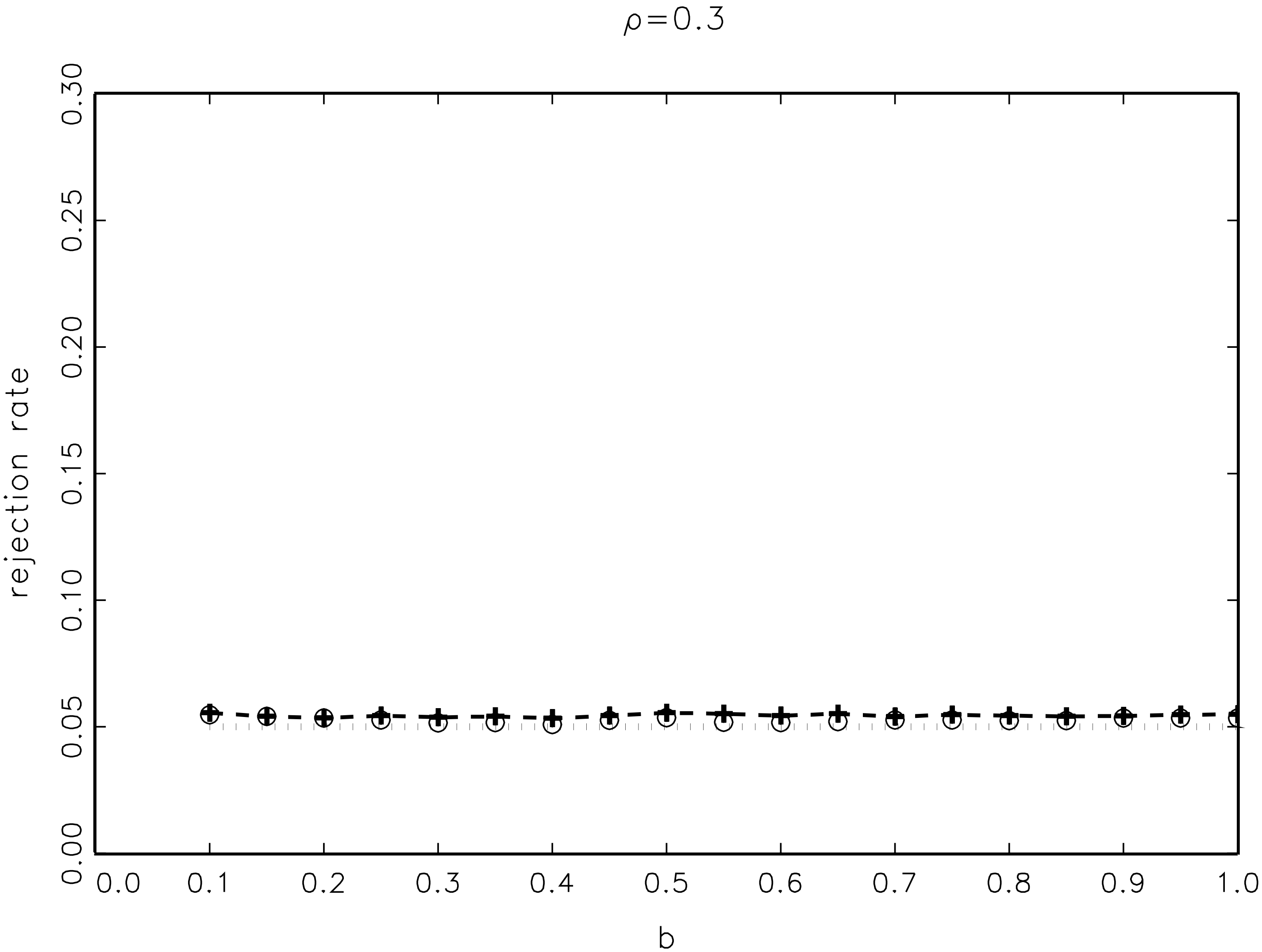


Figure 2.32: (cont'd)

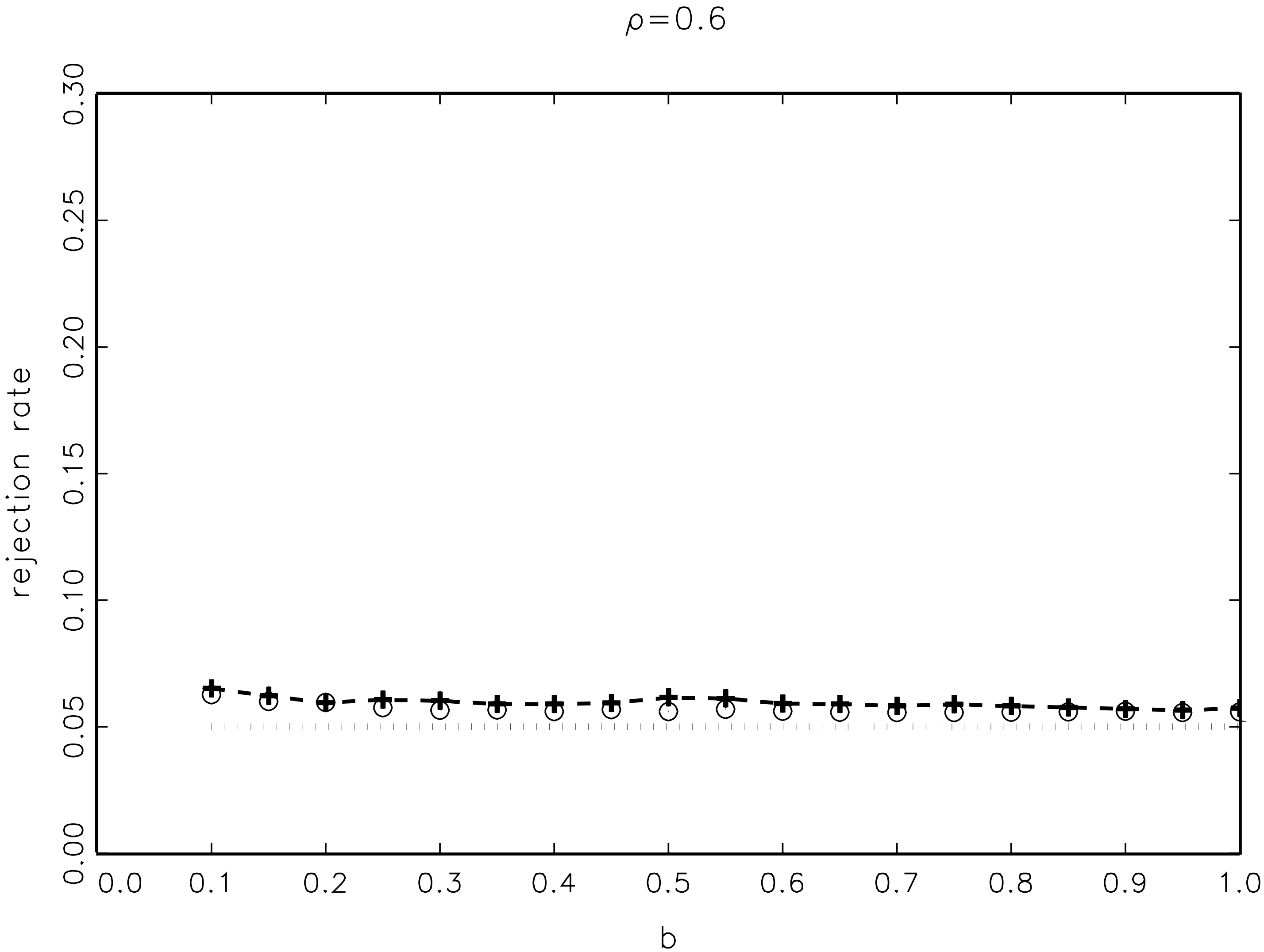


Figure 2.32: (cont'd)

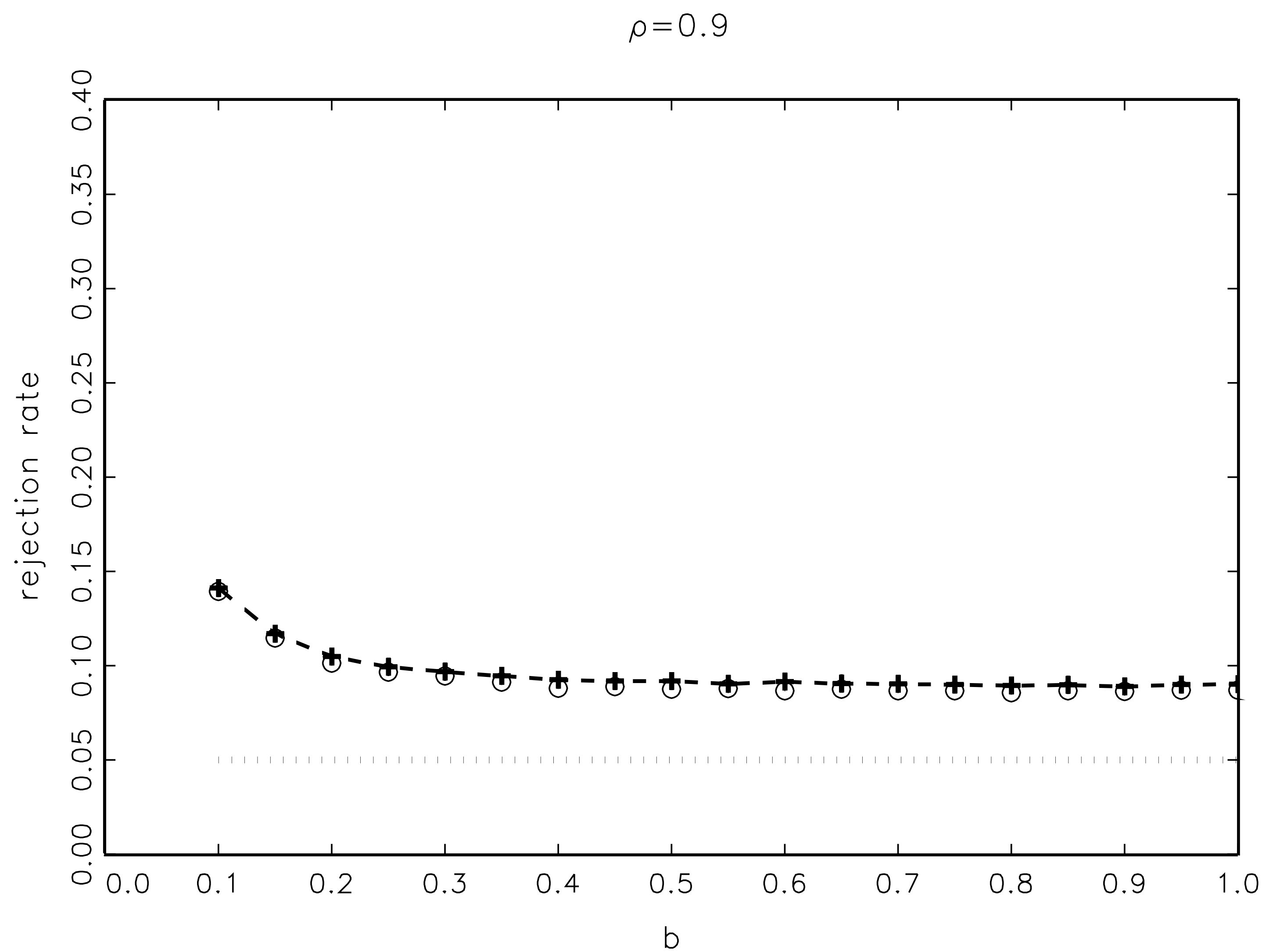


Figure 2.33: AM Series - Random Bernoulli ($p = 0.5$), Bartlett, $T = 200$

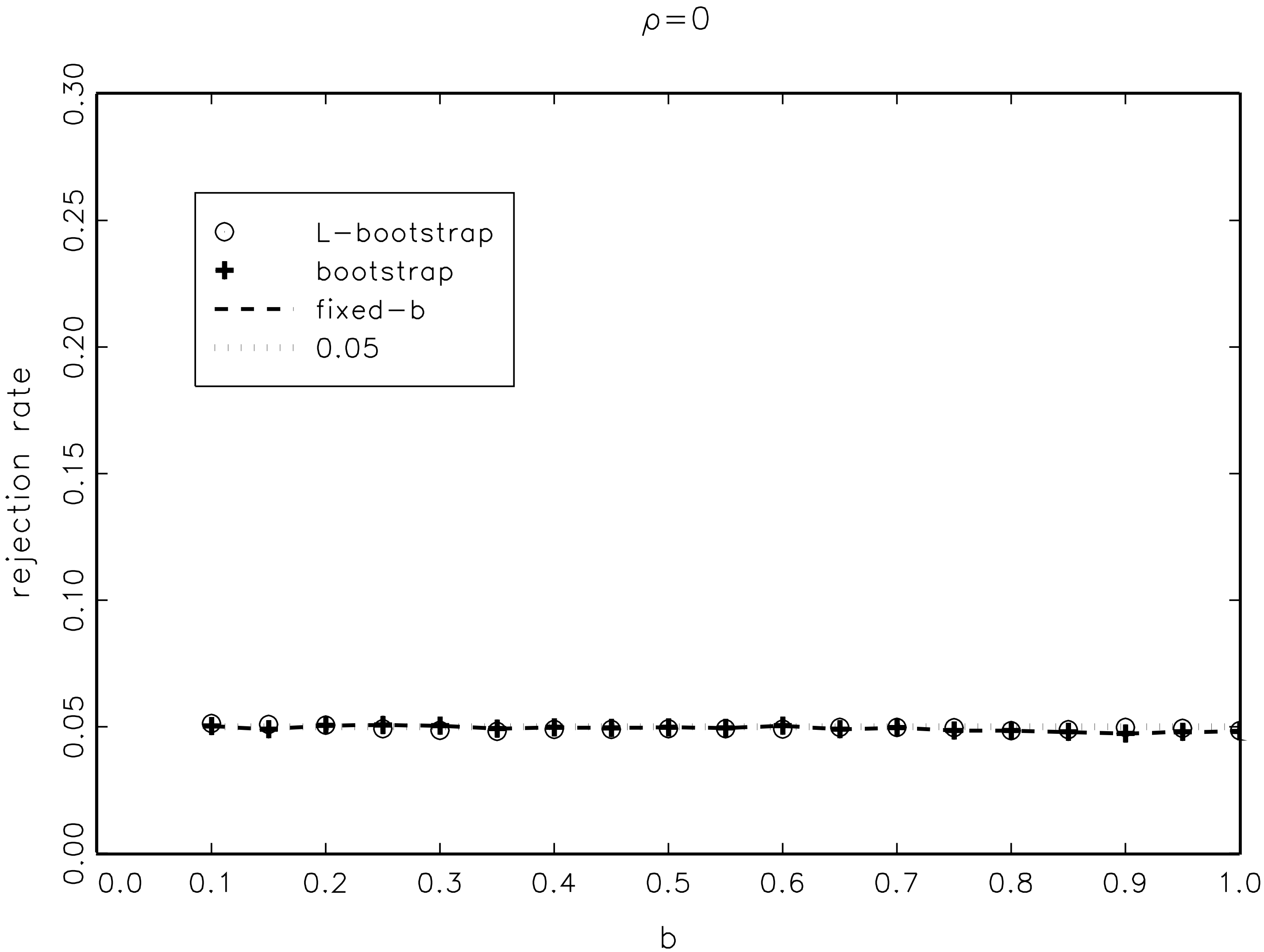


Figure 2.33: (cont'd)

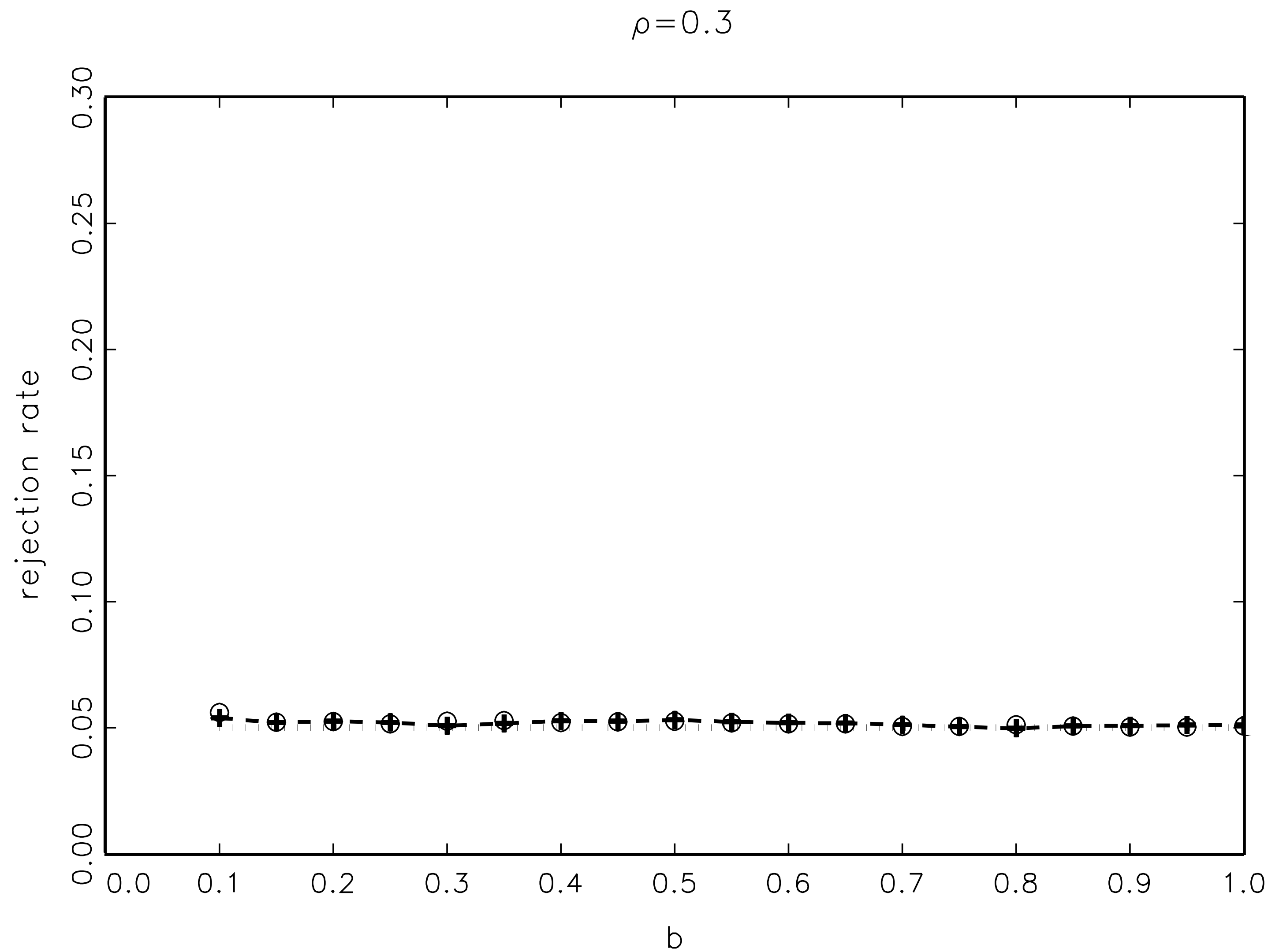


Figure 2.33: (cont'd)

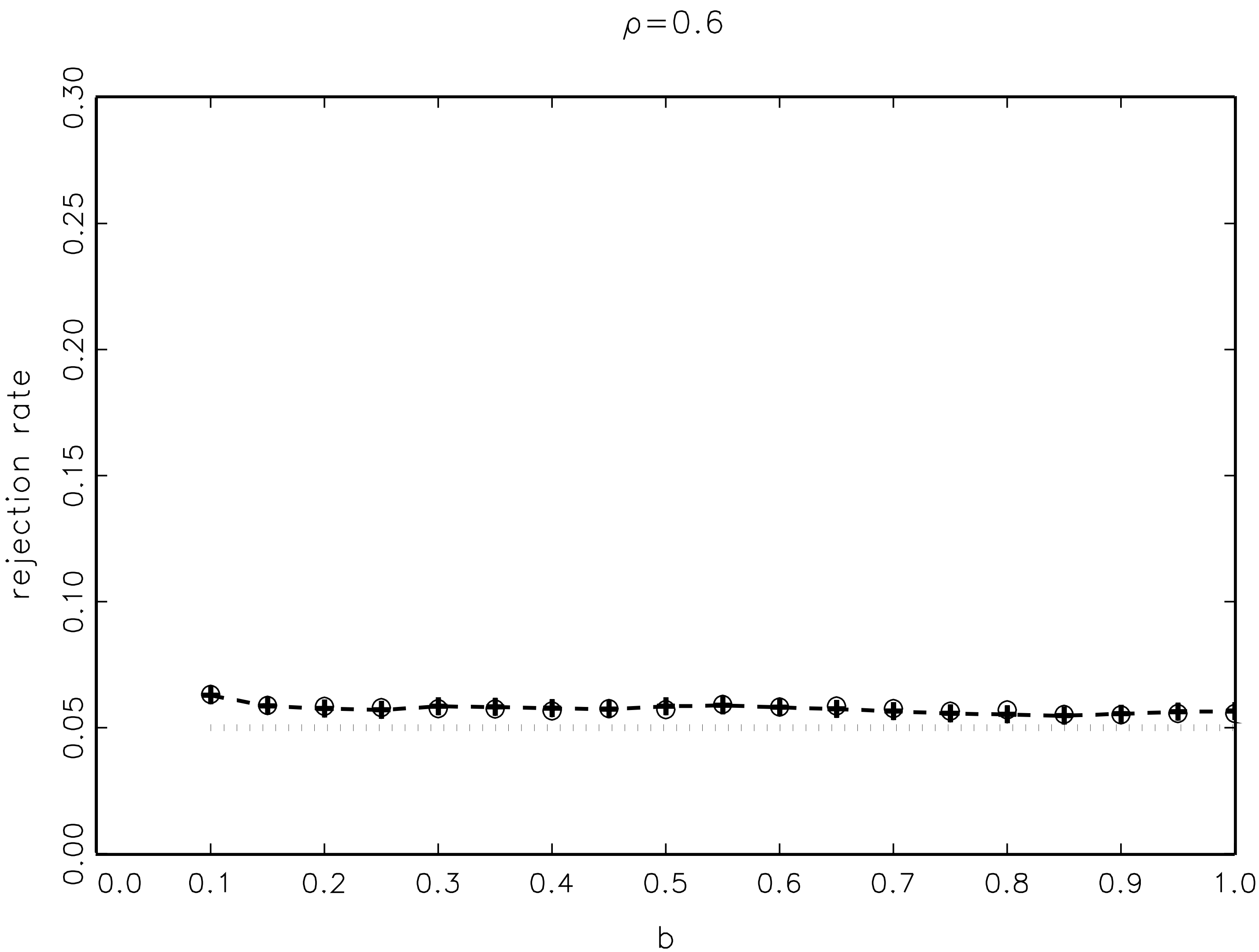


Figure 2.33: (cont'd)

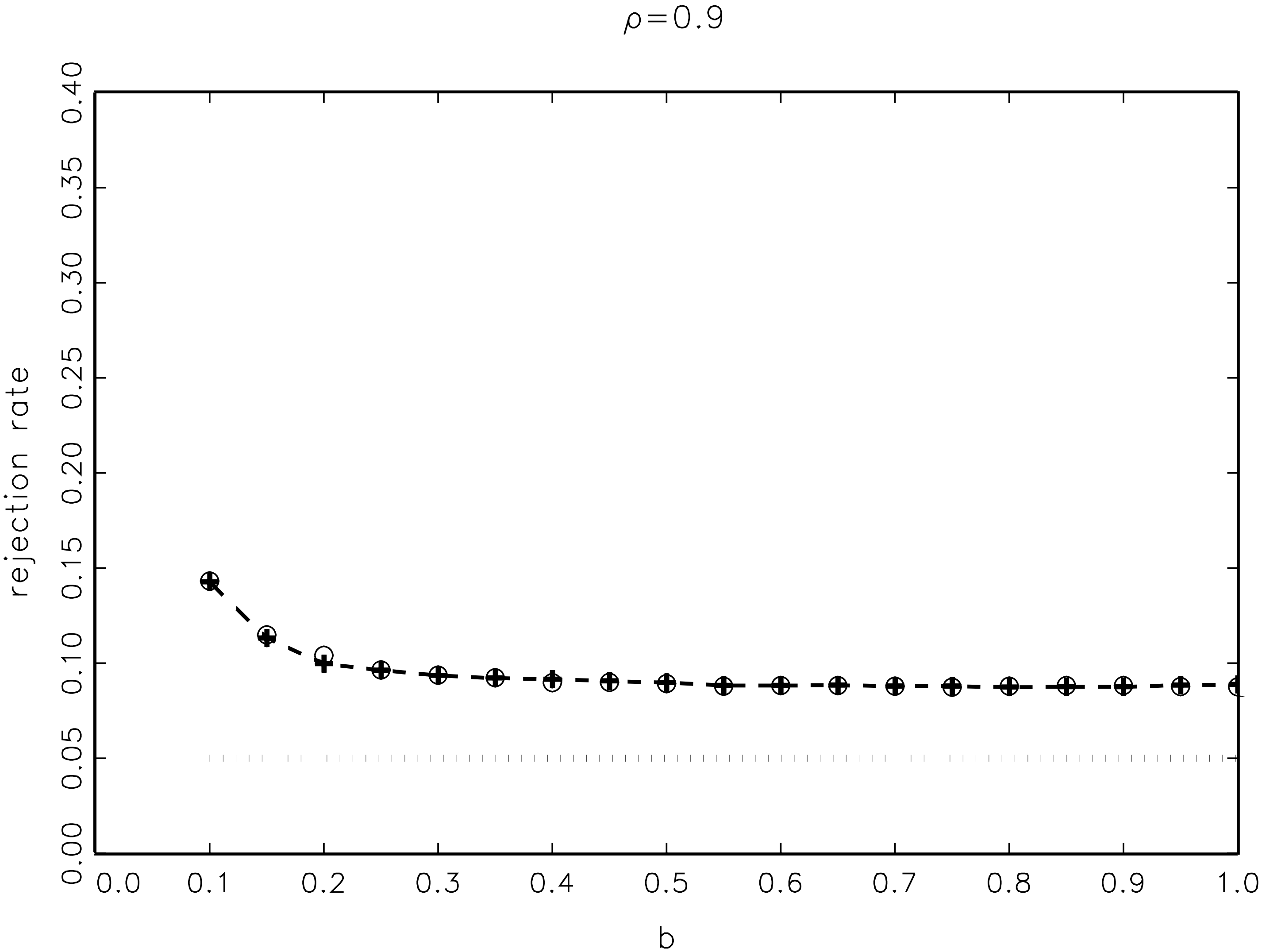


Figure 2.34: AM Series - Random Bernoulli ($p = 0.7$), Bartlett, $T = 200$

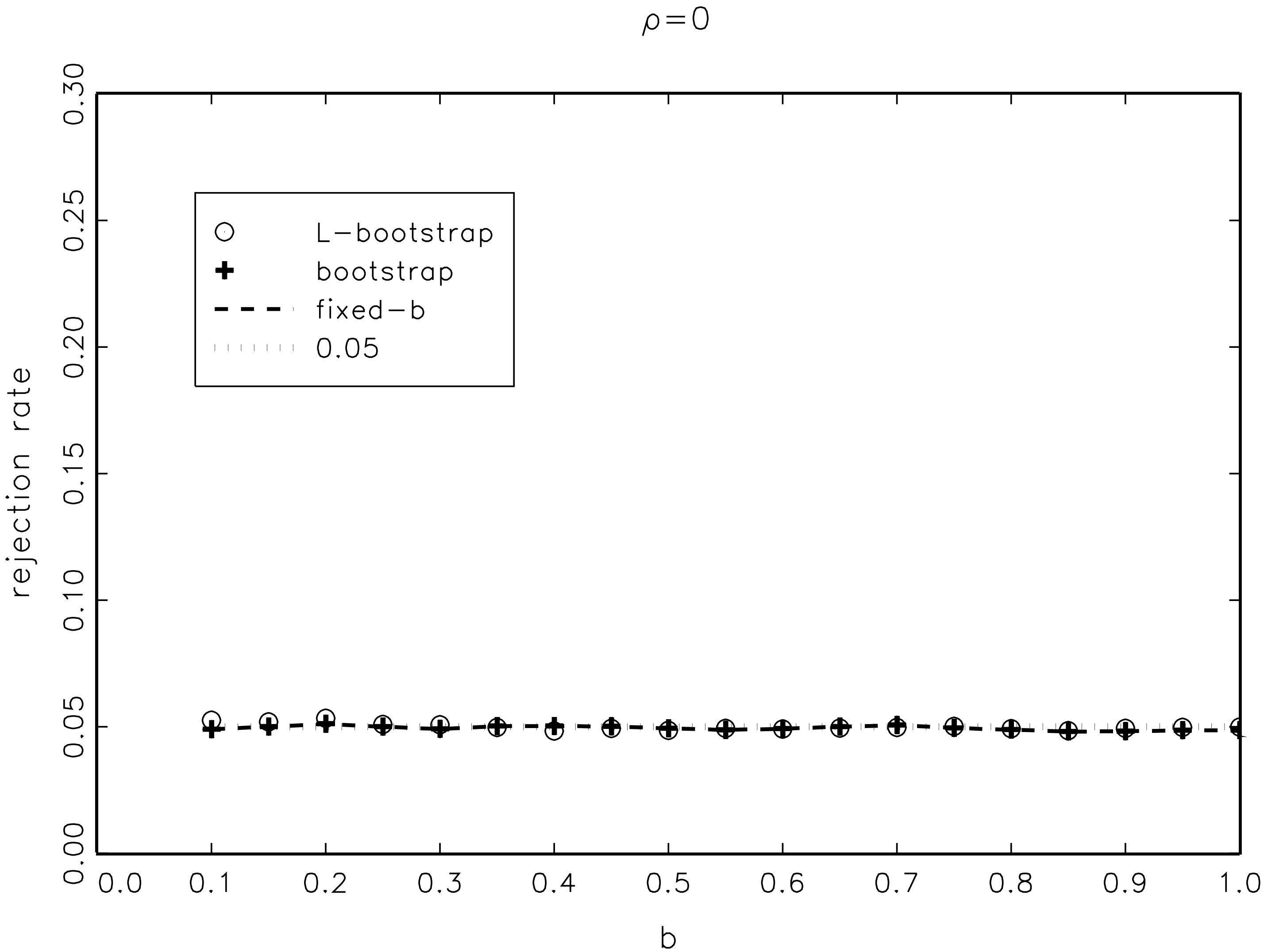


Figure 2.34: (cont'd)

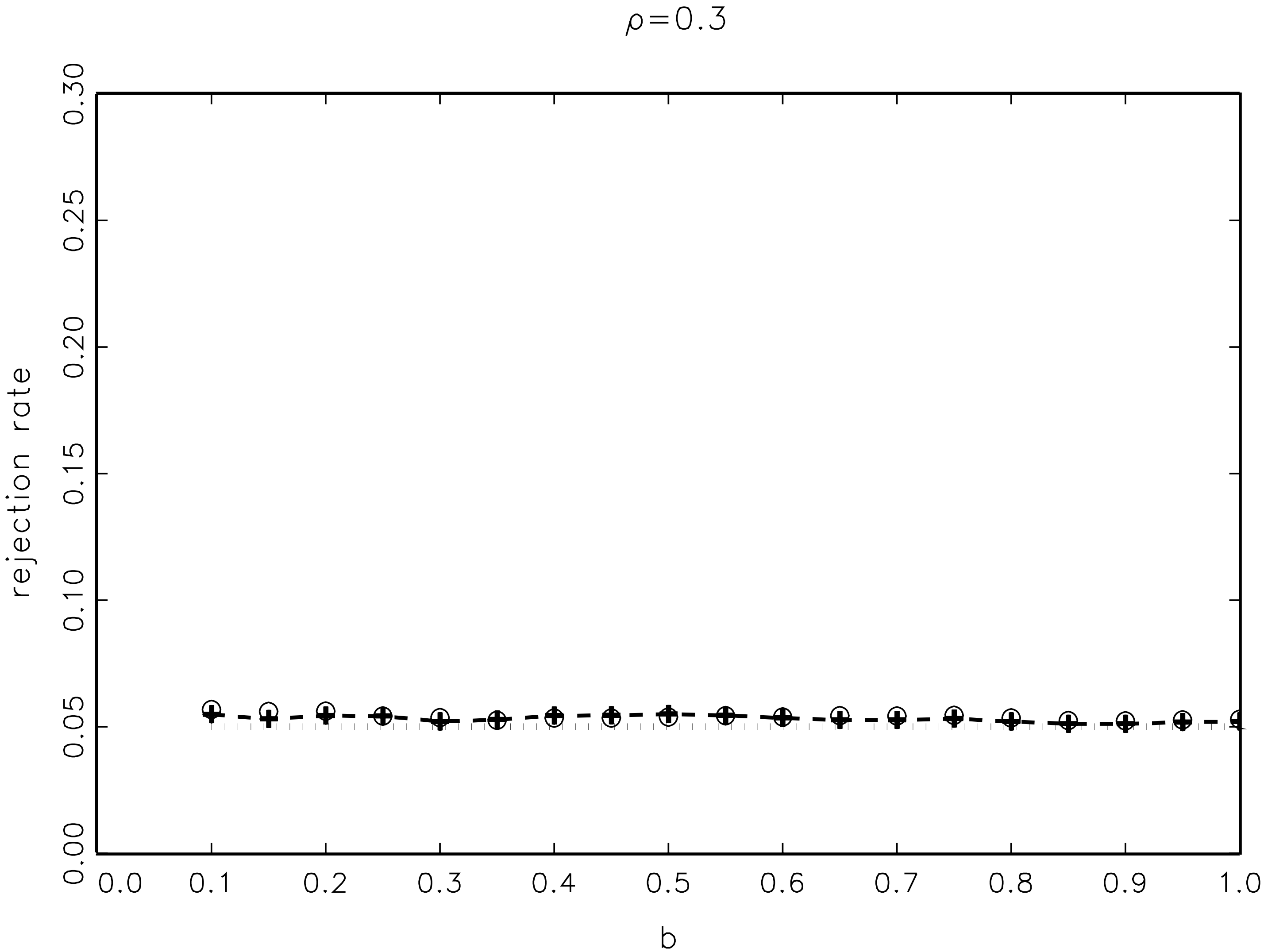


Figure 2.34: (cont'd)

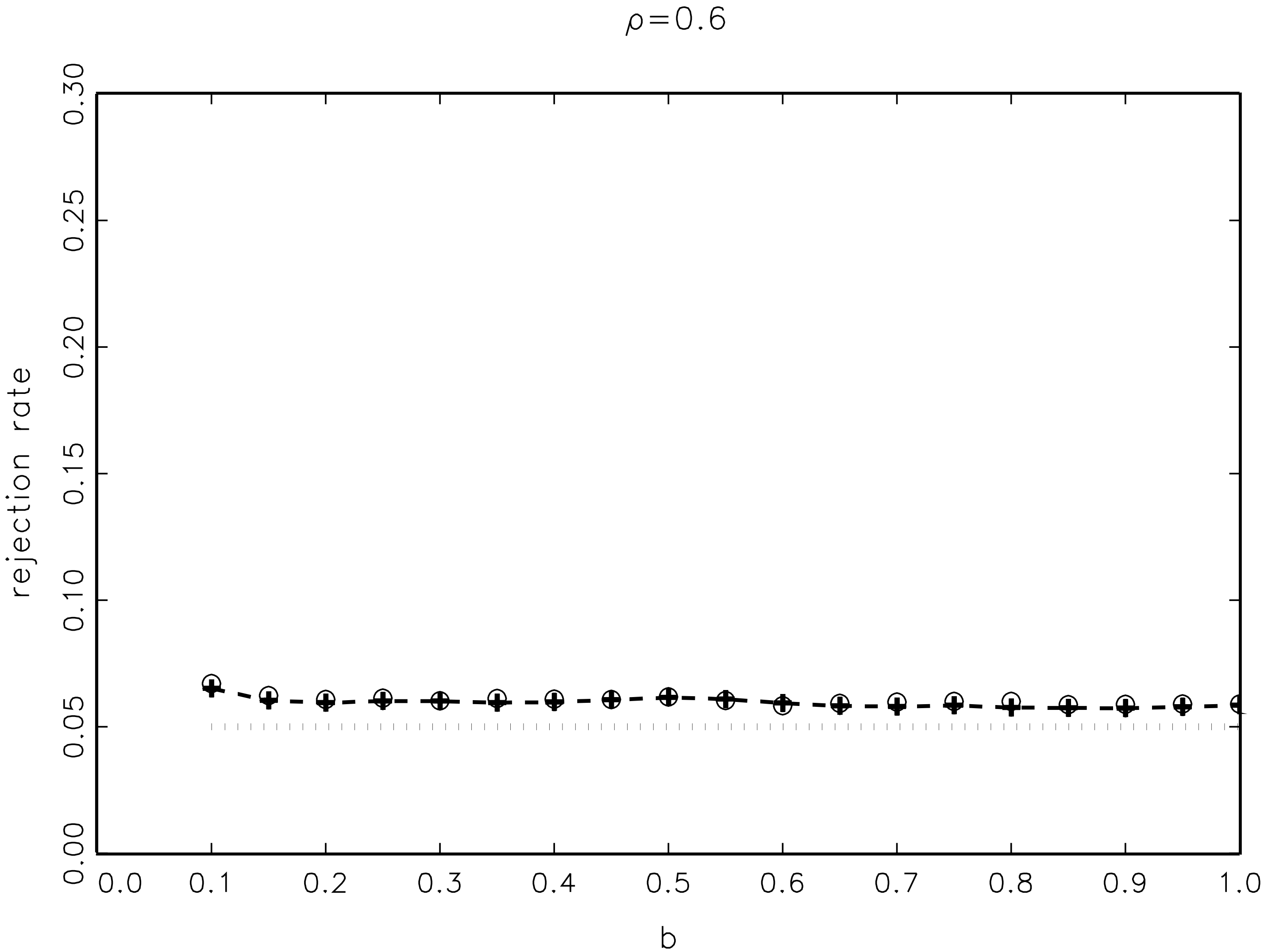


Figure 2.34: (cont'd)

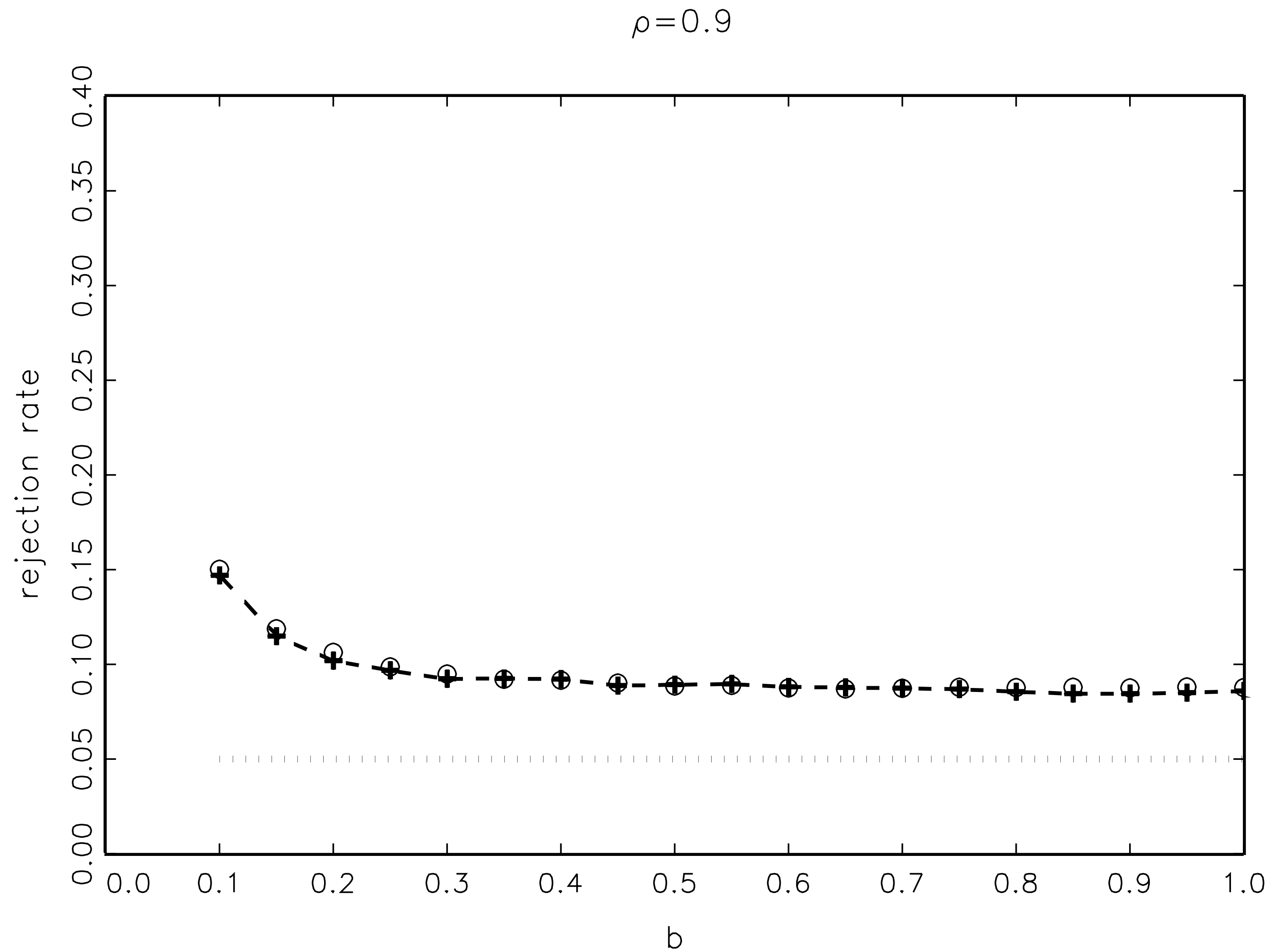


Figure 2.35: ES - World War (yearly), Bartlett, $T = 36$

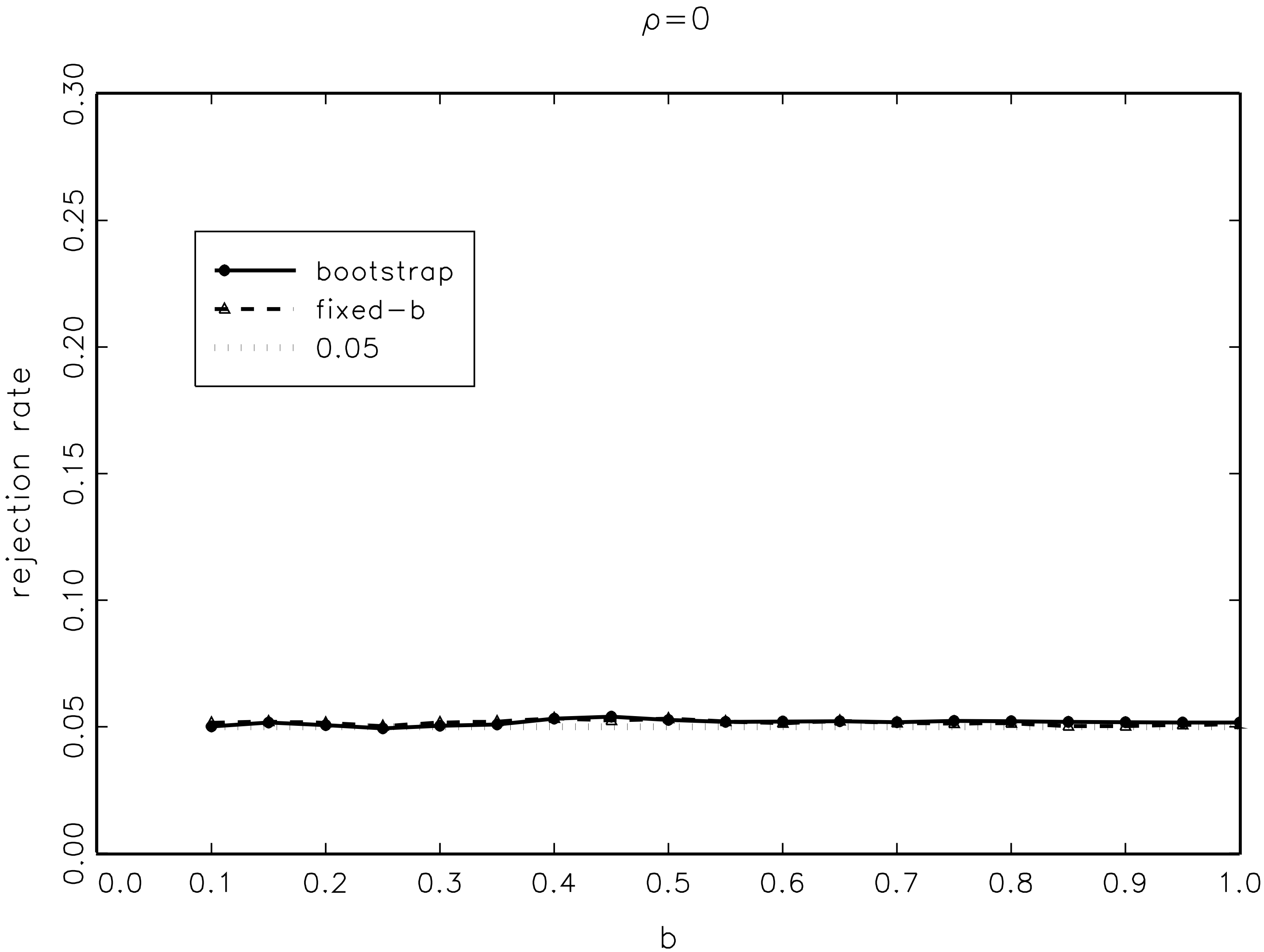


Figure 2.35: (cont'd)

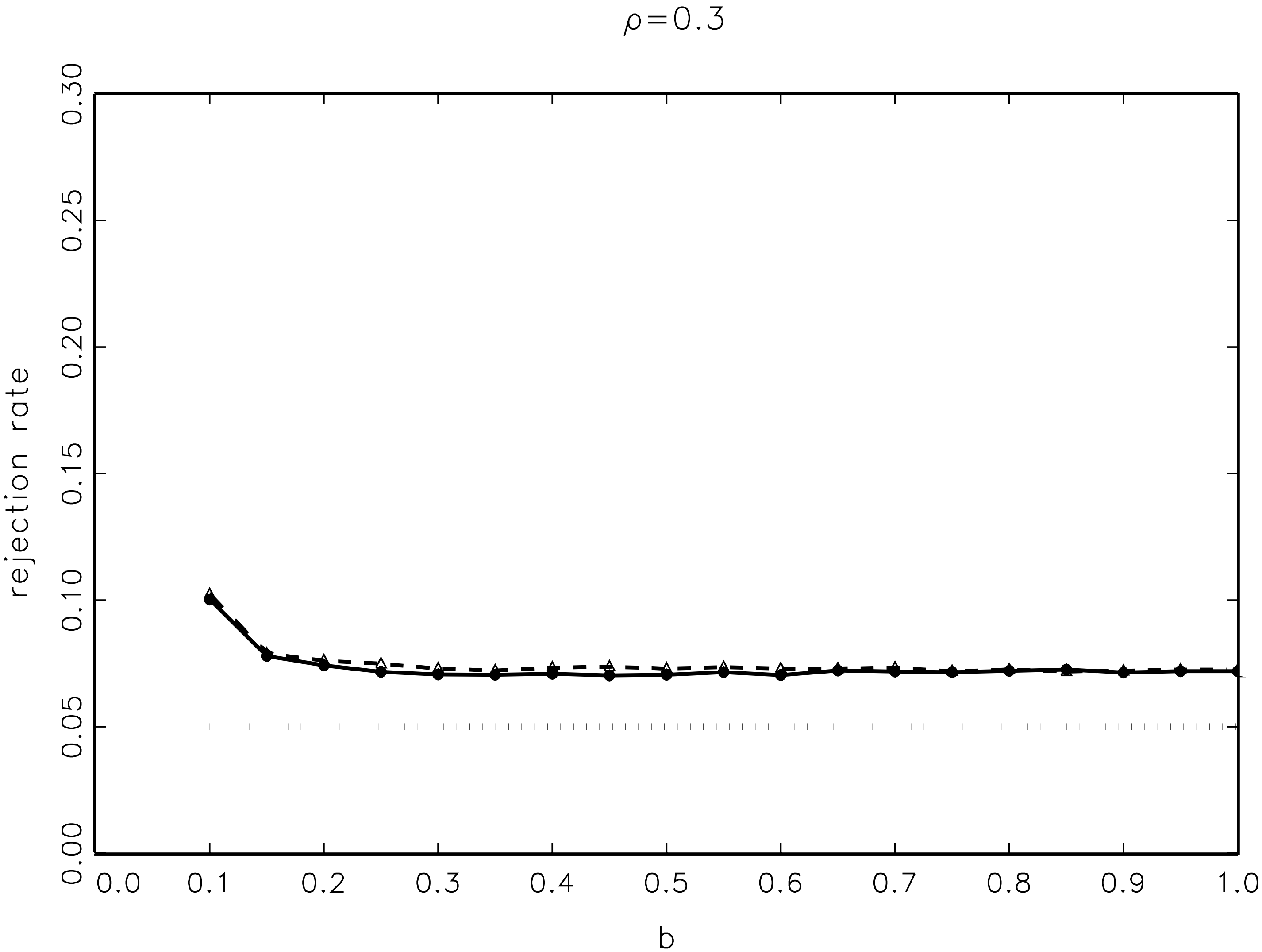


Figure 2.35: (cont'd)

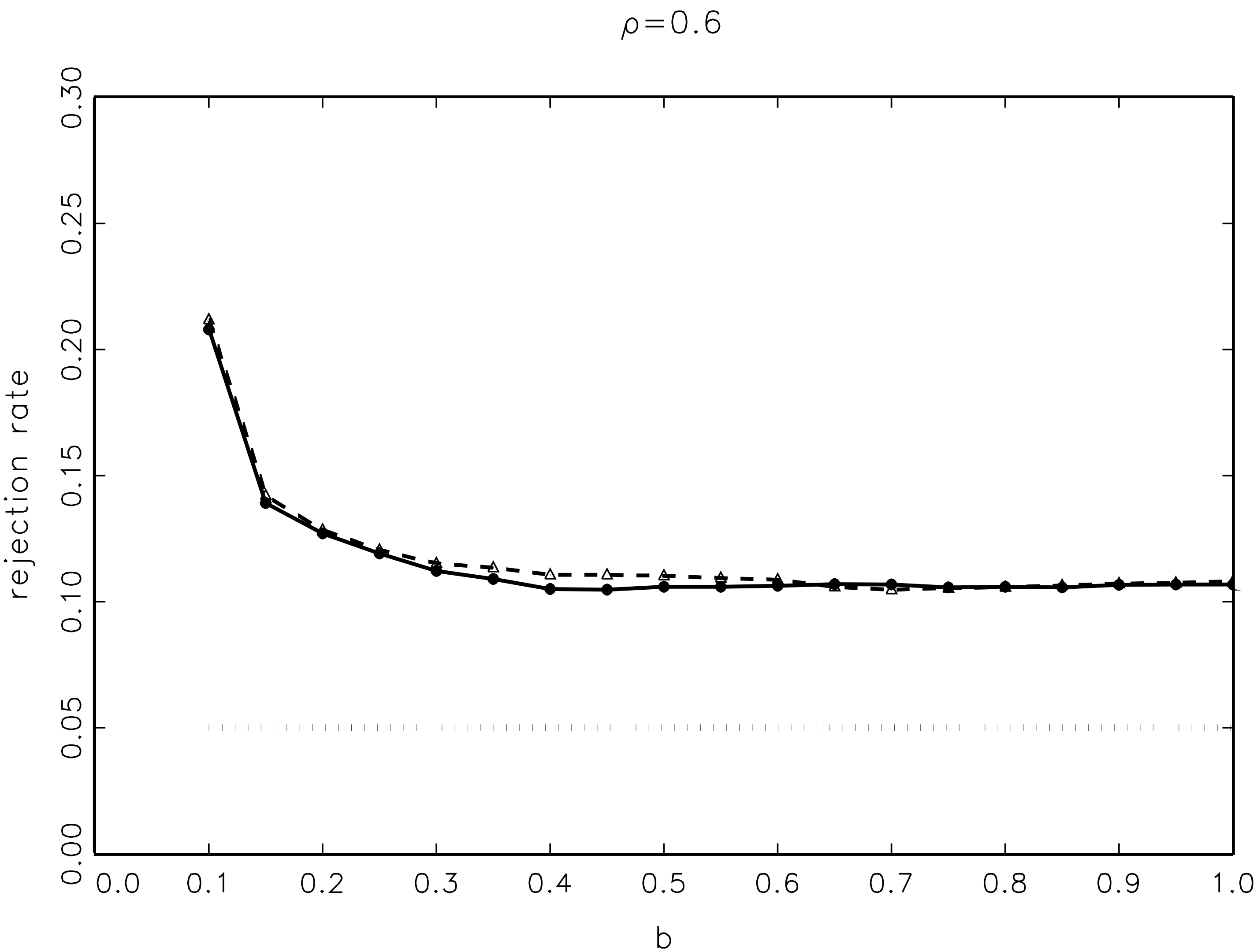


Figure 2.35: (cont'd)

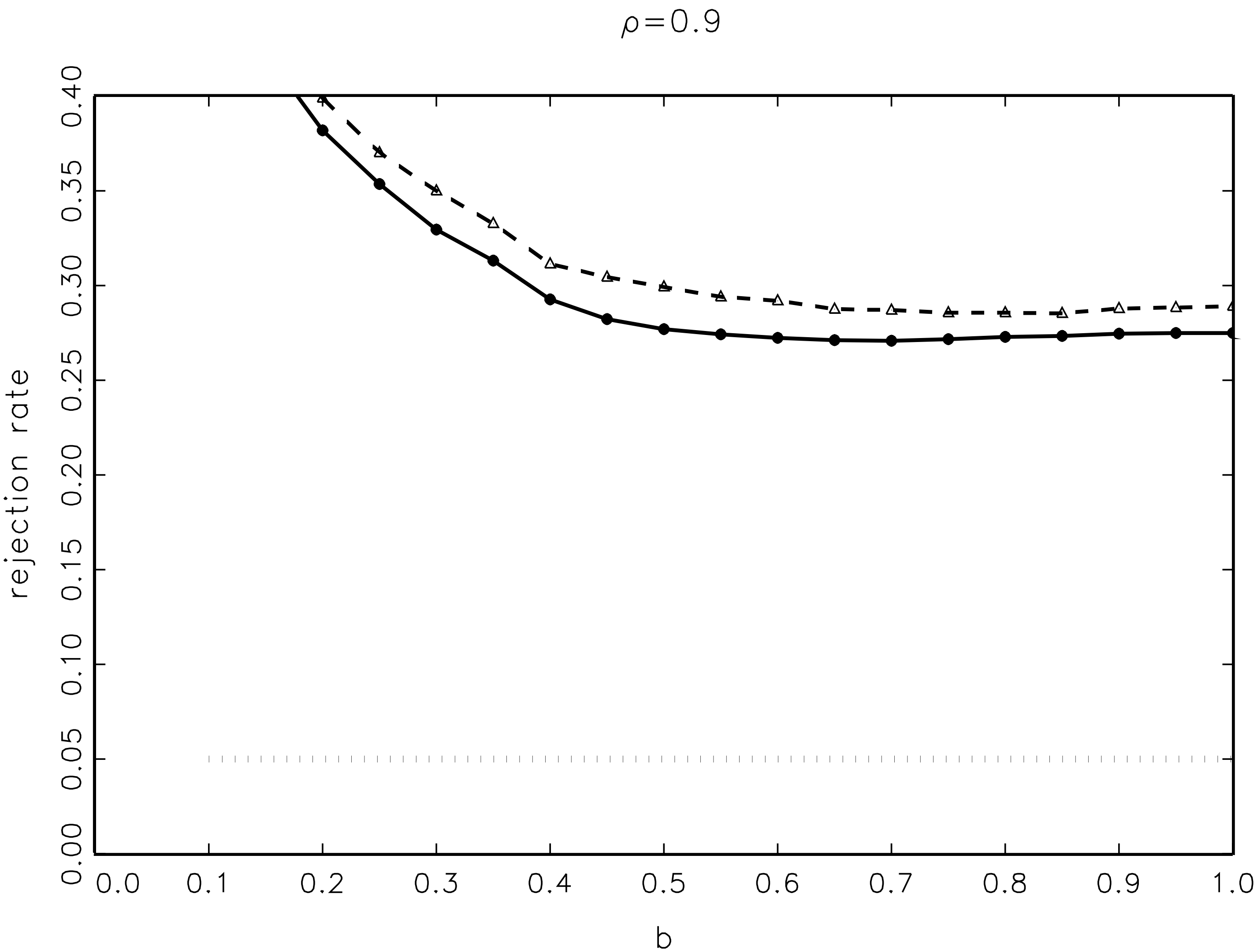


Figure 2.36: ES - World War (yearly), Bartlett, $T = 48$

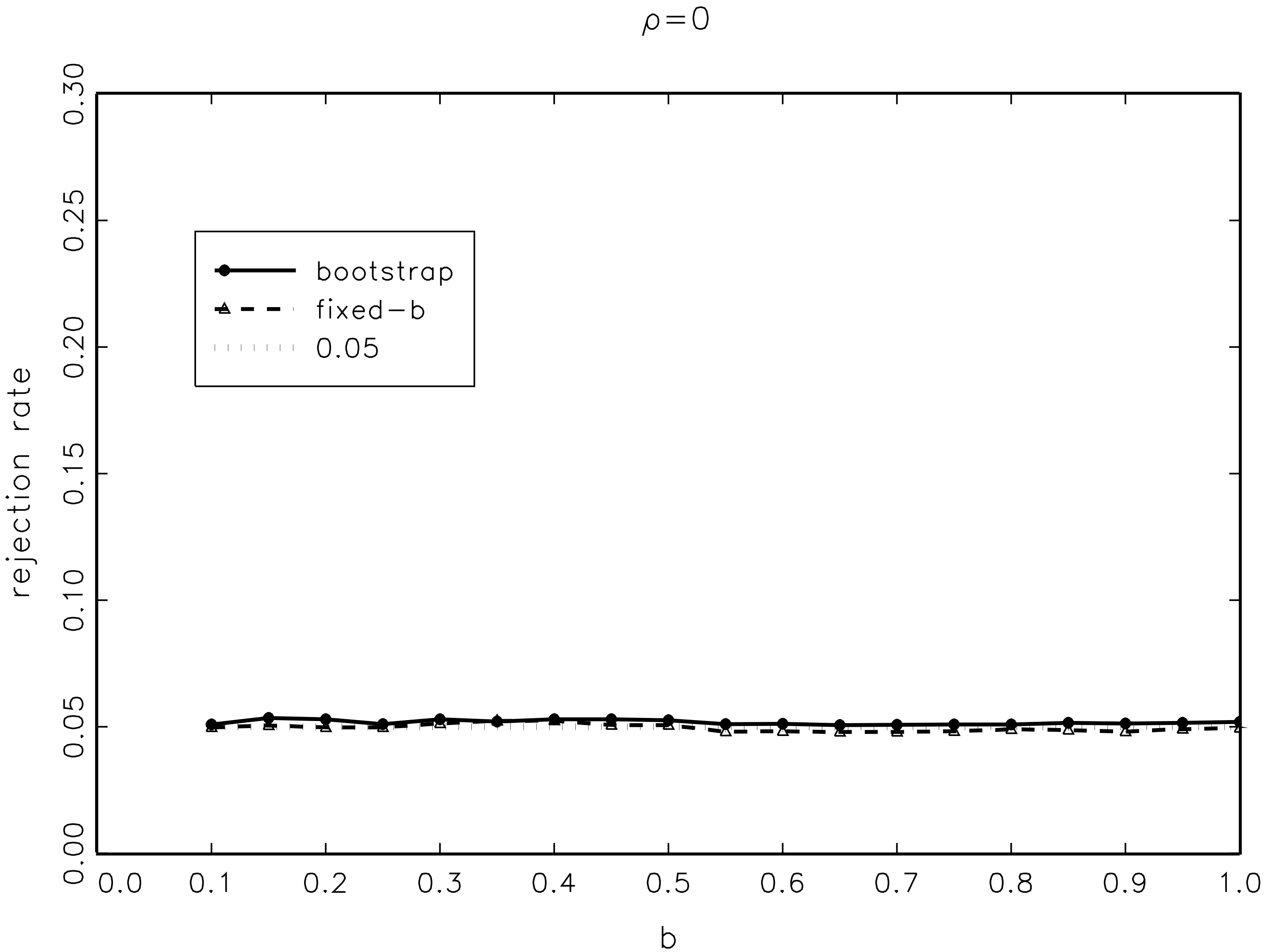


Figure 2.36: (cont'd)

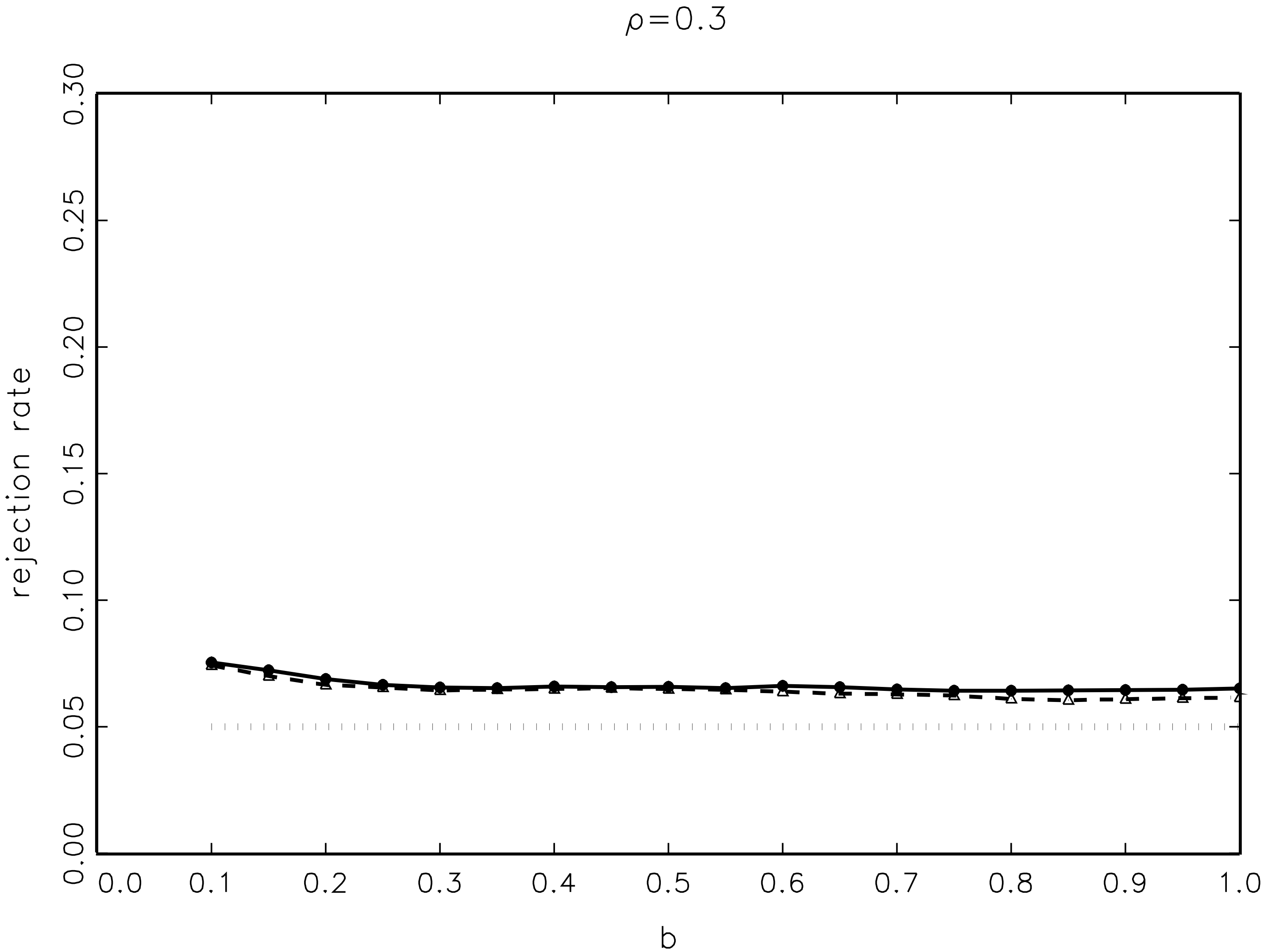


Figure 2.36: (cont'd)

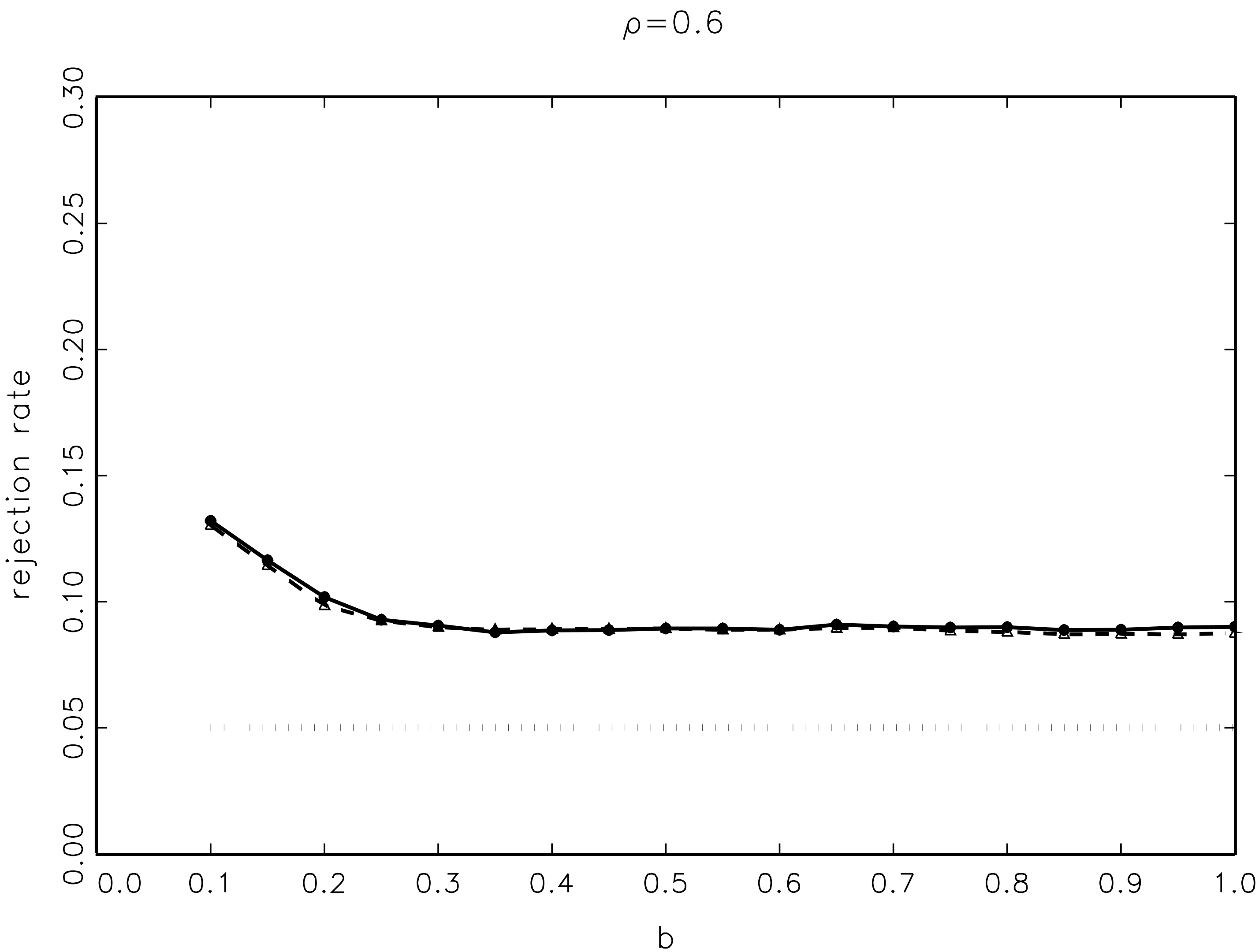


Figure 2.36: (cont'd)

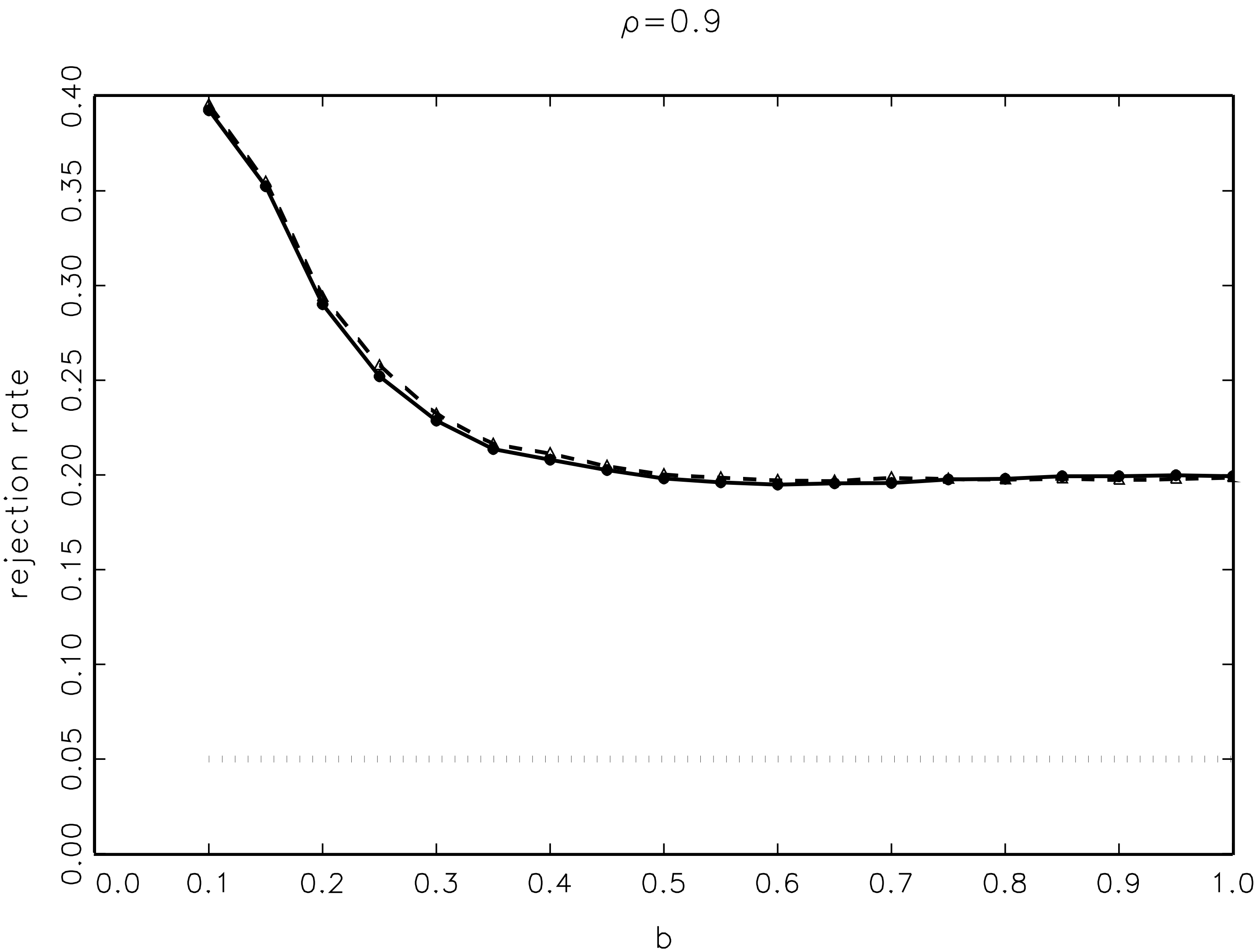


Figure 2.37: ES - World War (yearly), Bartlett, $T = 60$

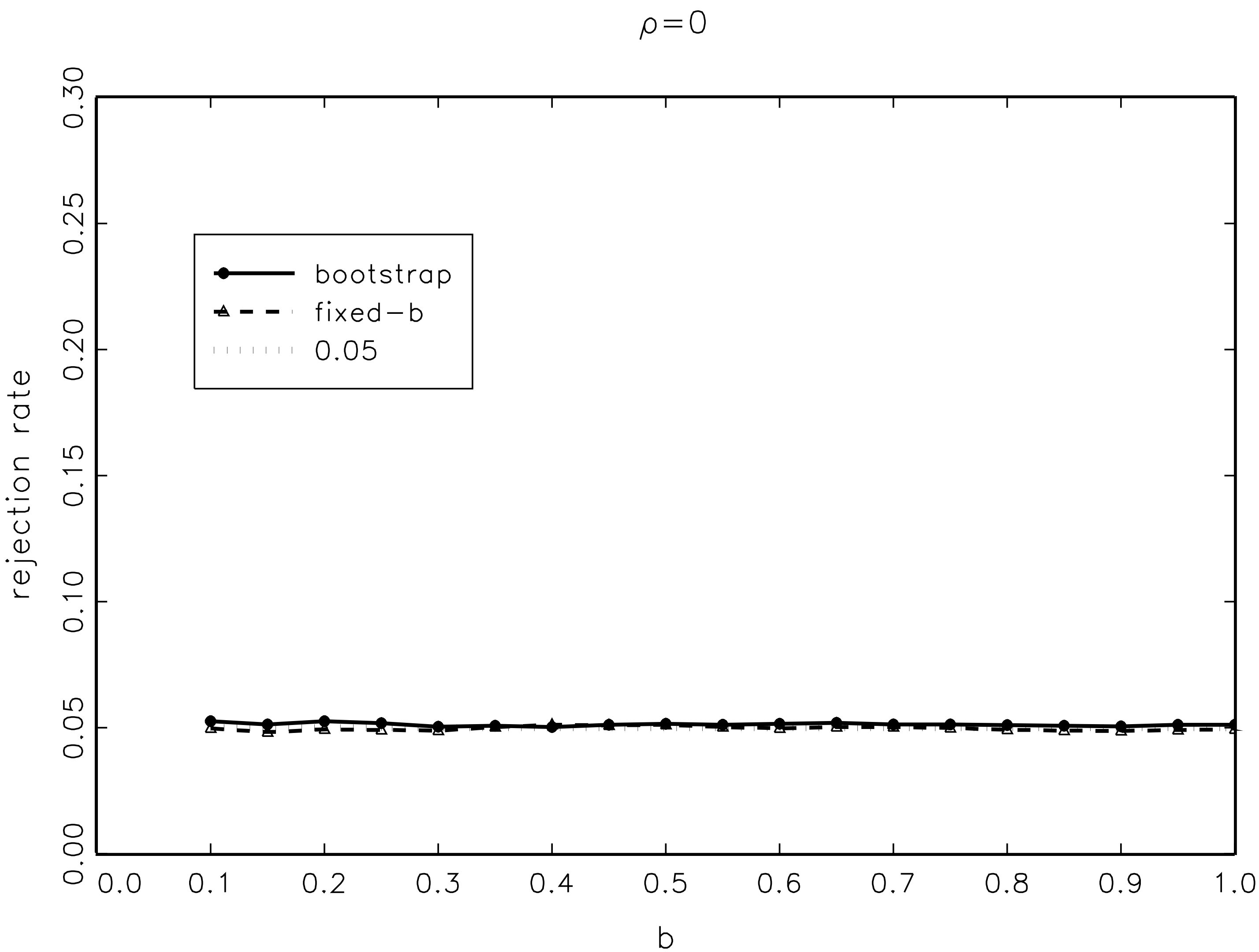


Figure 2.37: (cont'd)

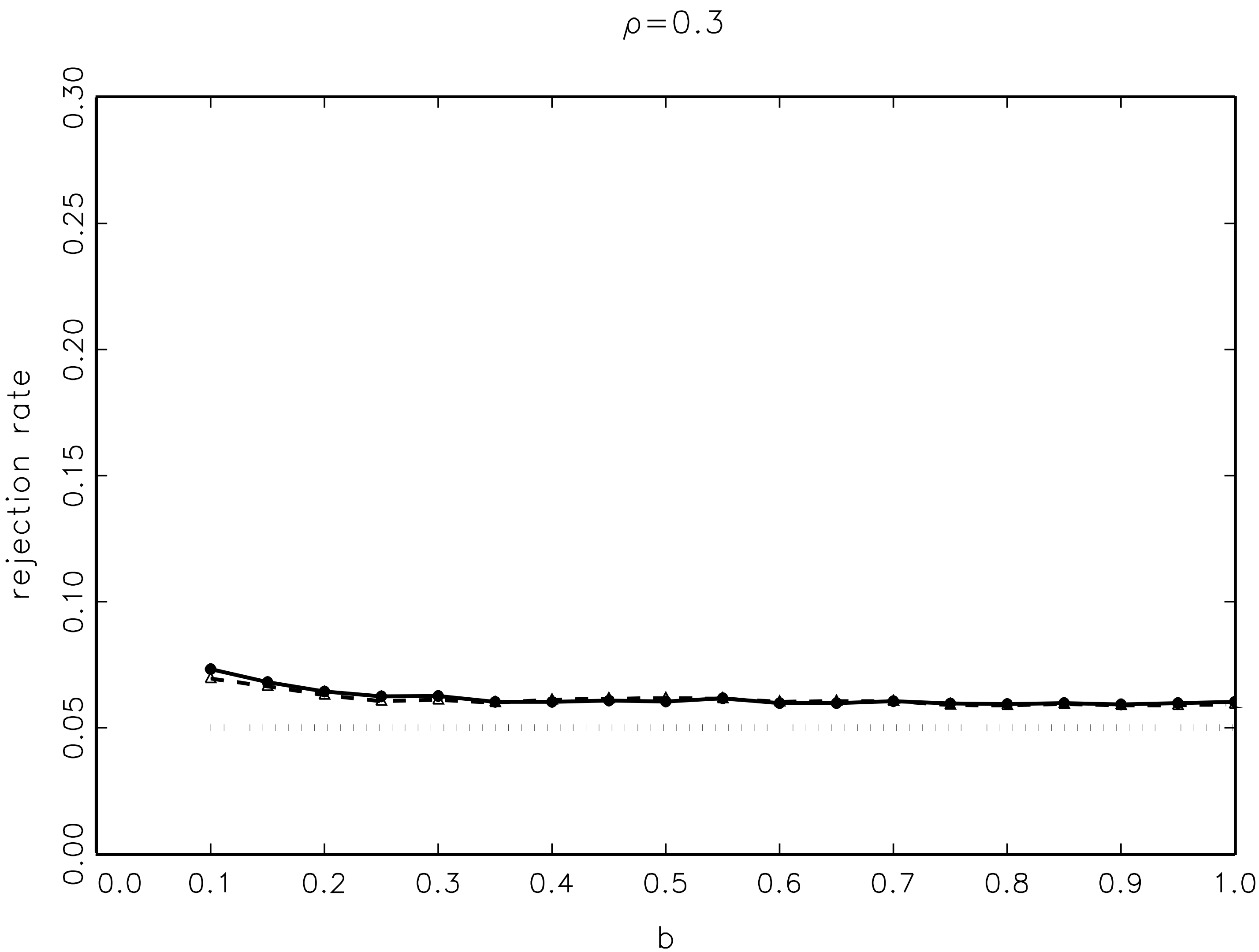


Figure 2.37: (cont'd)

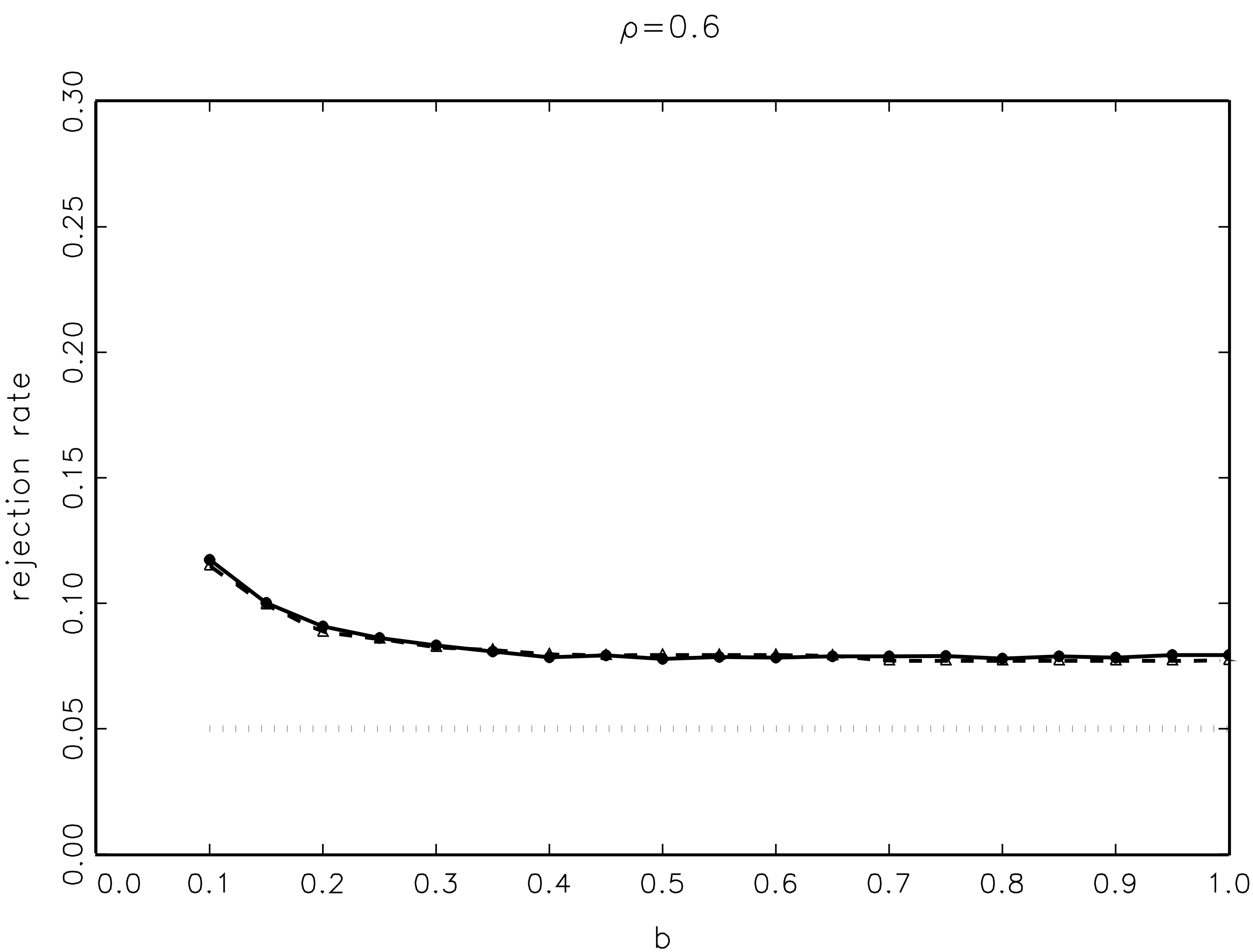


Figure 2.37: (cont'd)

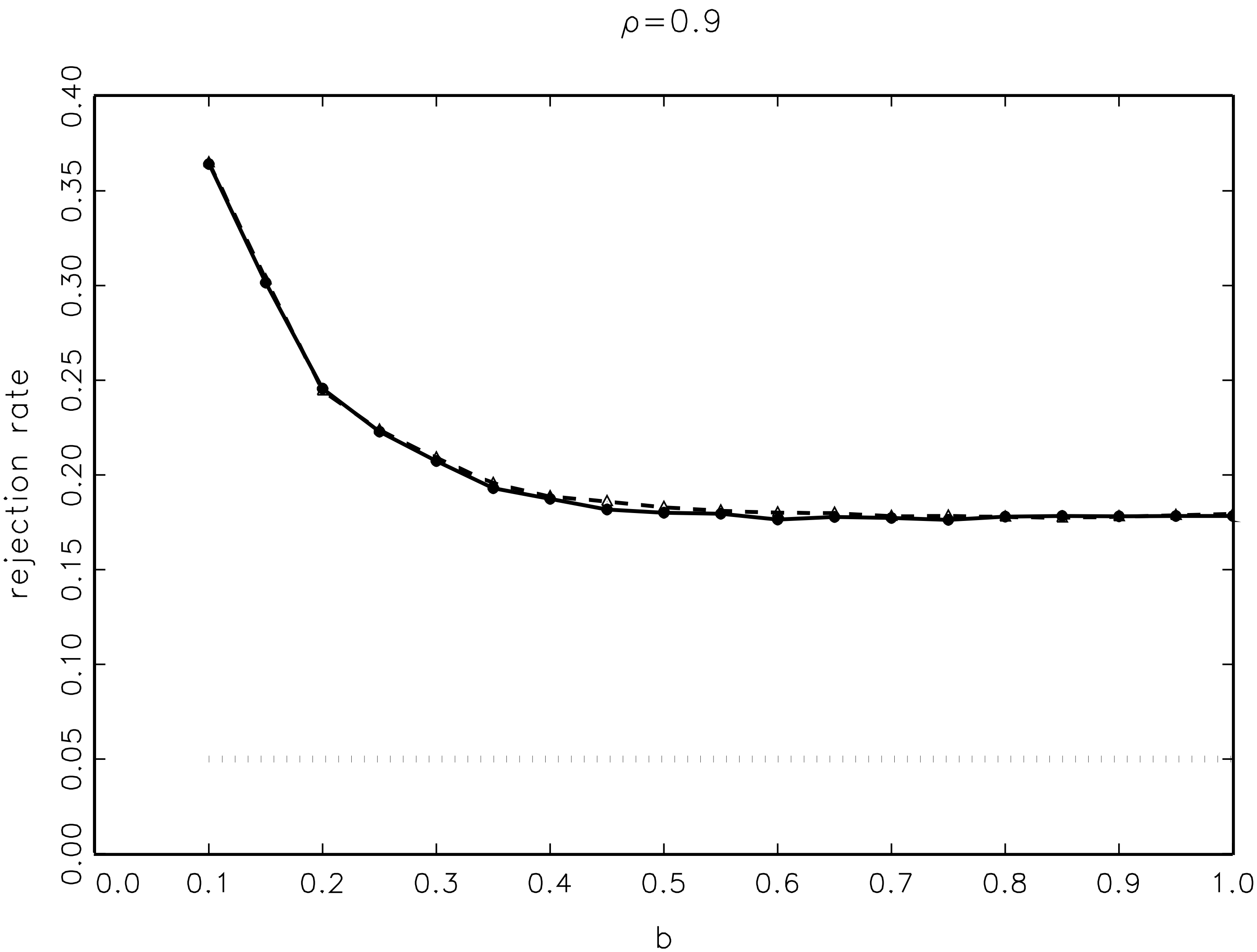


Figure 2.38: ES - World War (quarterly), Bartlett, $T = 144$

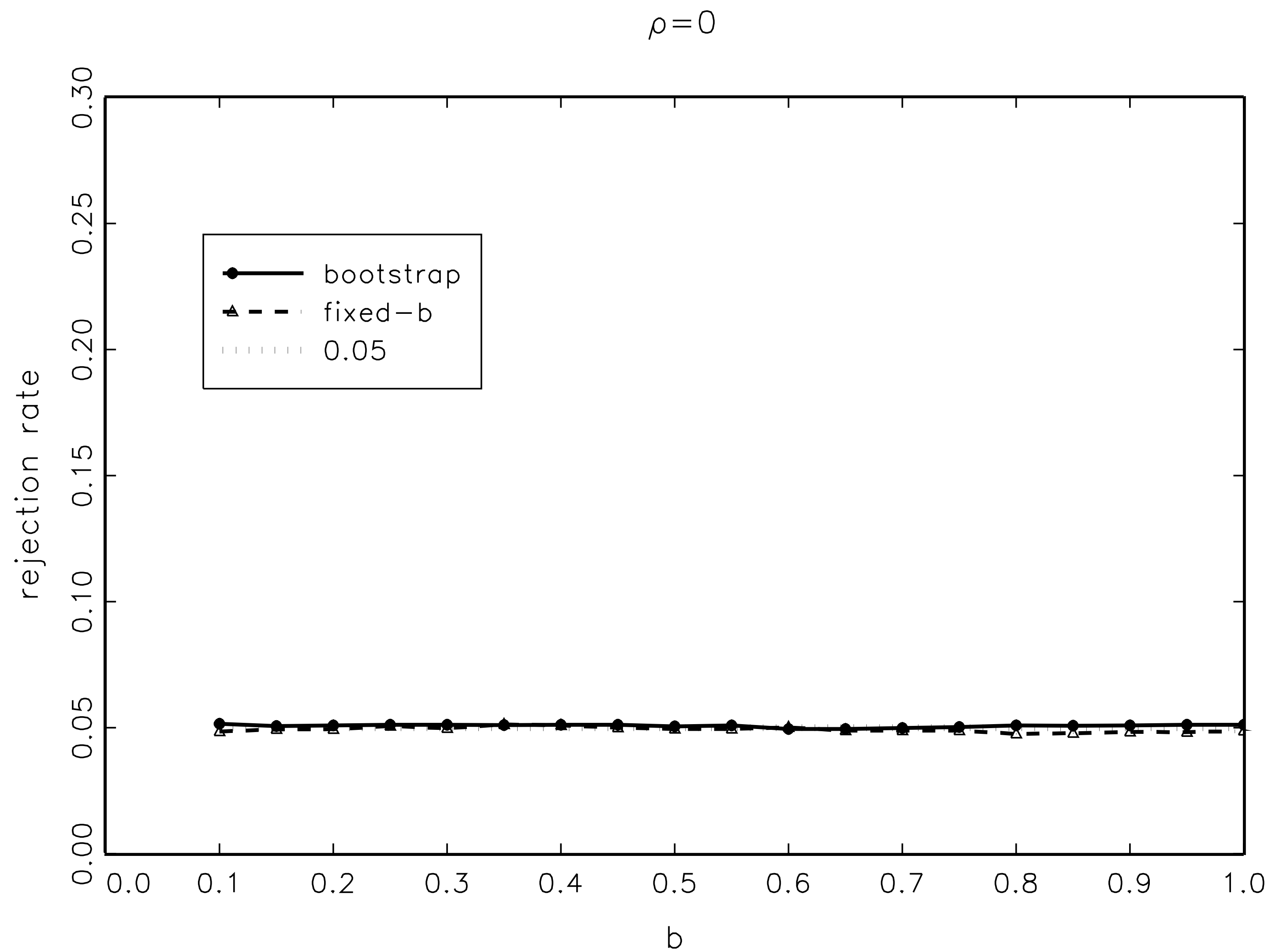


Figure 2.38: (cont'd)

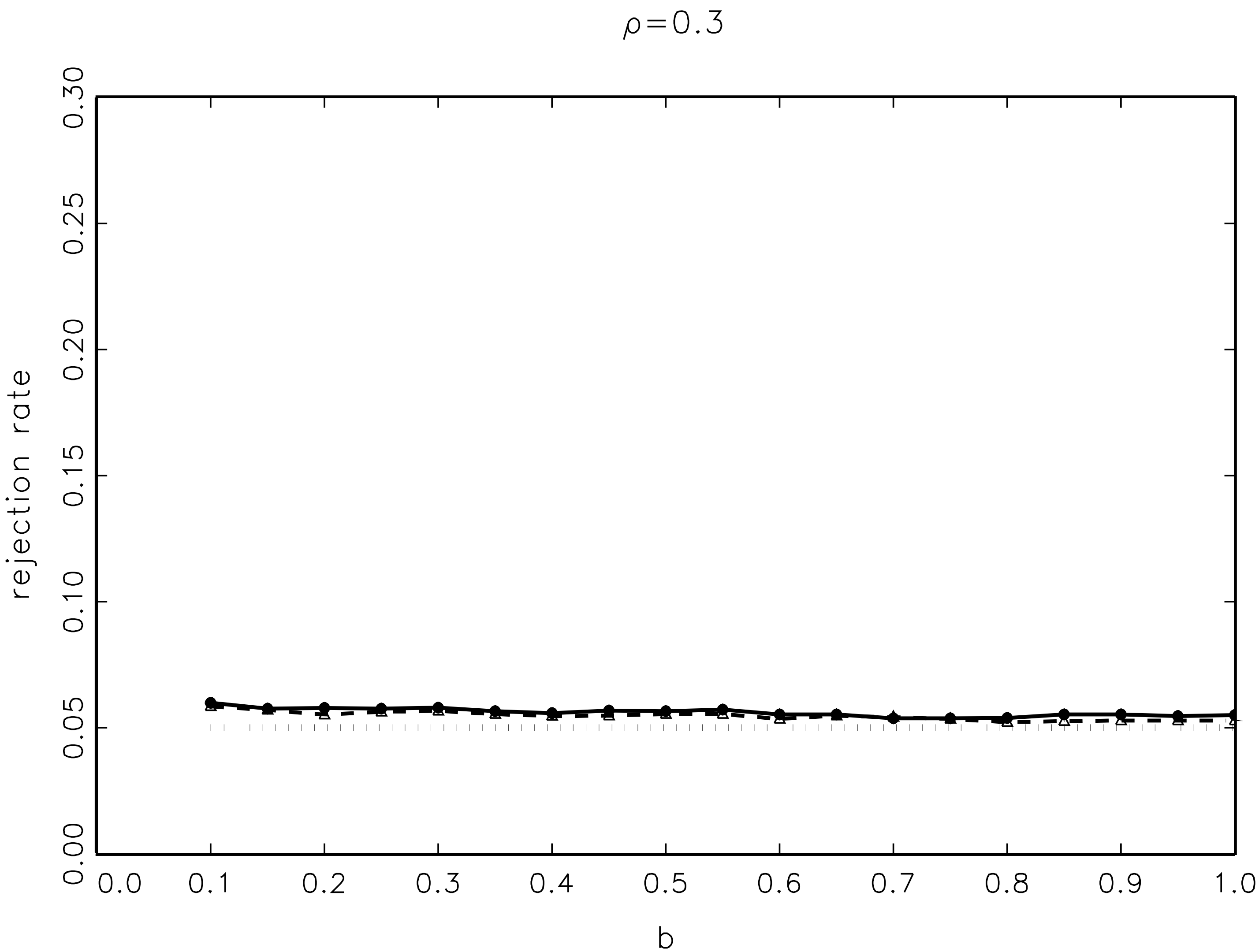


Figure 2.38: (cont'd)

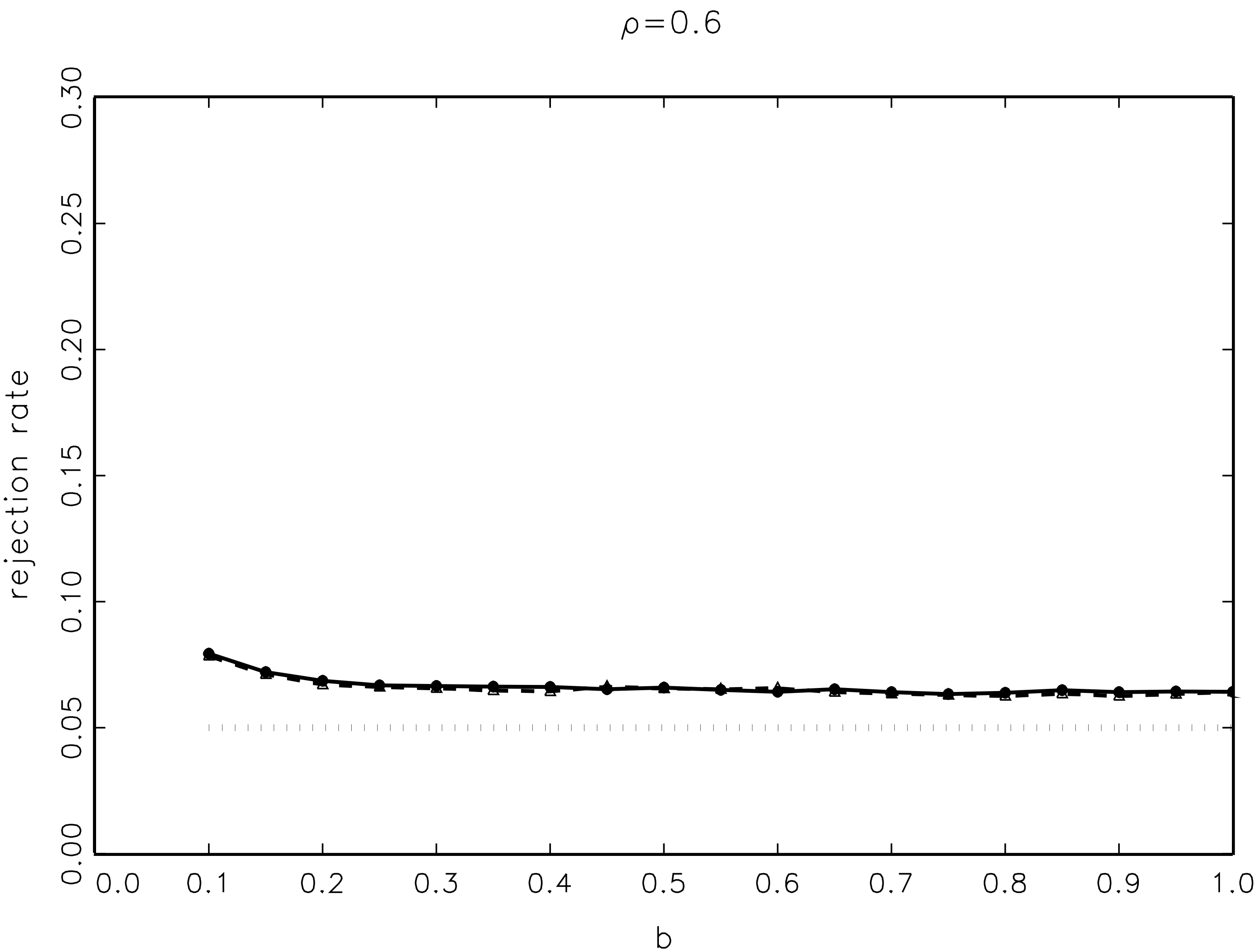


Figure 2.38: (cont'd)

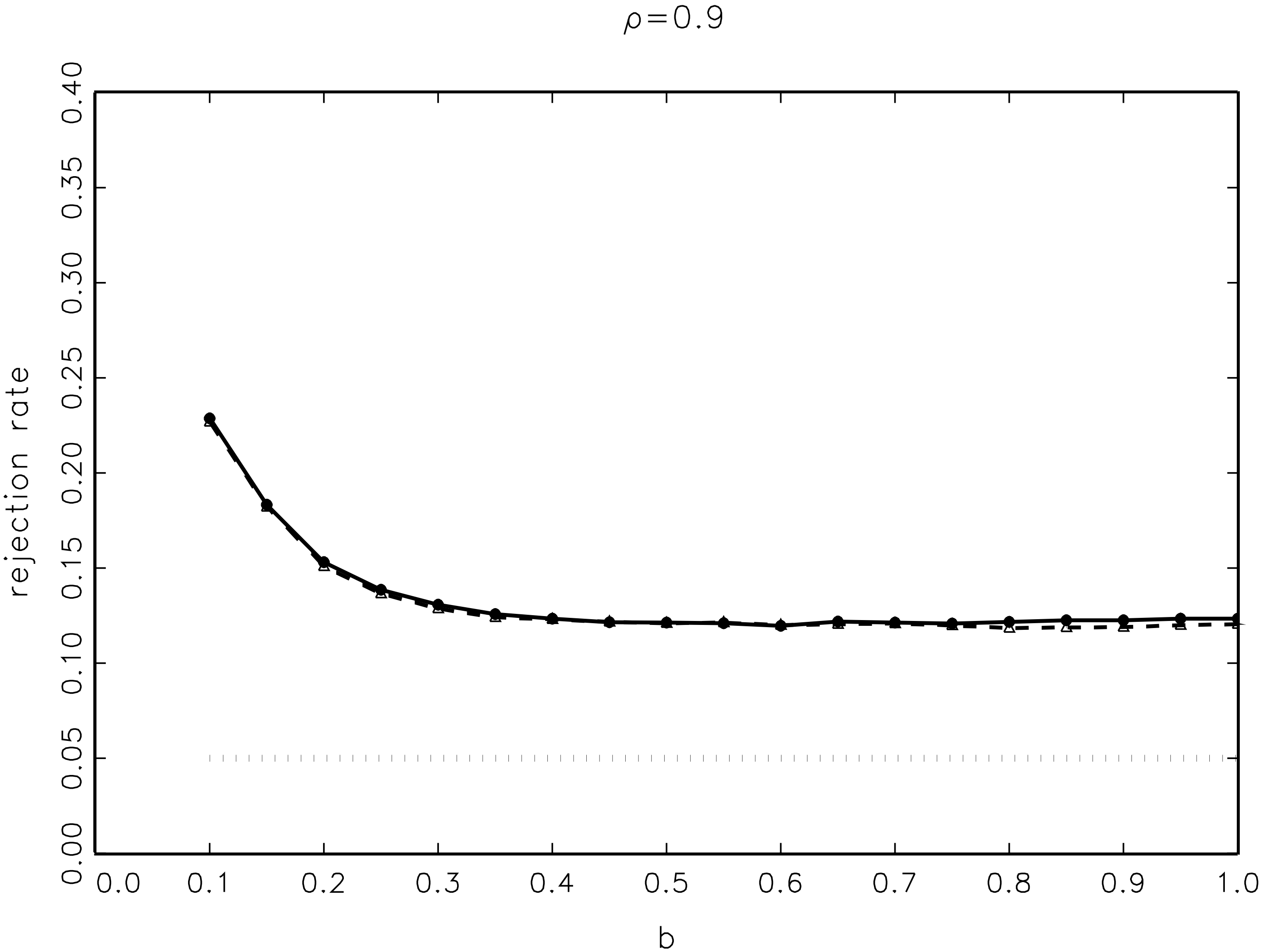


Figure 2.39: ES - World War (quarterly), Bartlett, $T = 192$

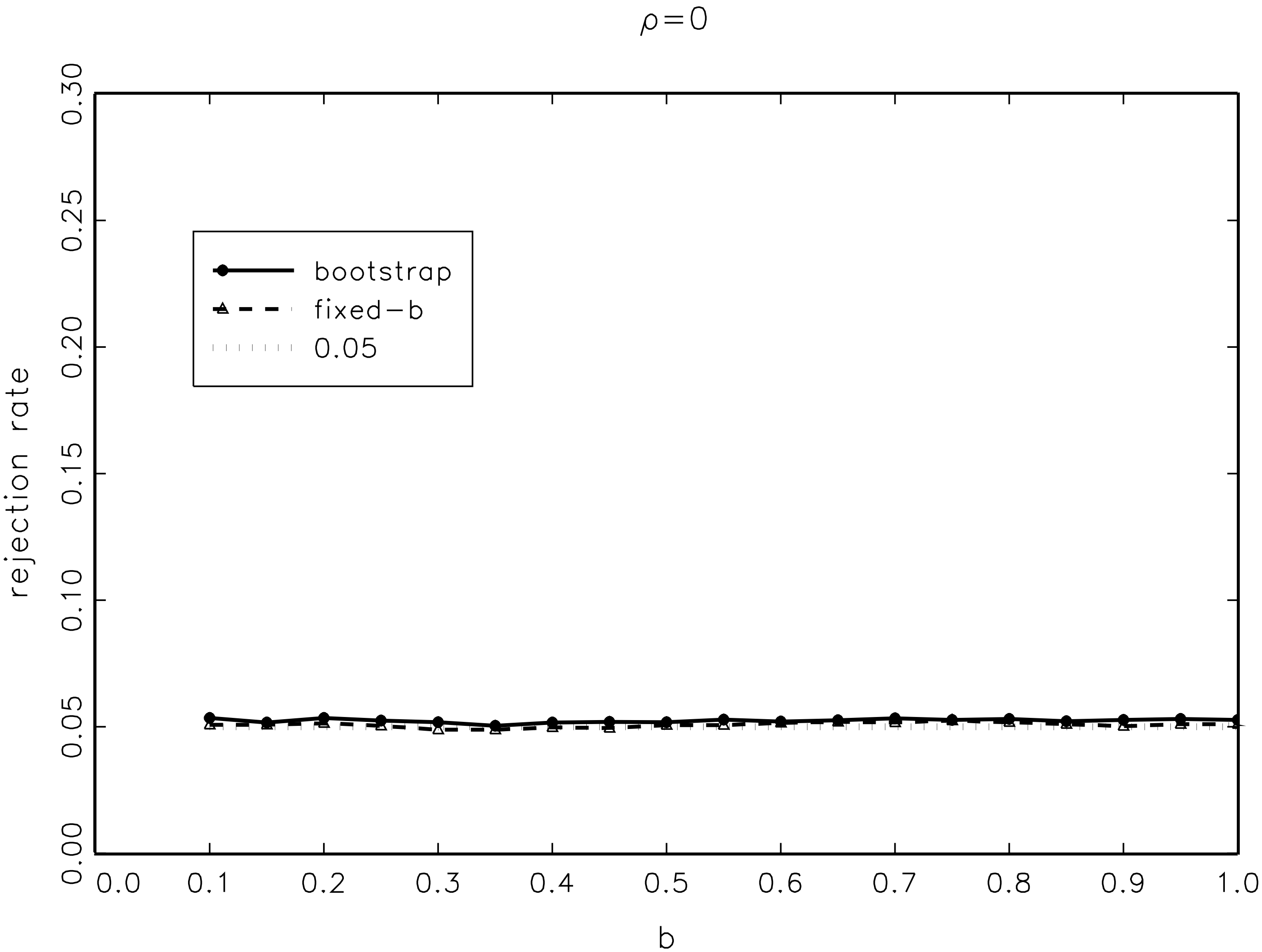


Figure 2.39: (cont'd)

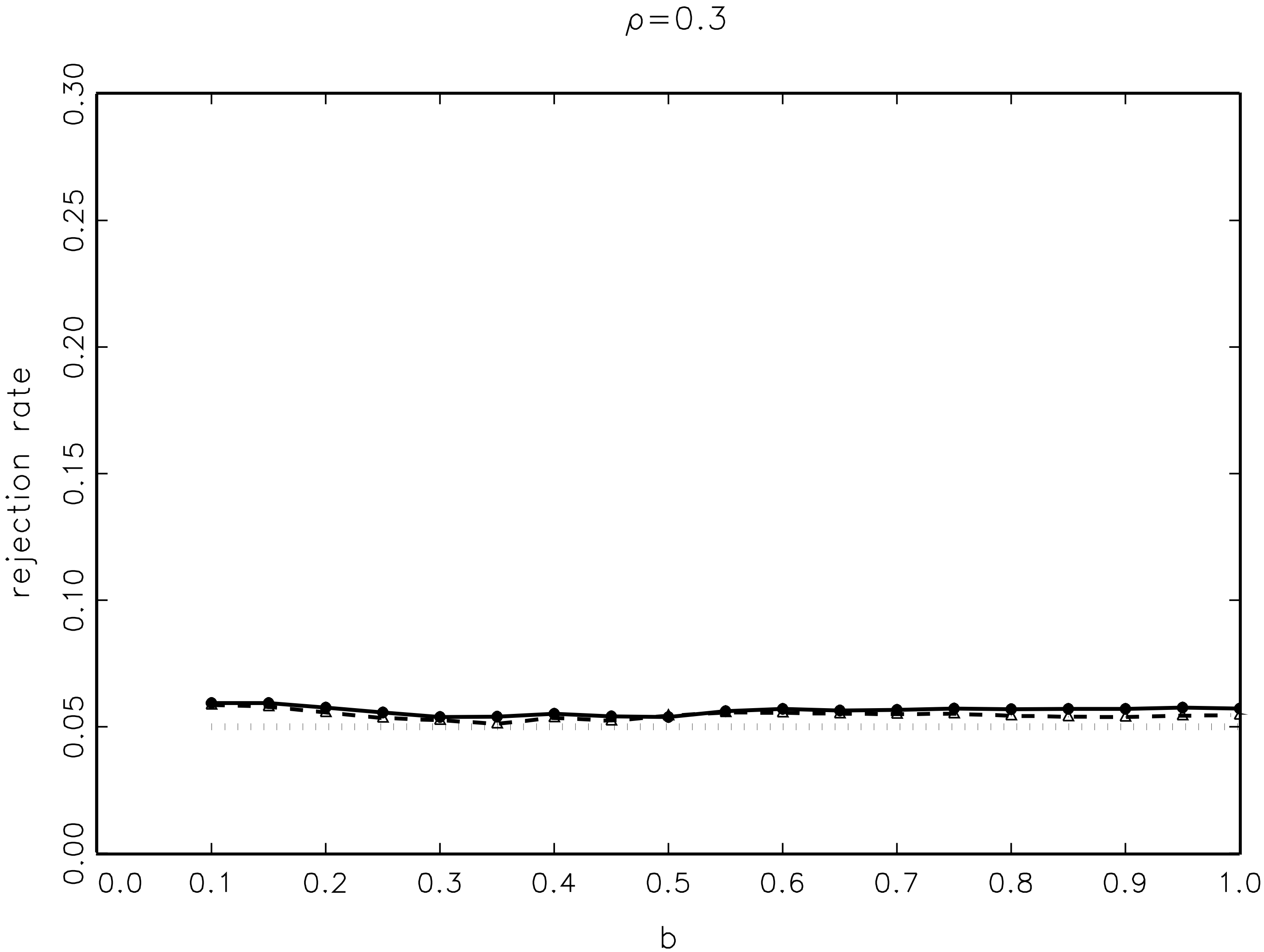


Figure 2.39: (cont'd)

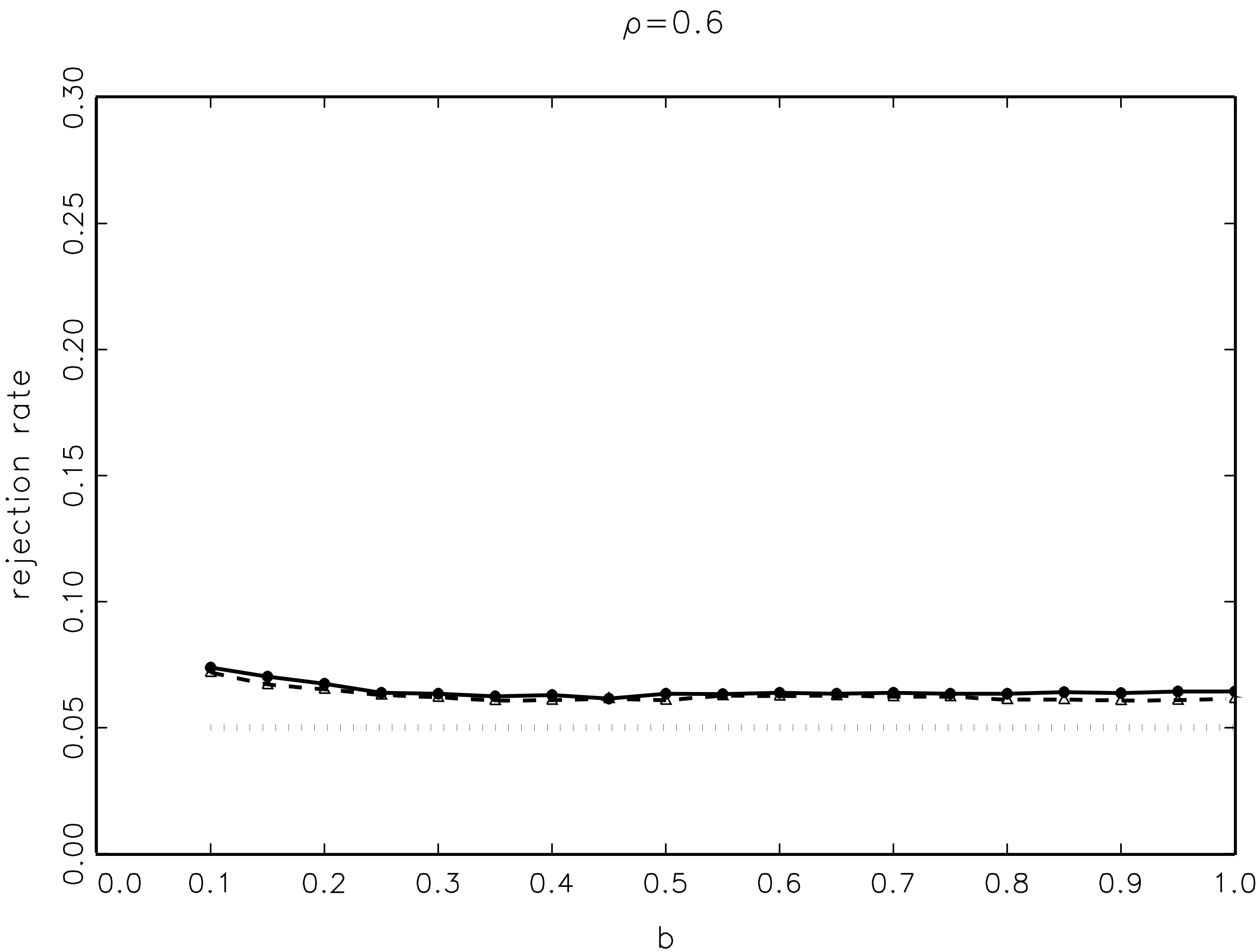


Figure 2.39: (cont'd)

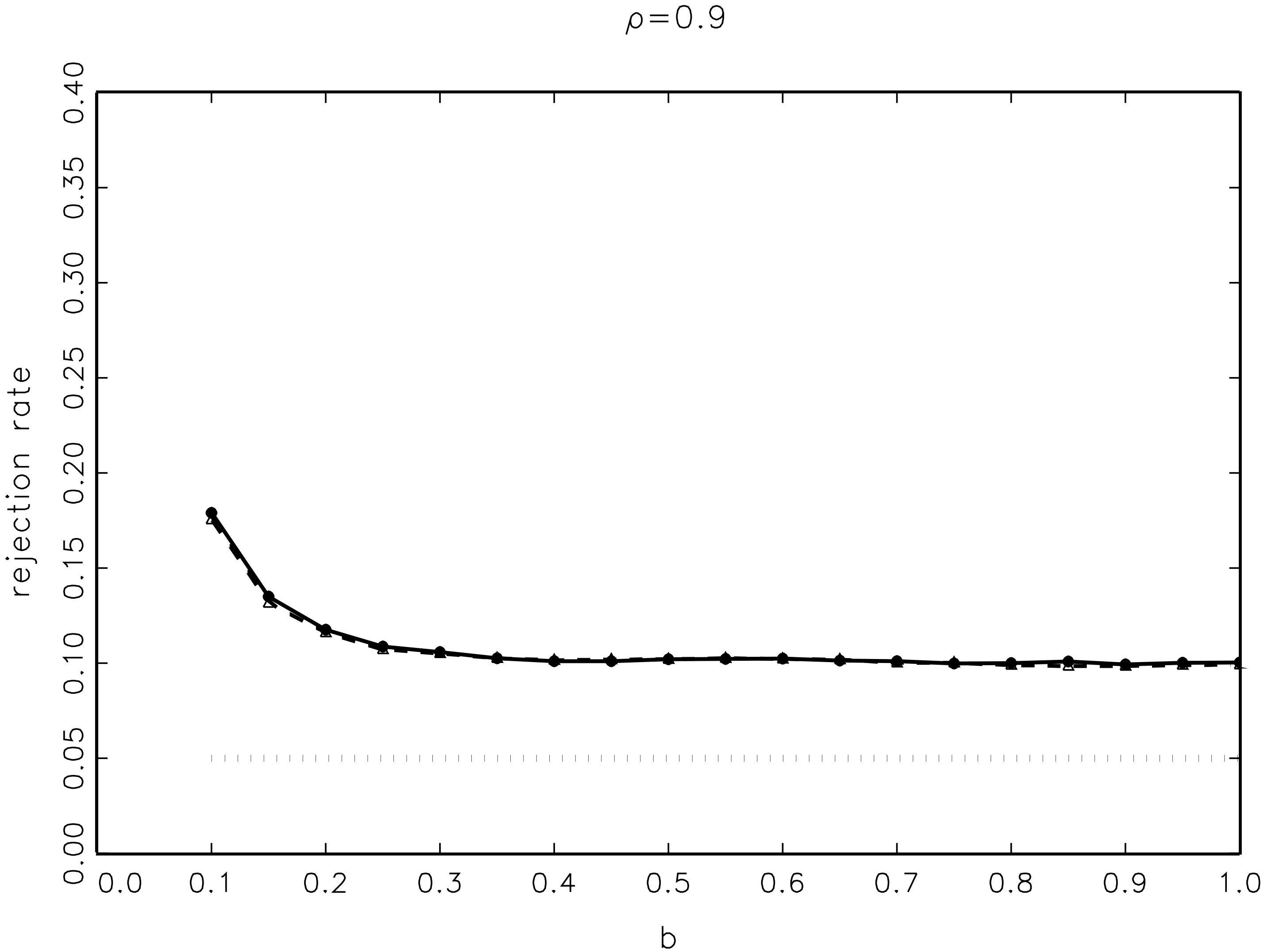


Figure 2.40: ES - World War (quarterly), Bartlett, $T = 240$

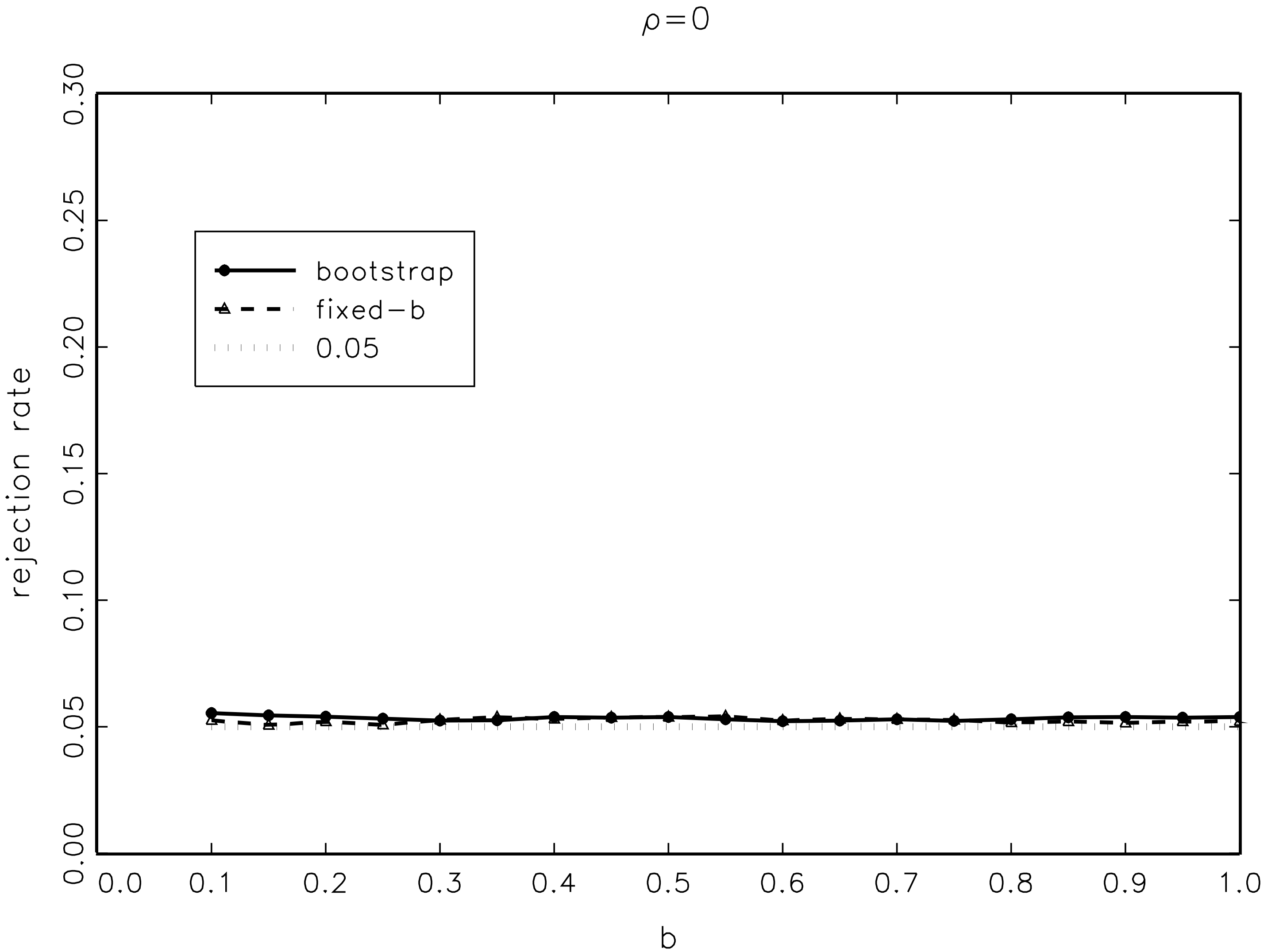


Figure 2.40: (cont'd)

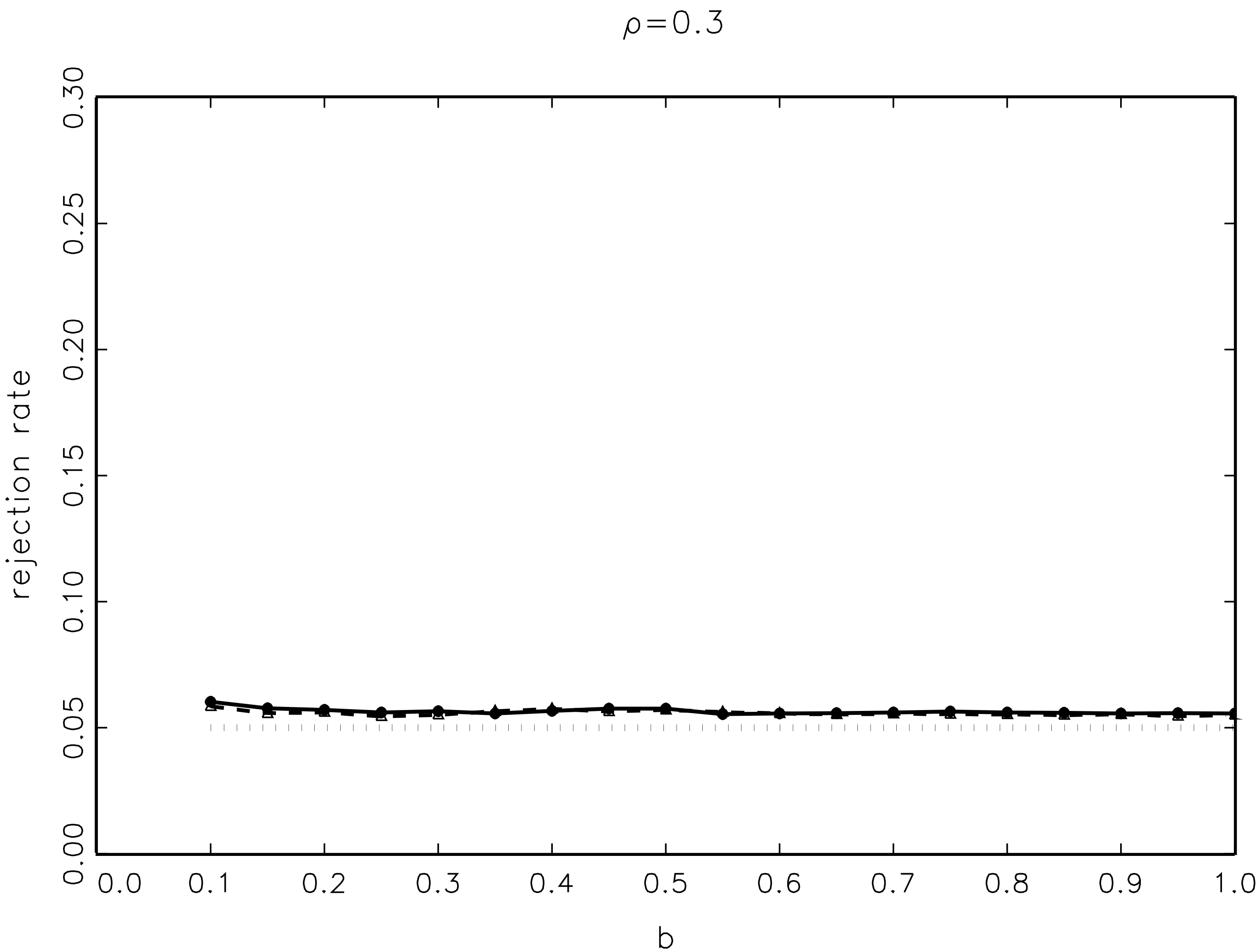


Figure 2.40: (cont'd)

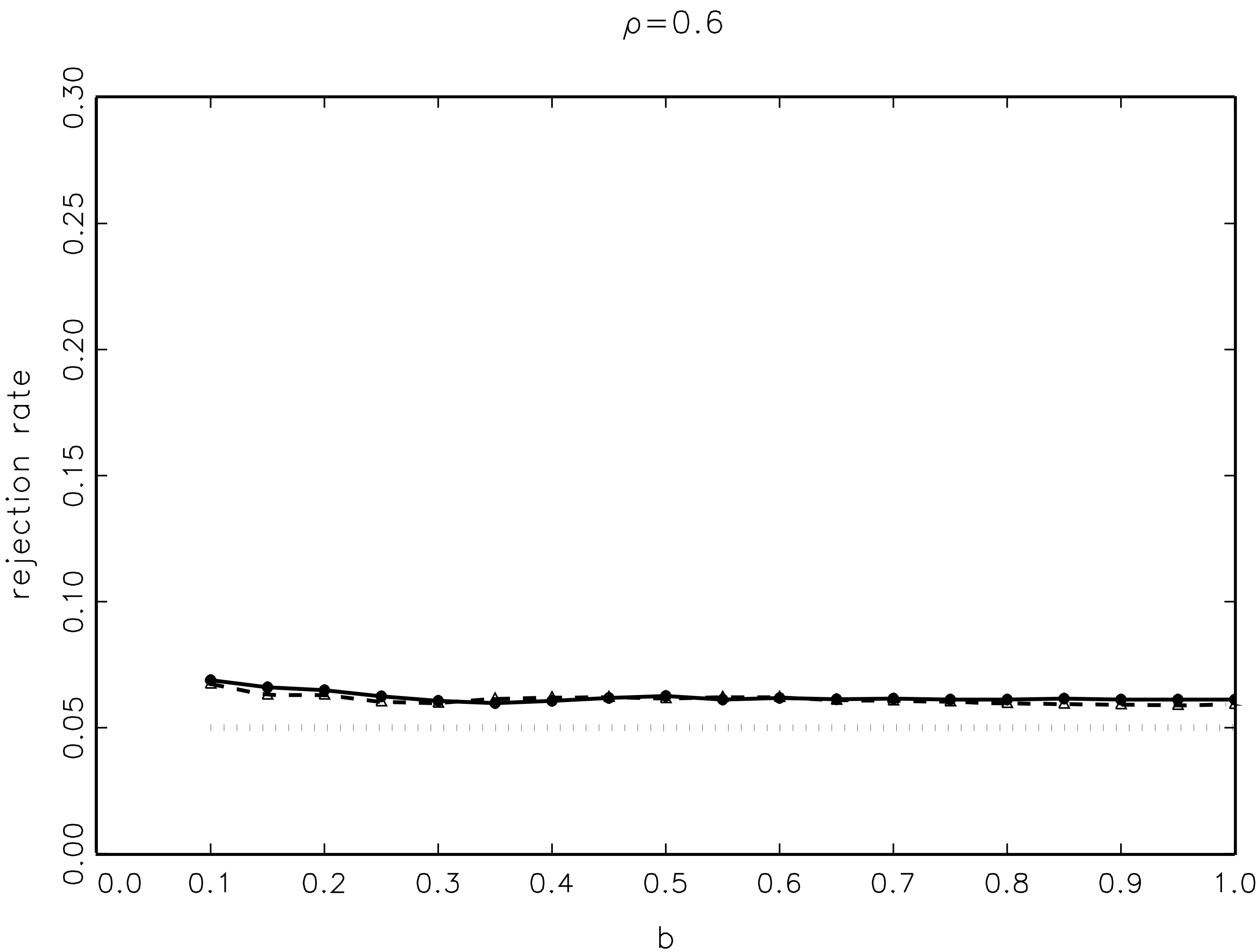


Figure 2.40: (cont'd)

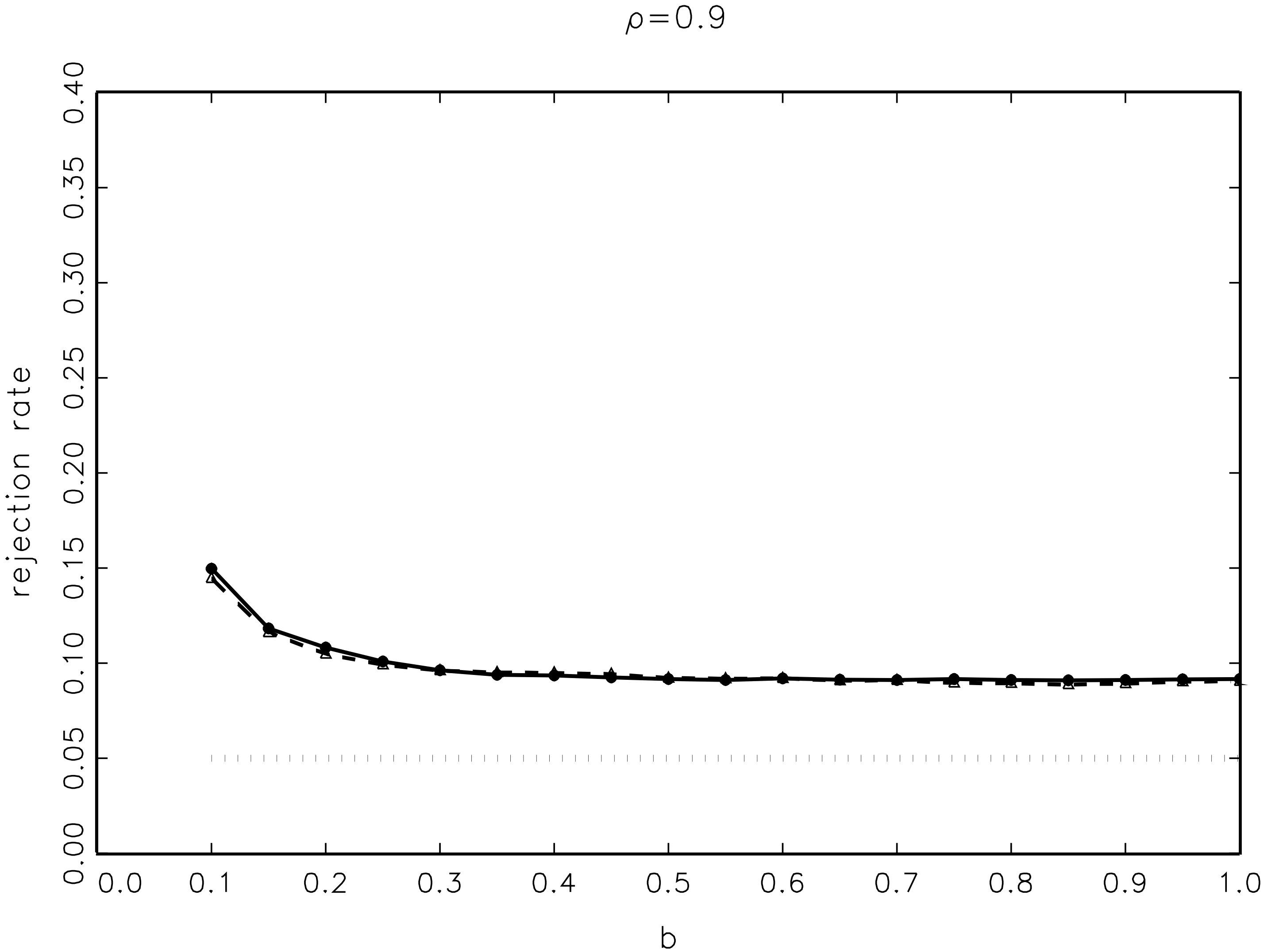


Figure 2.41: ES - Initially Scarce Data, Bartlett, $N_Q = 12$ $N_M = 12$

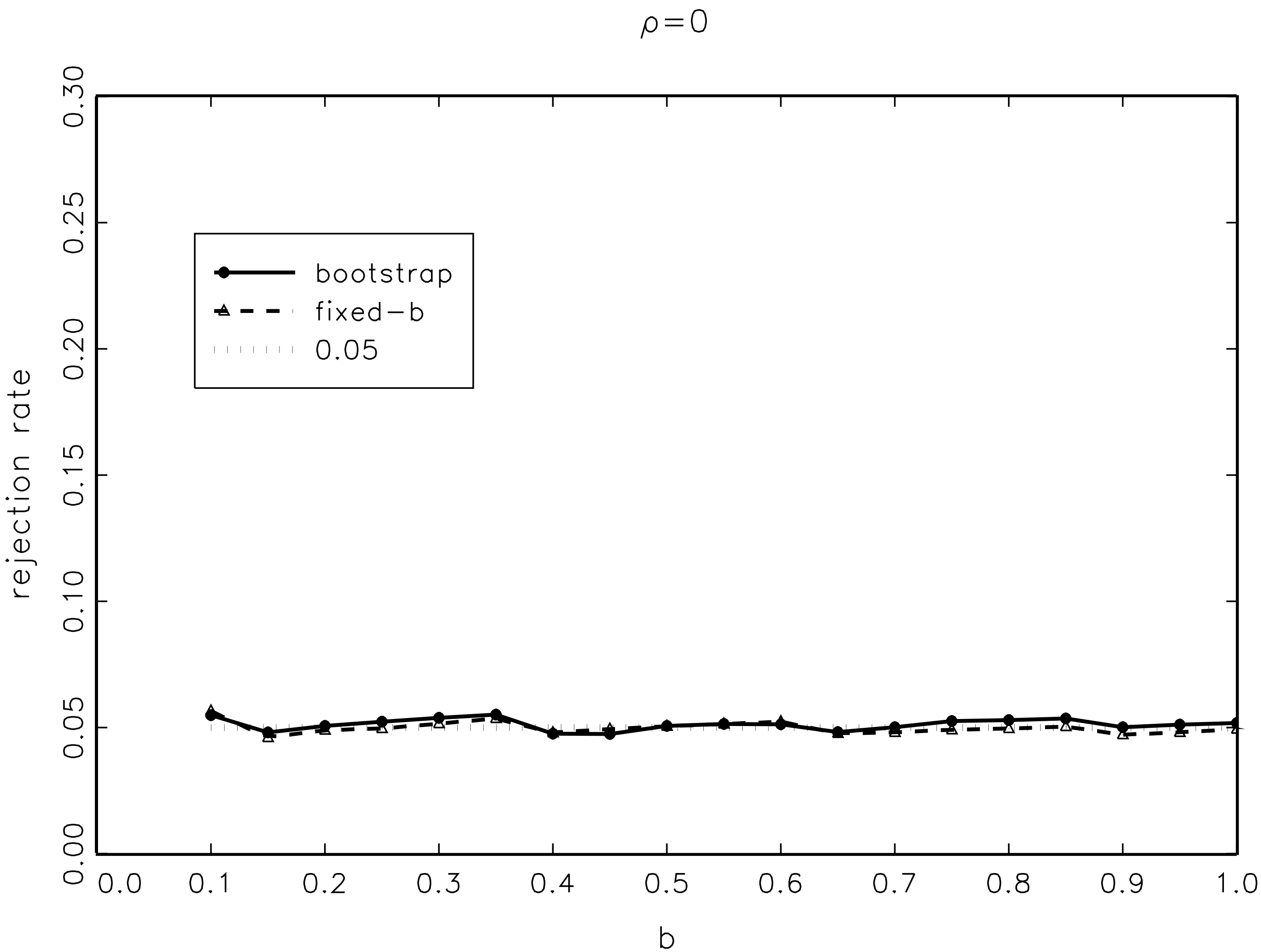


Figure 2.41: (cont'd)

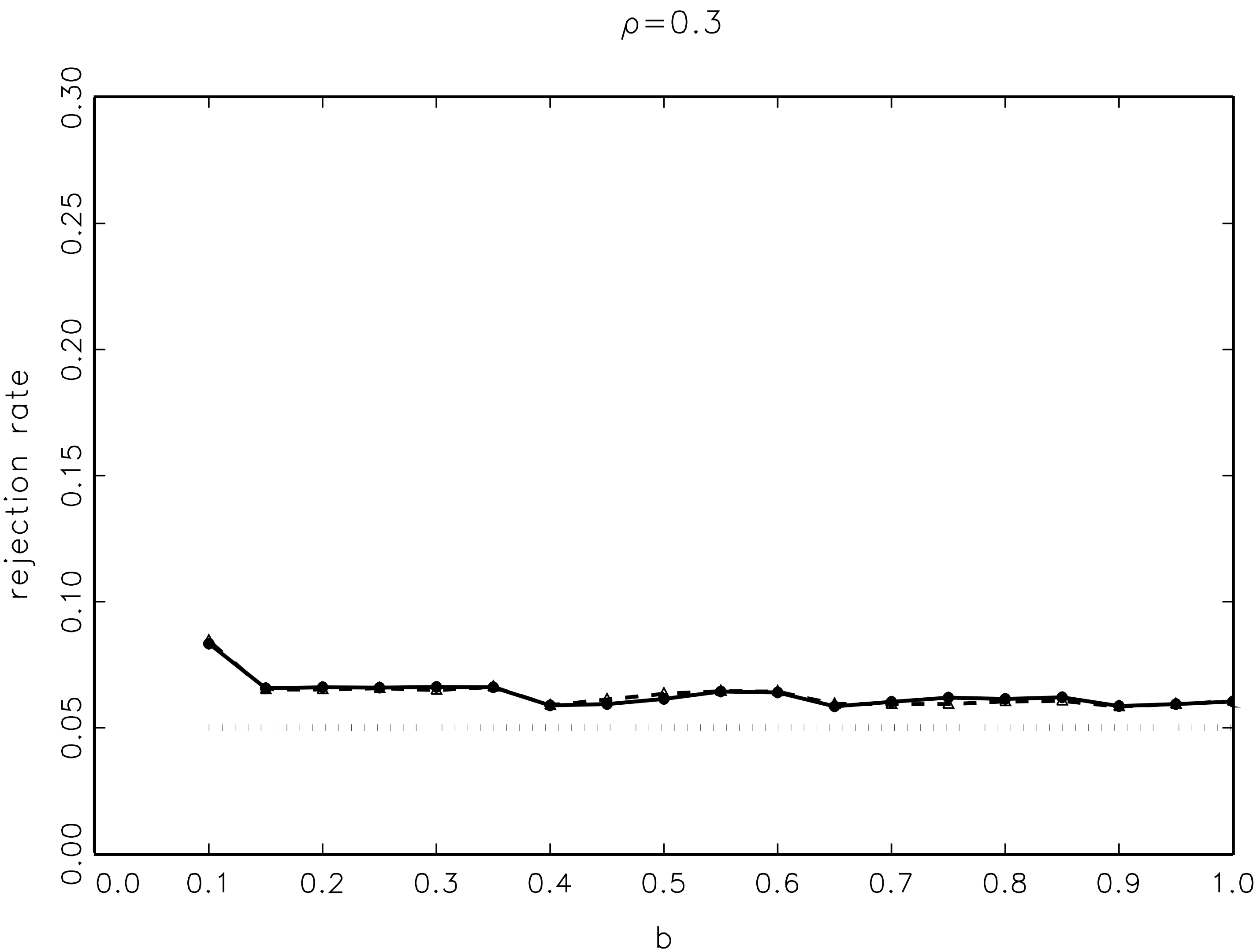


Figure 2.41: (cont'd)

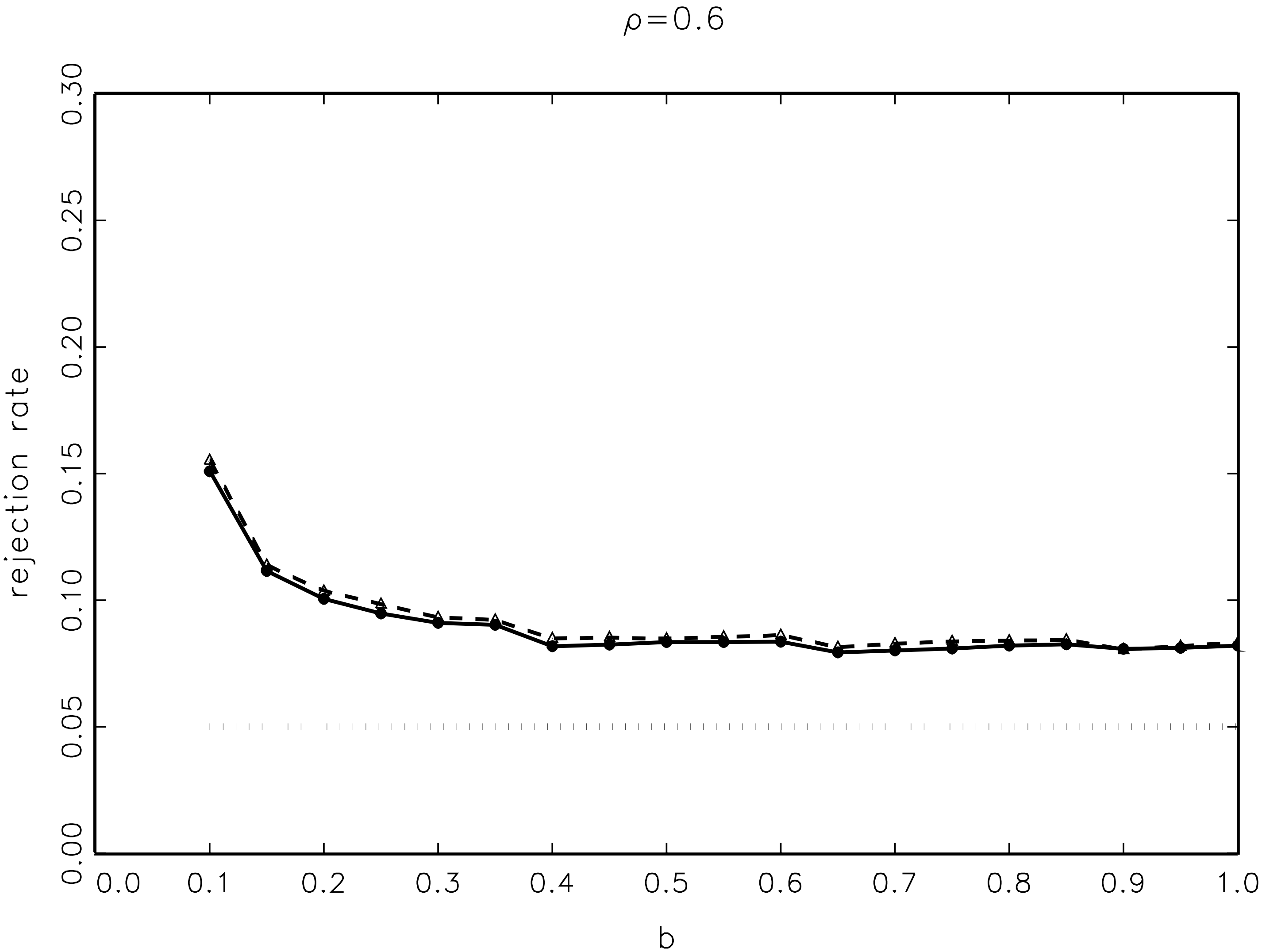


Figure 2.41: (cont'd)

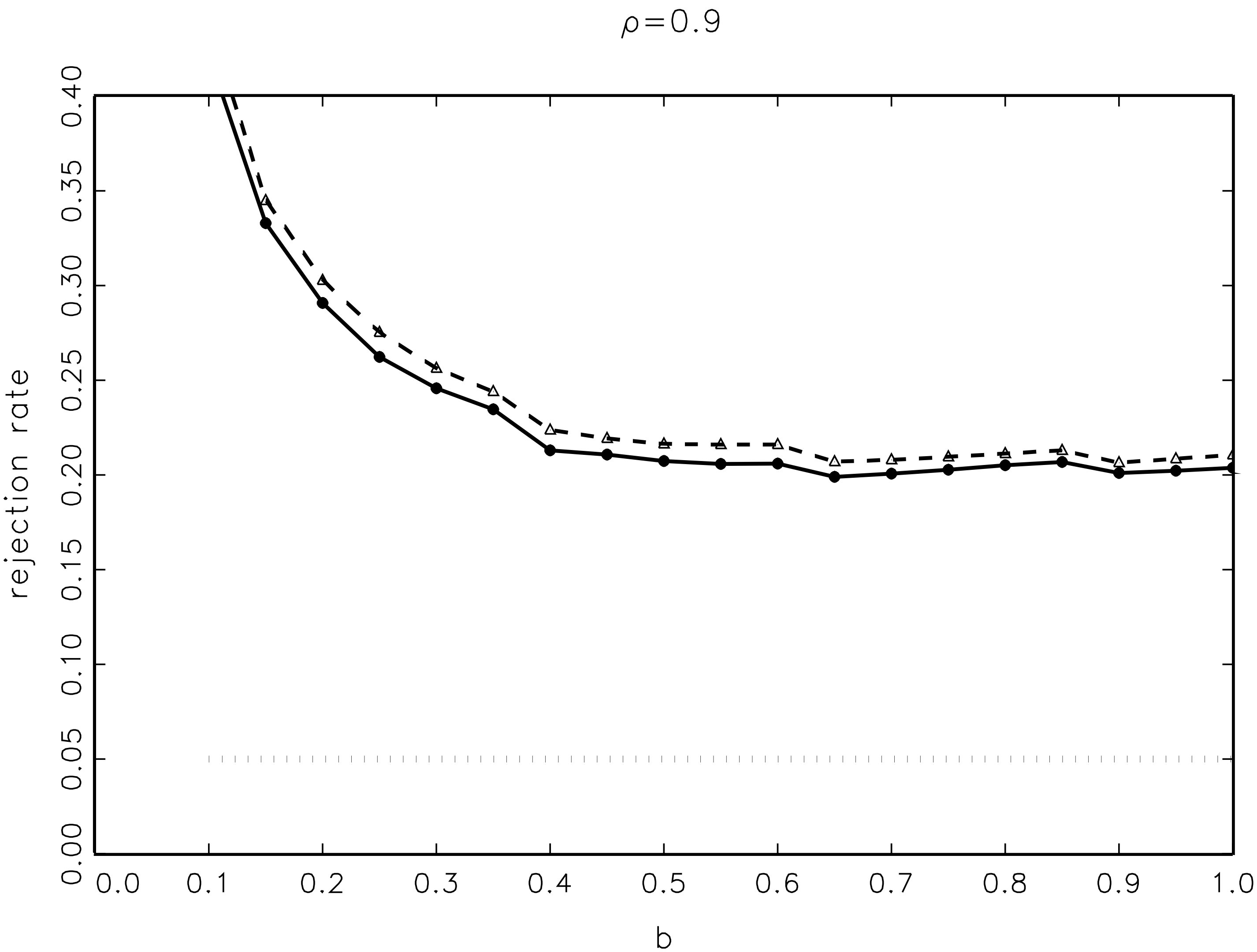


Figure 2.42: ES - Initially Scarce Data, Bartlett, $N_Q = 12$ $N_M = 24$

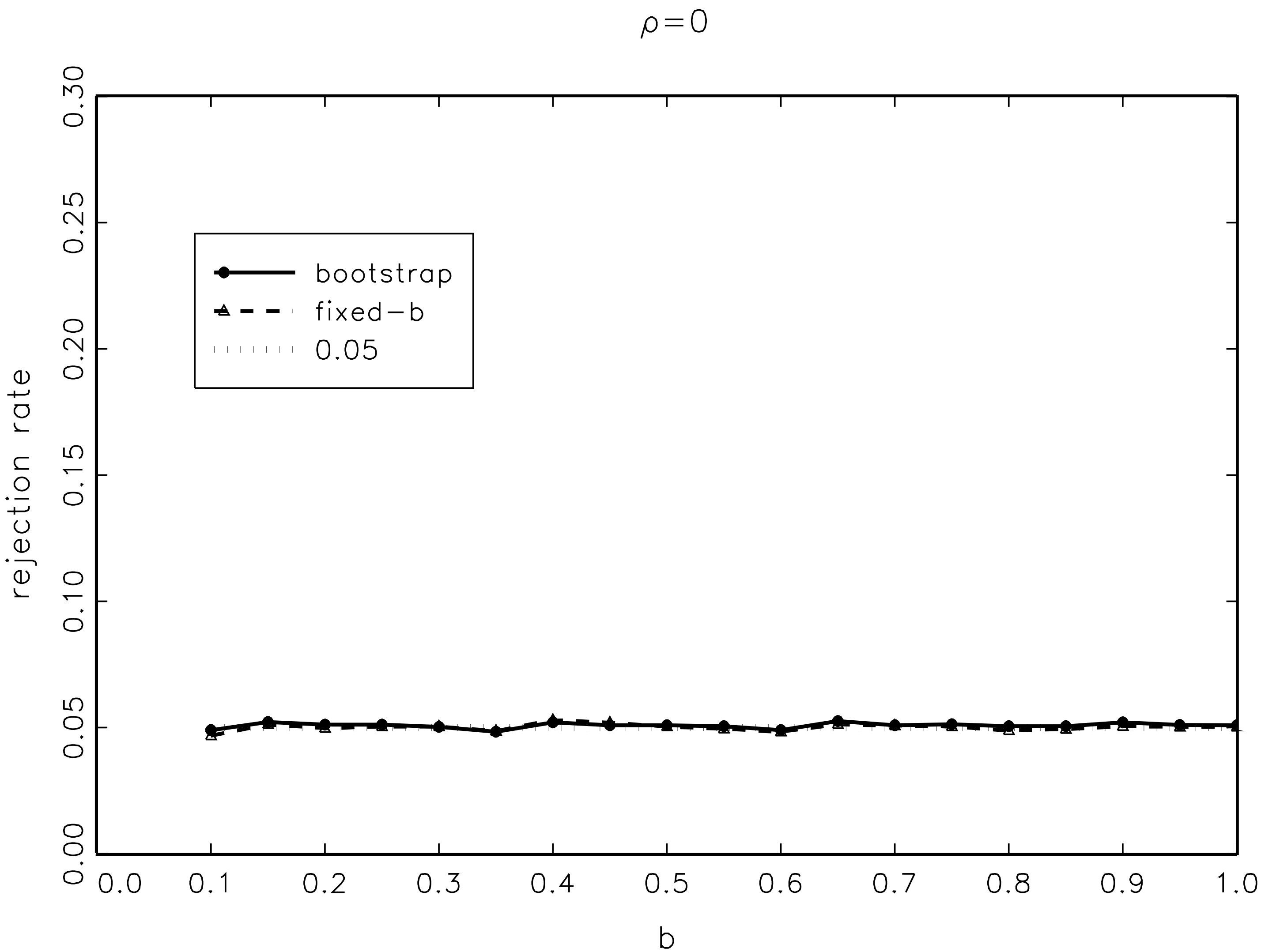


Figure 2.42: (cont'd)

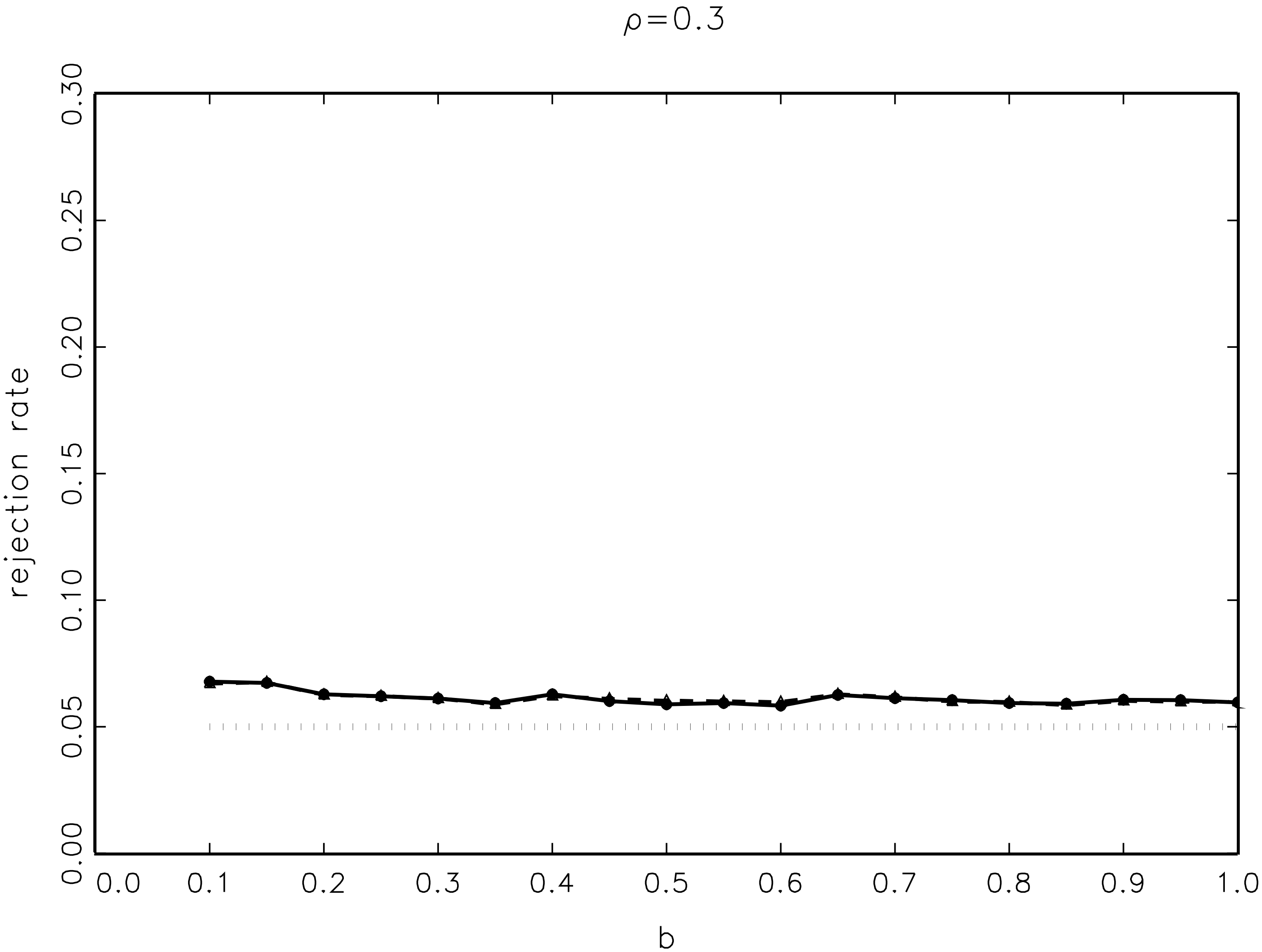


Figure 2.42: (cont'd)

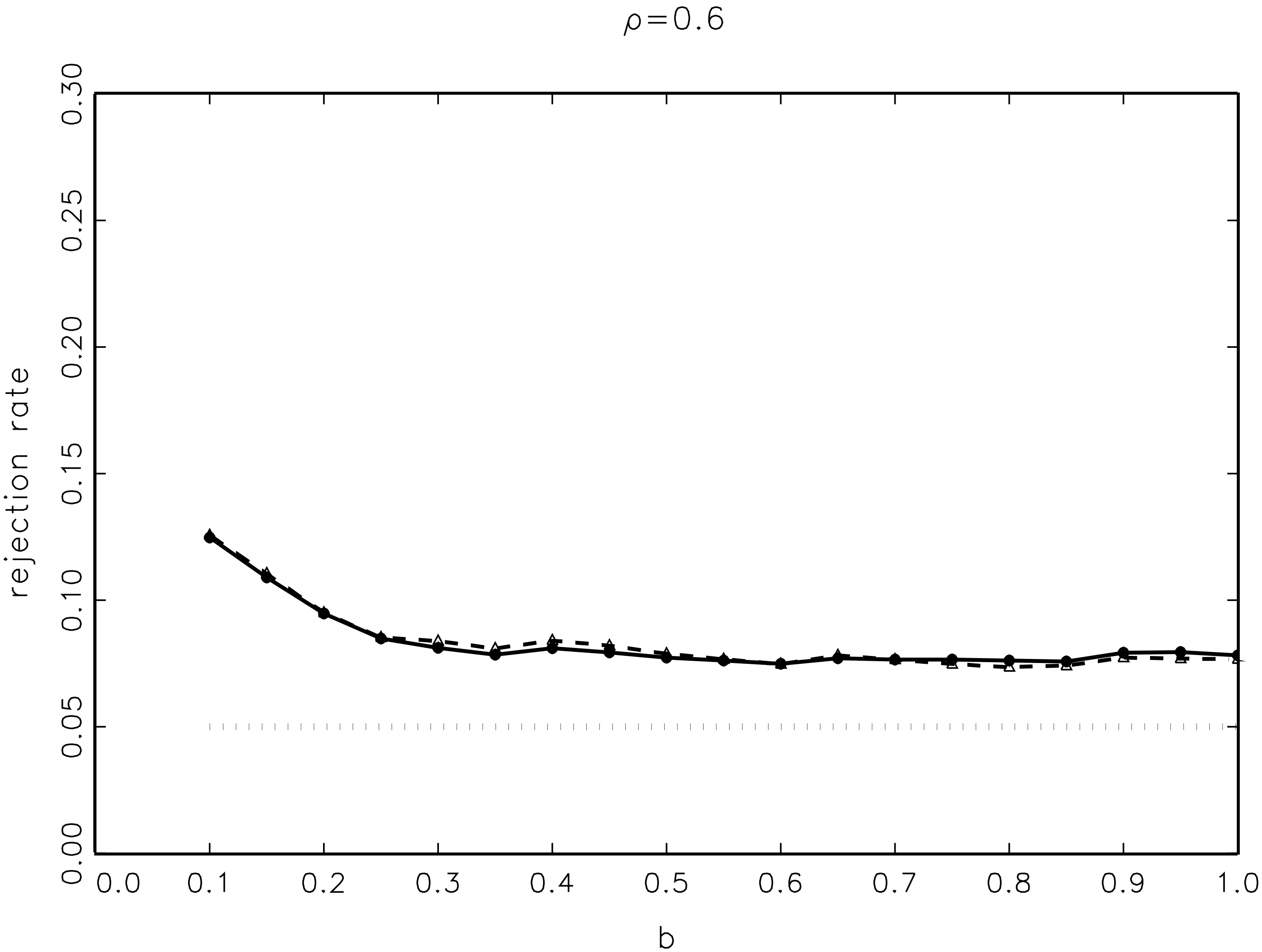


Figure 2.42: (cont'd)

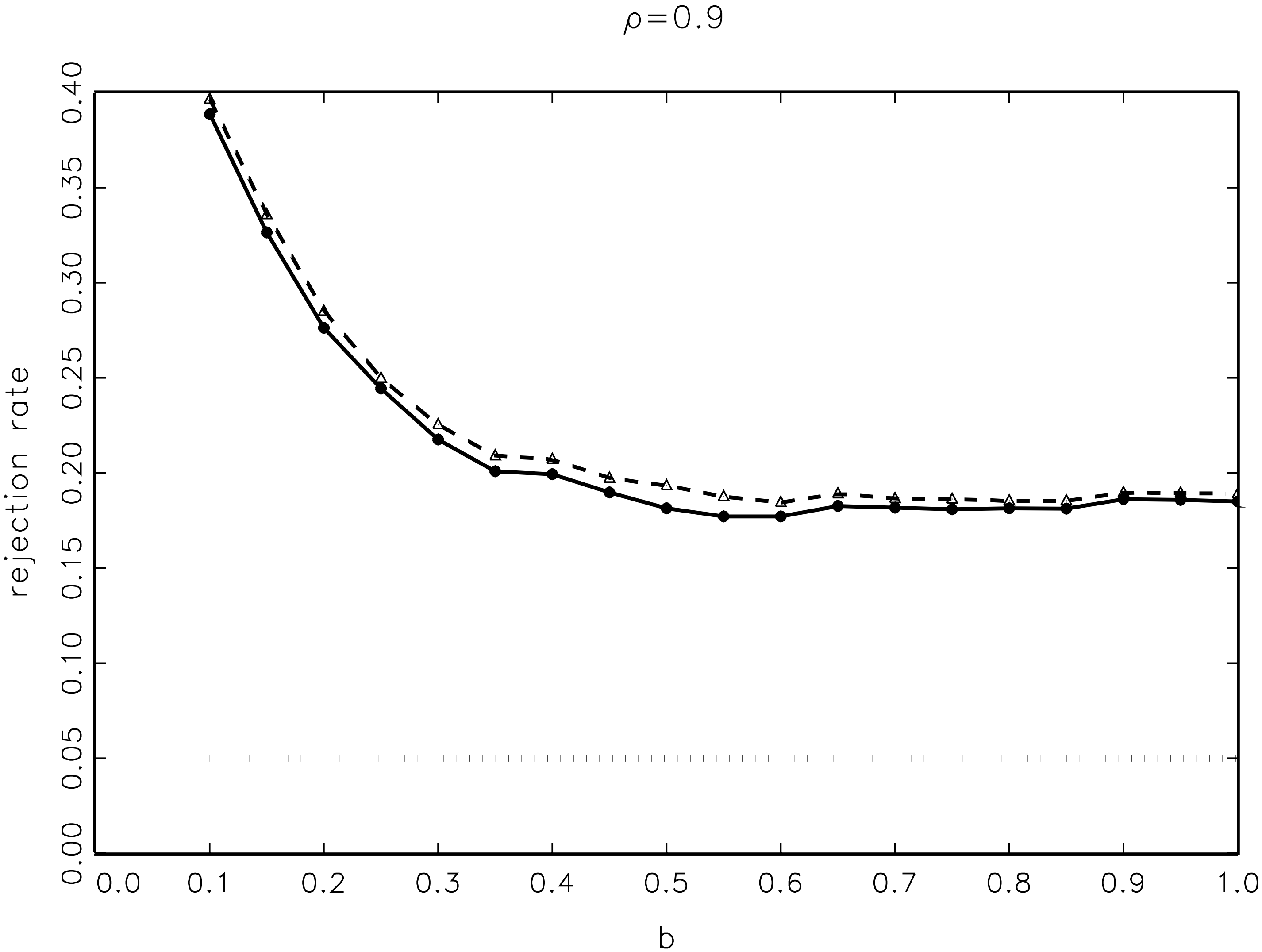


Figure 2.43: ES - Initially Scarce Data, Bartlett, $N_Q = 12$ $N_M = 48$

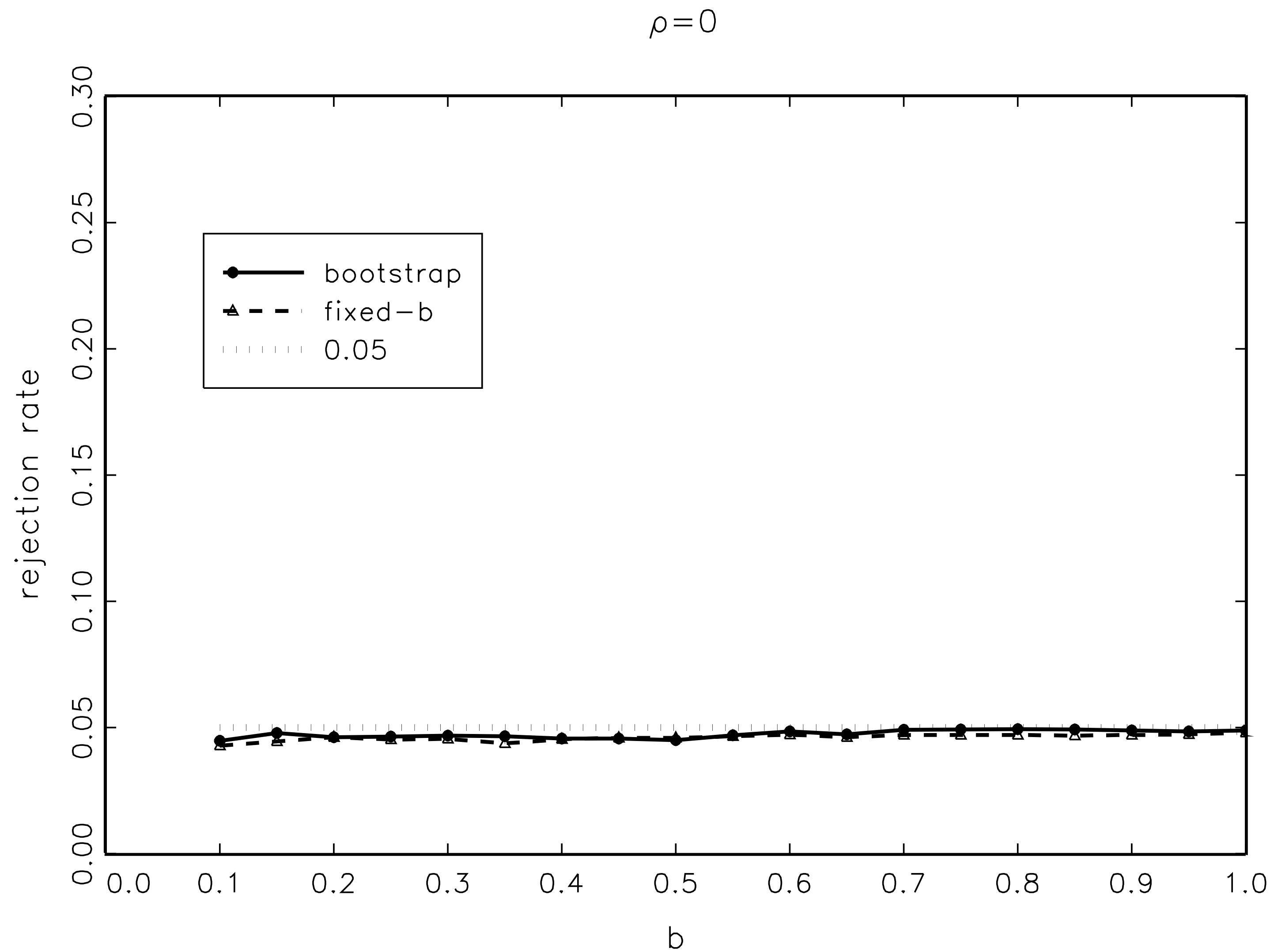


Figure 2.43: (cont'd)

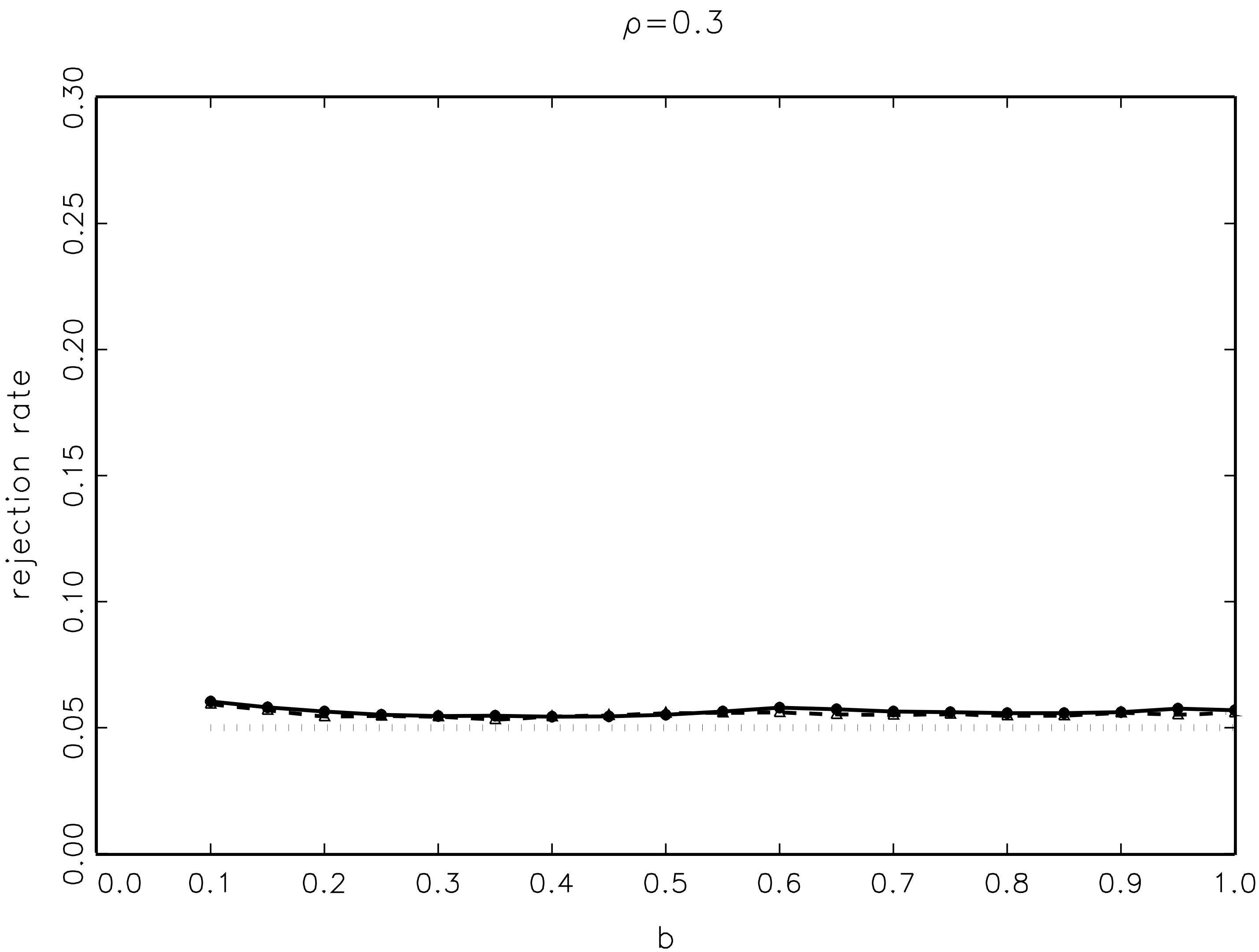


Figure 2.43: (cont'd)

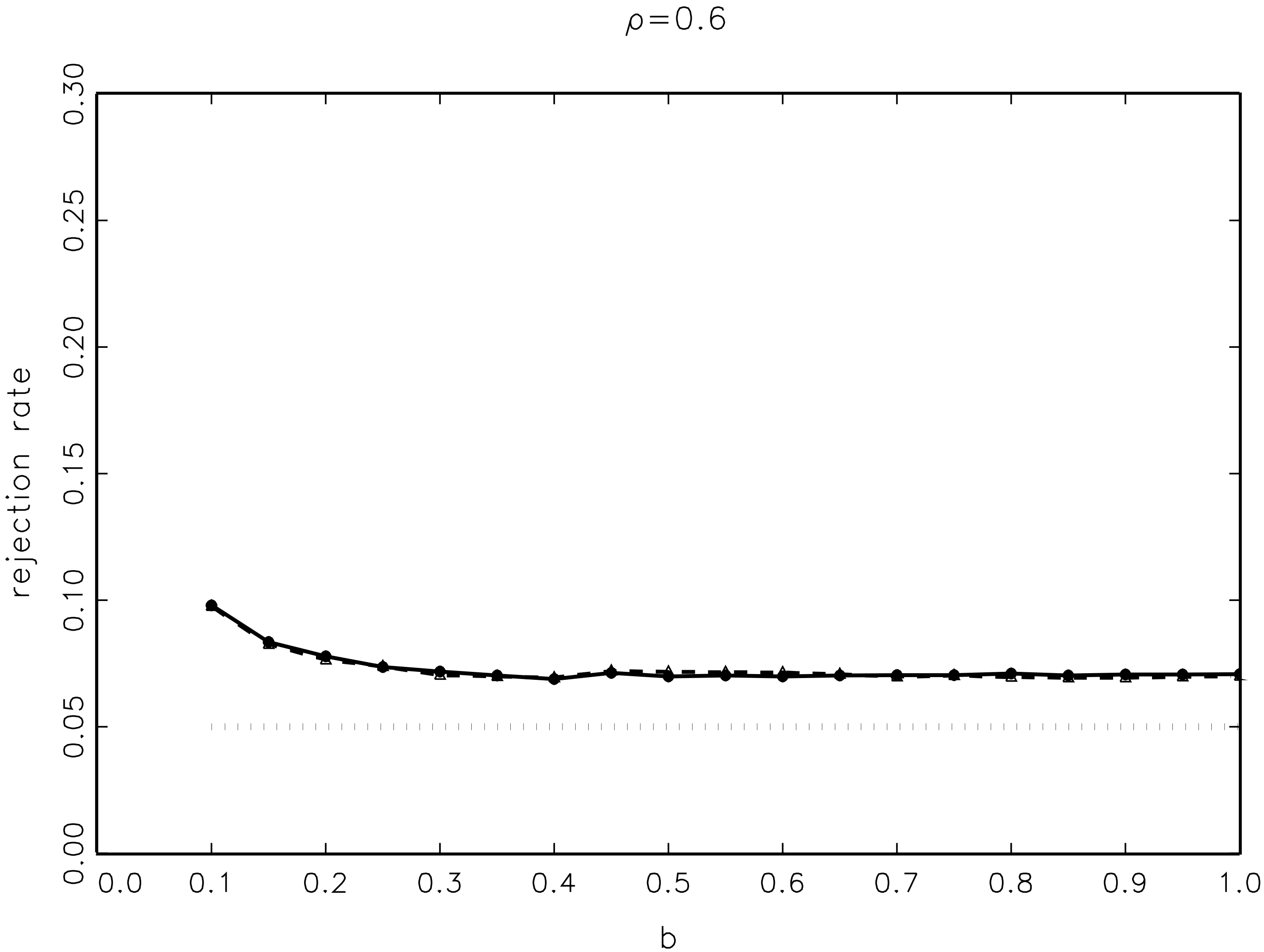


Figure 2.43: (cont'd)

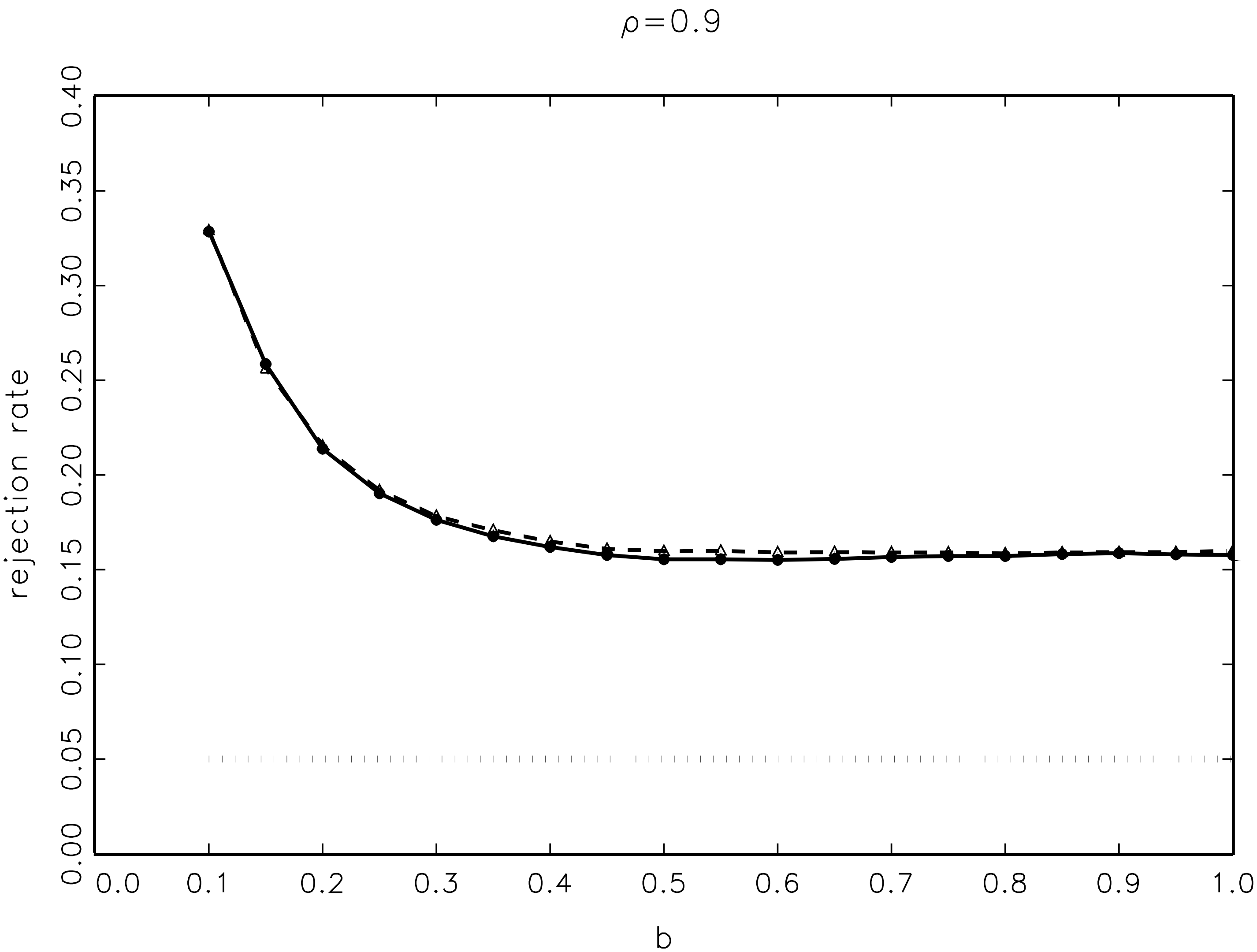


Figure 2.44: ES - Initially Scarce Data, Bartlett, $N_Q = 24$ $N_M = 12$

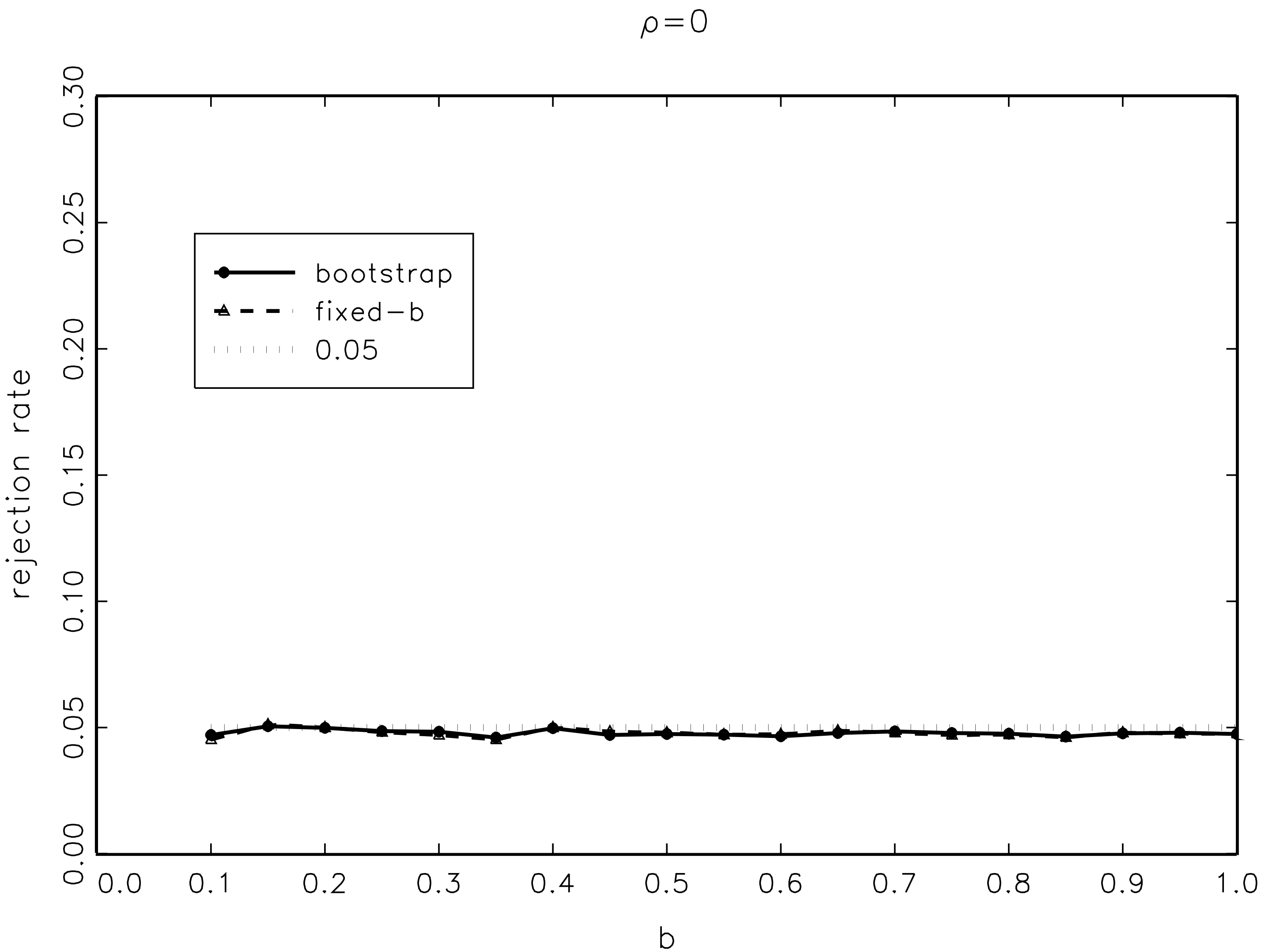


Figure 2.44: (cont'd)

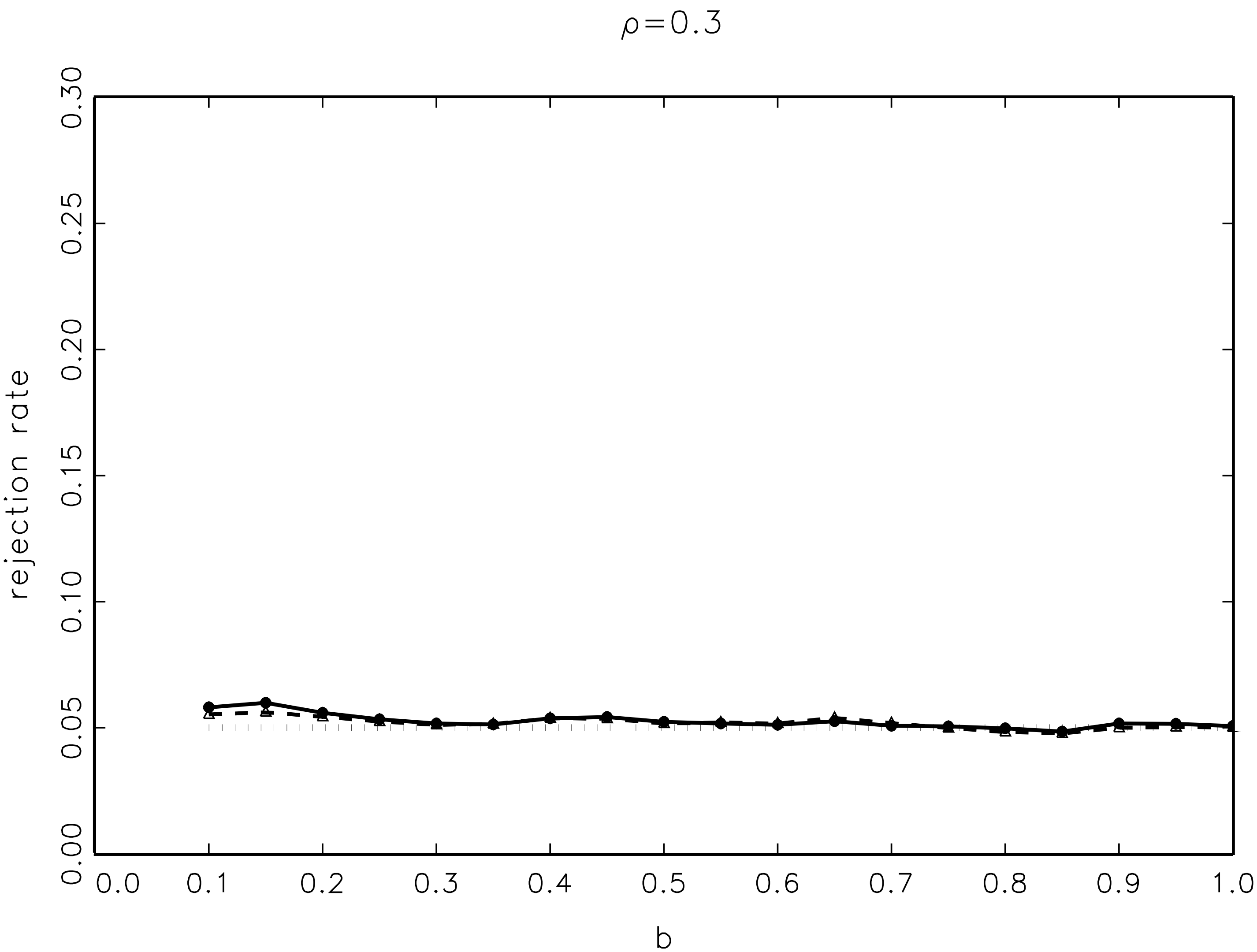


Figure 2.44: (cont'd)

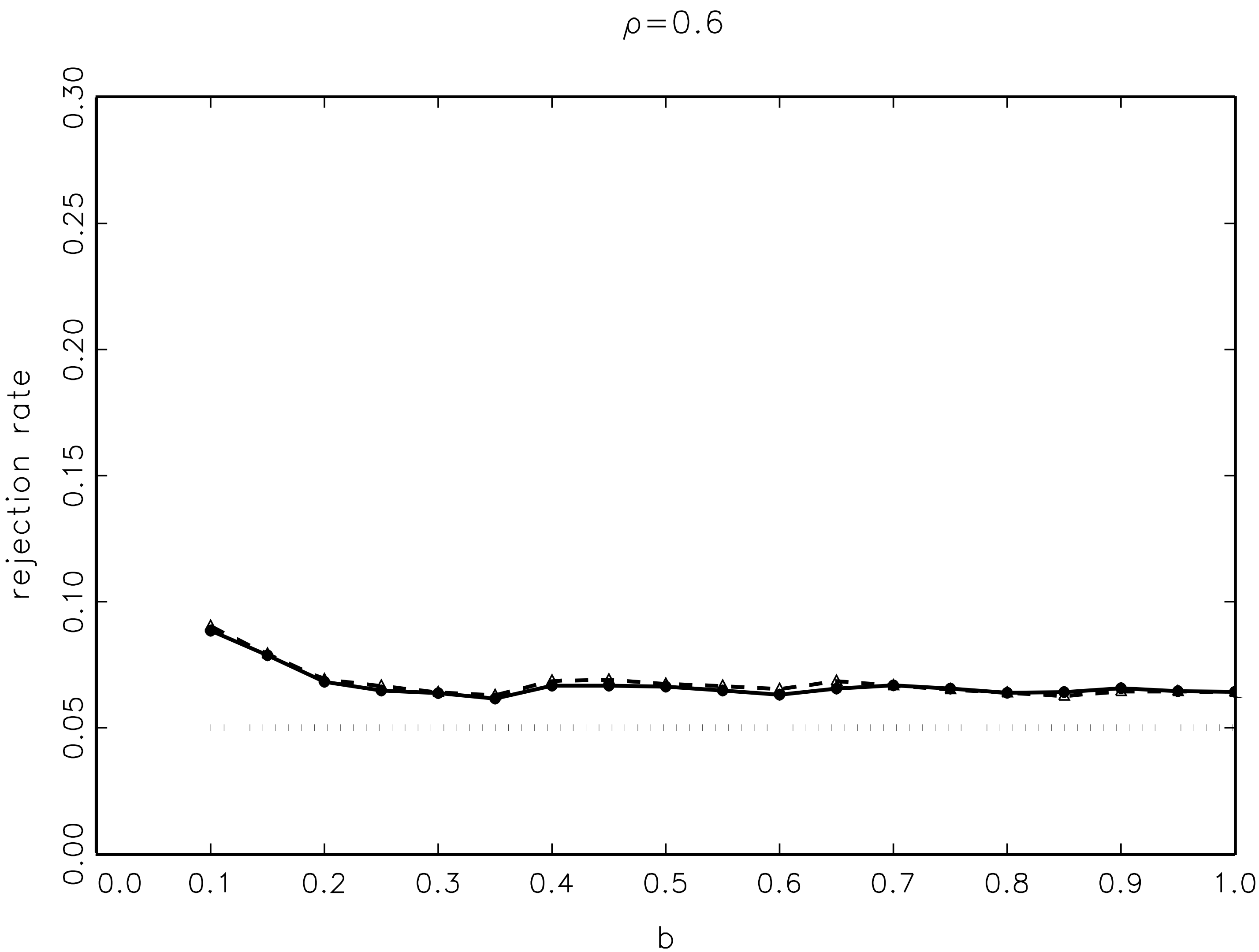


Figure 2.44: (cont'd)

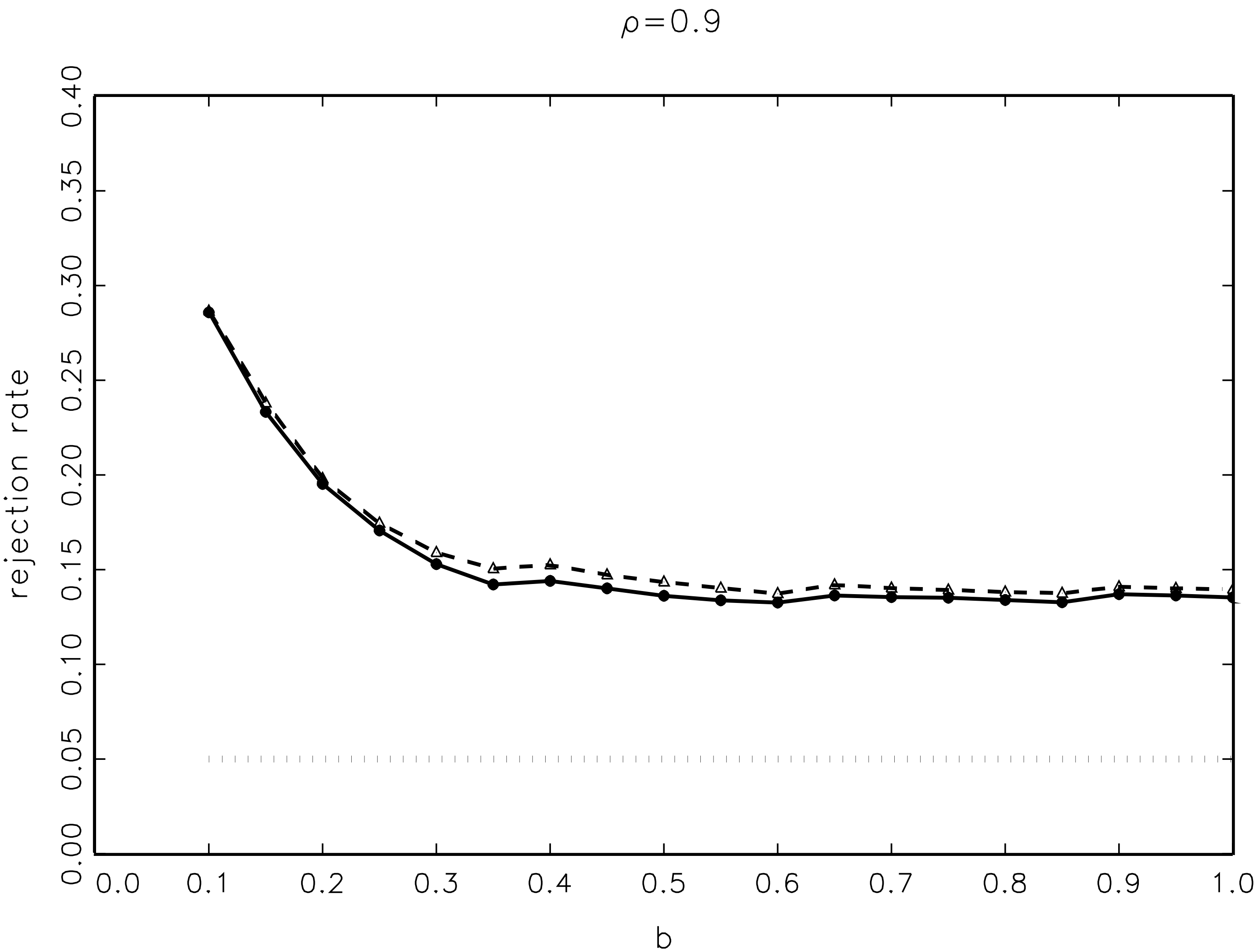


Figure 2.45: ES - Initially Scarce Data, Bartlett, $N_Q = 24$ $N_M = 24$

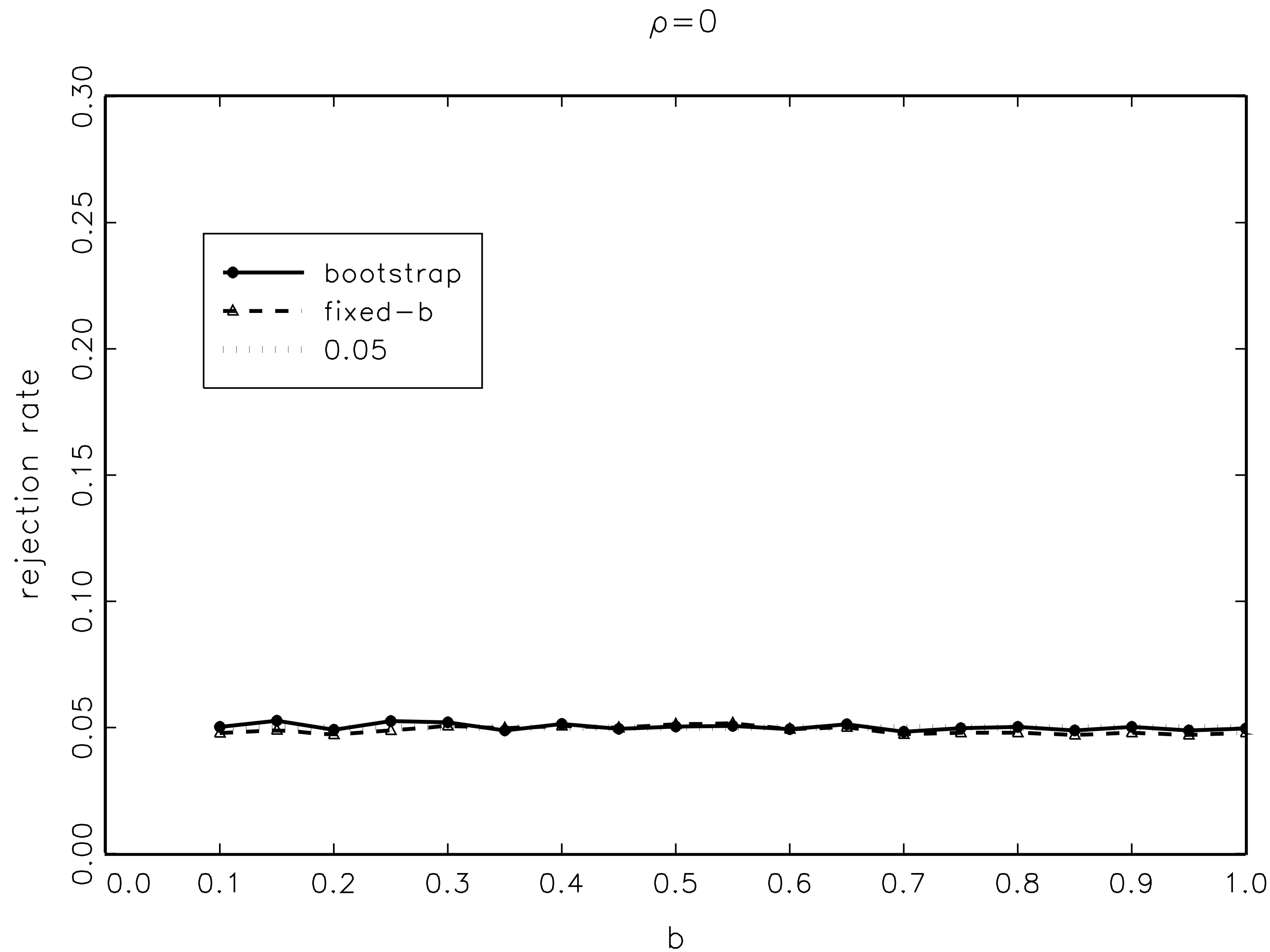


Figure 2.45: (cont'd)

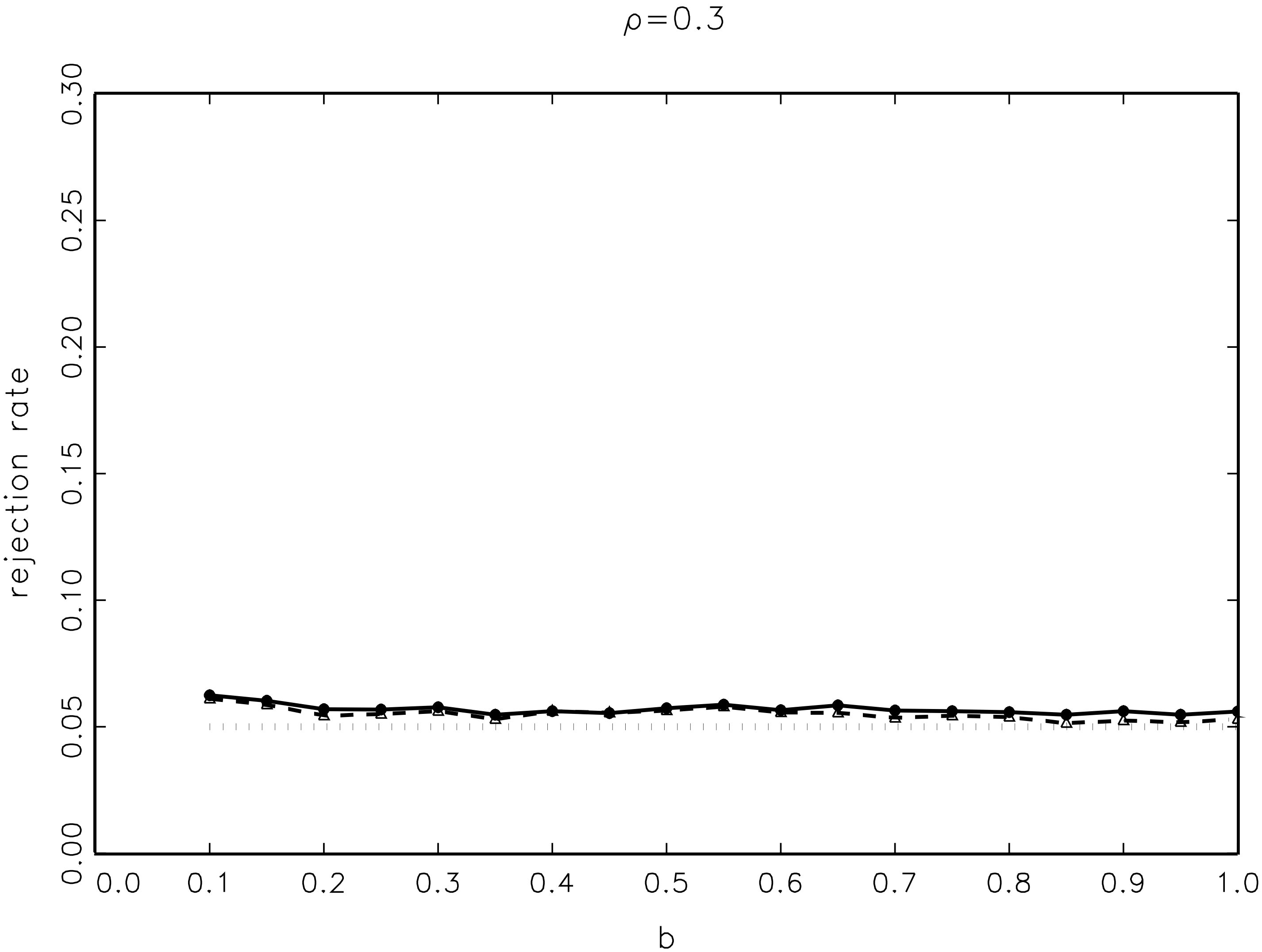


Figure 2.45: (cont'd)

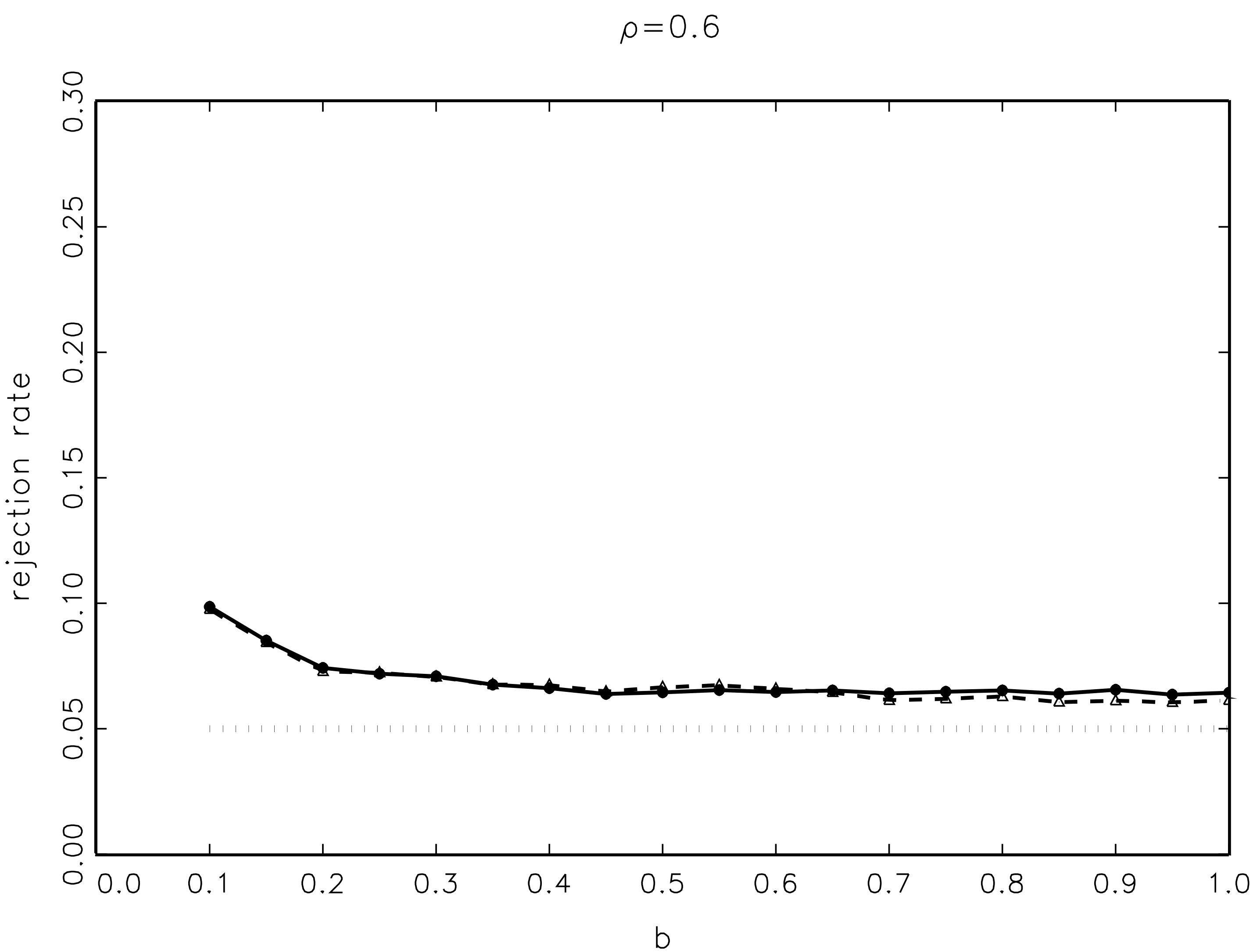


Figure 2.45: (cont'd)

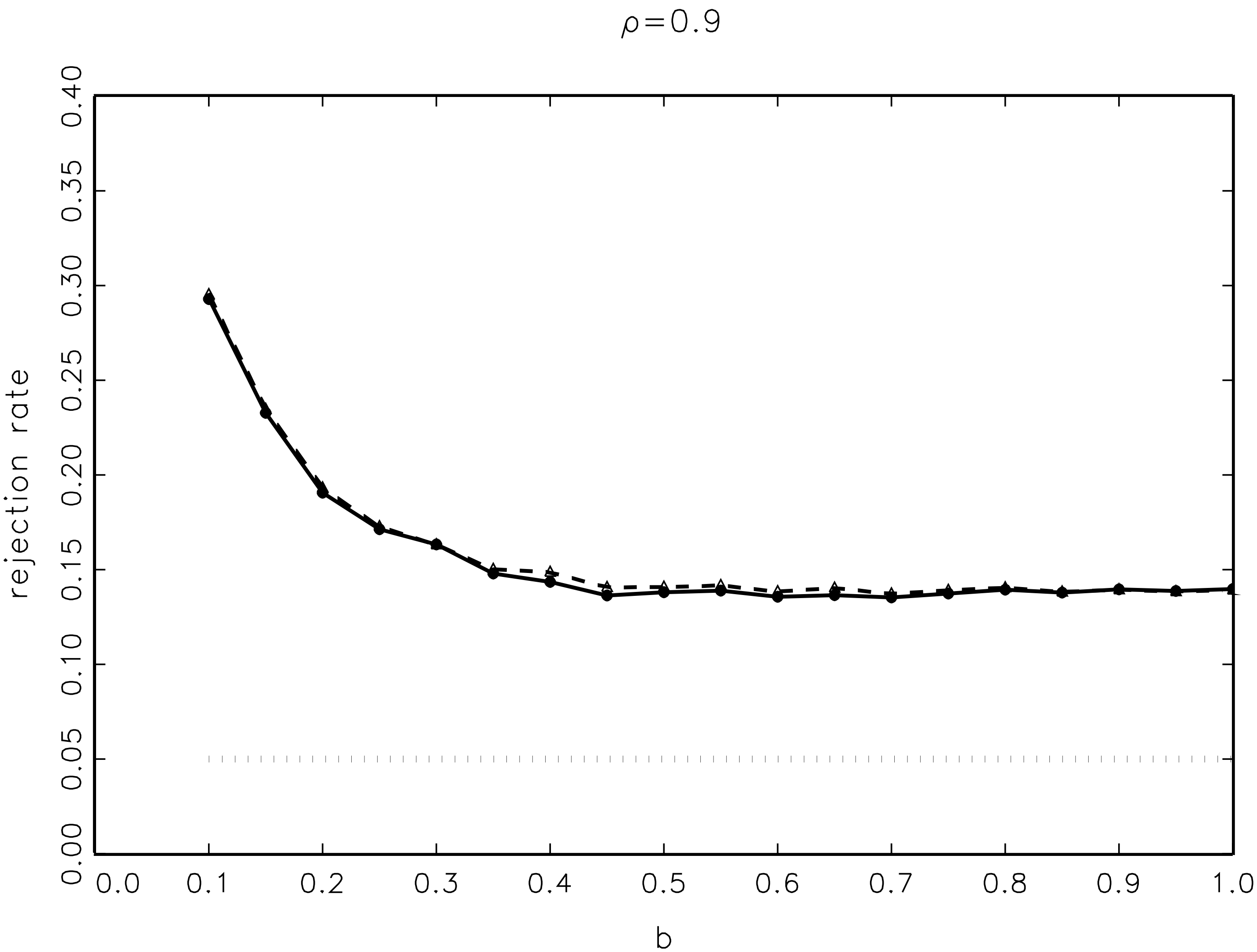


Figure 2.46: ES - Initially Scarce Data, Bartlett, $N_Q = 24$ $N_M = 48$

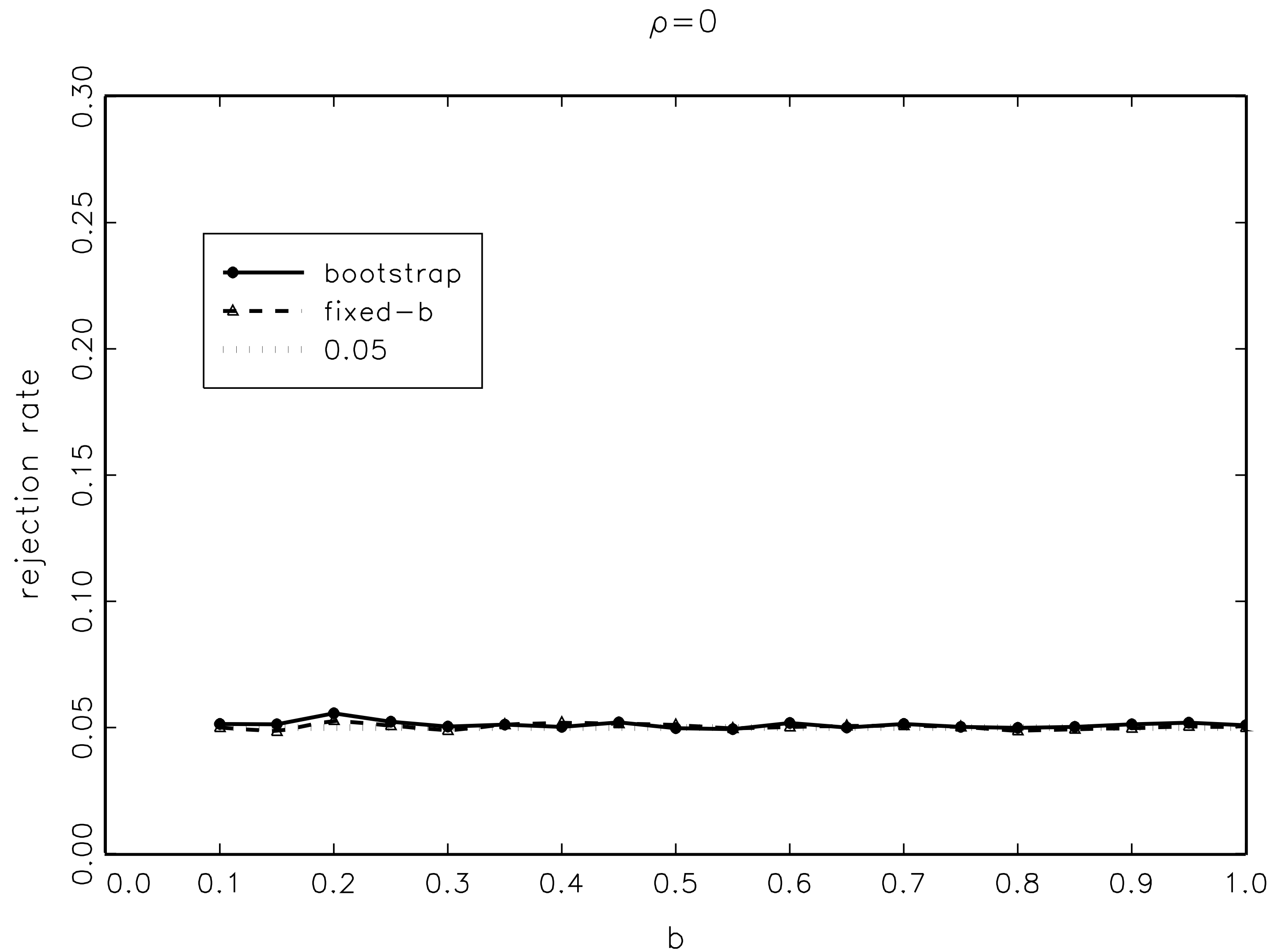


Figure 2.46: (cont'd)

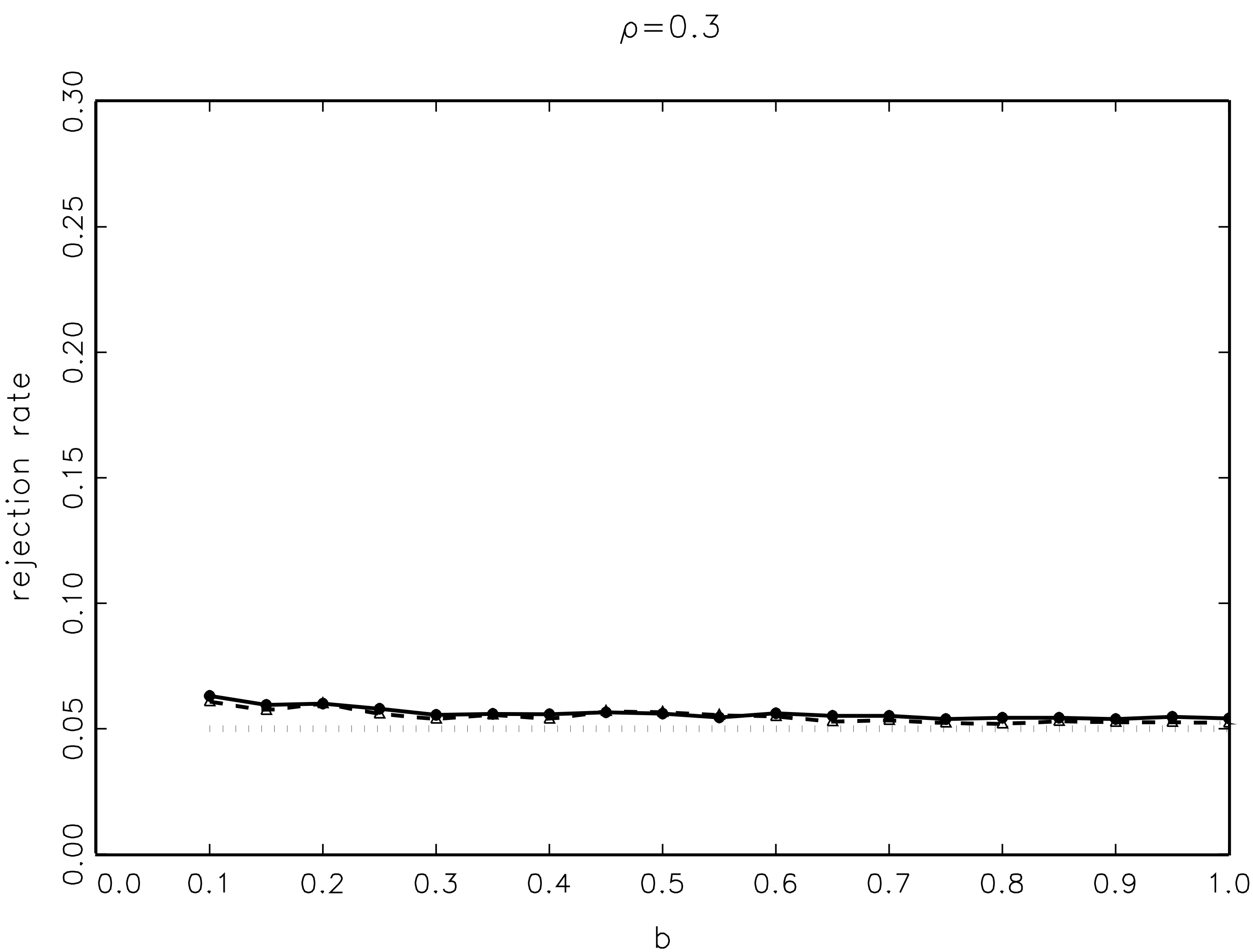


Figure 2.46: (cont'd)

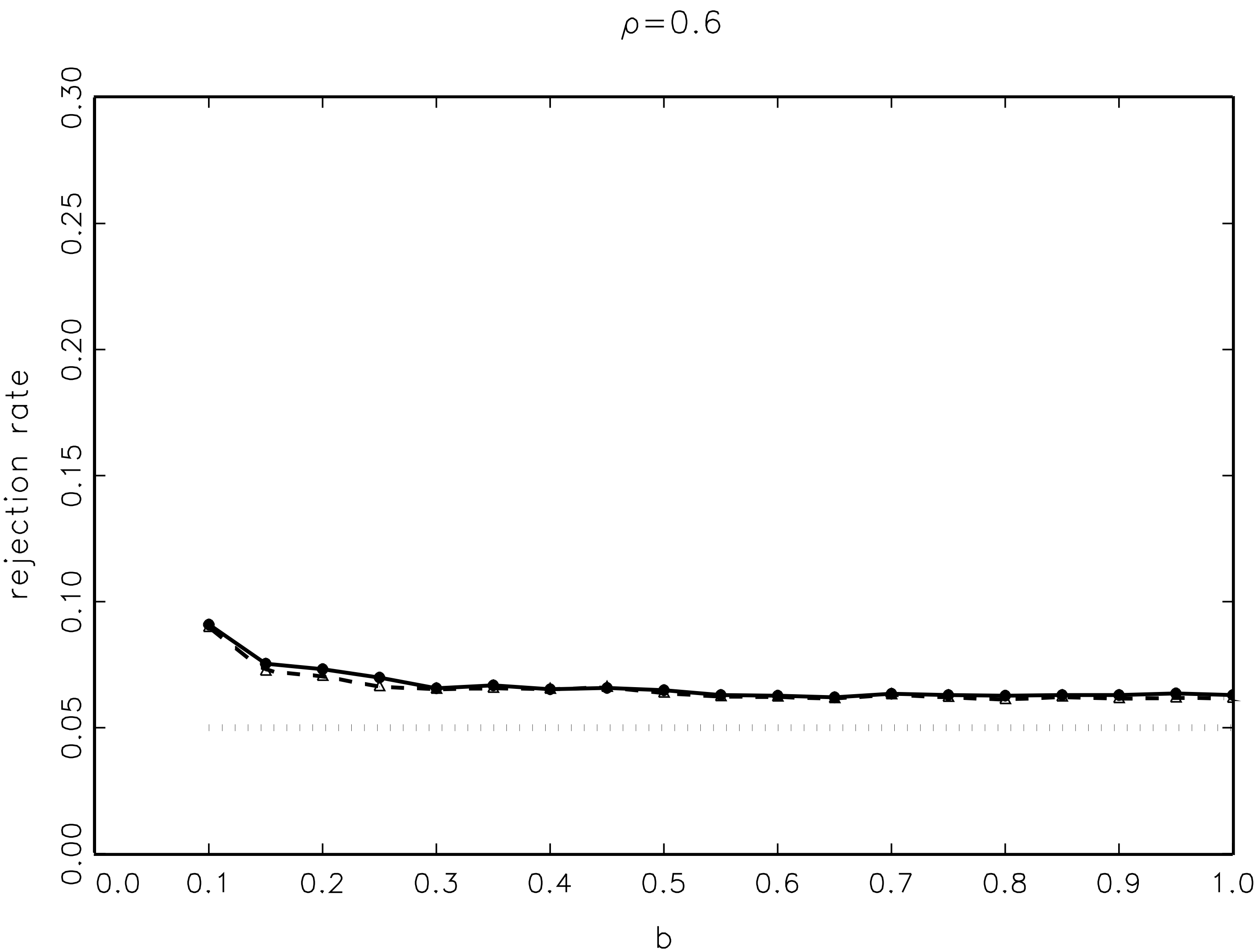


Figure 2.46: (cont'd)

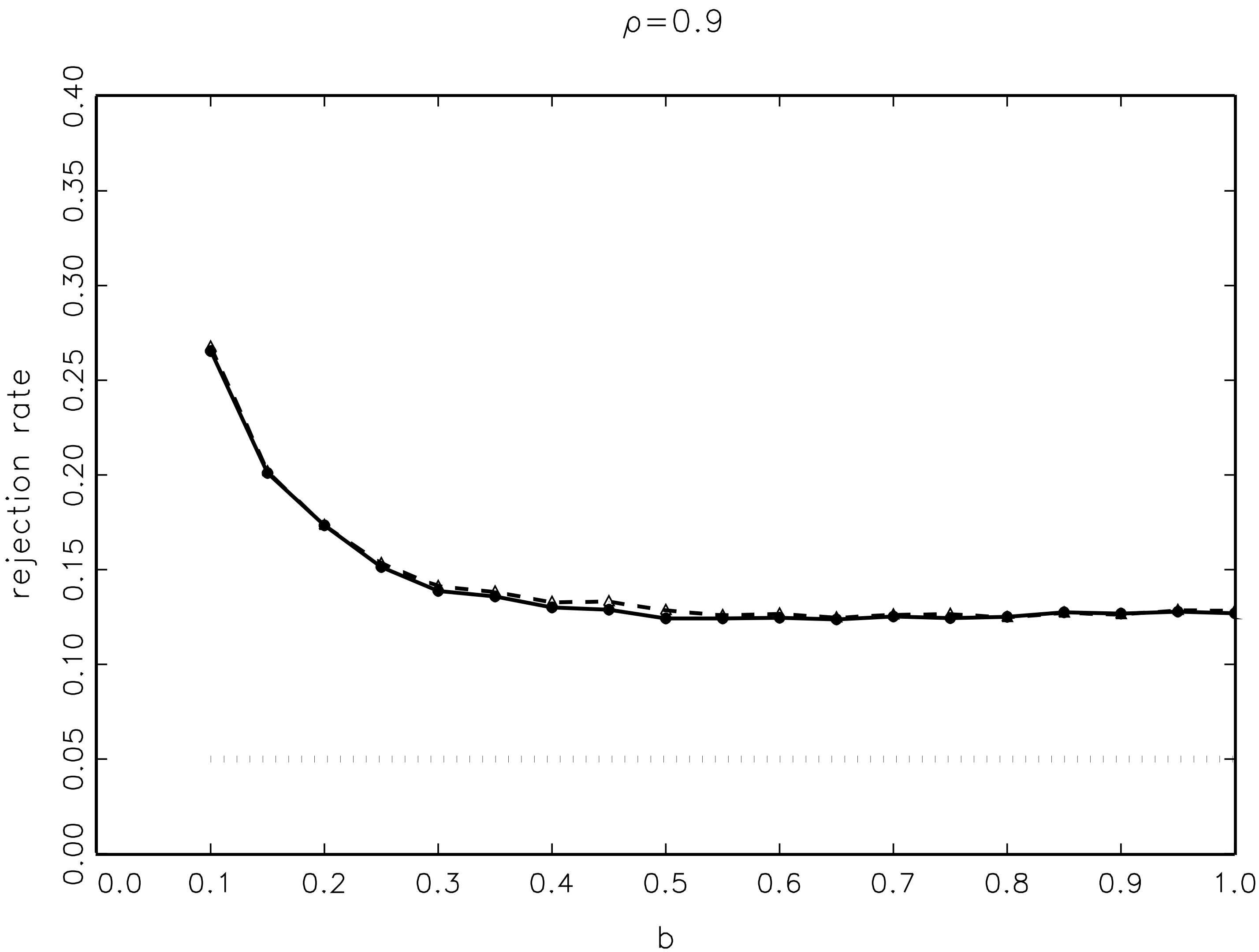


Figure 2.47: ES - Conditional Bernoulli ($p = 0.3$), Bartlett, $T = 50$

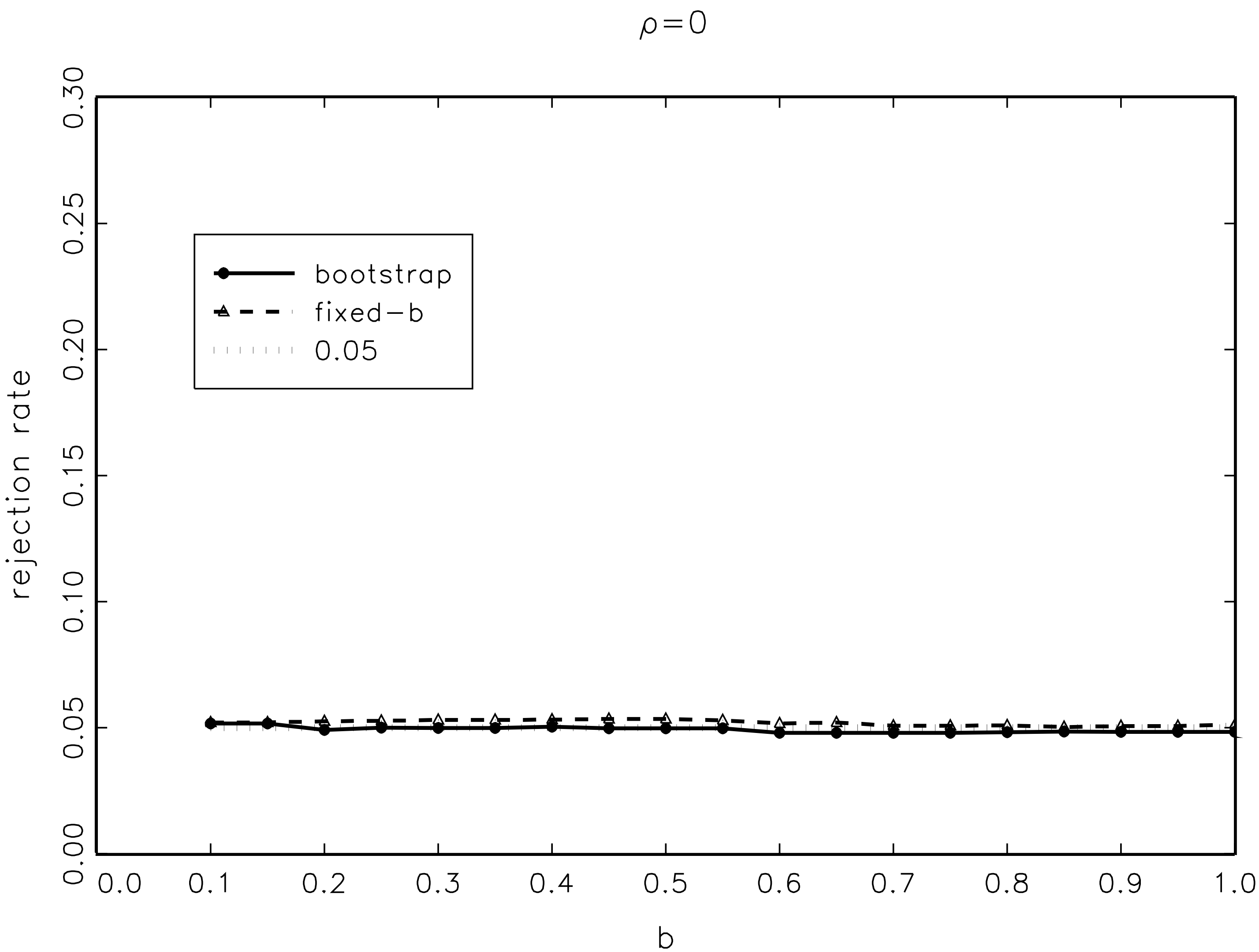


Figure 2.47: (cont'd)

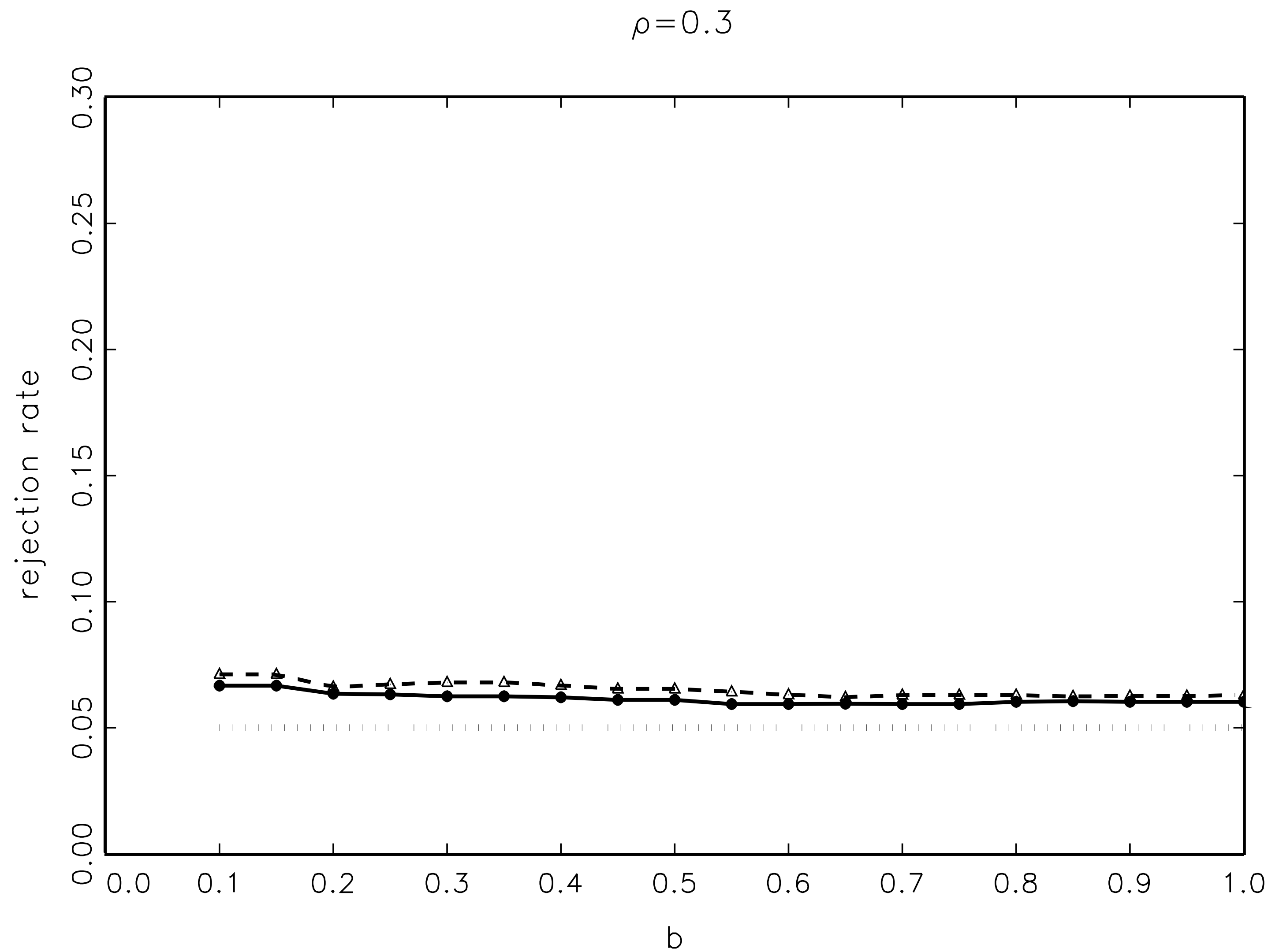


Figure 2.47: (cont'd)

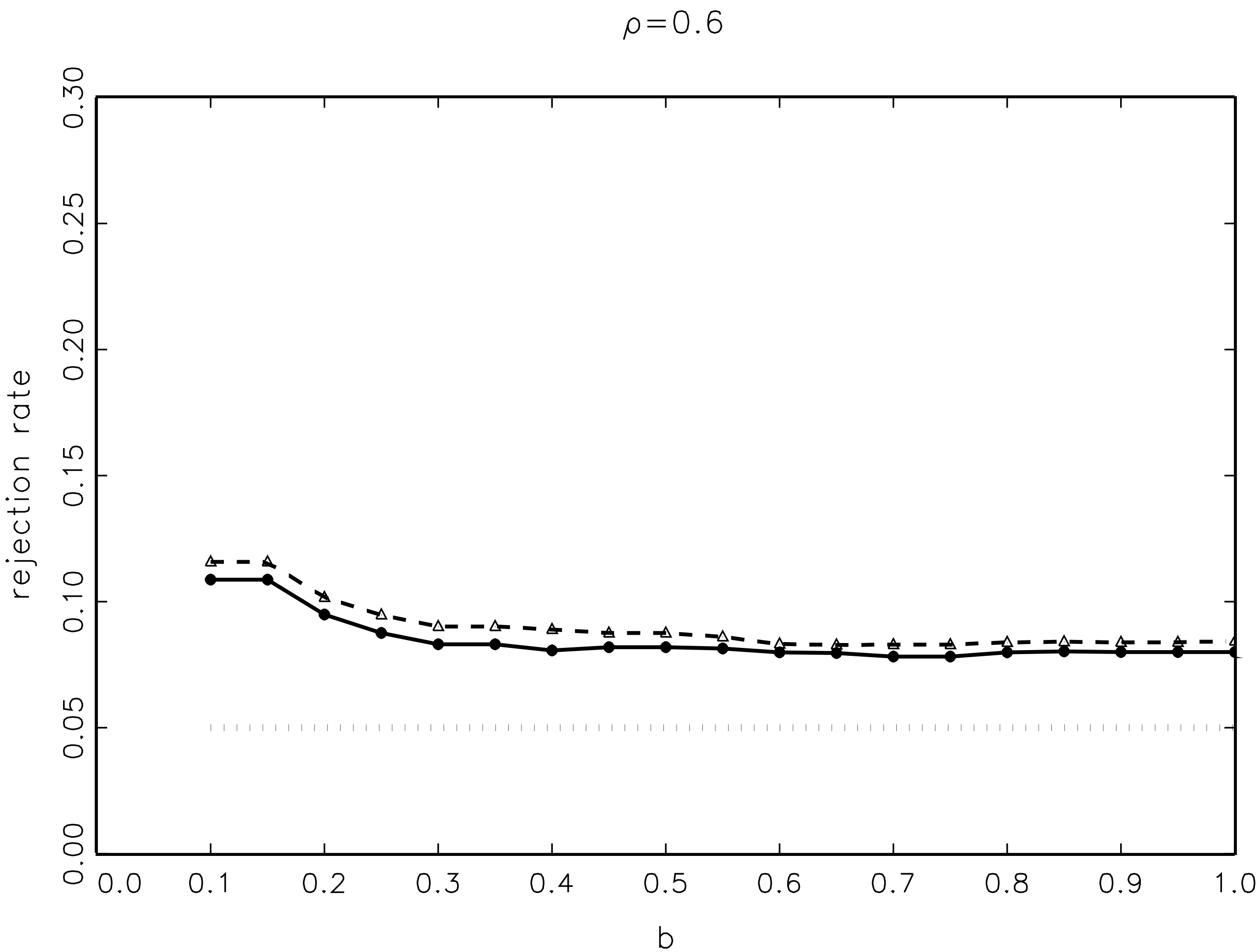


Figure 2.47: (cont'd)

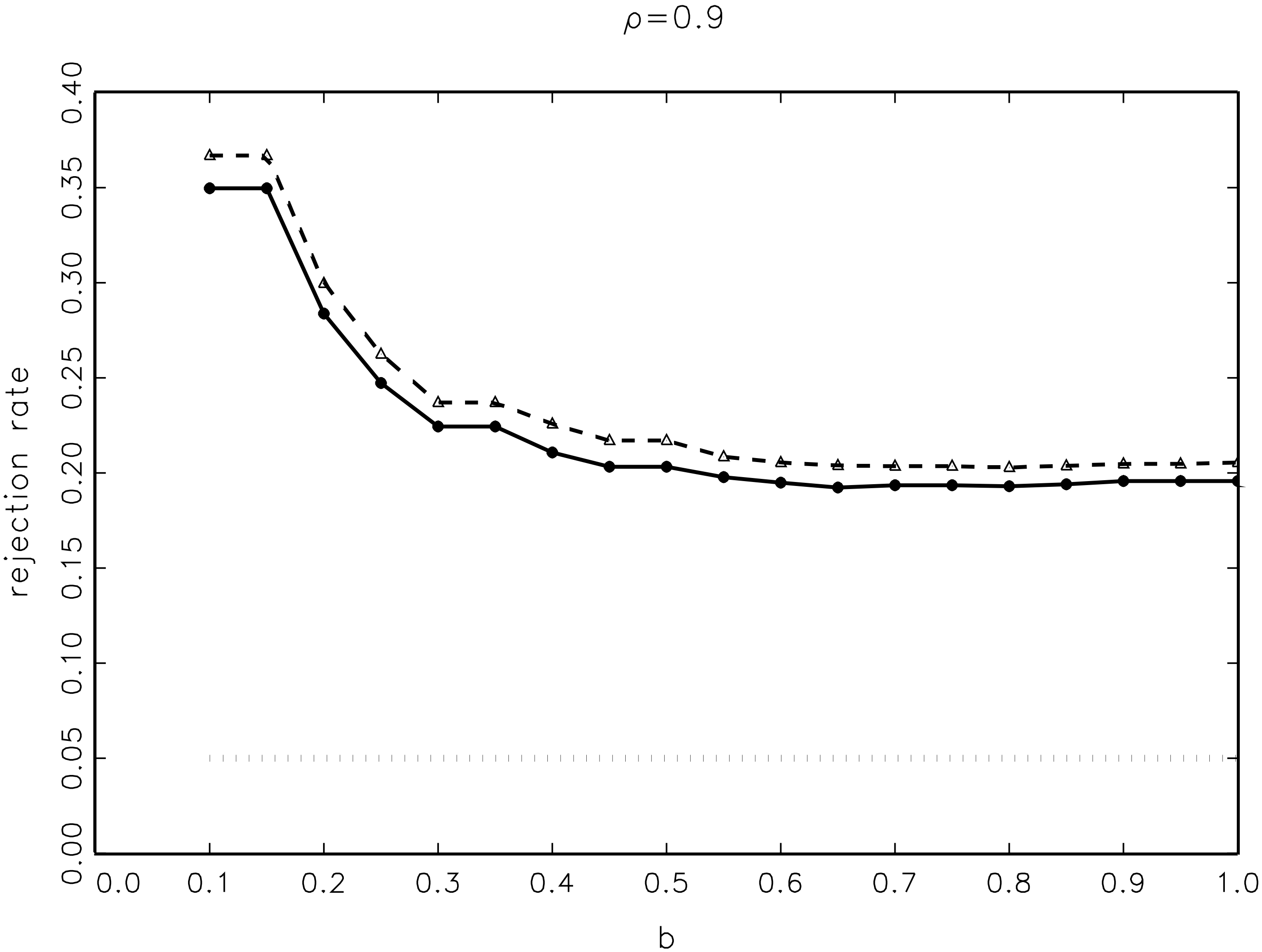


Figure 2.48: ES - Conditional Bernoulli ($p = 0.5$), Bartlett, $T = 50$

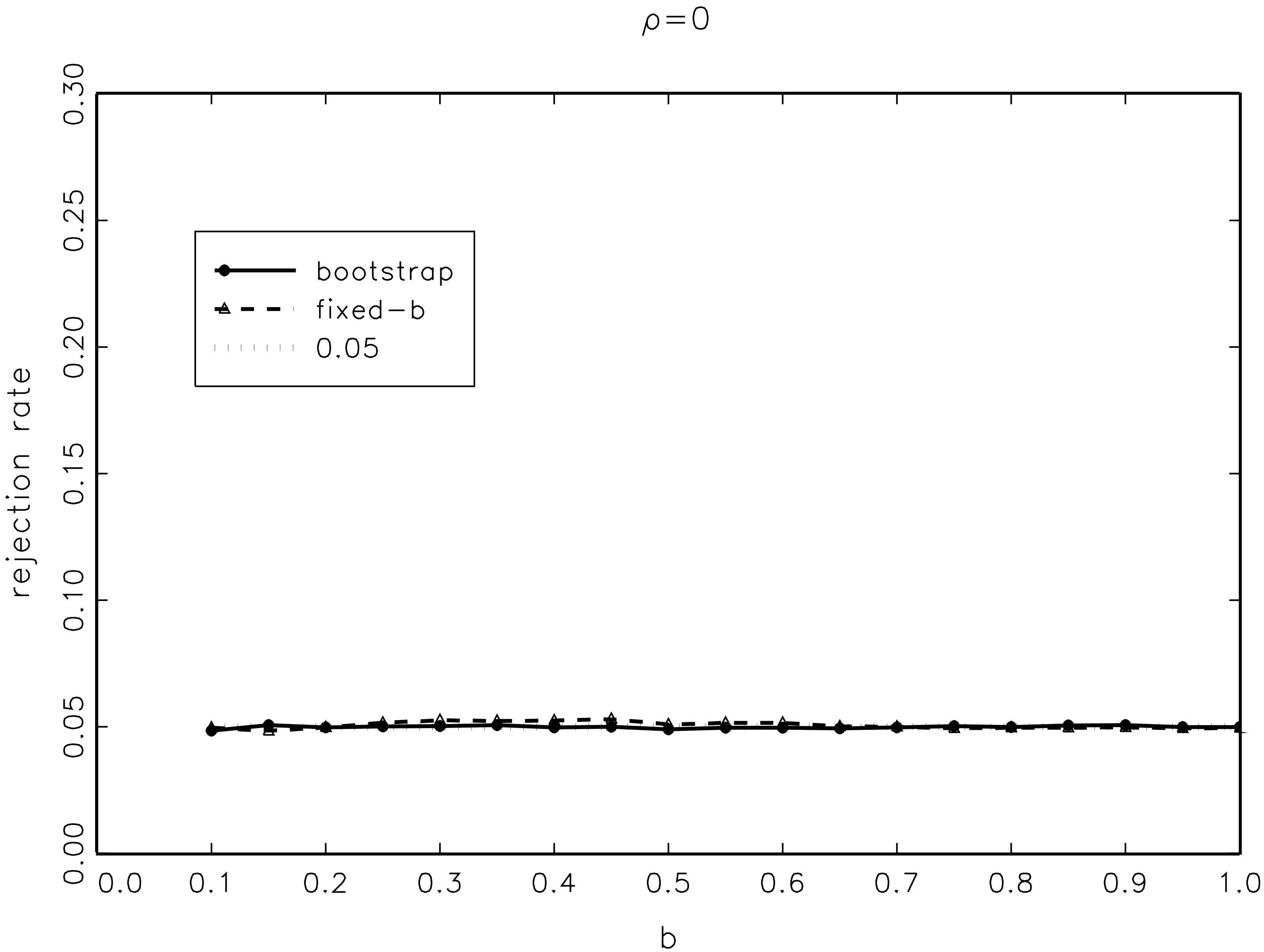


Figure 2.48: (cont'd)

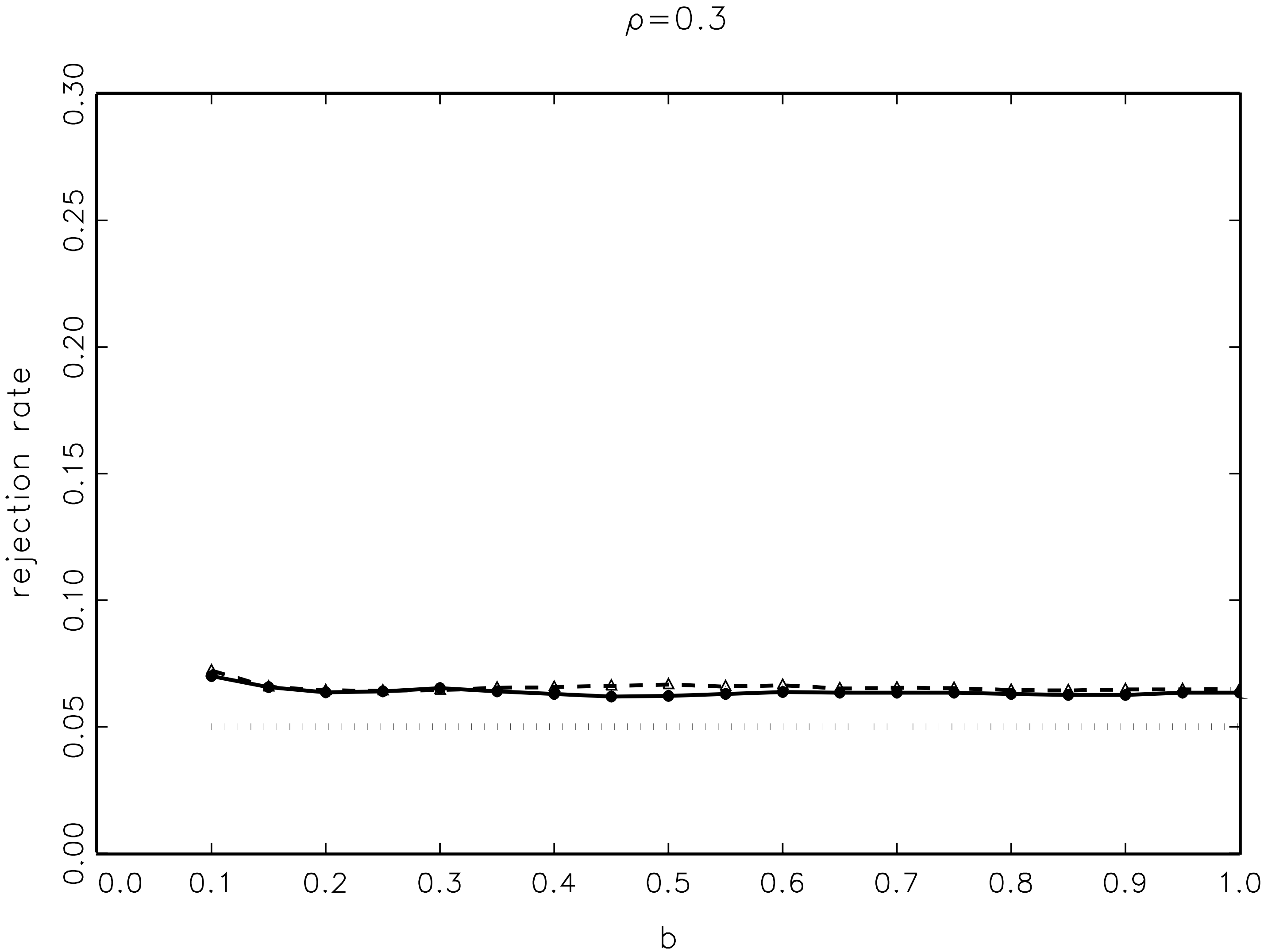


Figure 2.48: (cont'd)

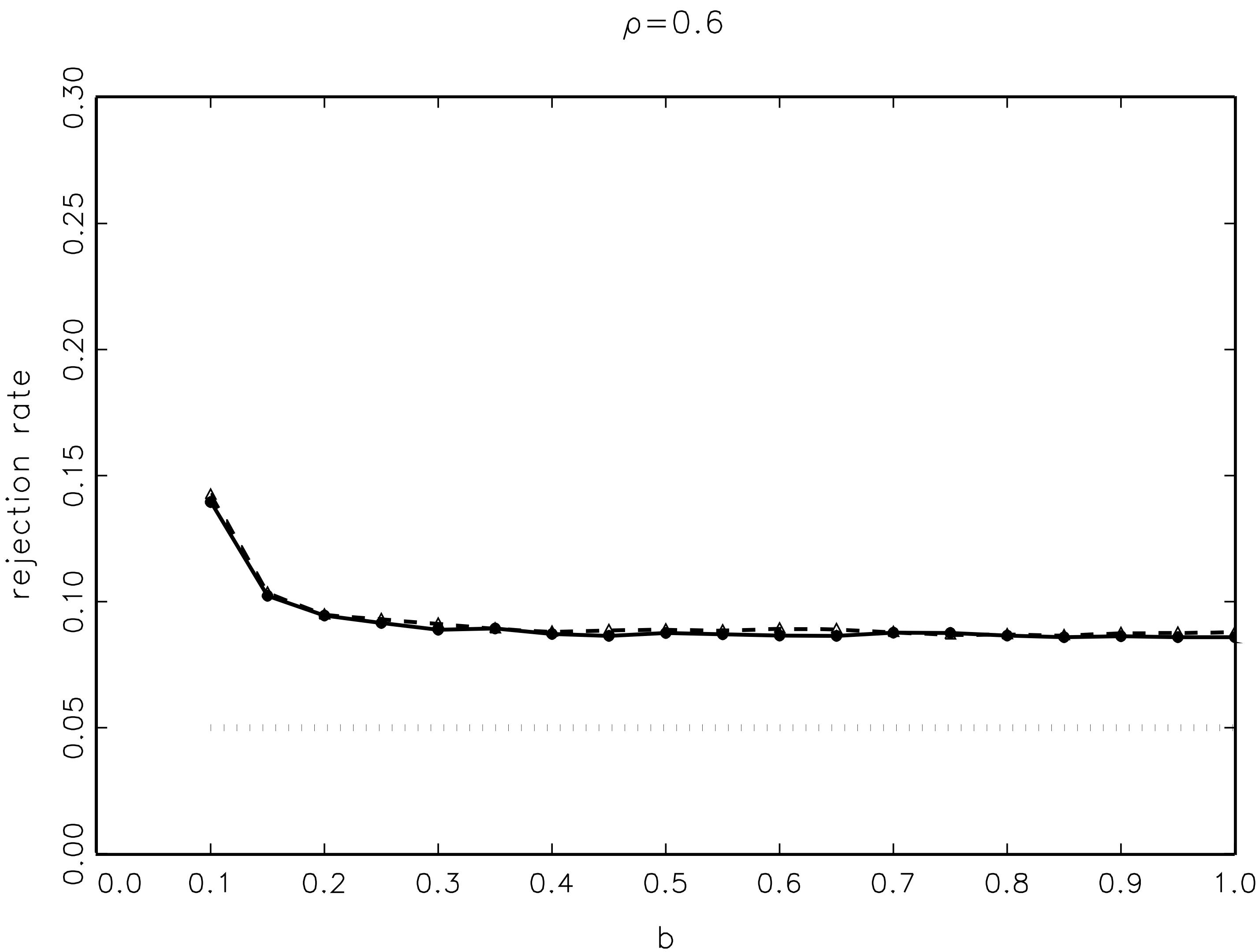


Figure 2.48: (cont'd)

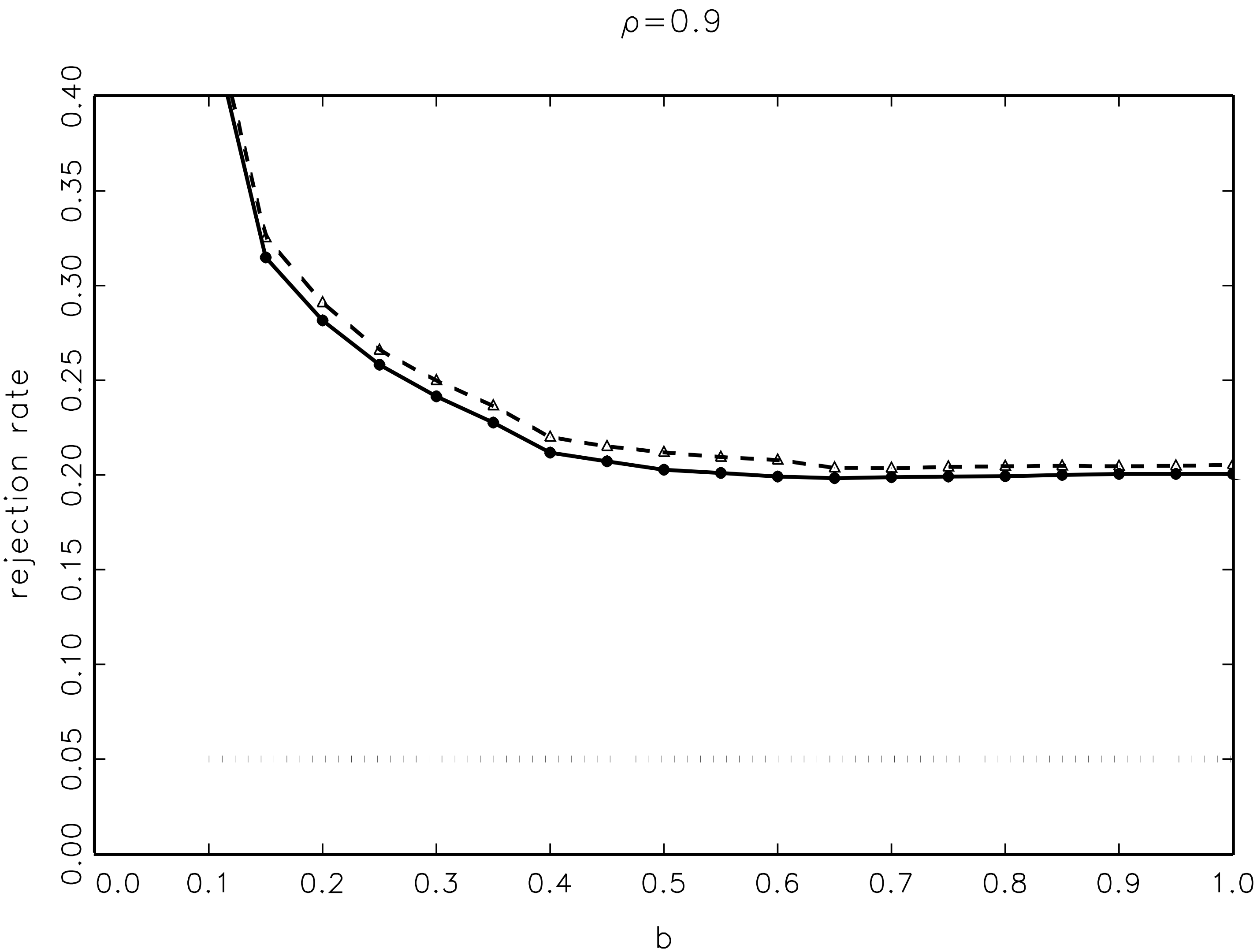


Figure 2.49: ES - Conditional Bernoulli ($p = 0.7$), Bartlett, $T = 50$

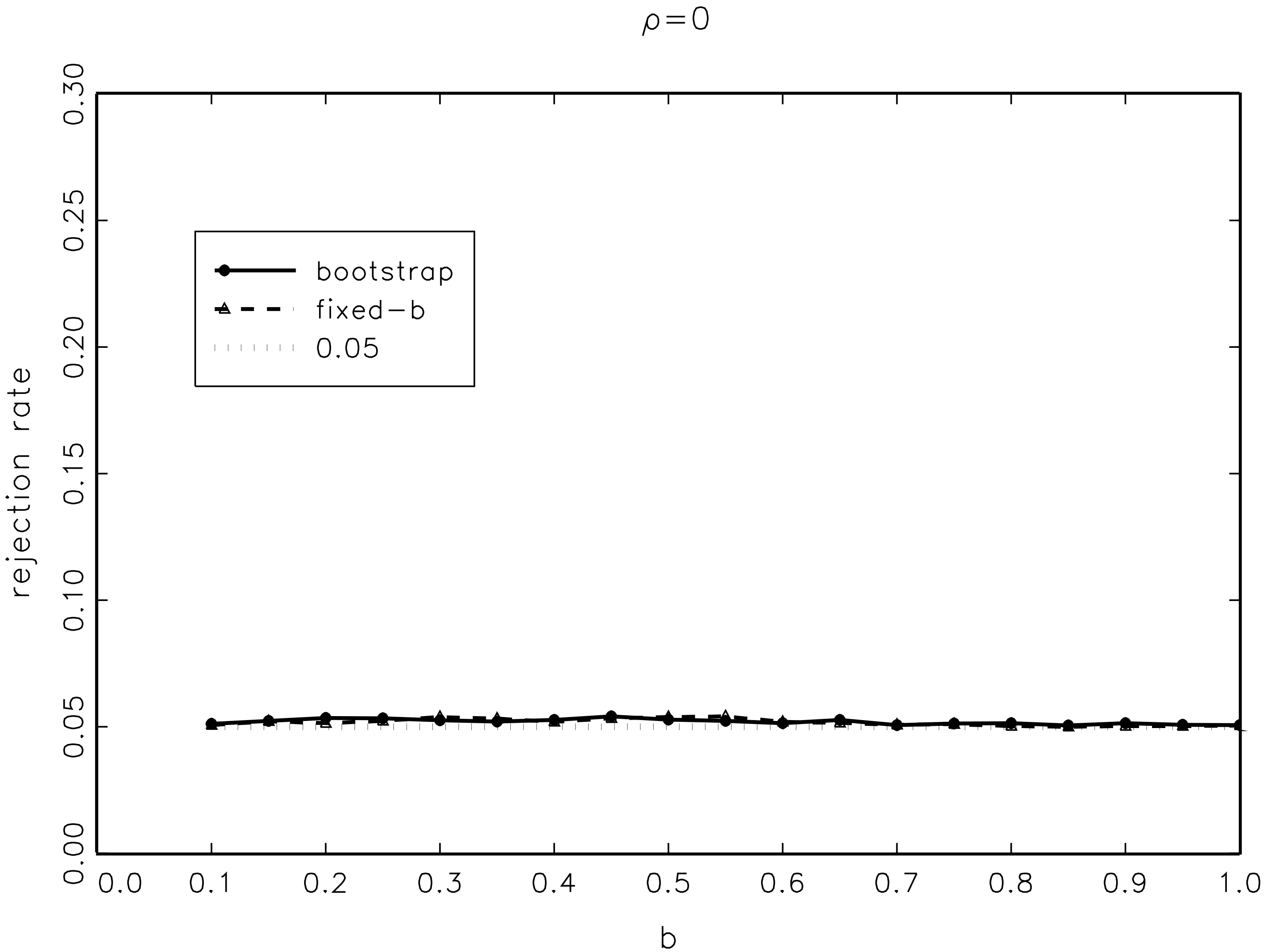


Figure 2.49: (cont'd)

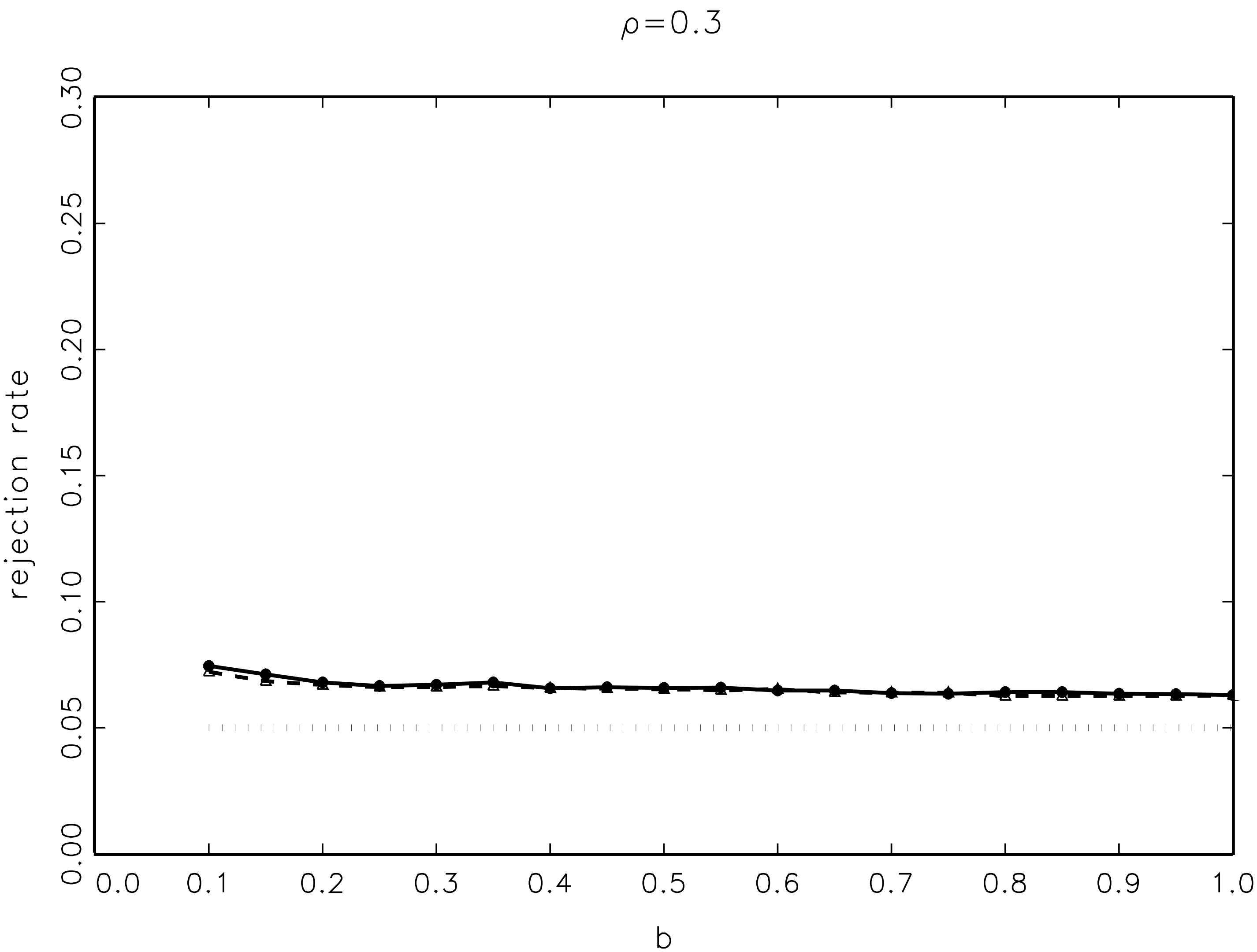


Figure 2.49: (cont'd)

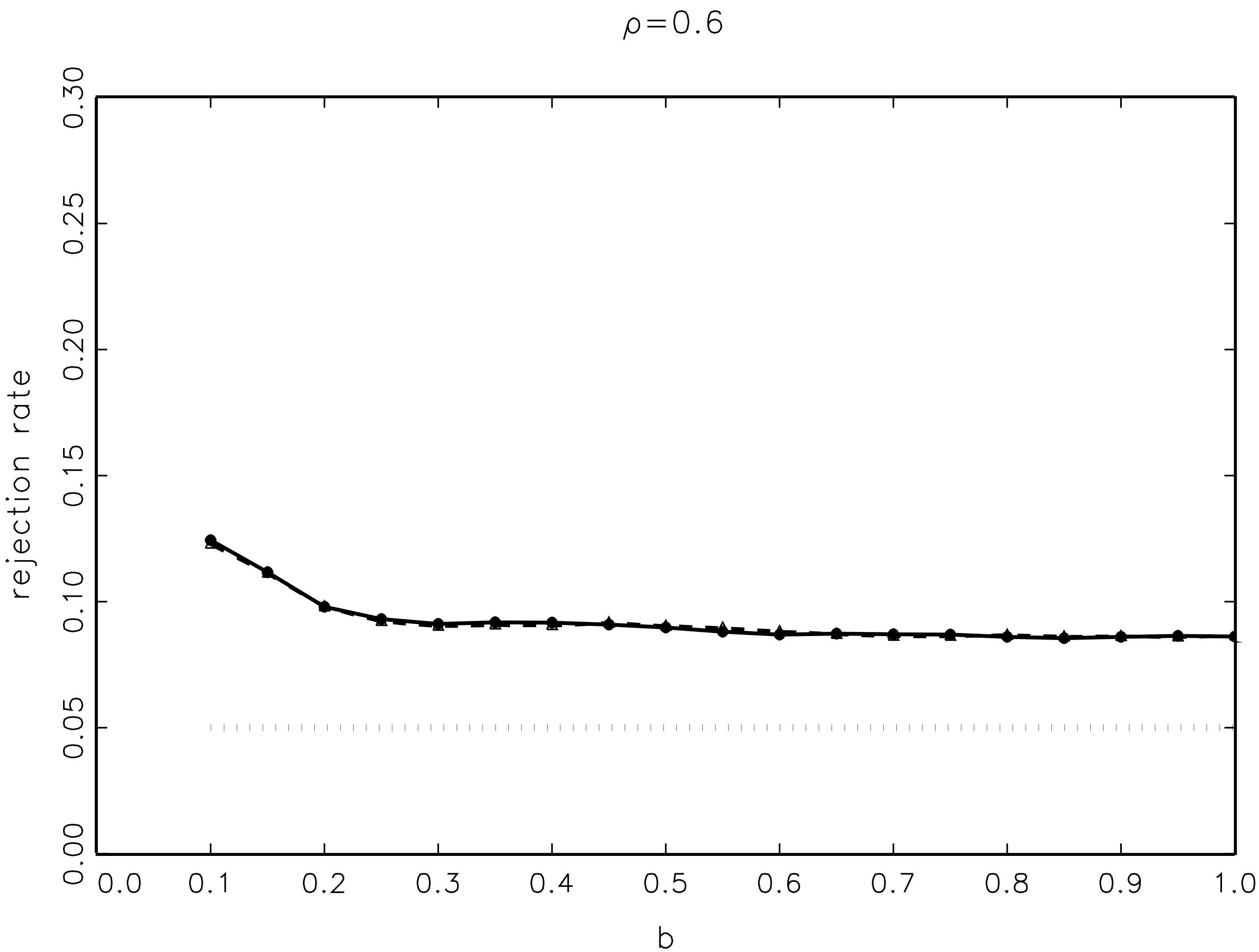


Figure 2.49: (cont'd)

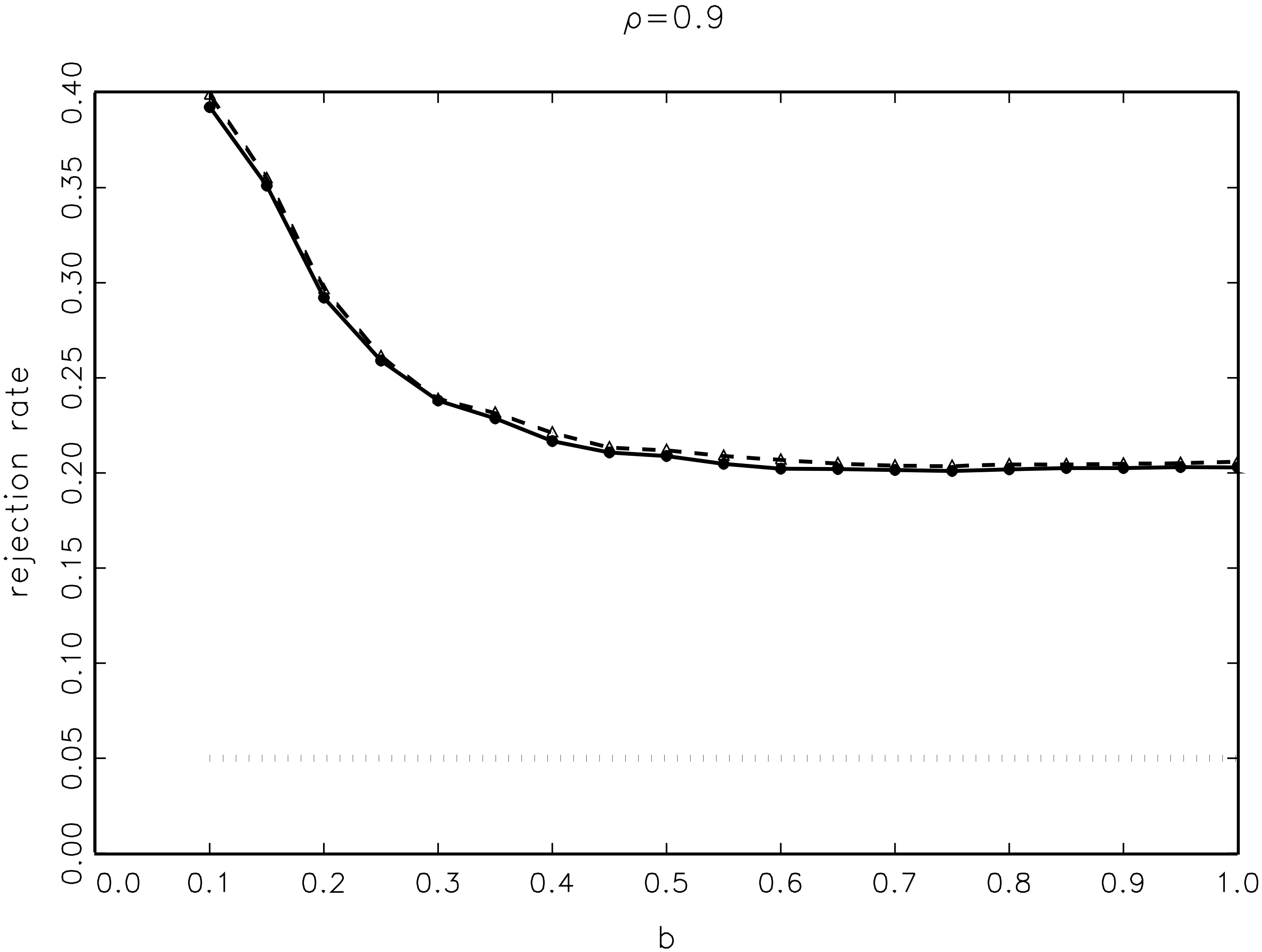


Figure 2.50: ES - Conditional Bernoulli ($p = 0.3$), Bartlett, $T = 100$

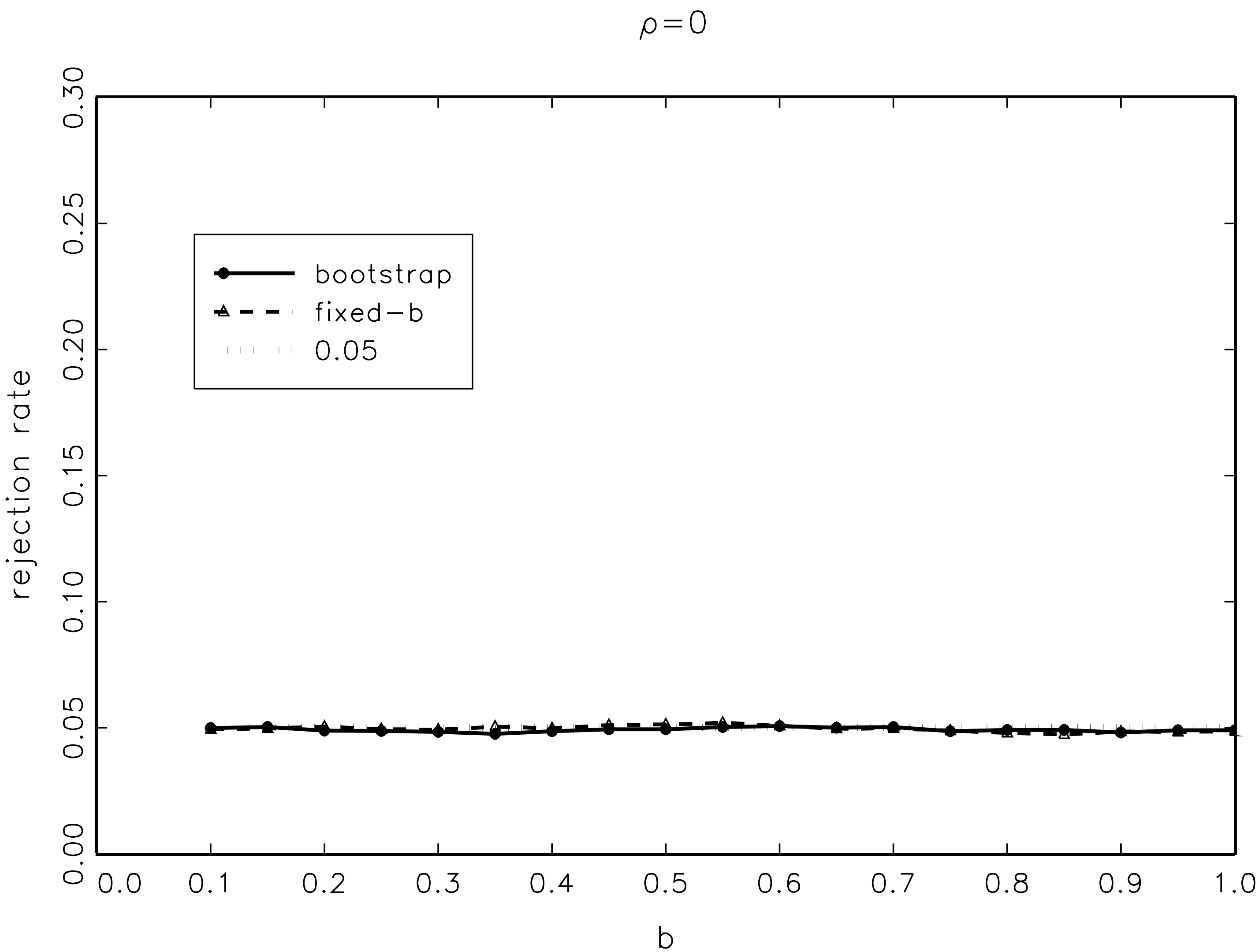


Figure 2.50: (cont'd)

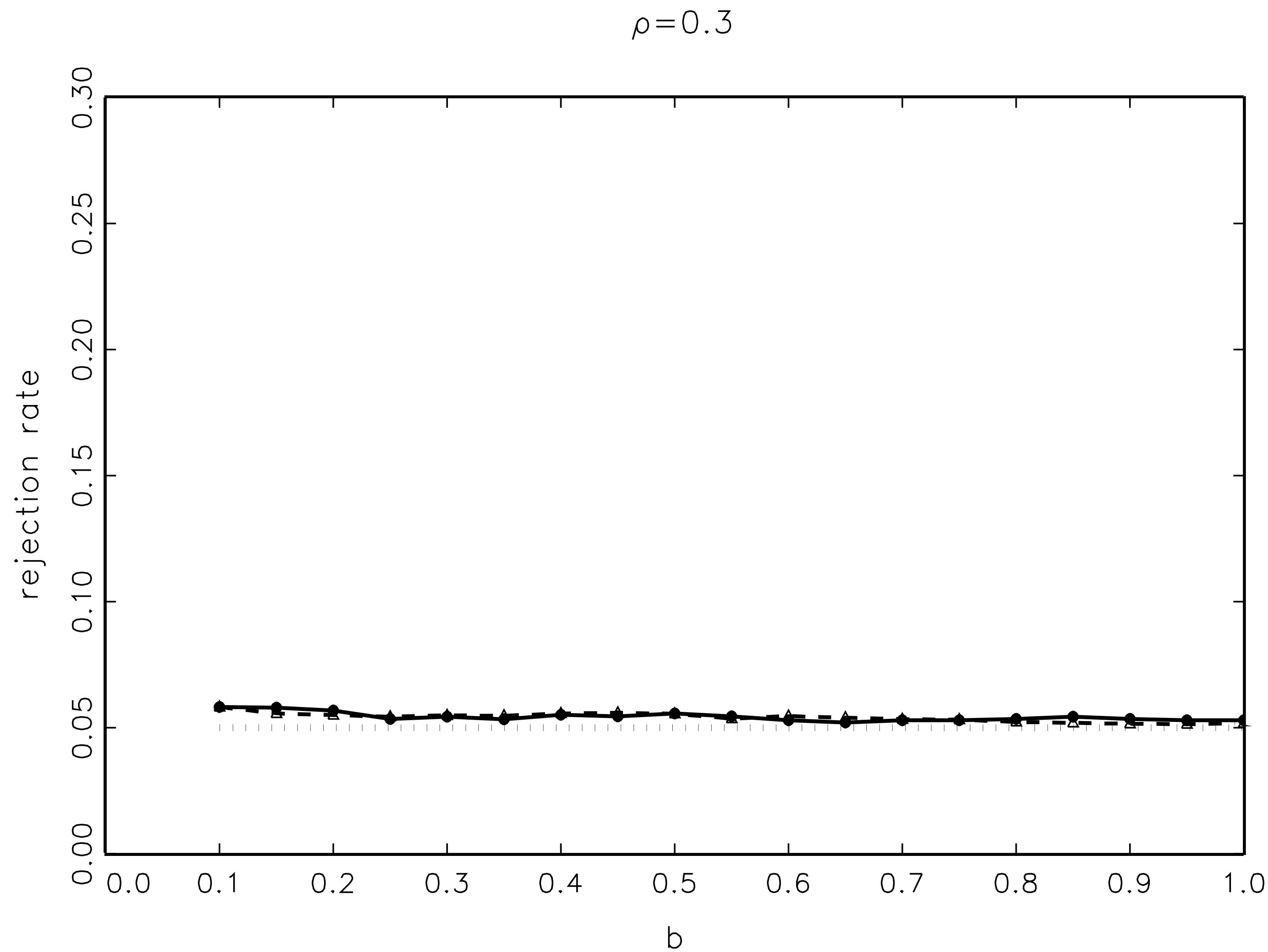


Figure 2.50: (cont'd)

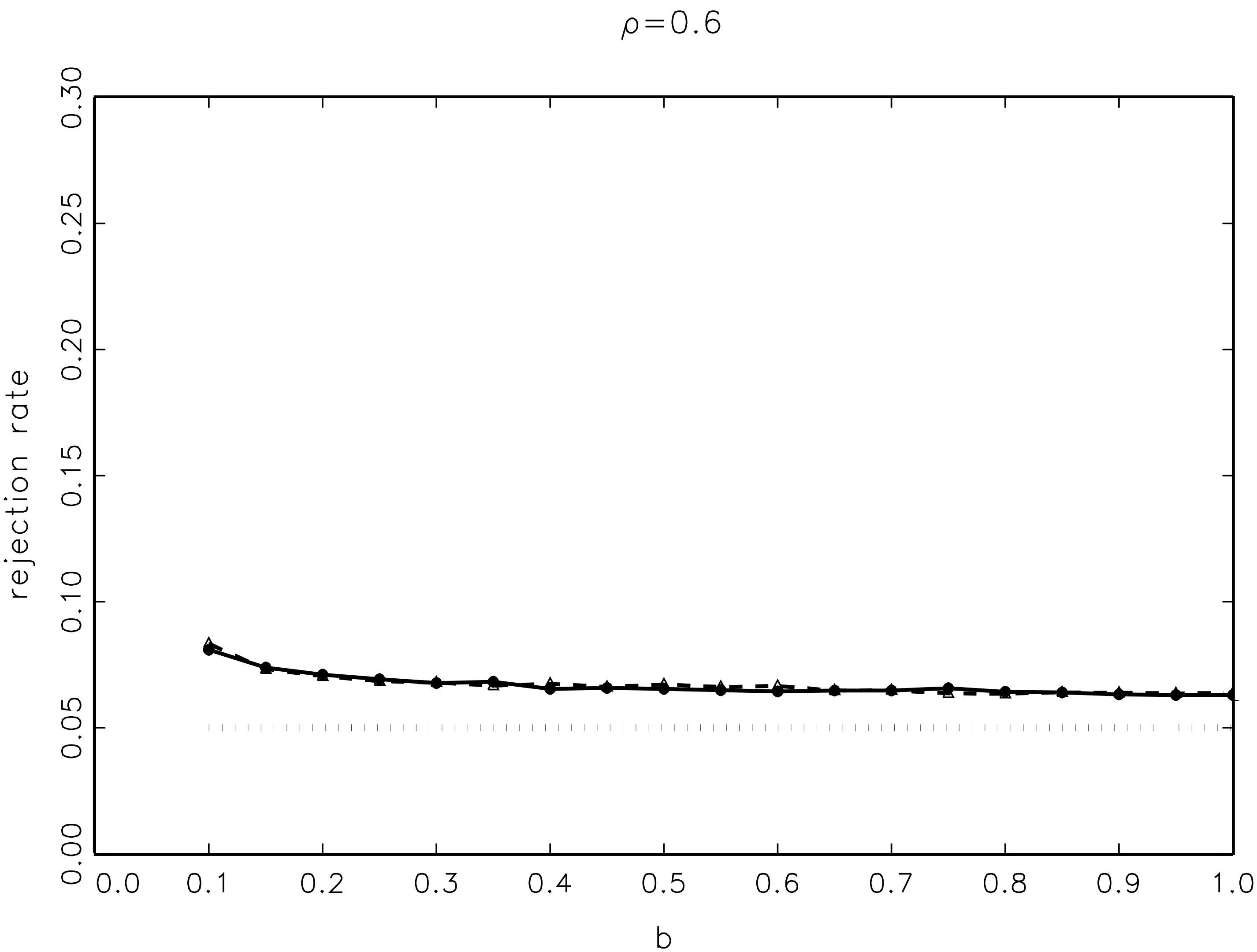


Figure 2.50: (cont'd)

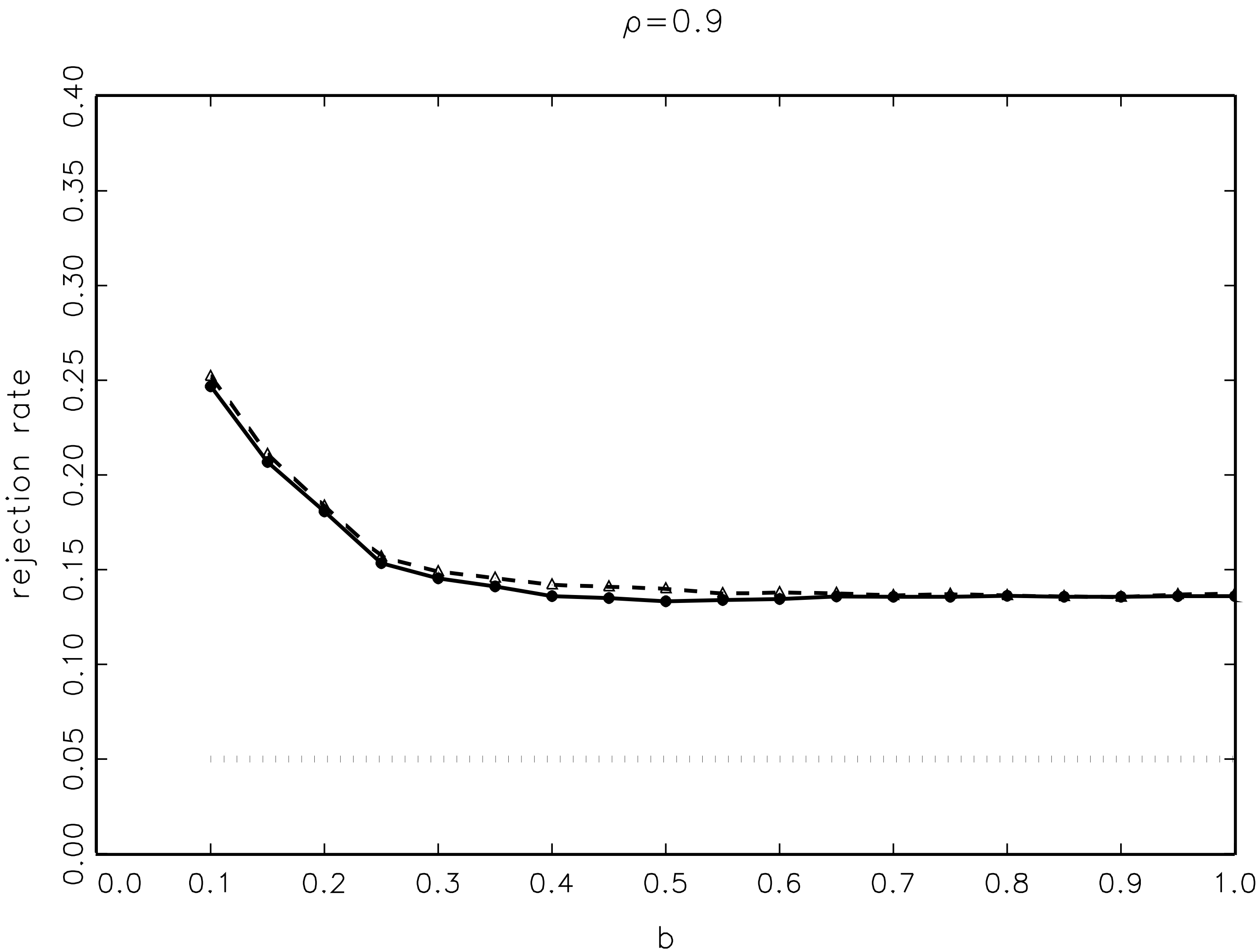


Figure 2.51: ES - Conditional Bernoulli ($p = 0.5$), Bartlett, $T = 100$

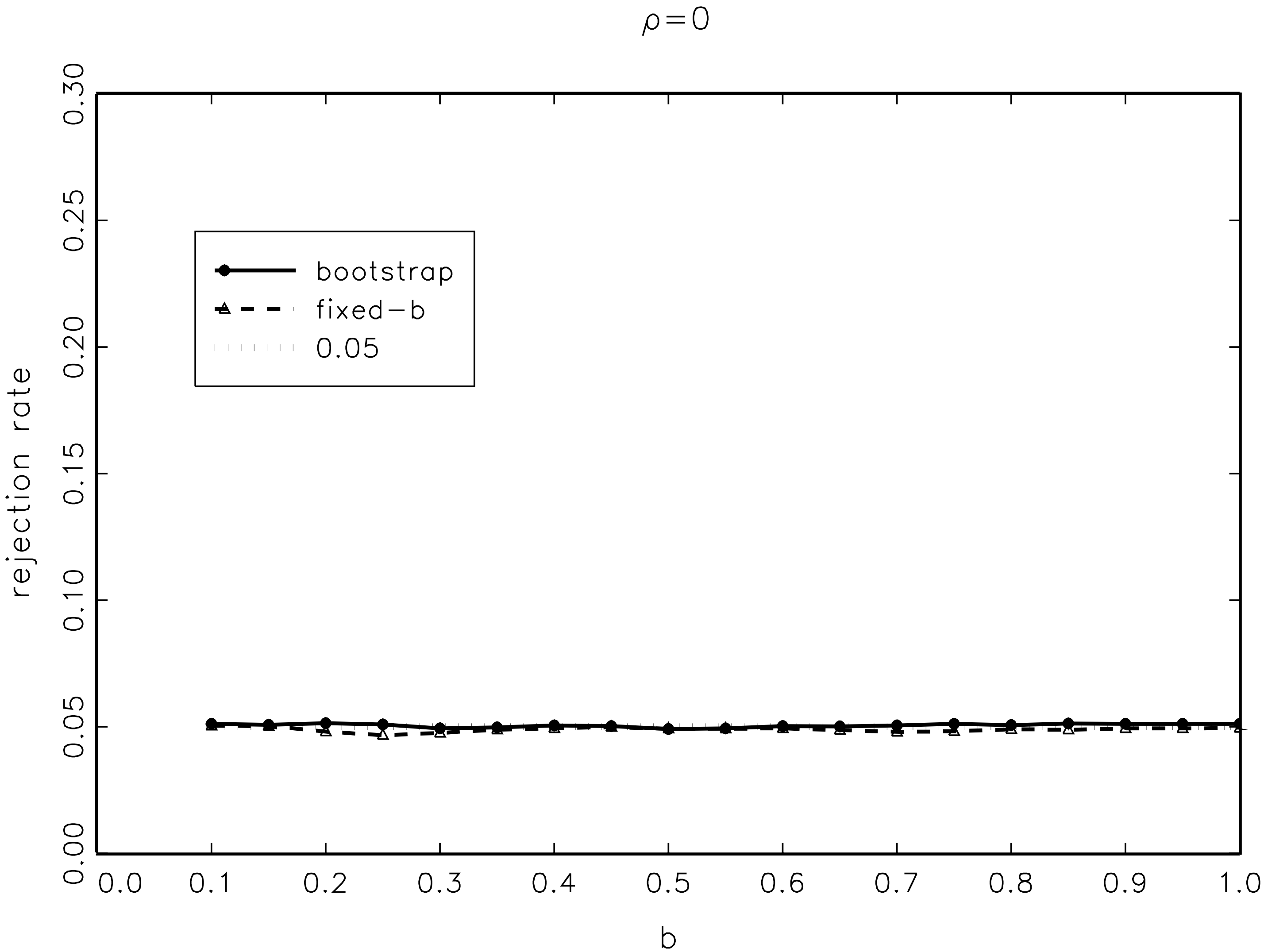


Figure 2.51: (cont'd)

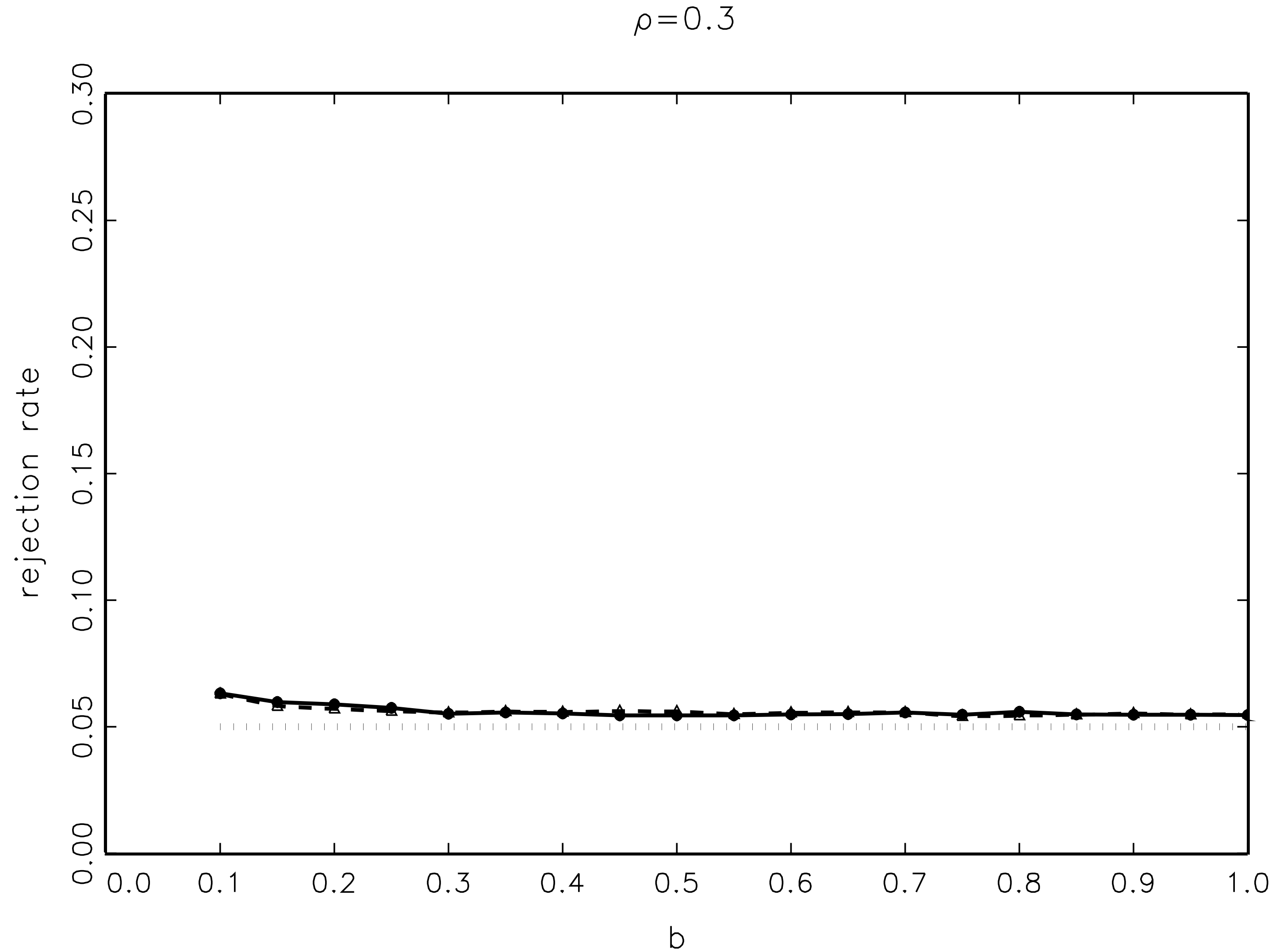


Figure 2.51: (cont'd)

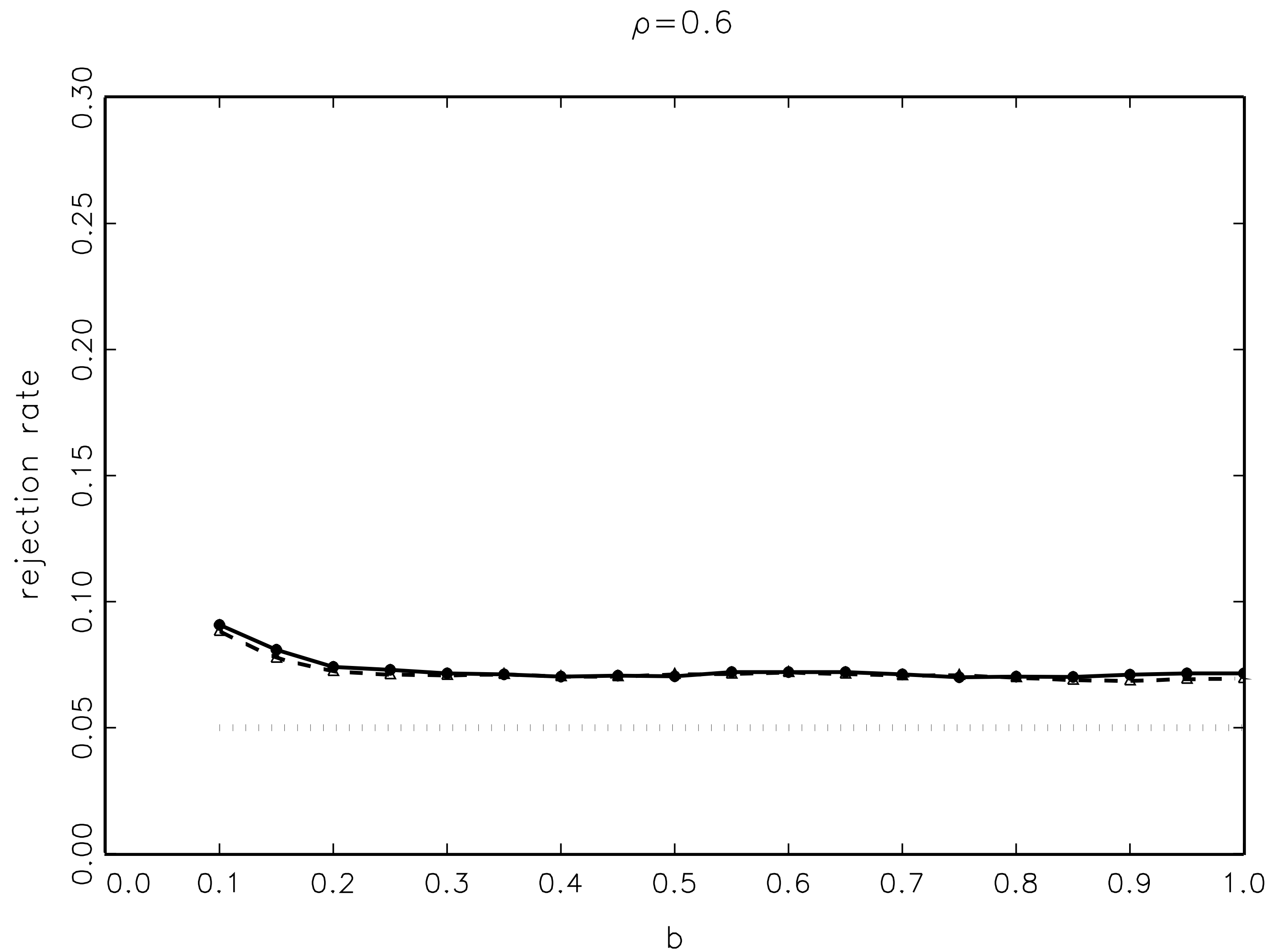


Figure 2.51: (cont'd)

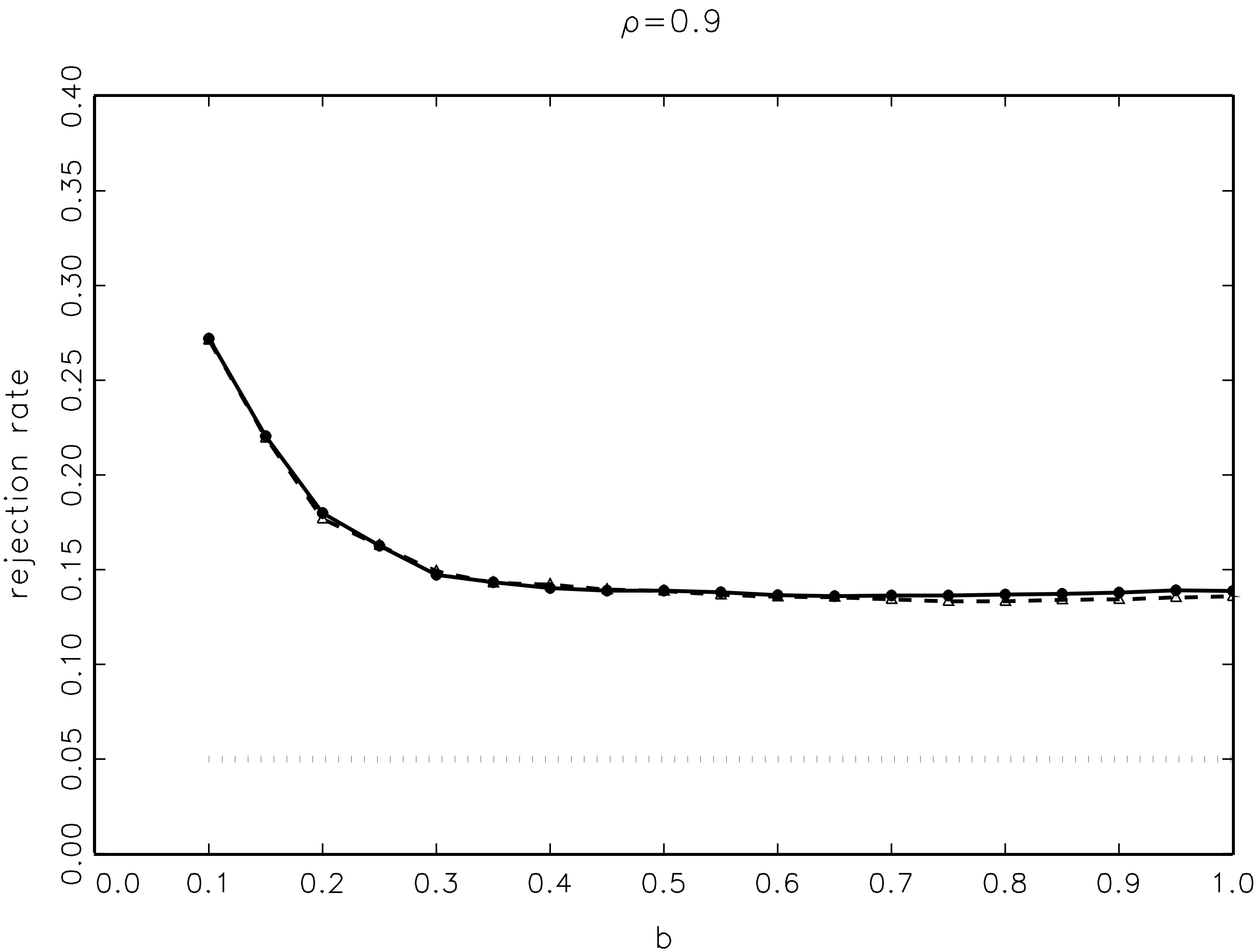


Figure 2.52: ES - Conditional Bernoulli ($p = 0.7$), Bartlett, $T = 100$

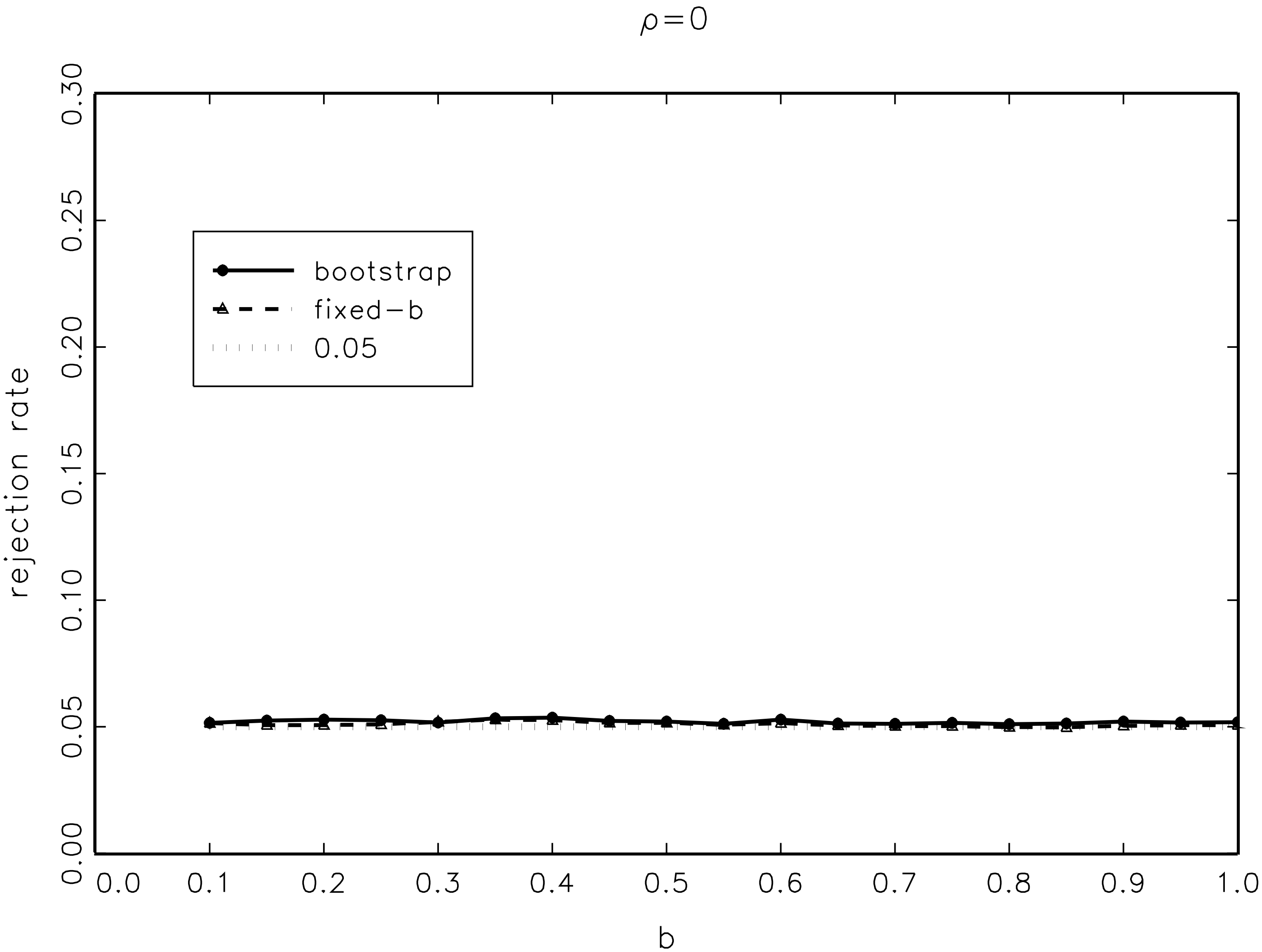


Figure 2.52: (cont'd)

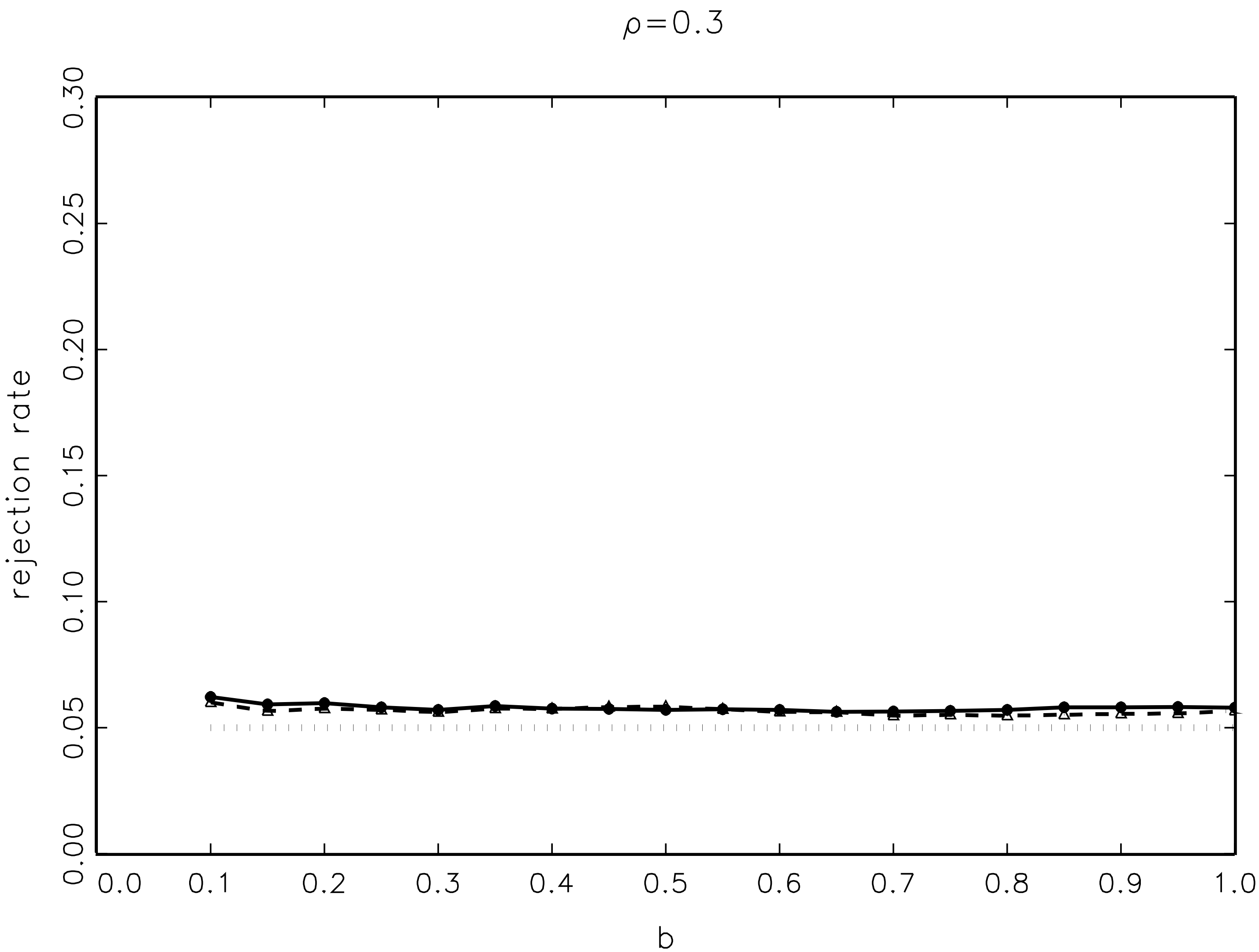


Figure 2.52: (cont'd)

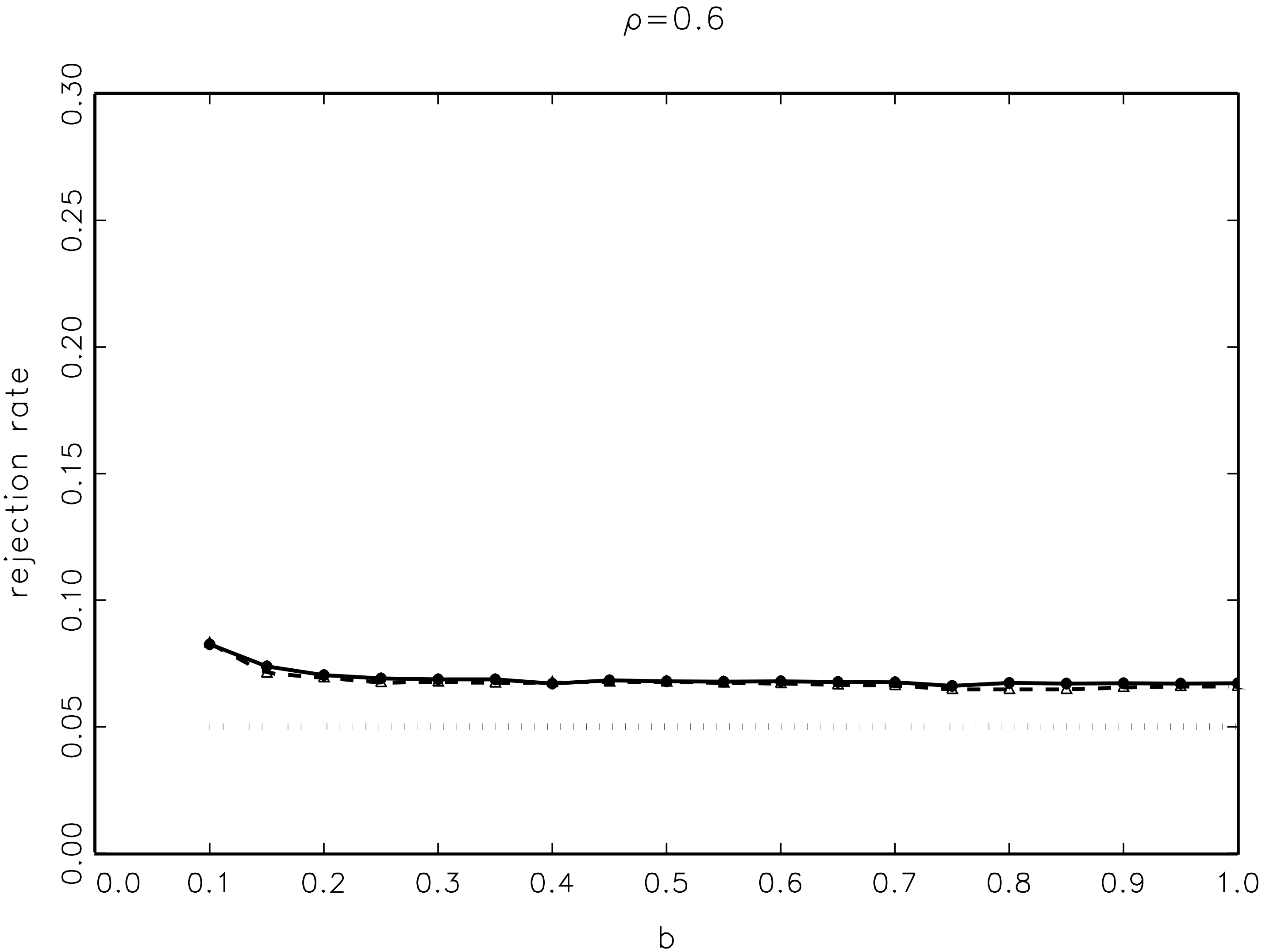


Figure 2.52: (cont'd)

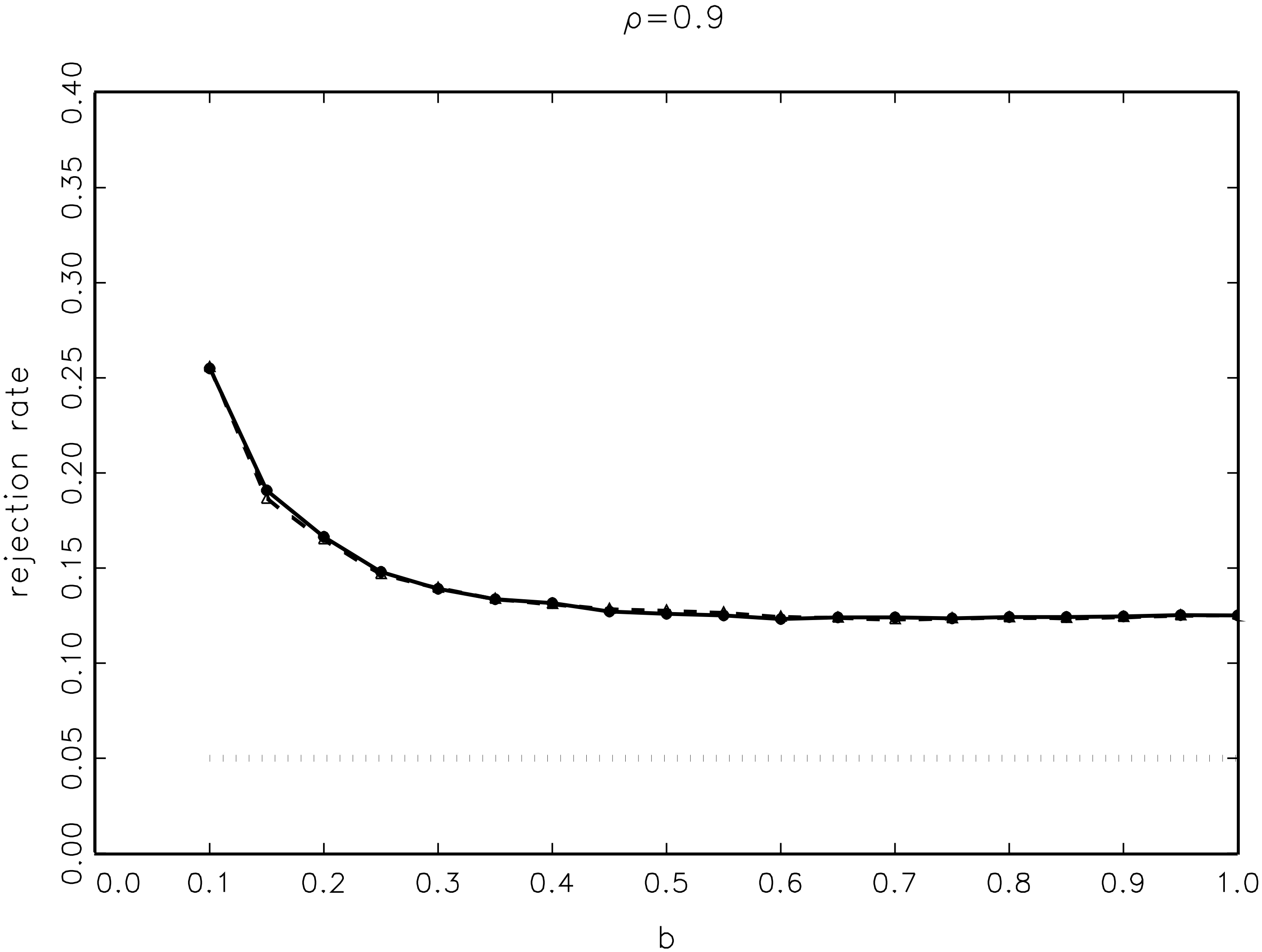


Figure 2.53: ES - Conditional Bernoulli ($p = 0.3$), Bartlett, $T = 200$

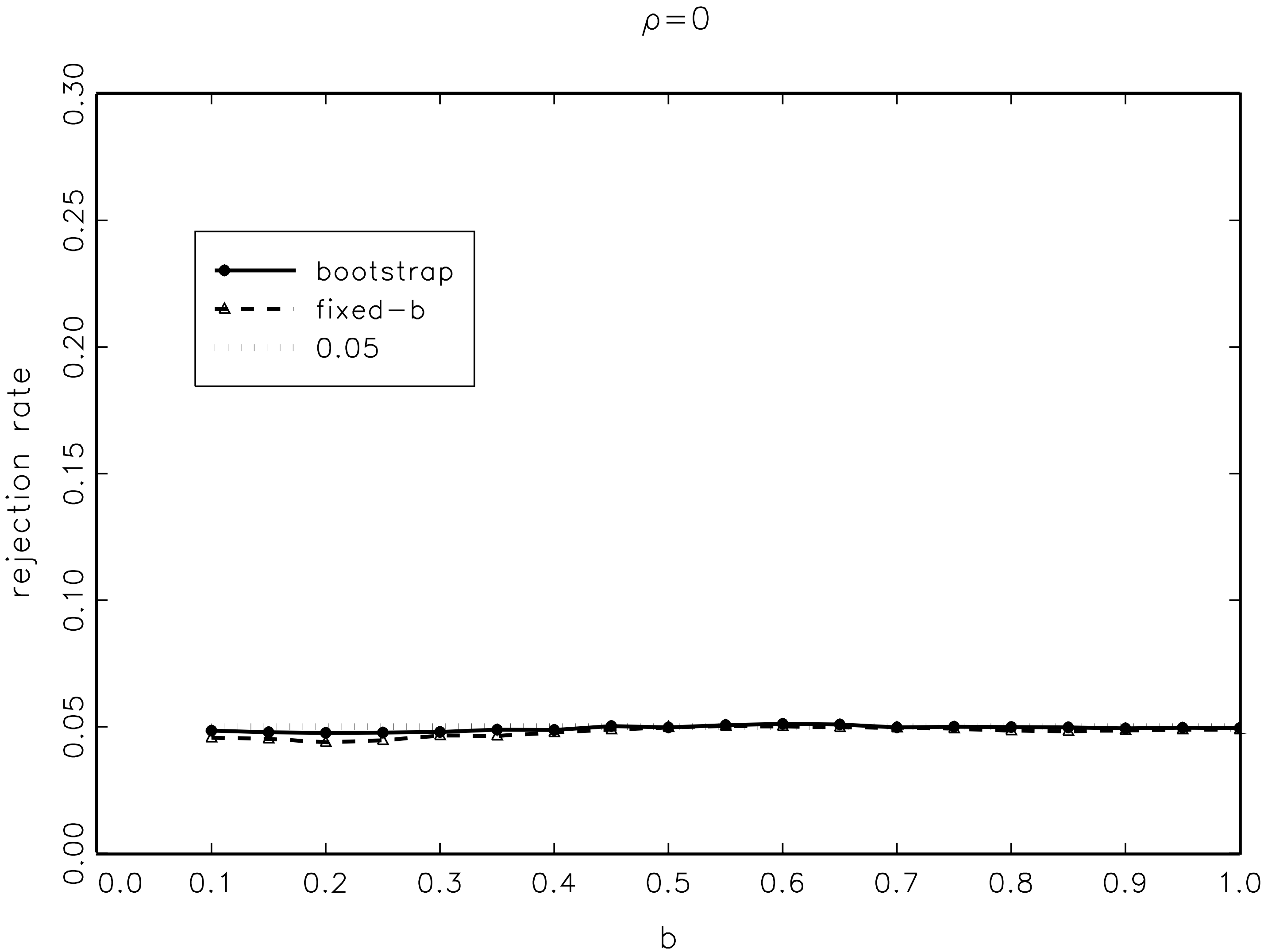


Figure 2.53: (cont'd)

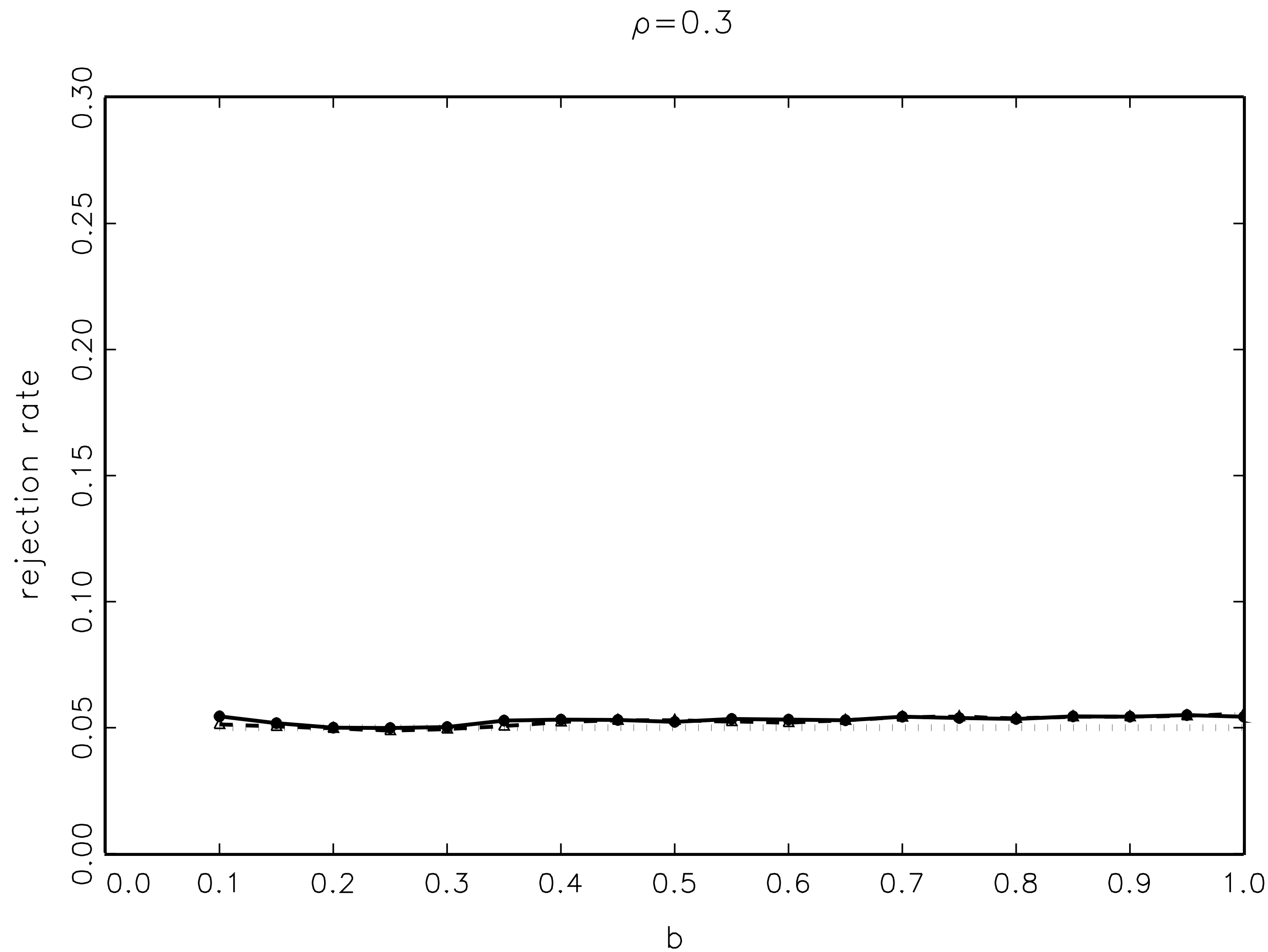


Figure 2.53: (cont'd)

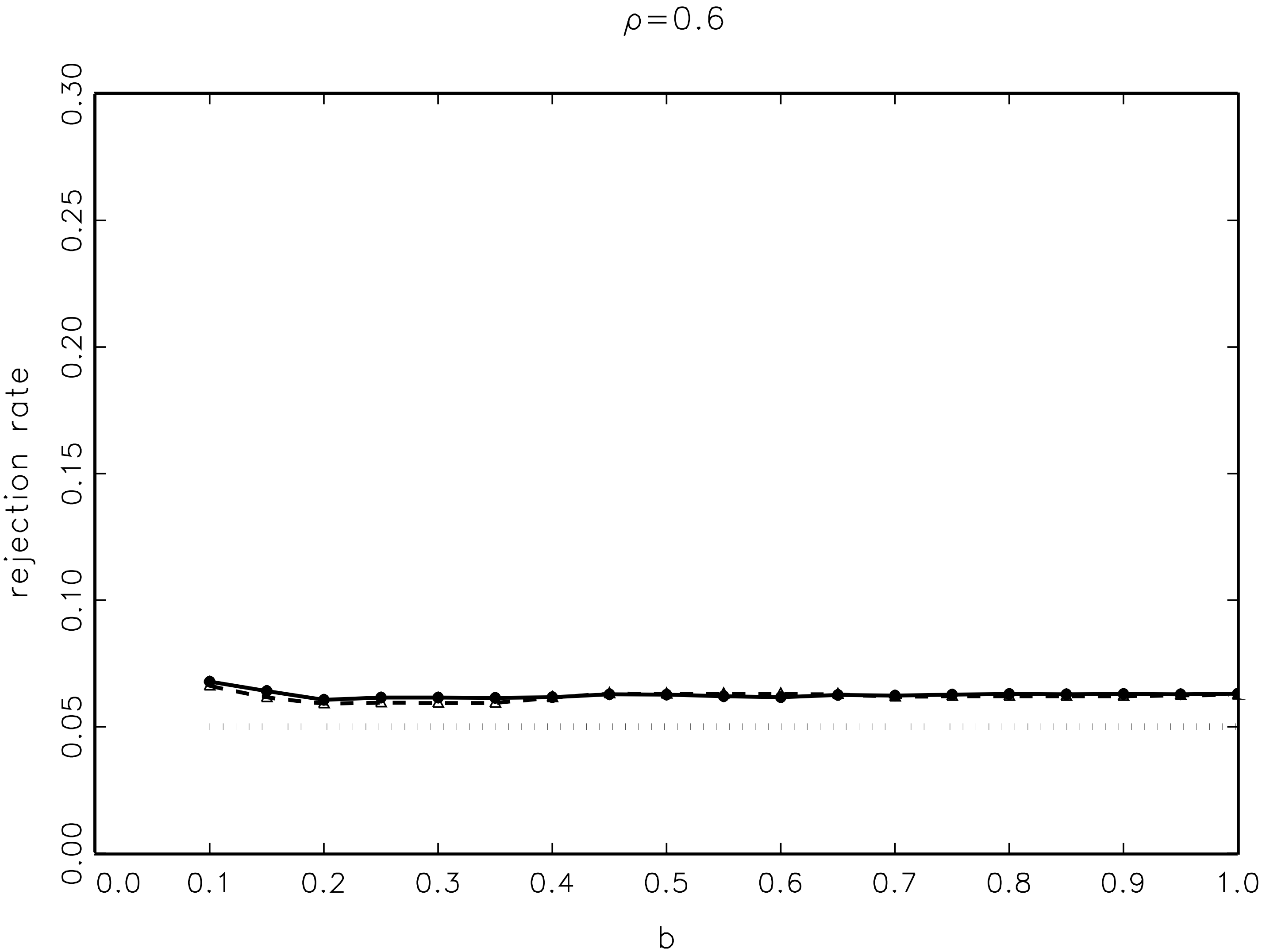


Figure 2.53: (cont'd)

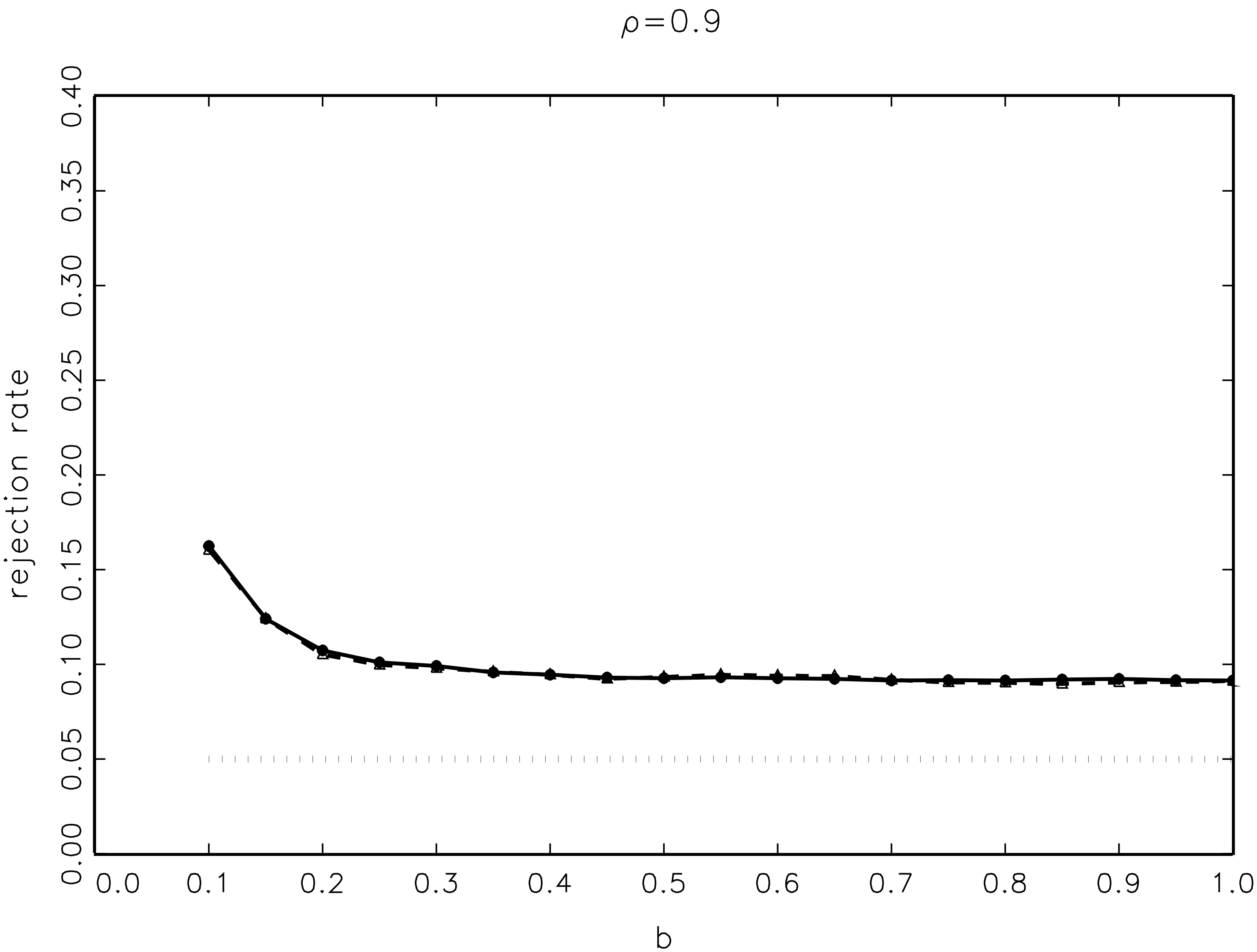


Figure 2.54: ES - Conditional Bernoulli ($p = 0.5$), Bartlett, $T = 200$

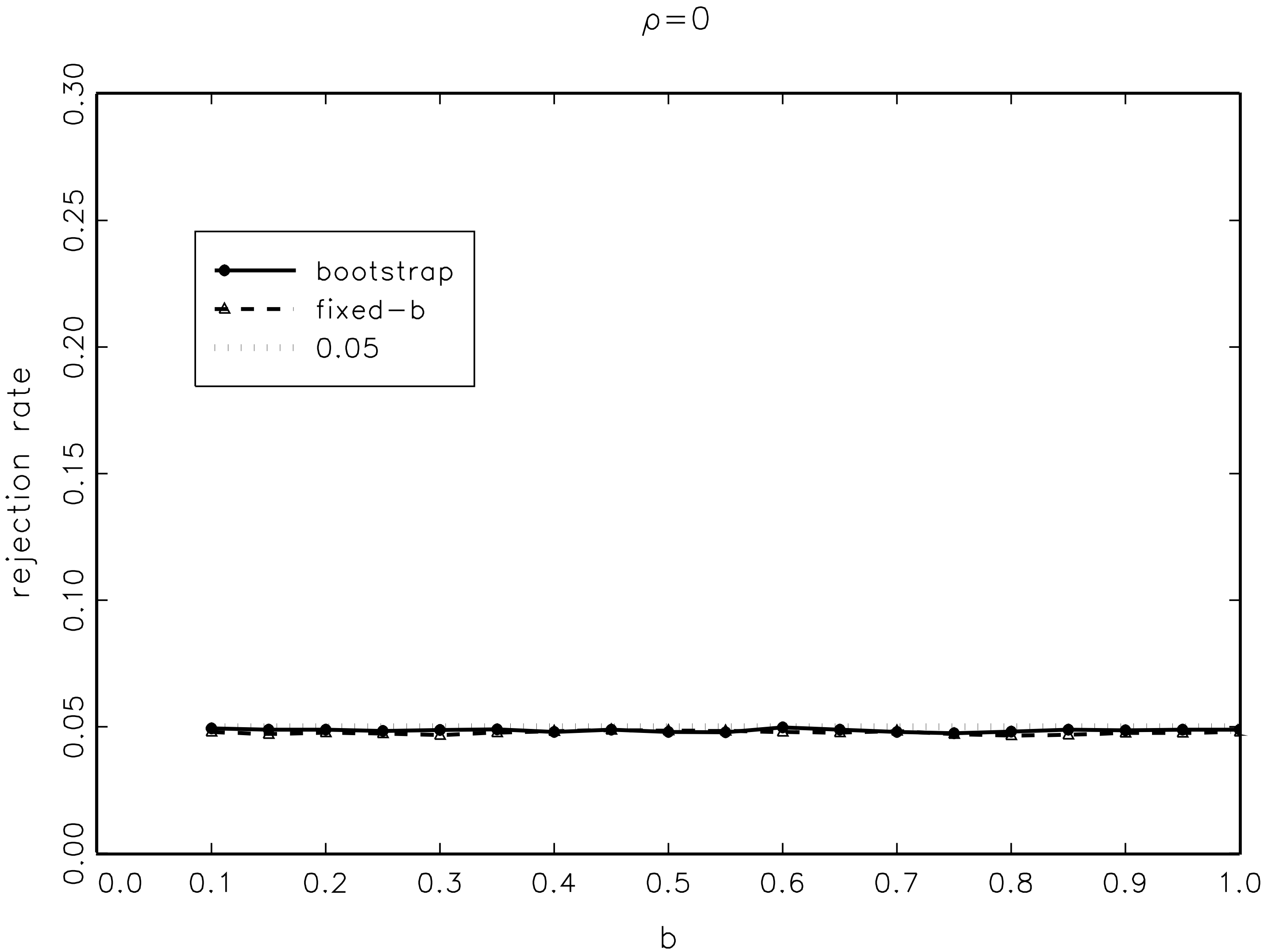


Figure 2.54: (cont'd)

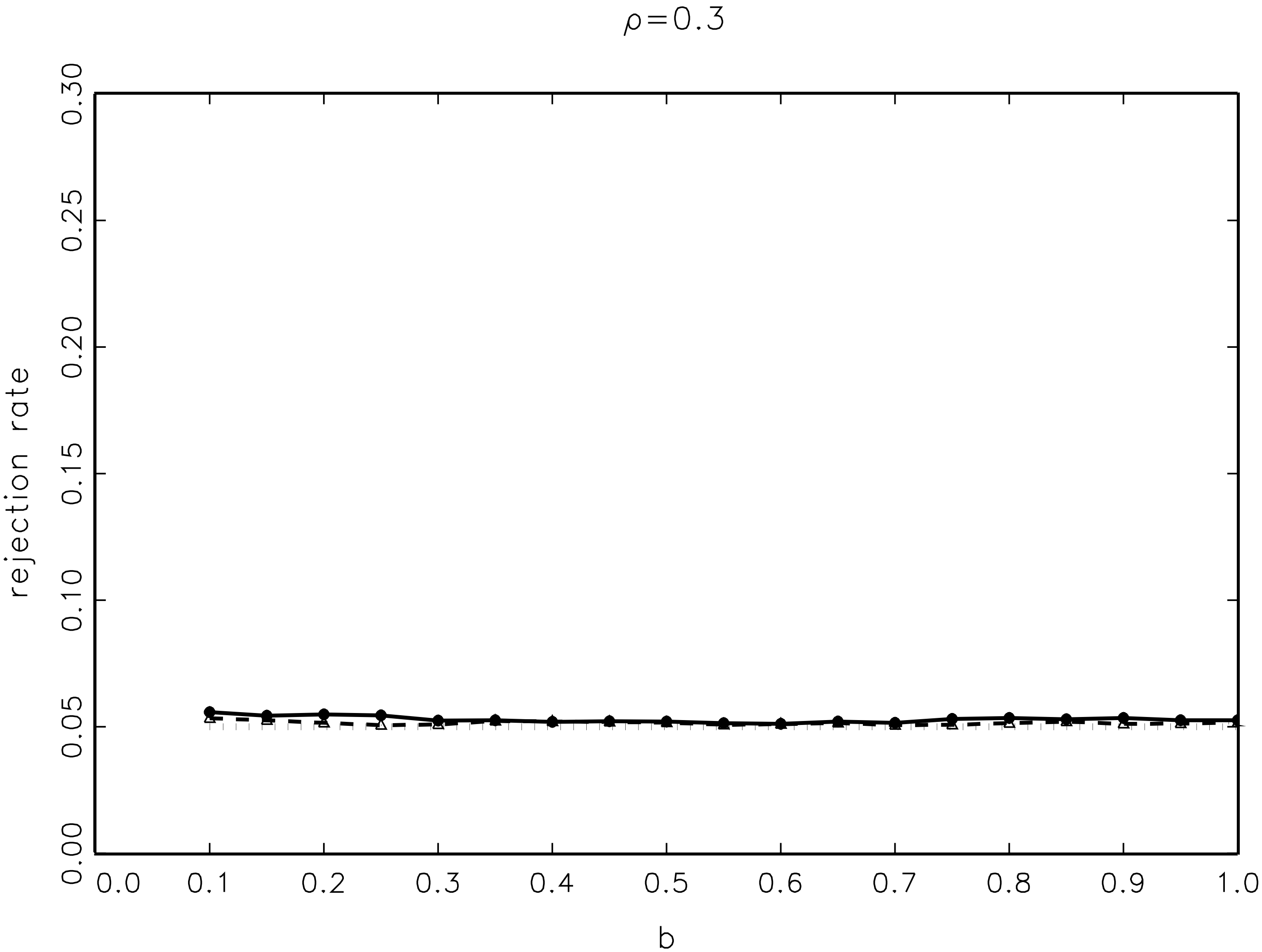


Figure 2.54: (cont'd)

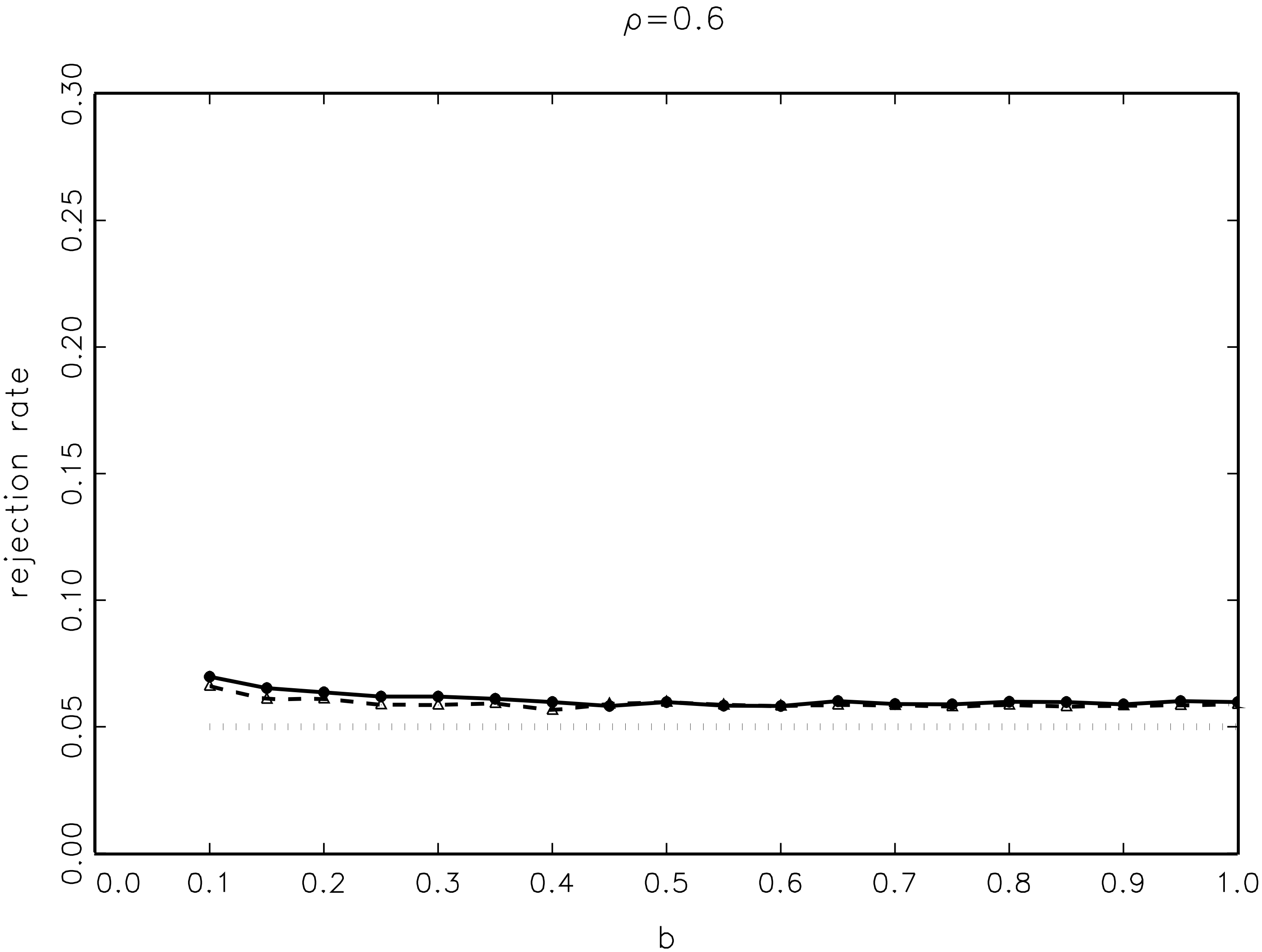


Figure 2.54: (cont'd)

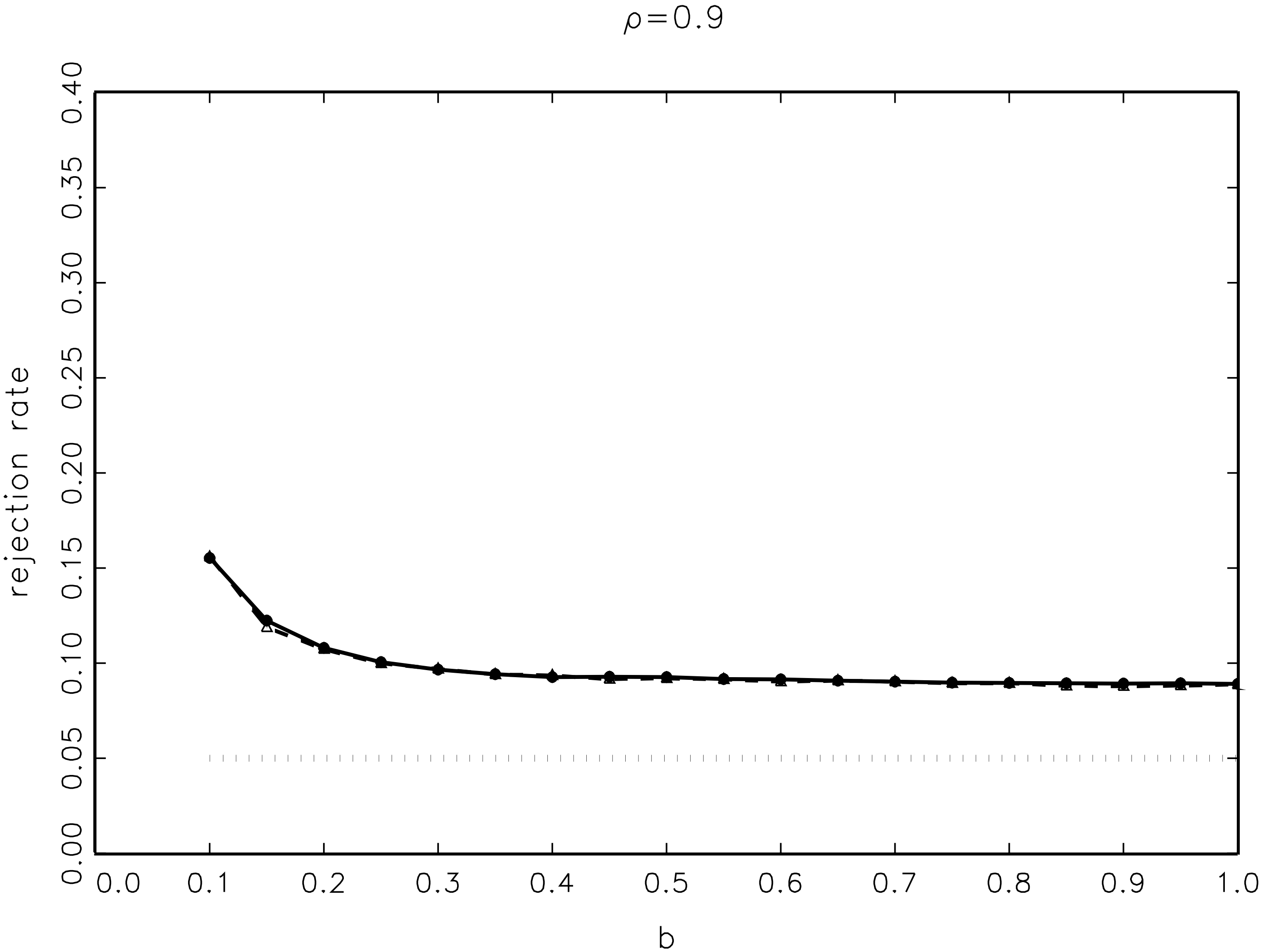


Figure 2.55: ES - Conditional Bernoulli ($p = 0.7$), Bartlett, $T = 200$

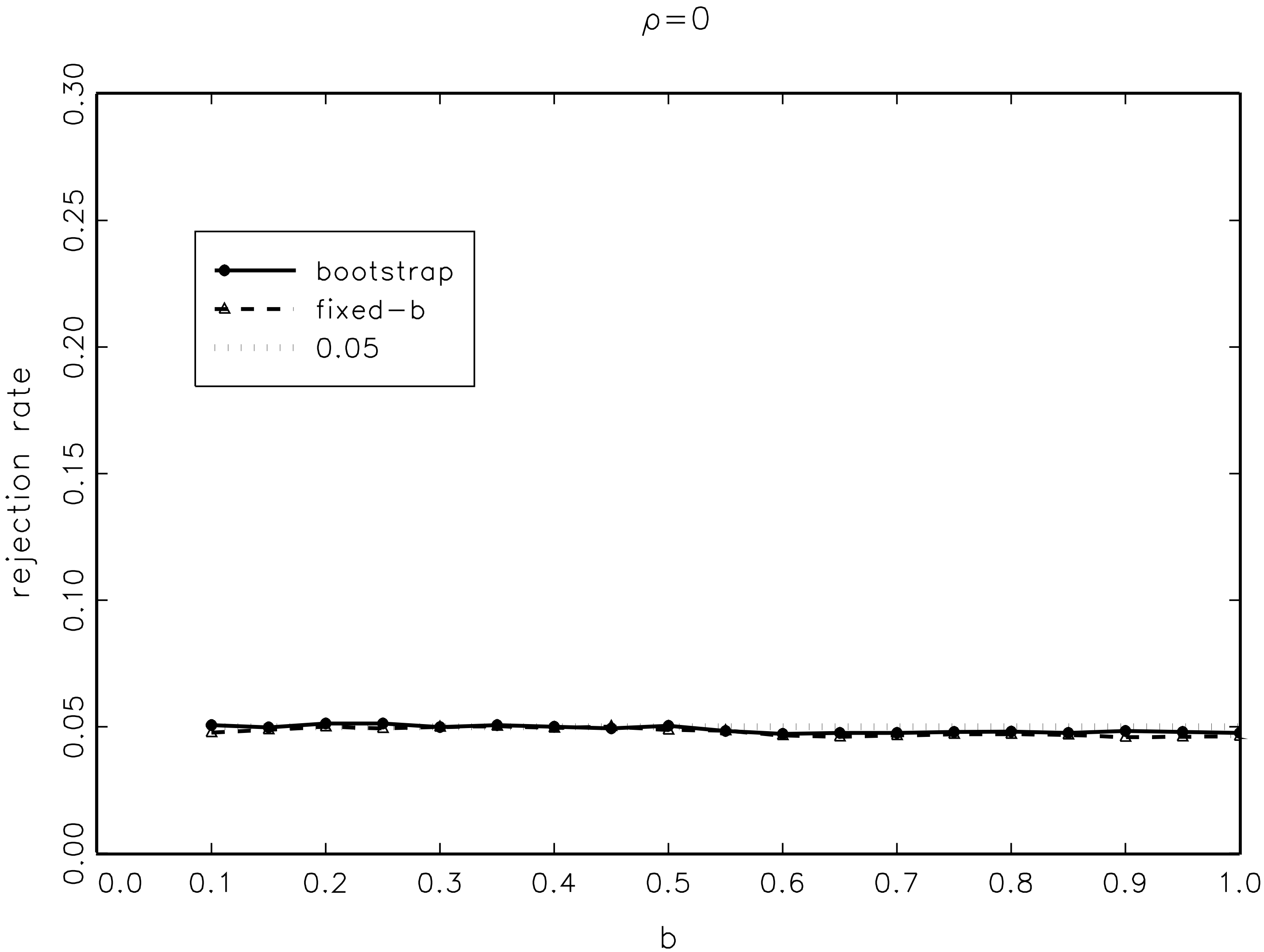


Figure 2.55: (cont'd)

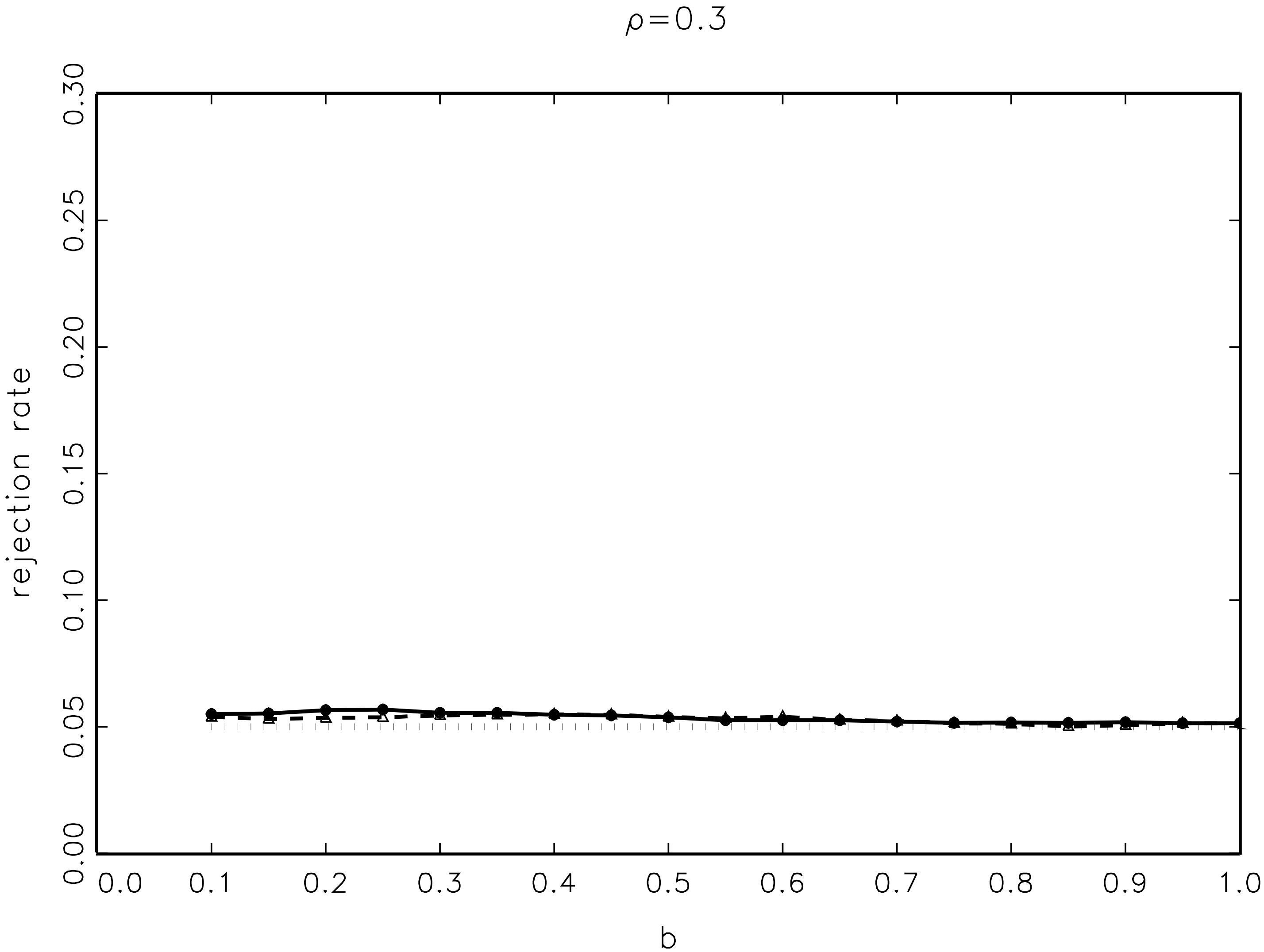


Figure 2.55: (cont'd)

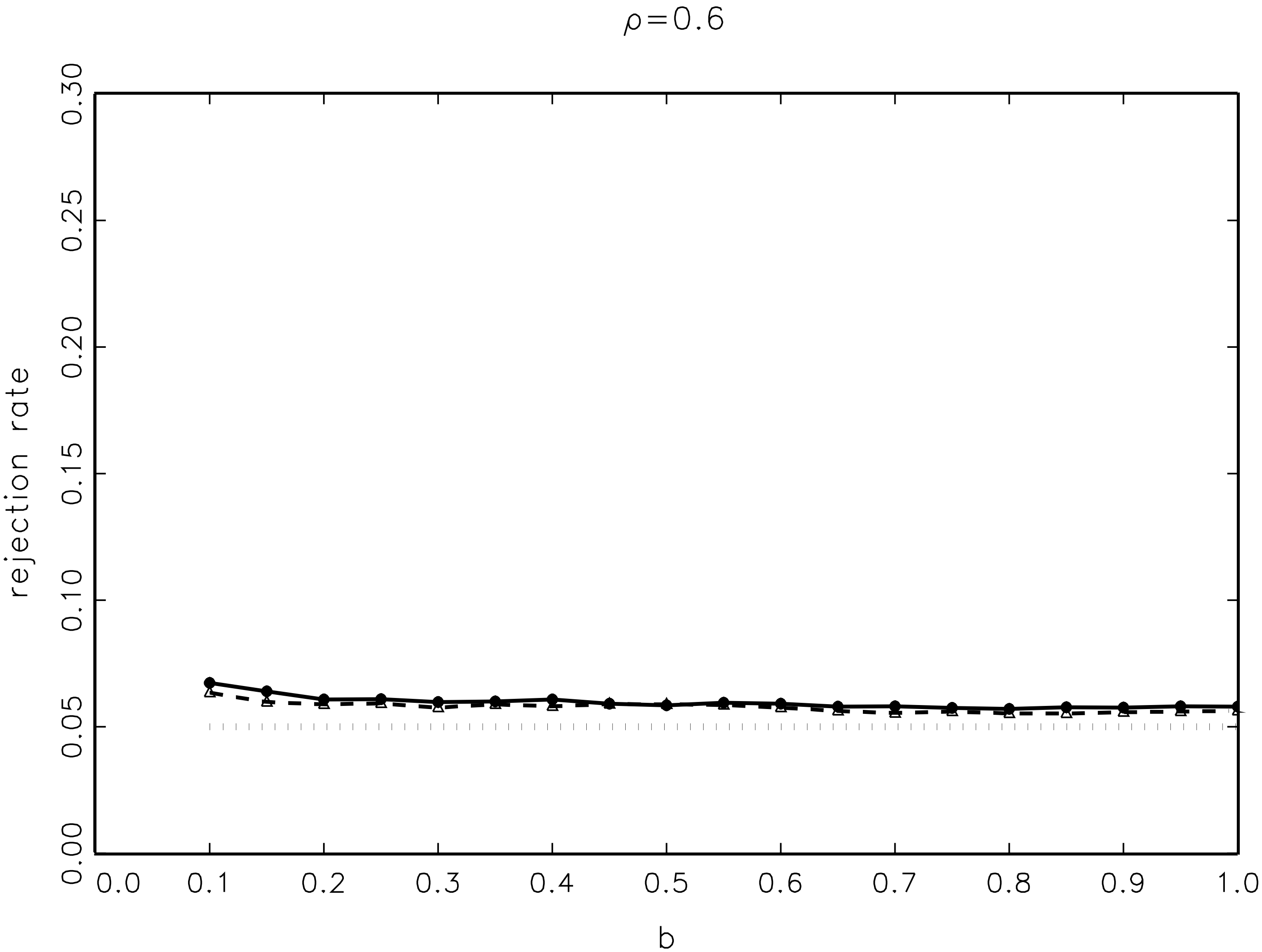


Figure 2.55: (cont'd)

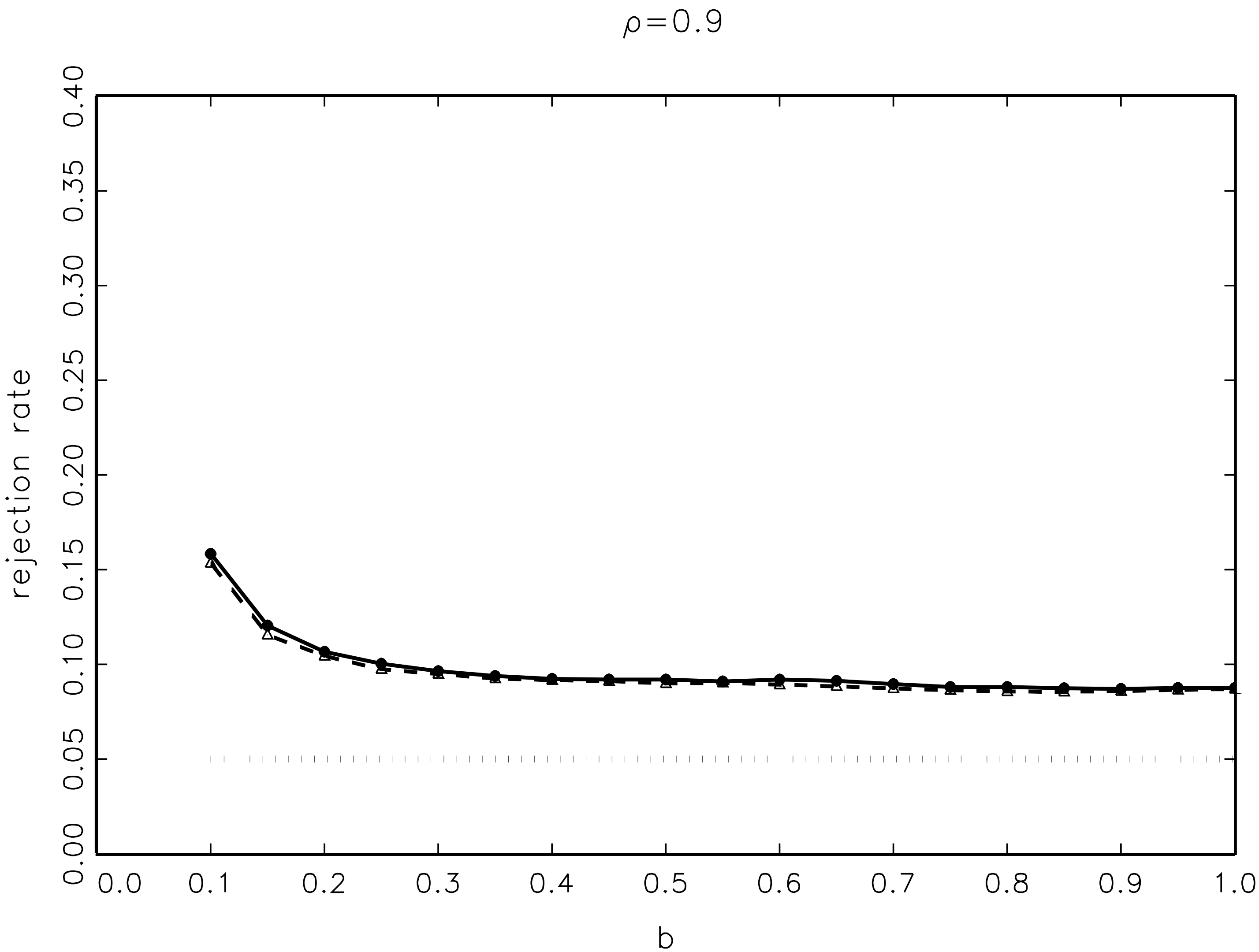


Figure 2.56: ES - Random Bernoulli ($p = 0.3$), Bartlett, $T = 50$

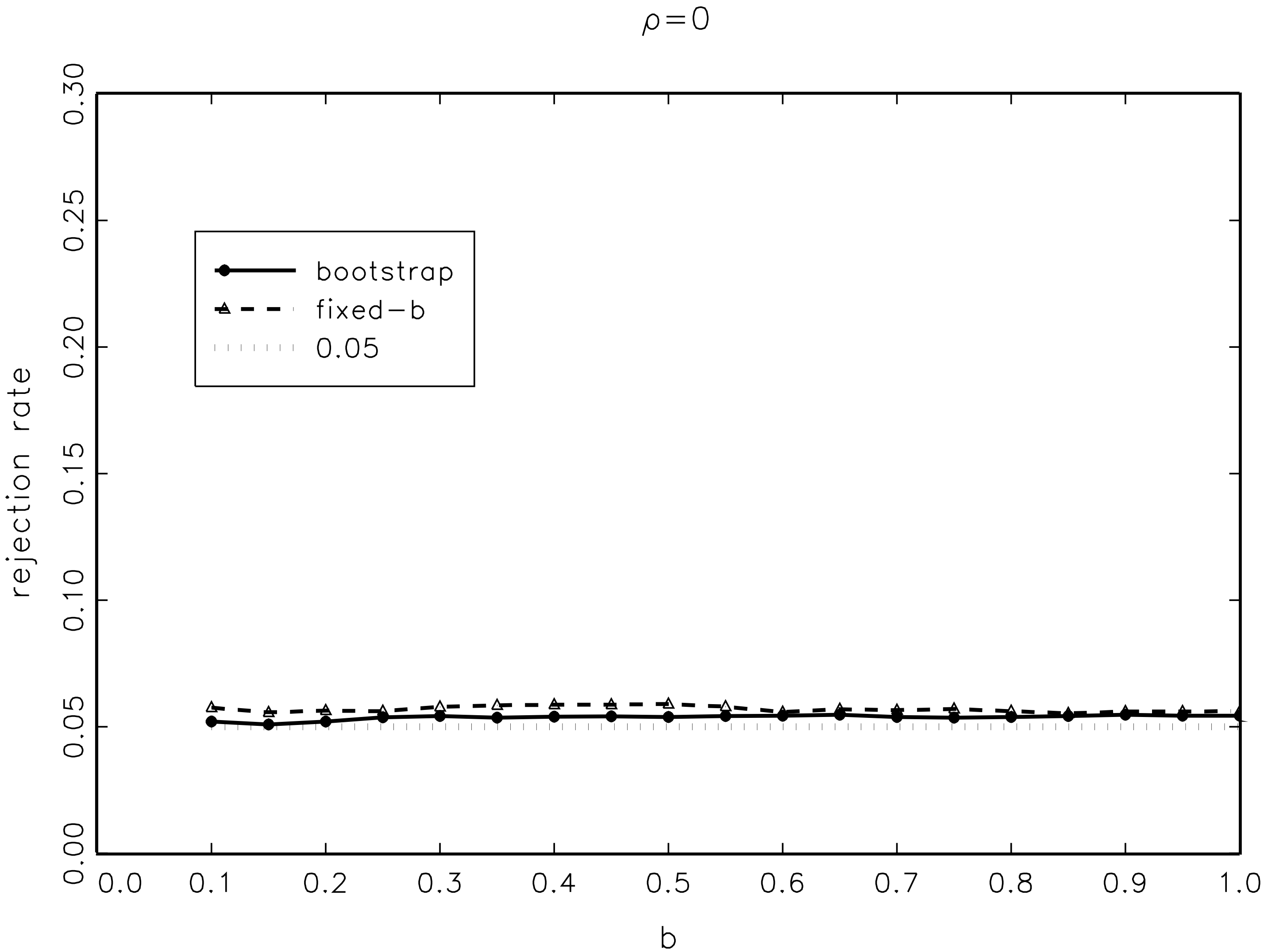


Figure 2.56: (cont'd)

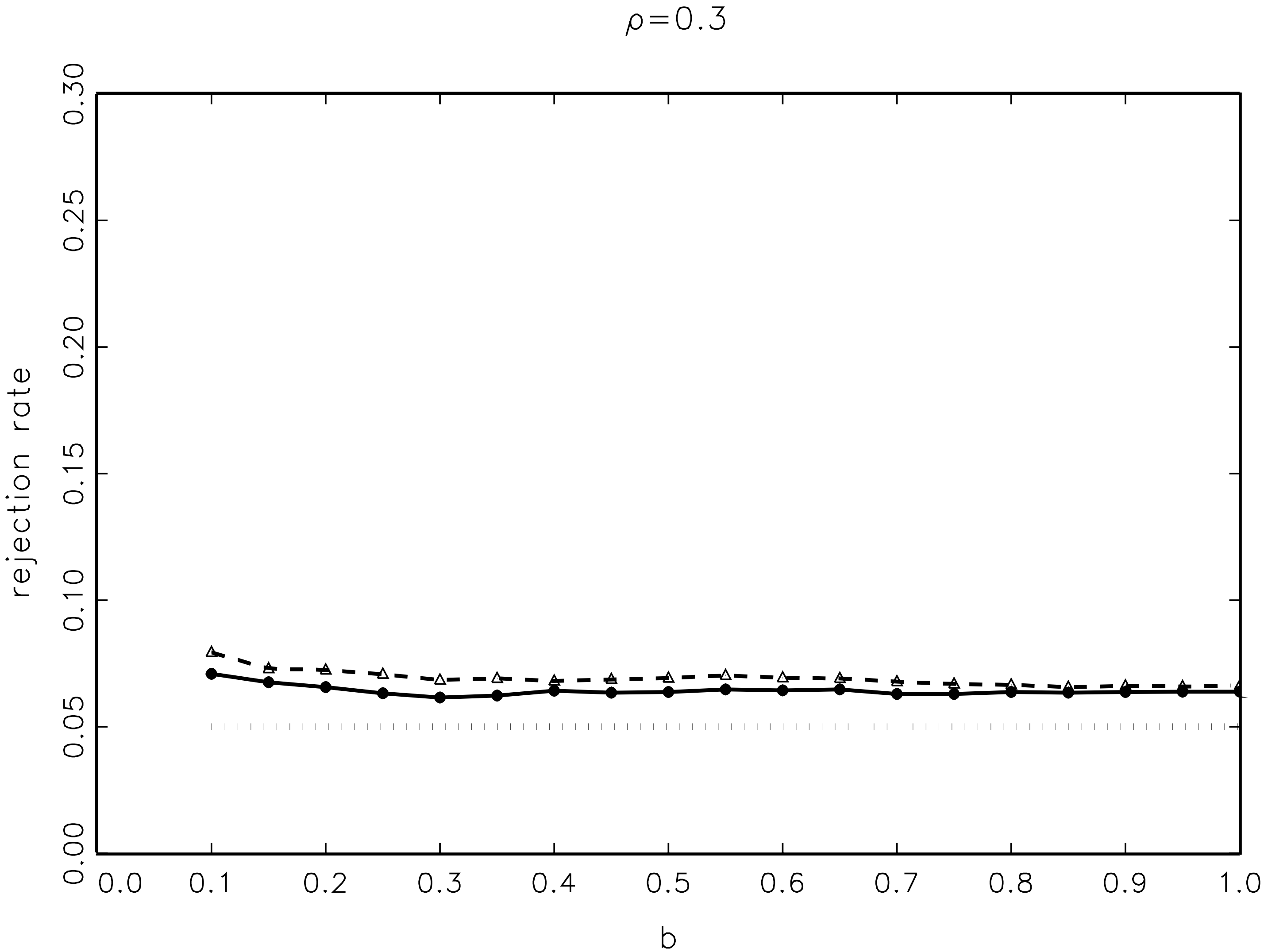


Figure 2.56: (cont'd)

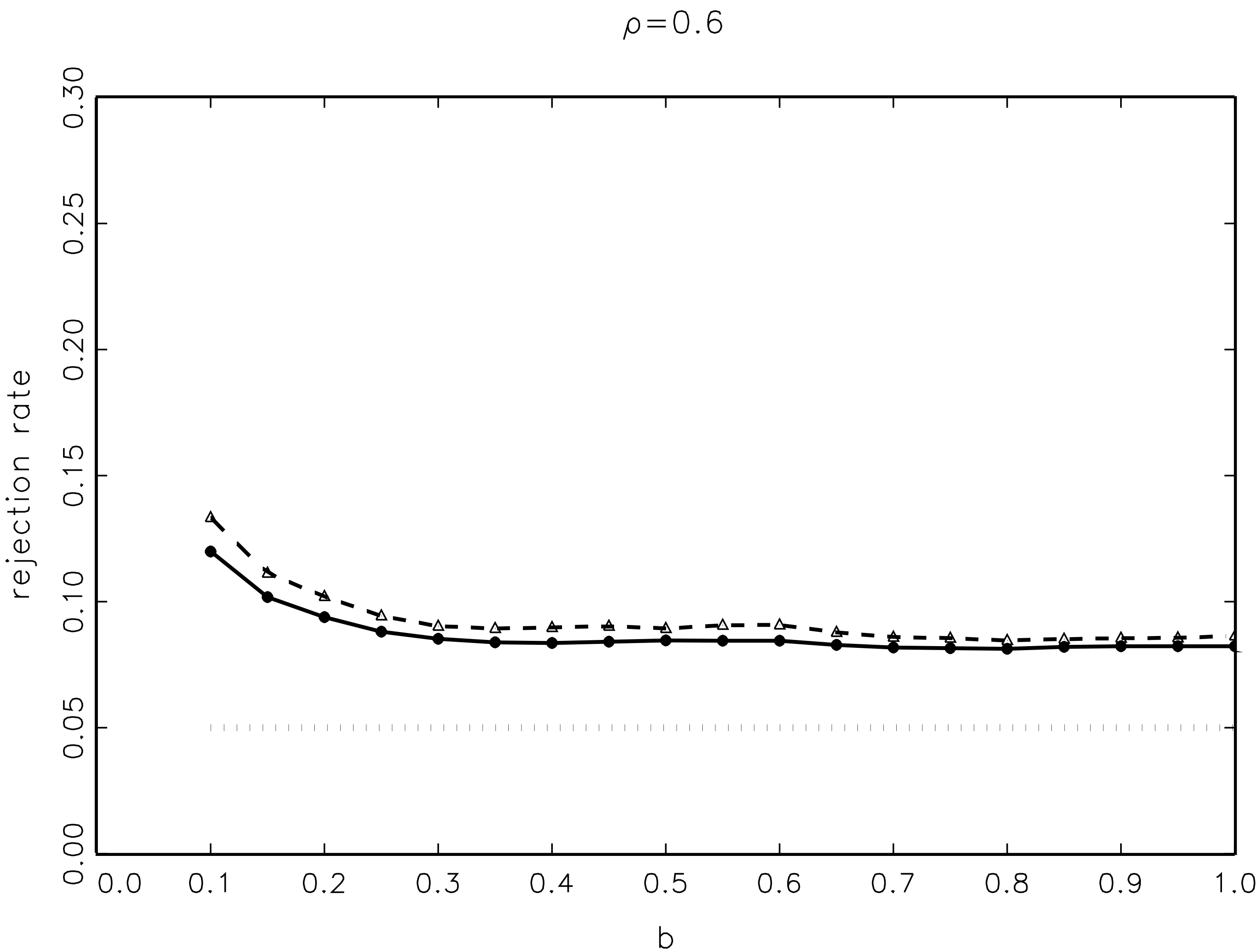


Figure 2.56: (cont'd)

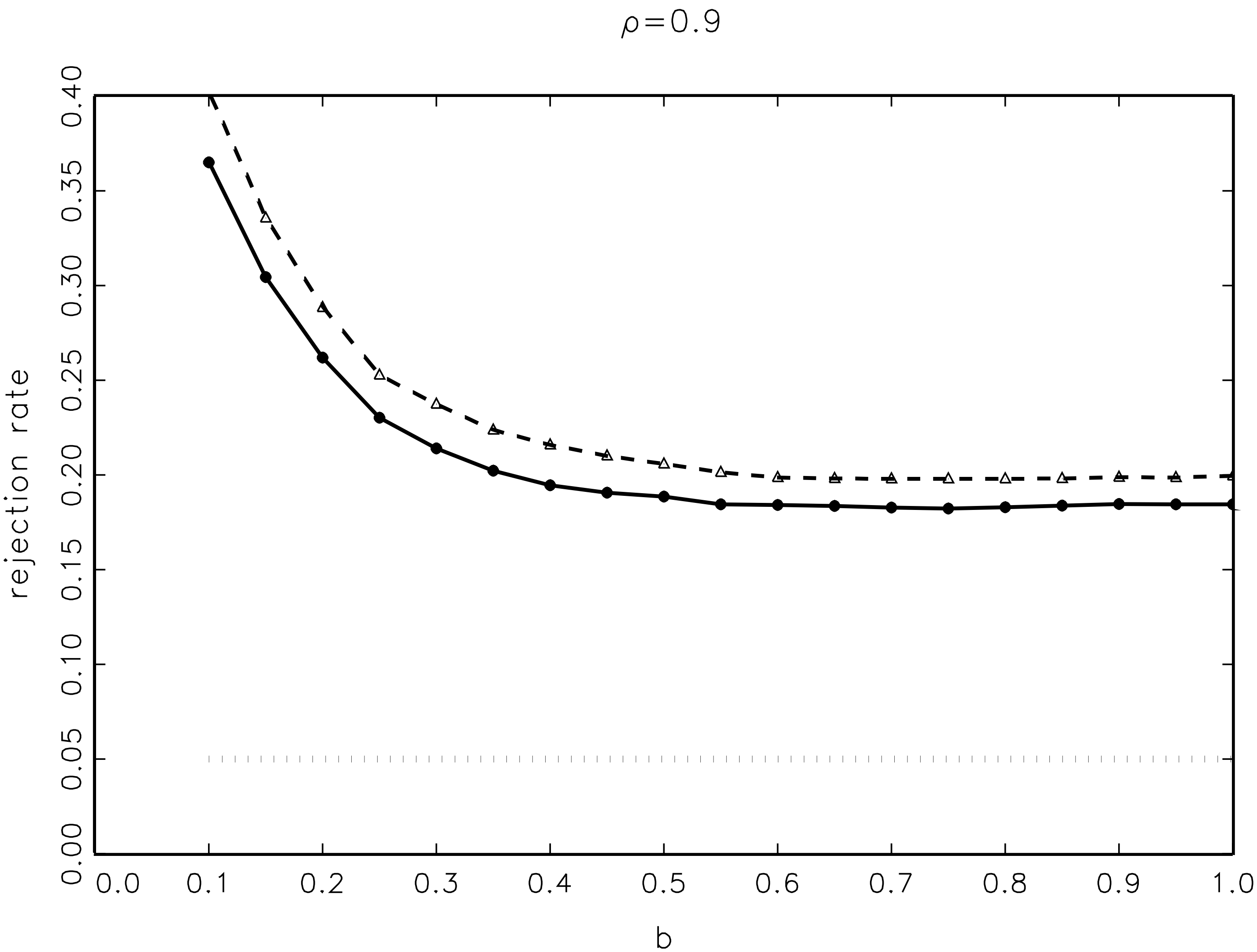


Figure 2.57: ES - Random Bernoulli ($p = 0.5$), Bartlett, $T = 50$

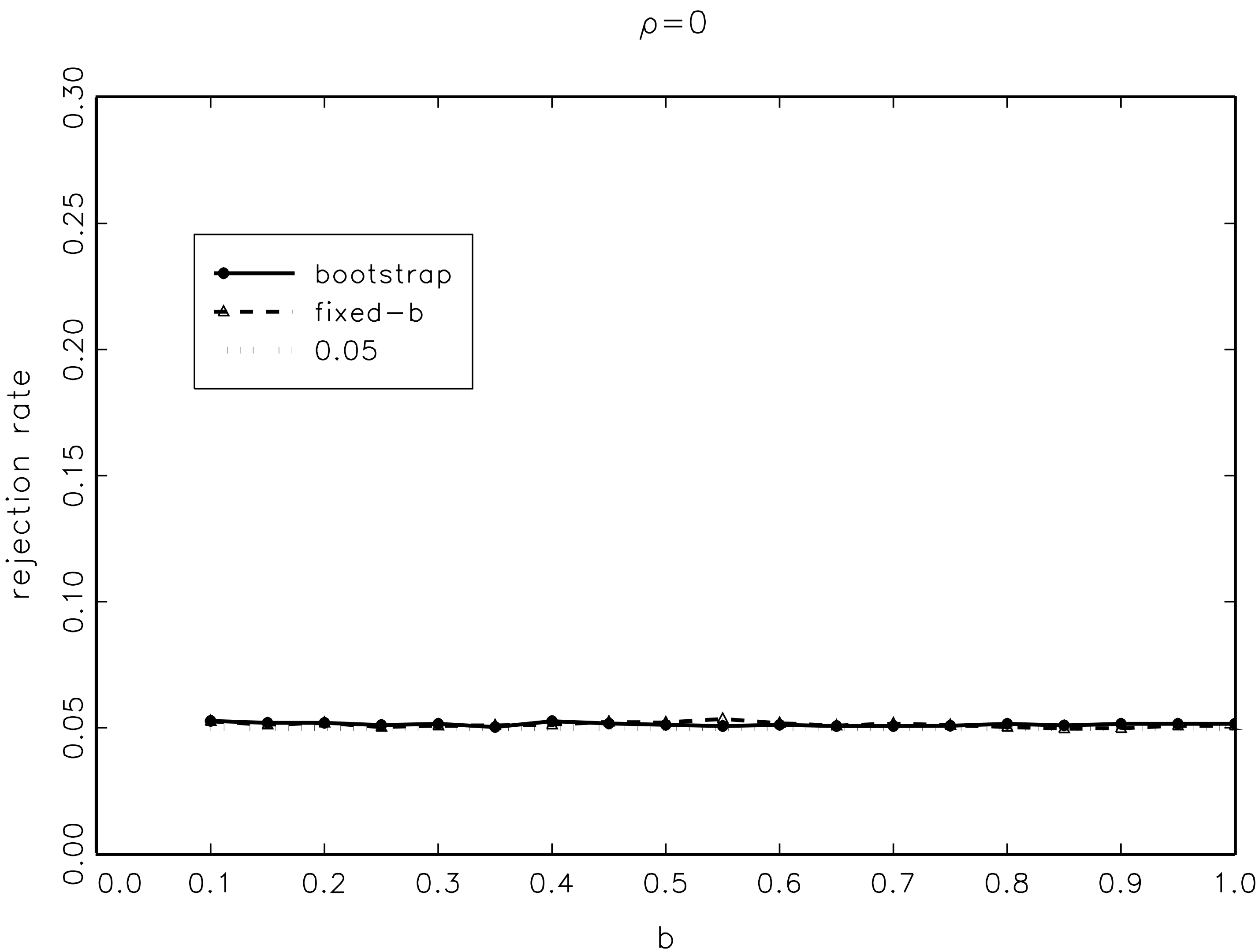


Figure 2.57: (cont'd)

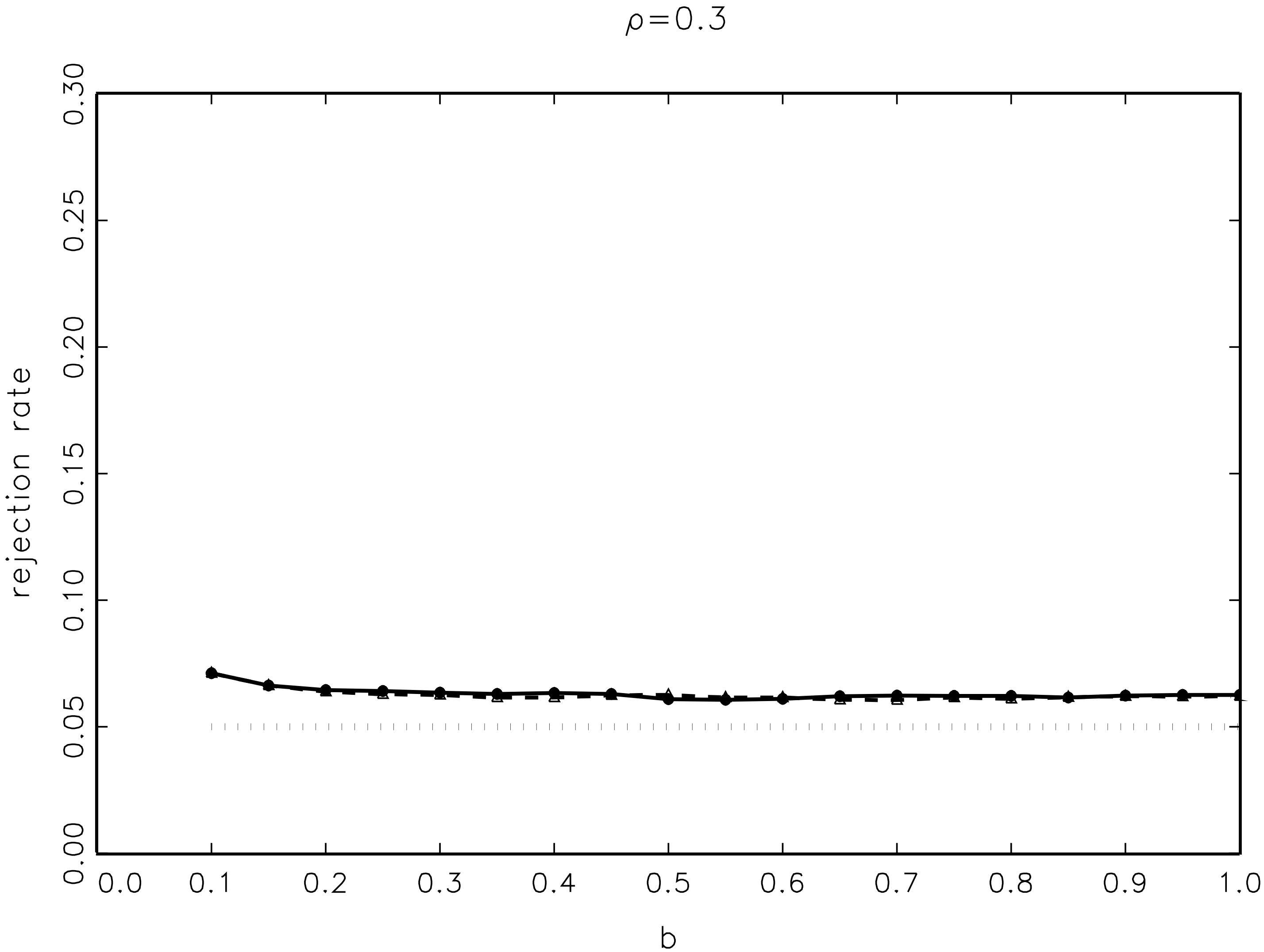


Figure 2.57: (cont'd)

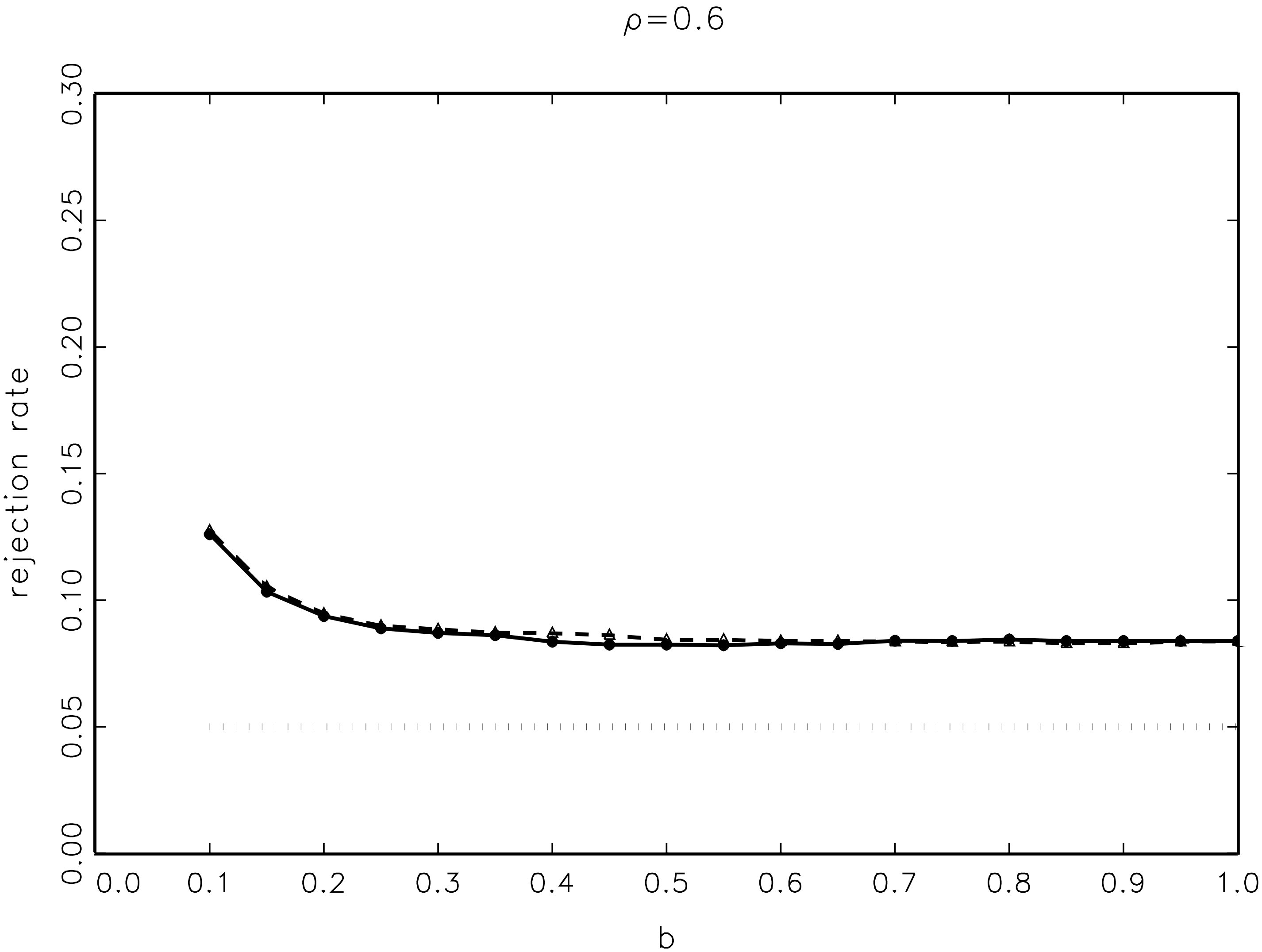


Figure 2.57: (cont'd)

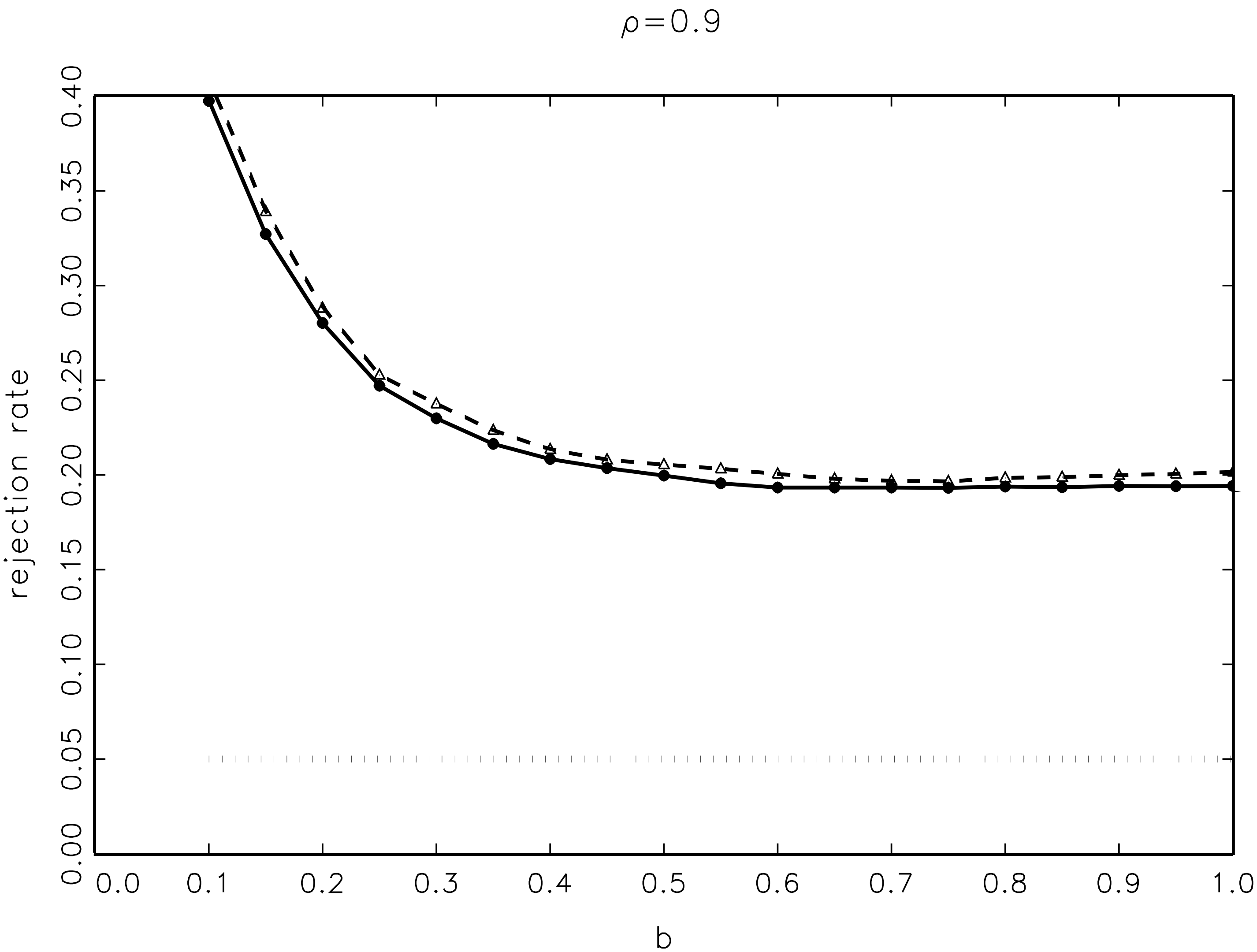


Figure 2.58: ES - Random Bernoulli ($p = 0.7$), Bartlett, $T = 50$

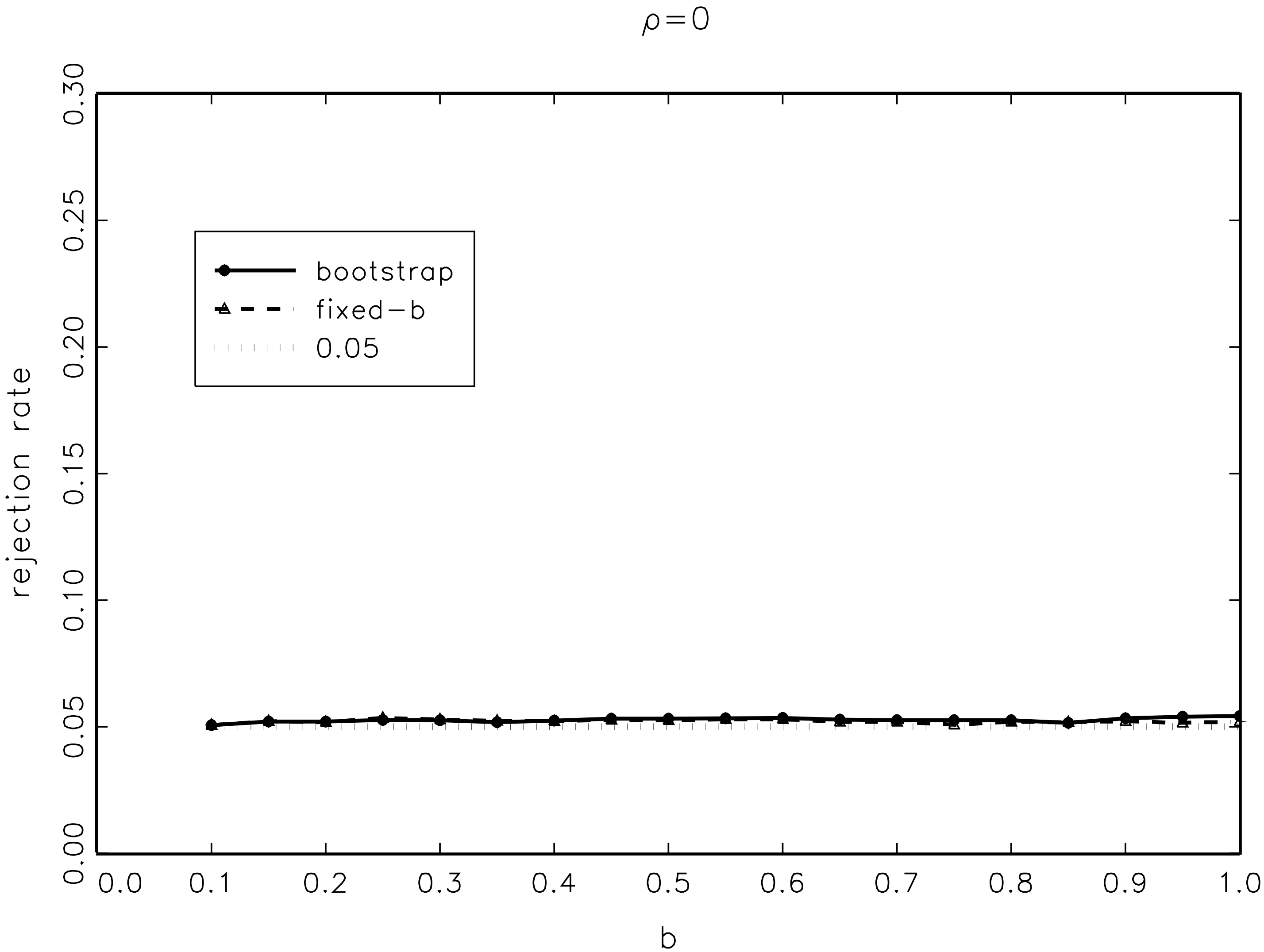


Figure 2.58: (cont'd)

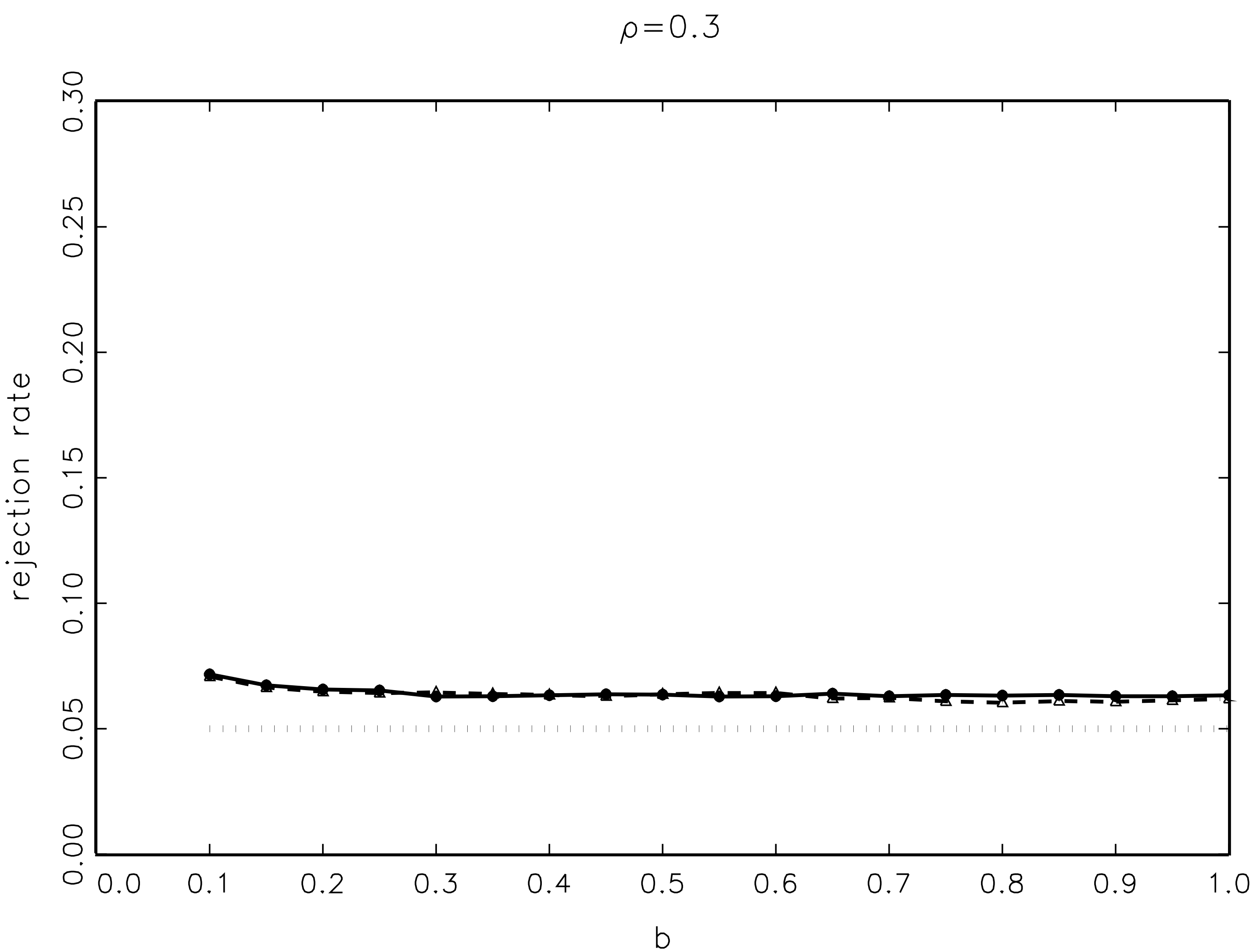


Figure 2.58: (cont'd)

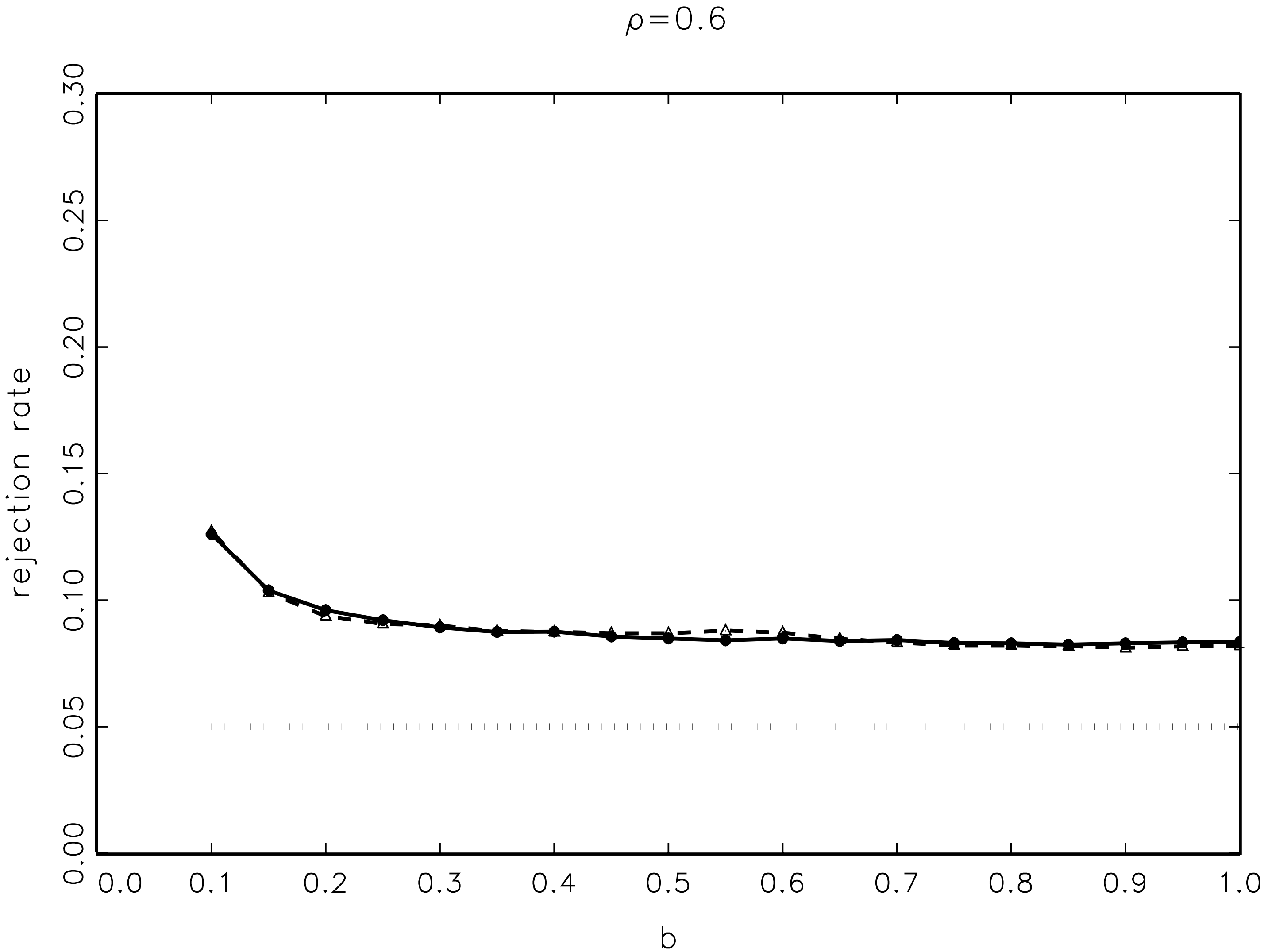


Figure 2.58: (cont'd)

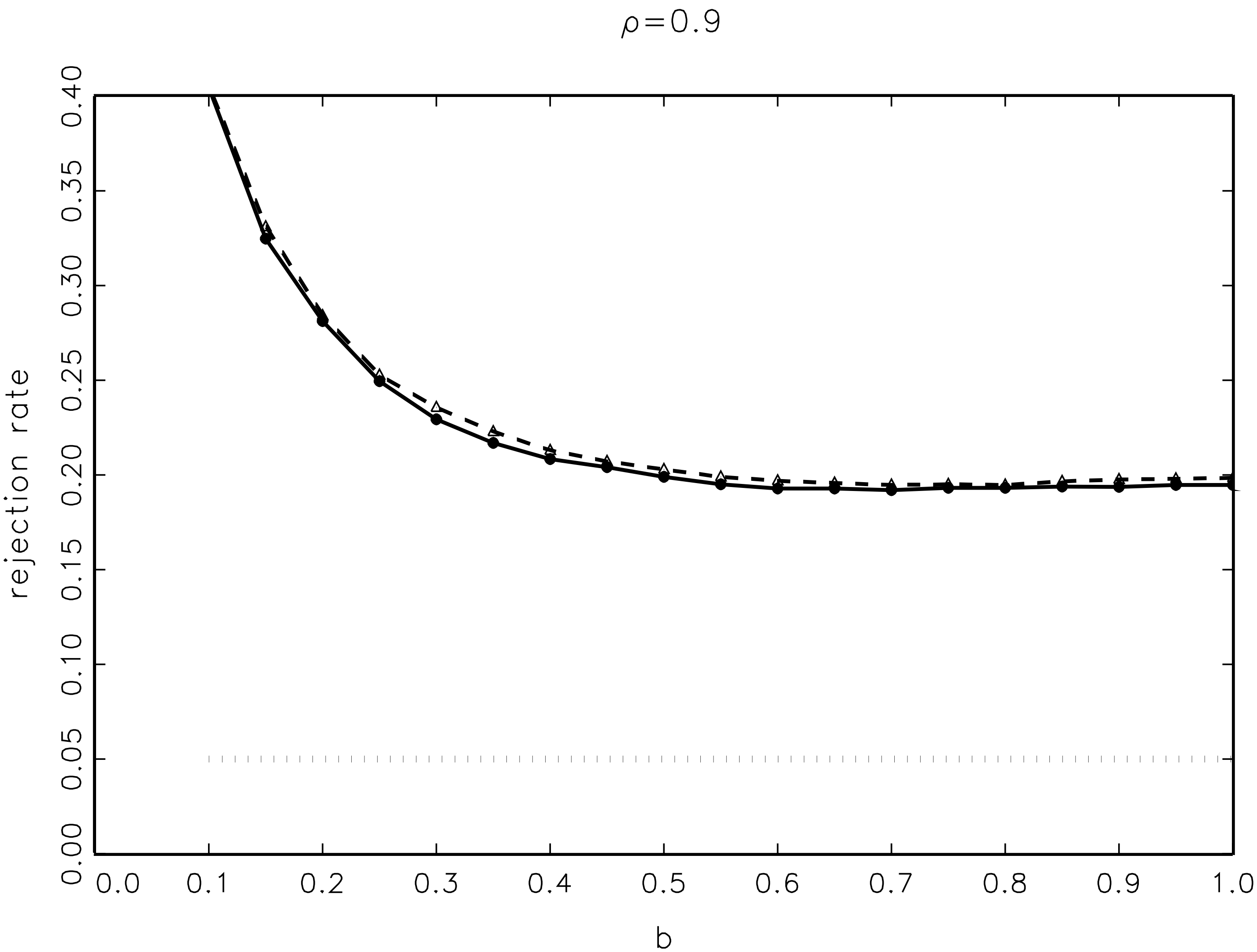


Figure 2.59: ES - Random Bernoulli ($p = 0.3$), Bartlett, $T = 100$

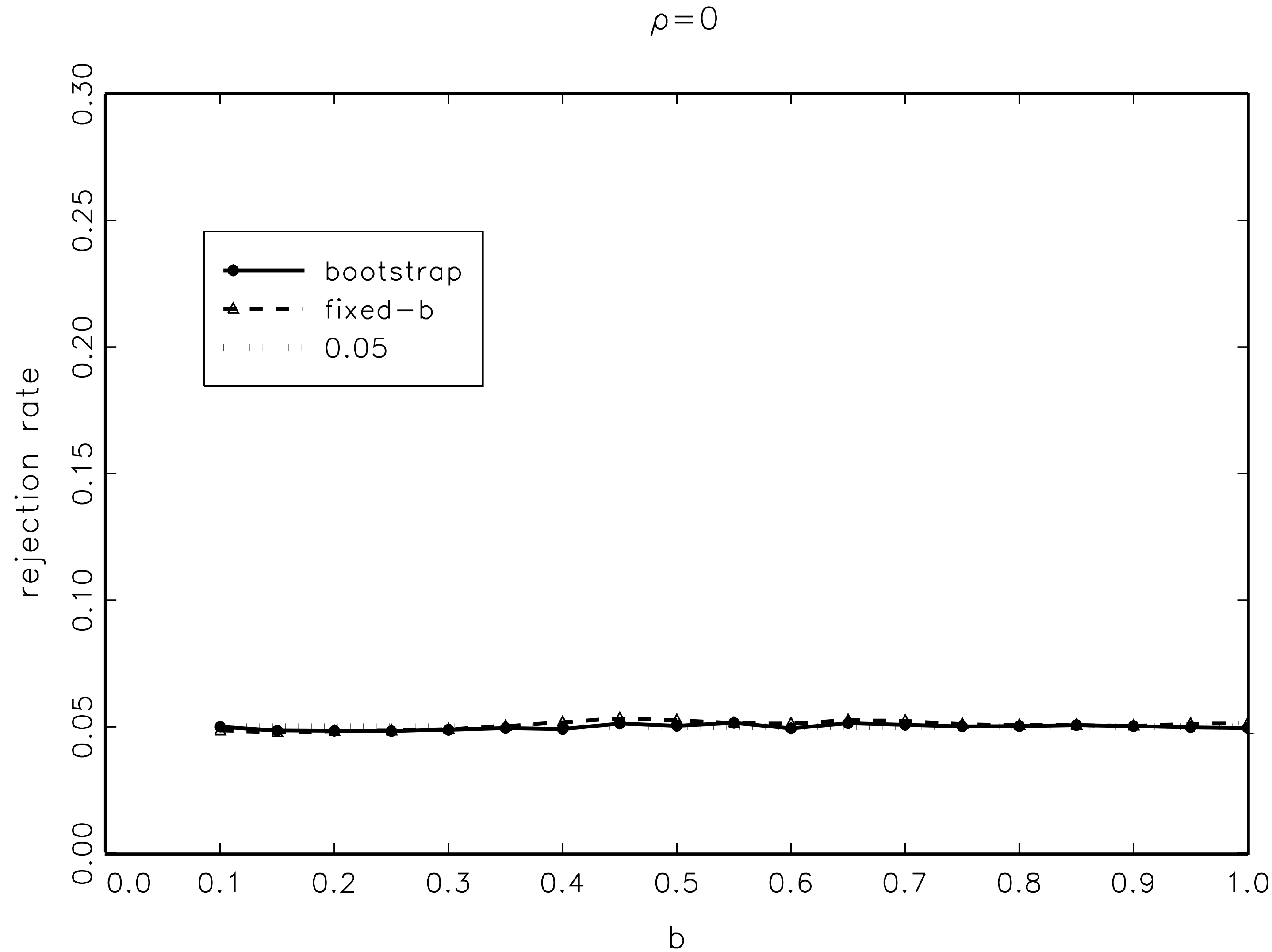


Figure 2.59: (cont'd)

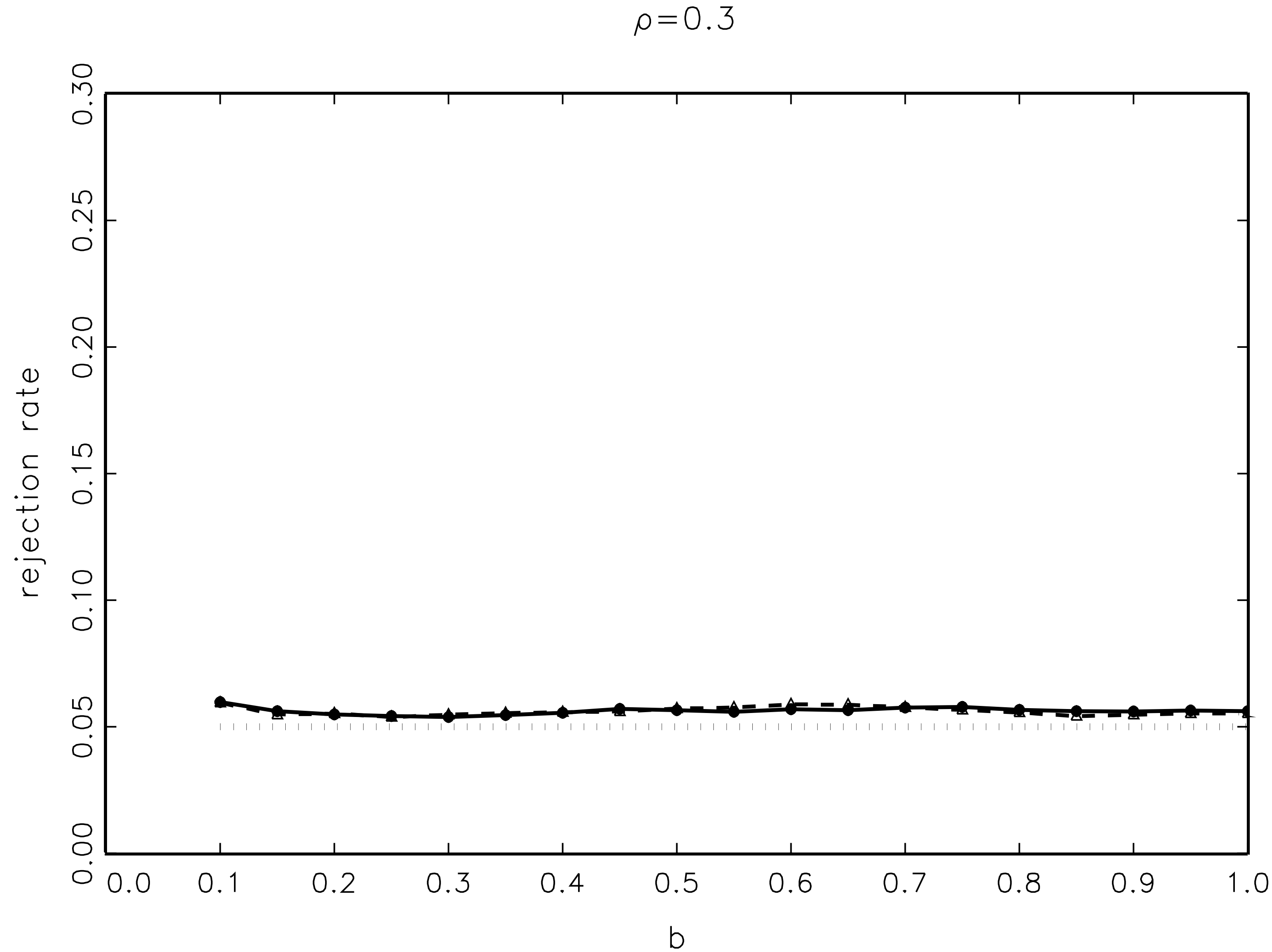


Figure 2.59: (cont'd)

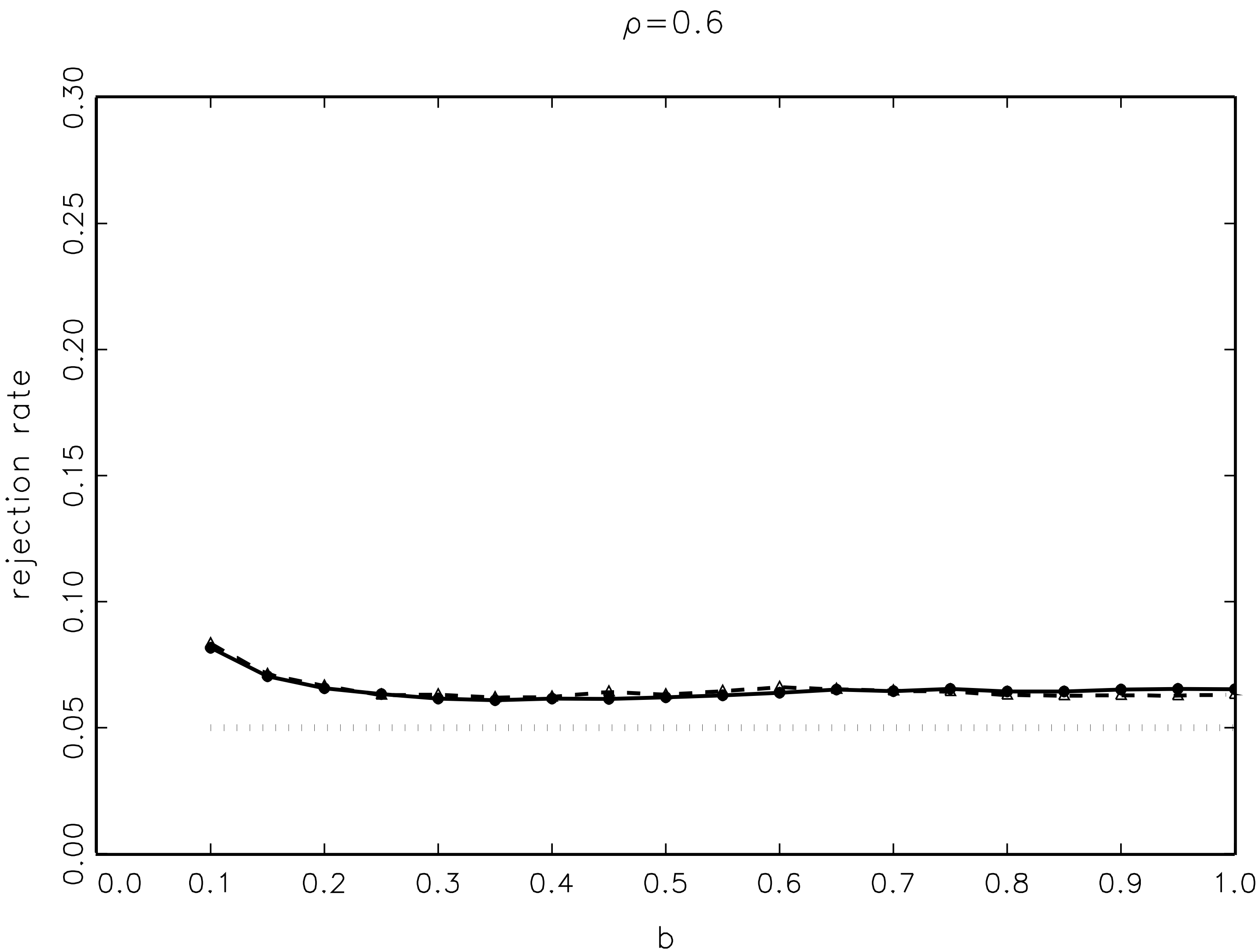


Figure 2.59: (cont'd)

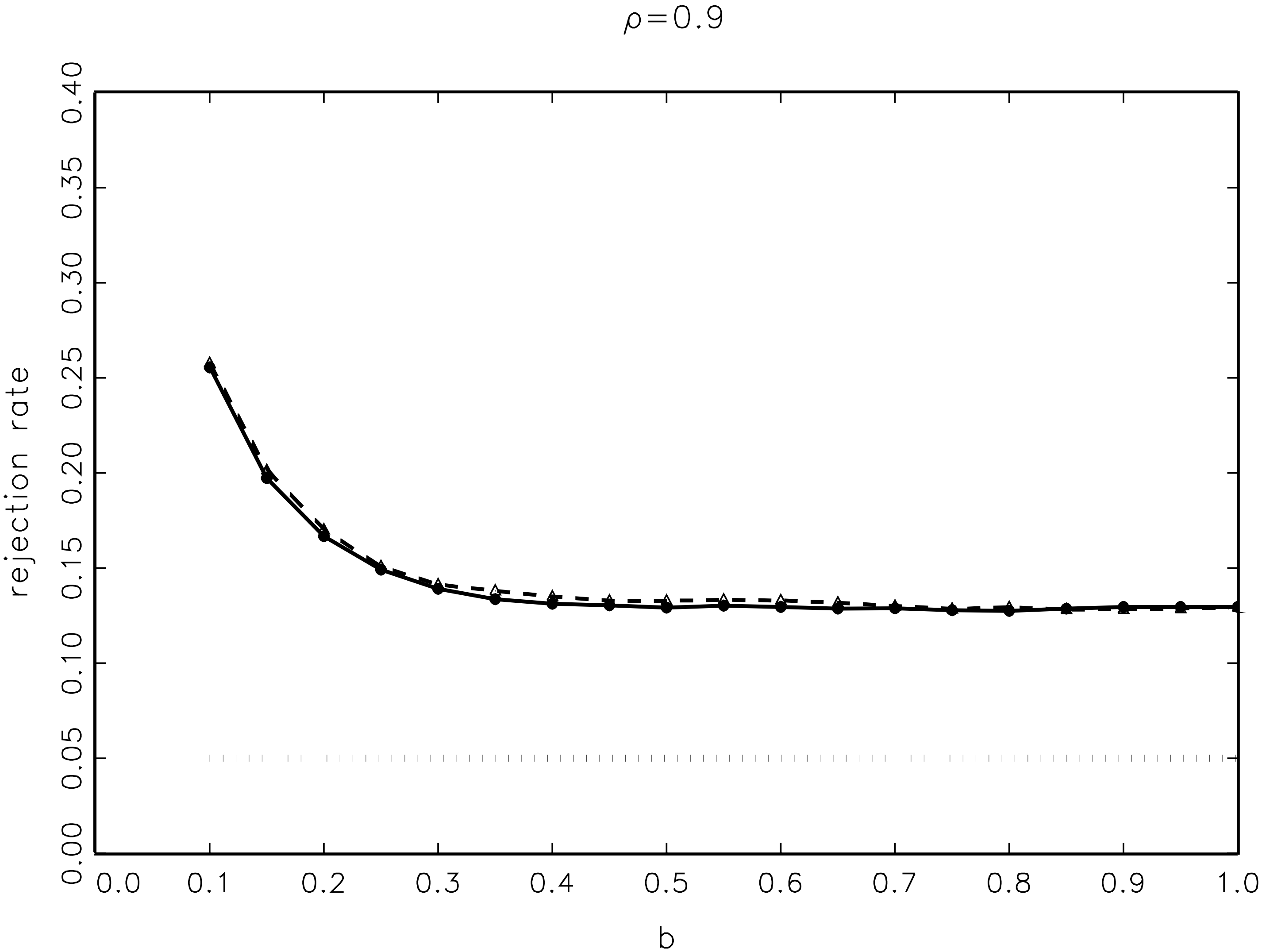


Figure 2.60: ES - Random Bernoulli ($p = 0.5$), Bartlett, $T = 100$

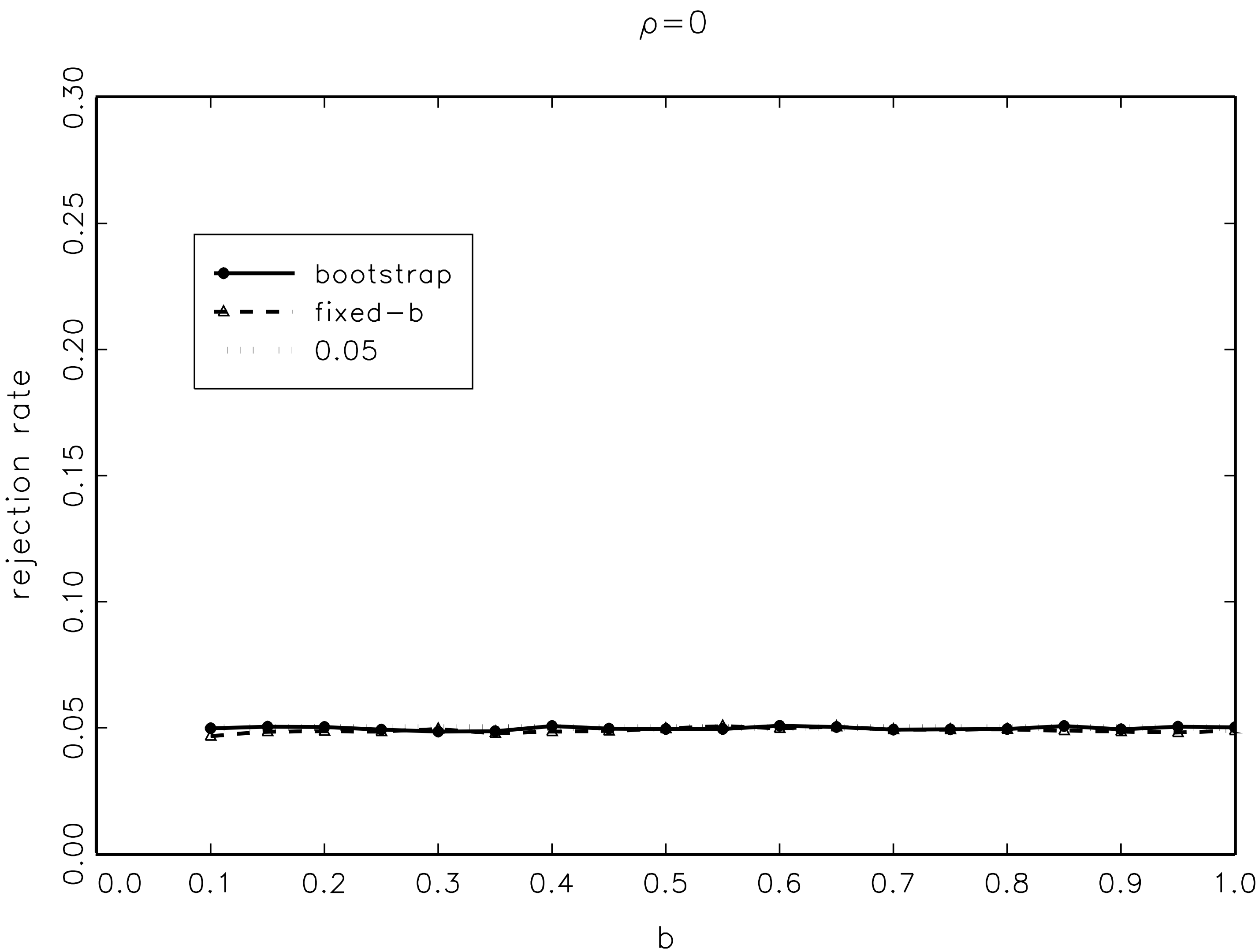


Figure 2.60: (cont'd)

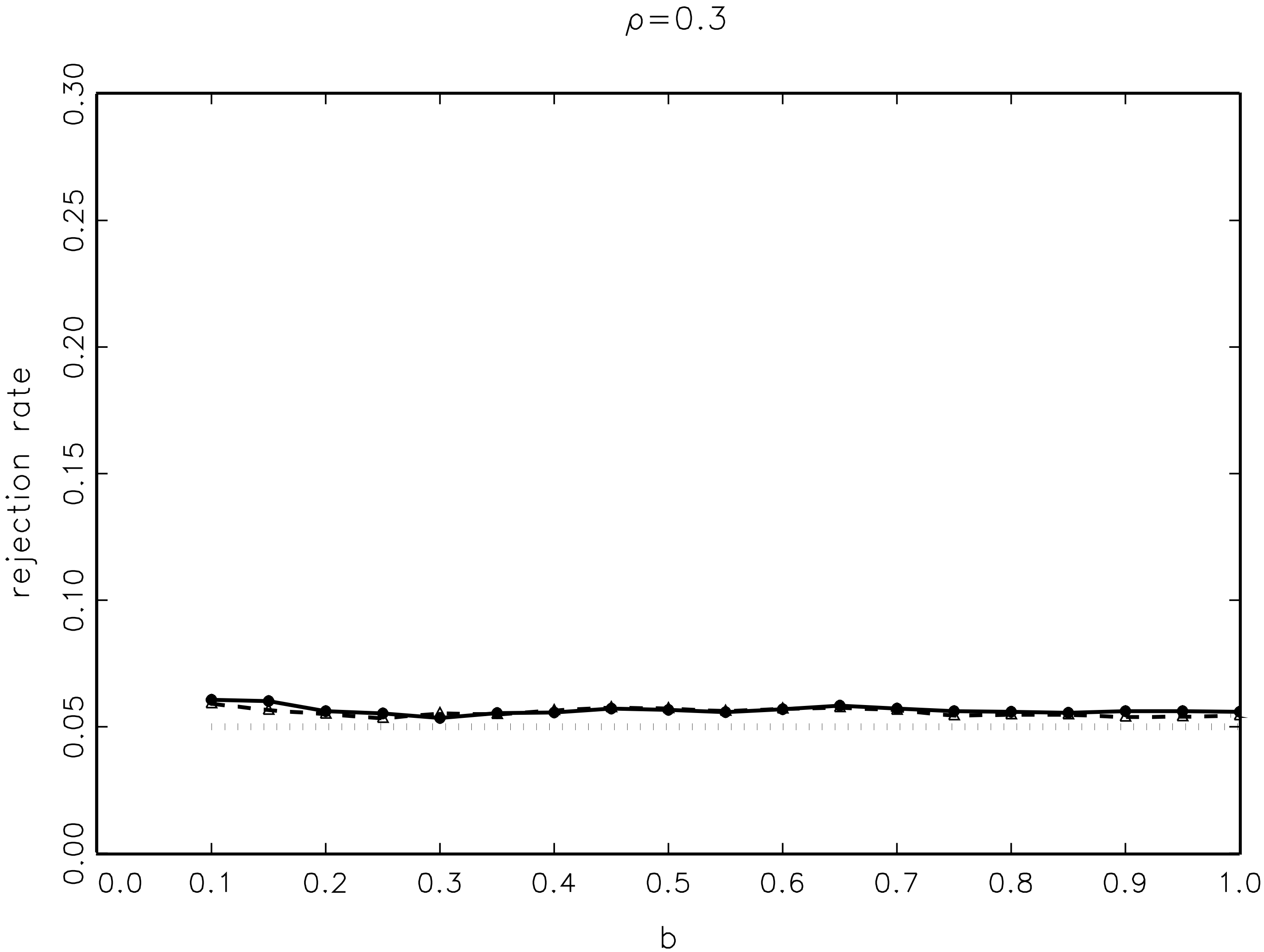


Figure 2.60: (cont'd)

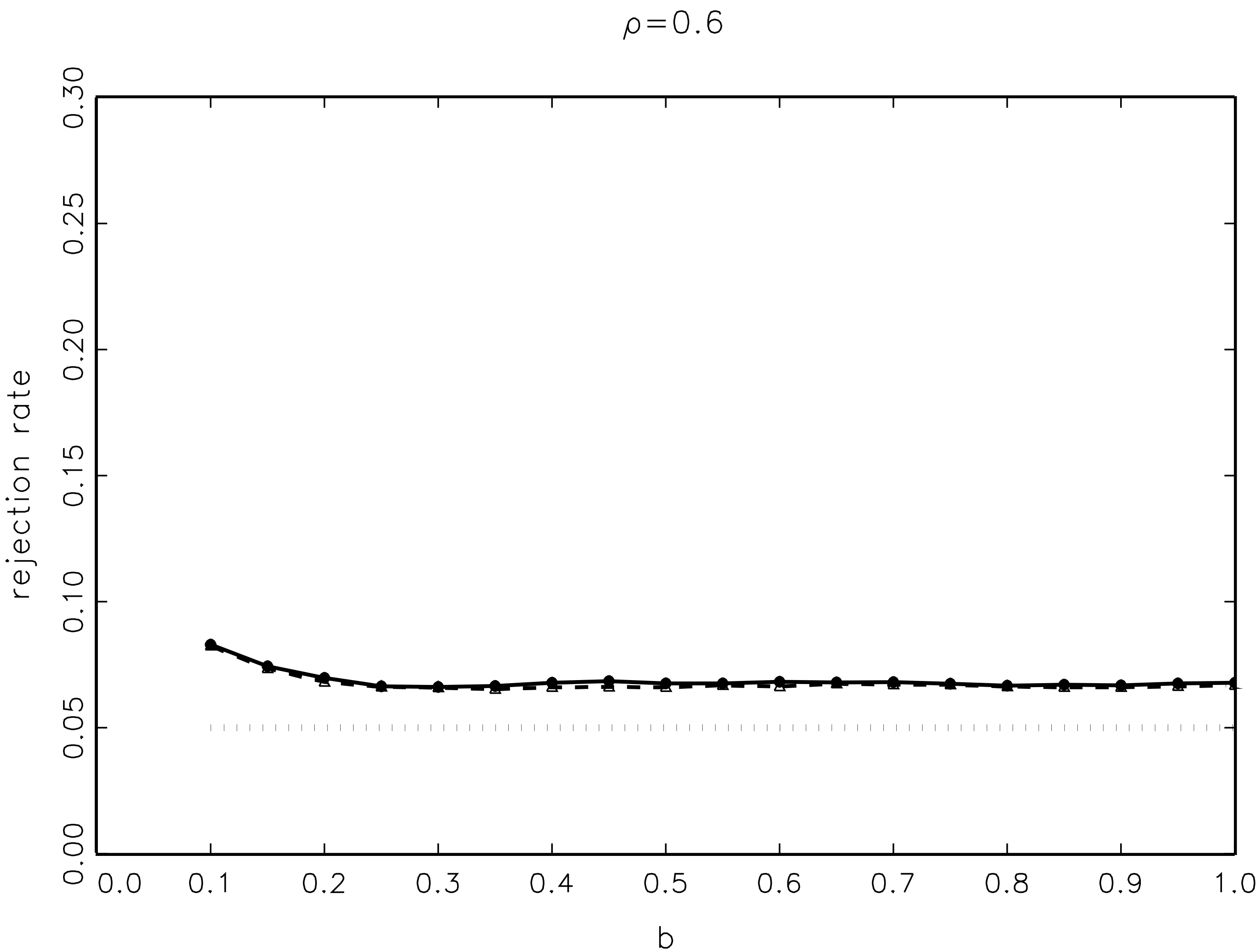


Figure 2.60: (cont'd)

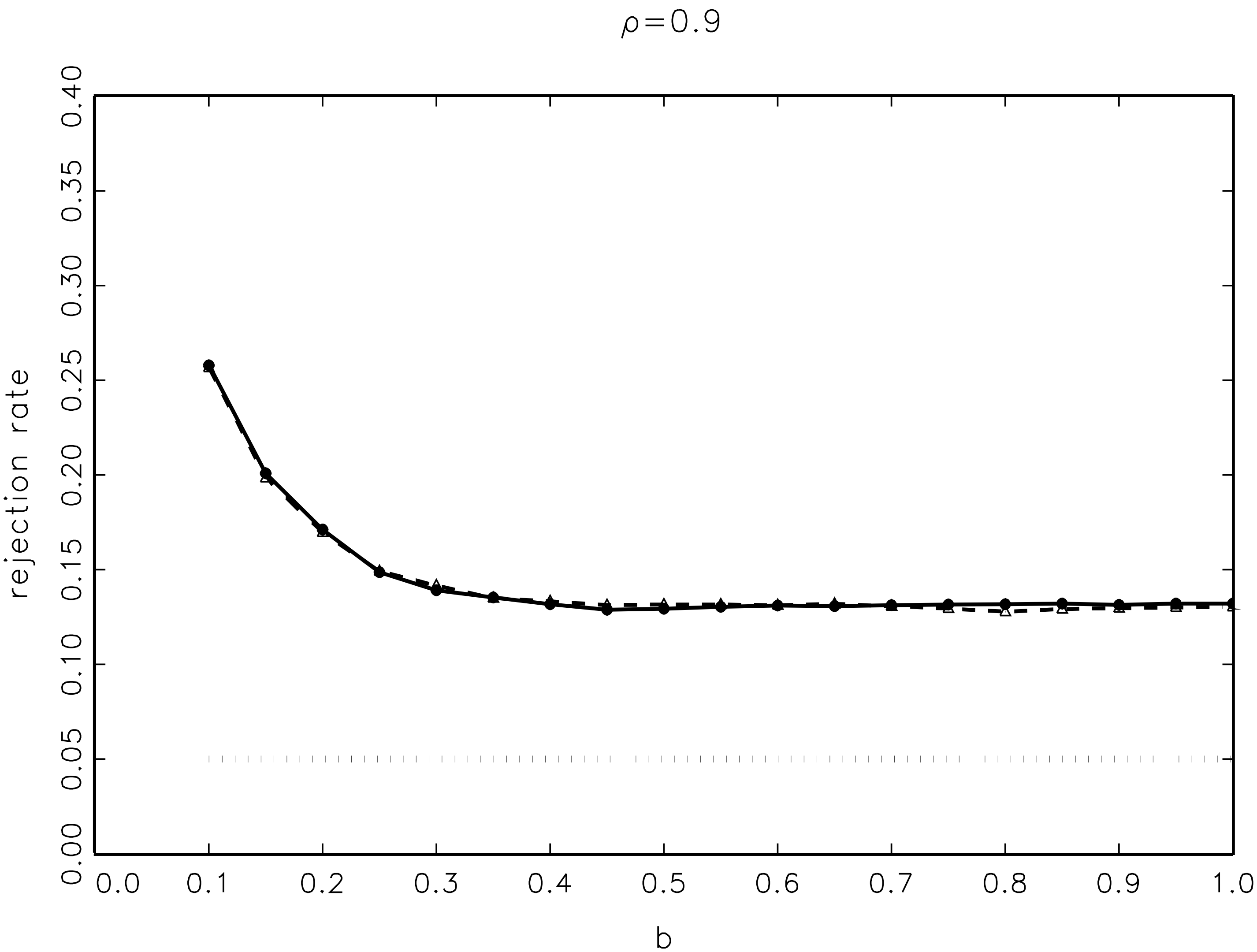


Figure 2.61: ES - Random Bernoulli ($p = 0.7$), Bartlett, $T = 100$

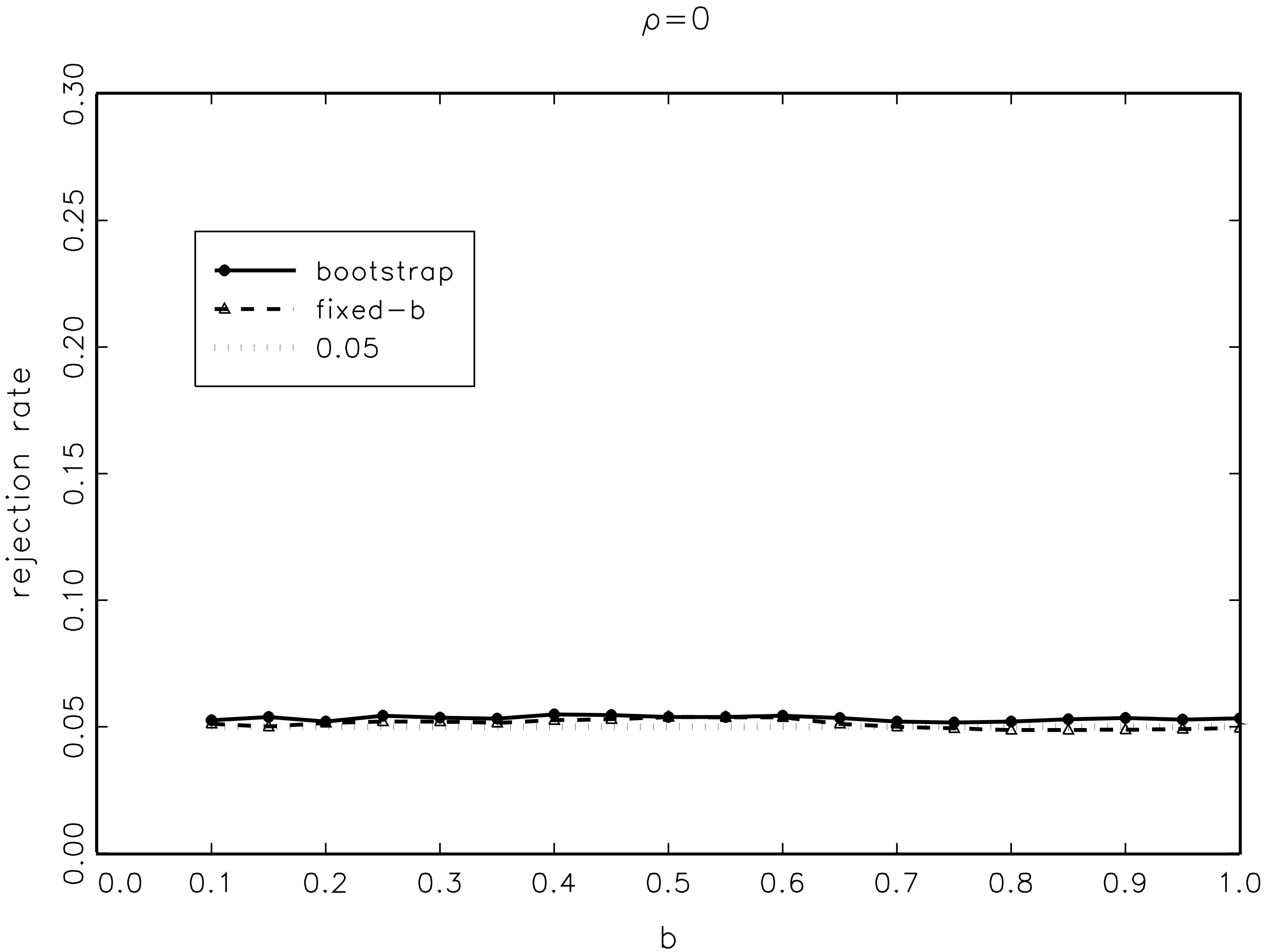


Figure 2.61: (cont'd)

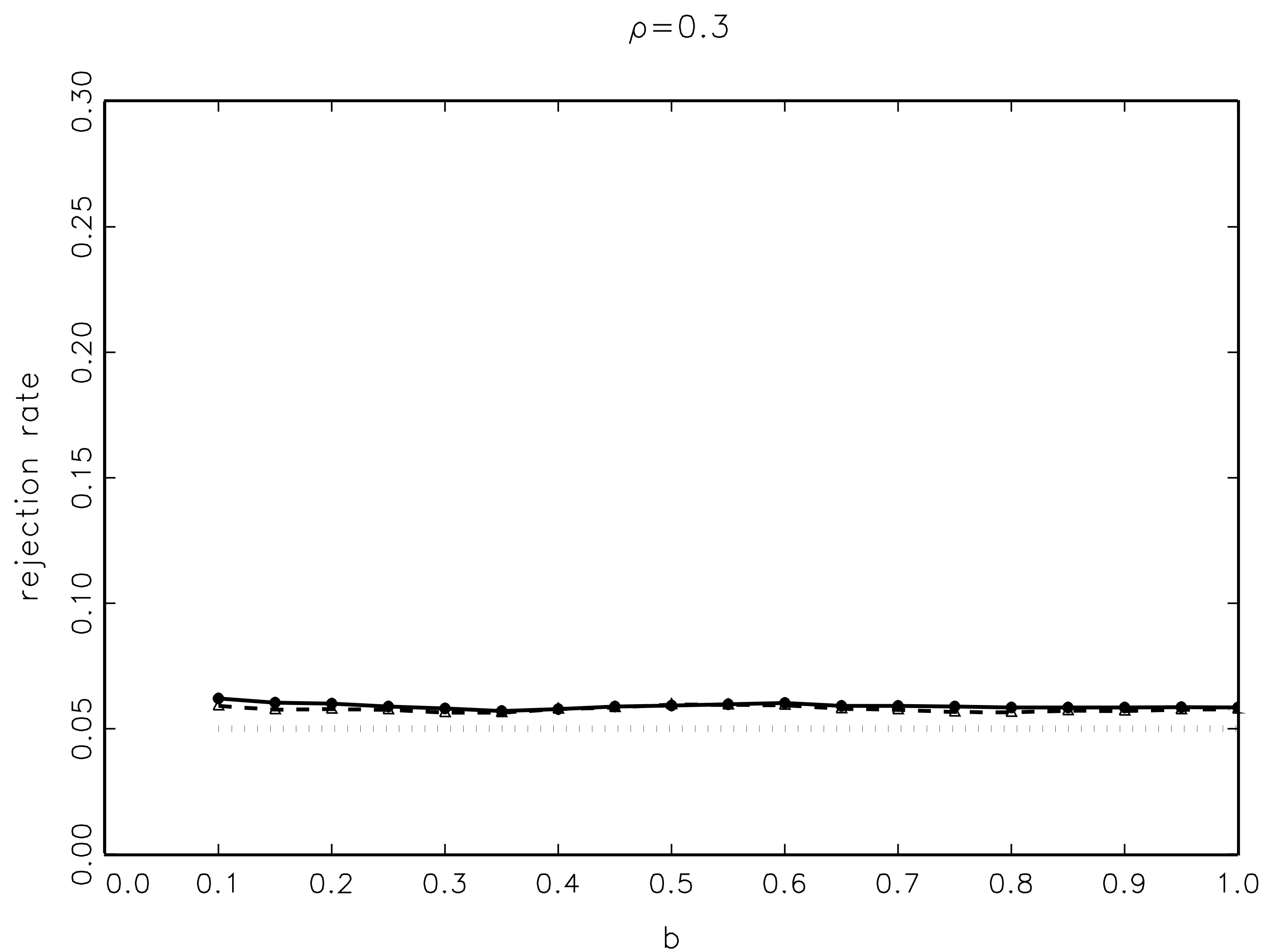


Figure 2.61: (cont'd)

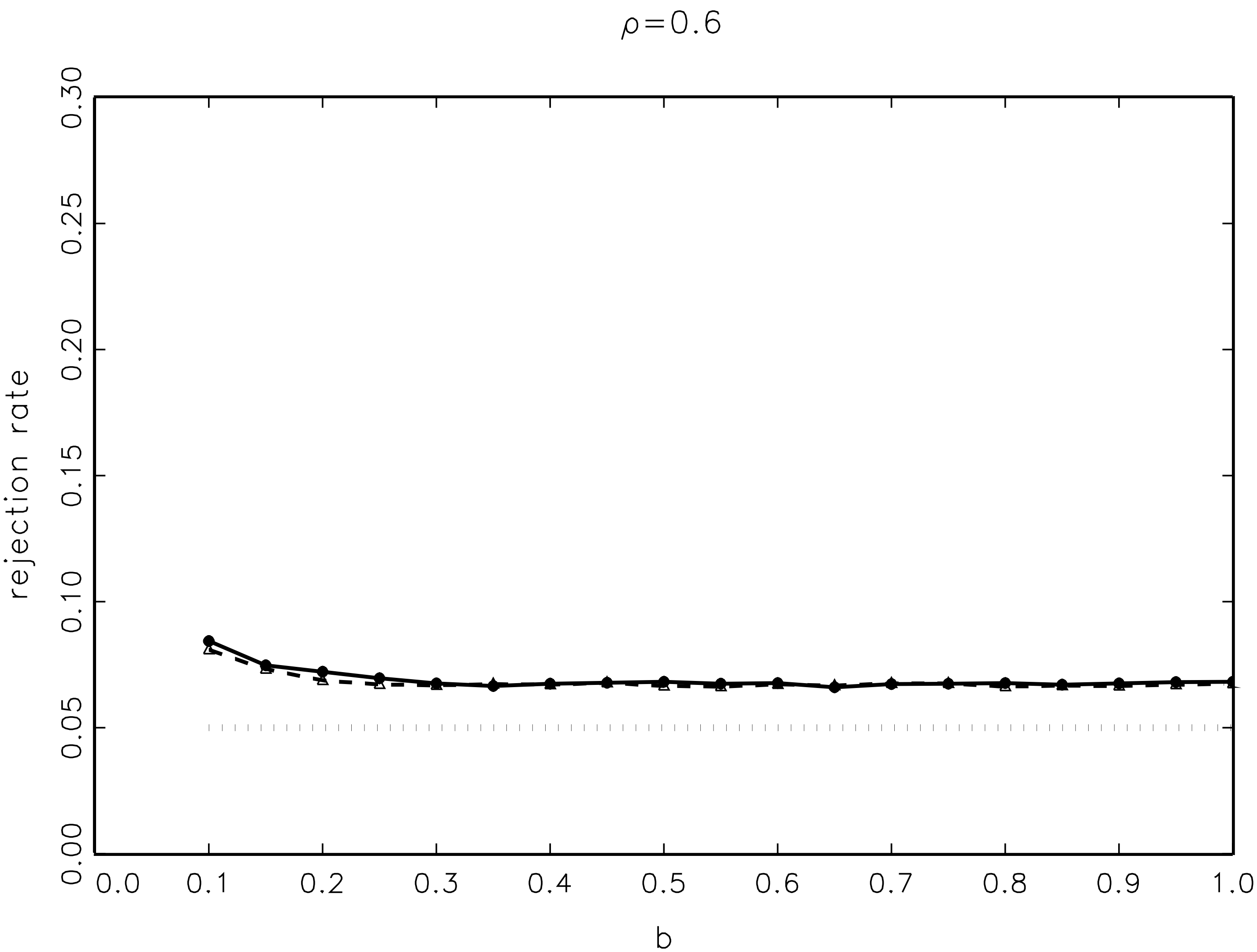


Figure 2.61: (cont'd)

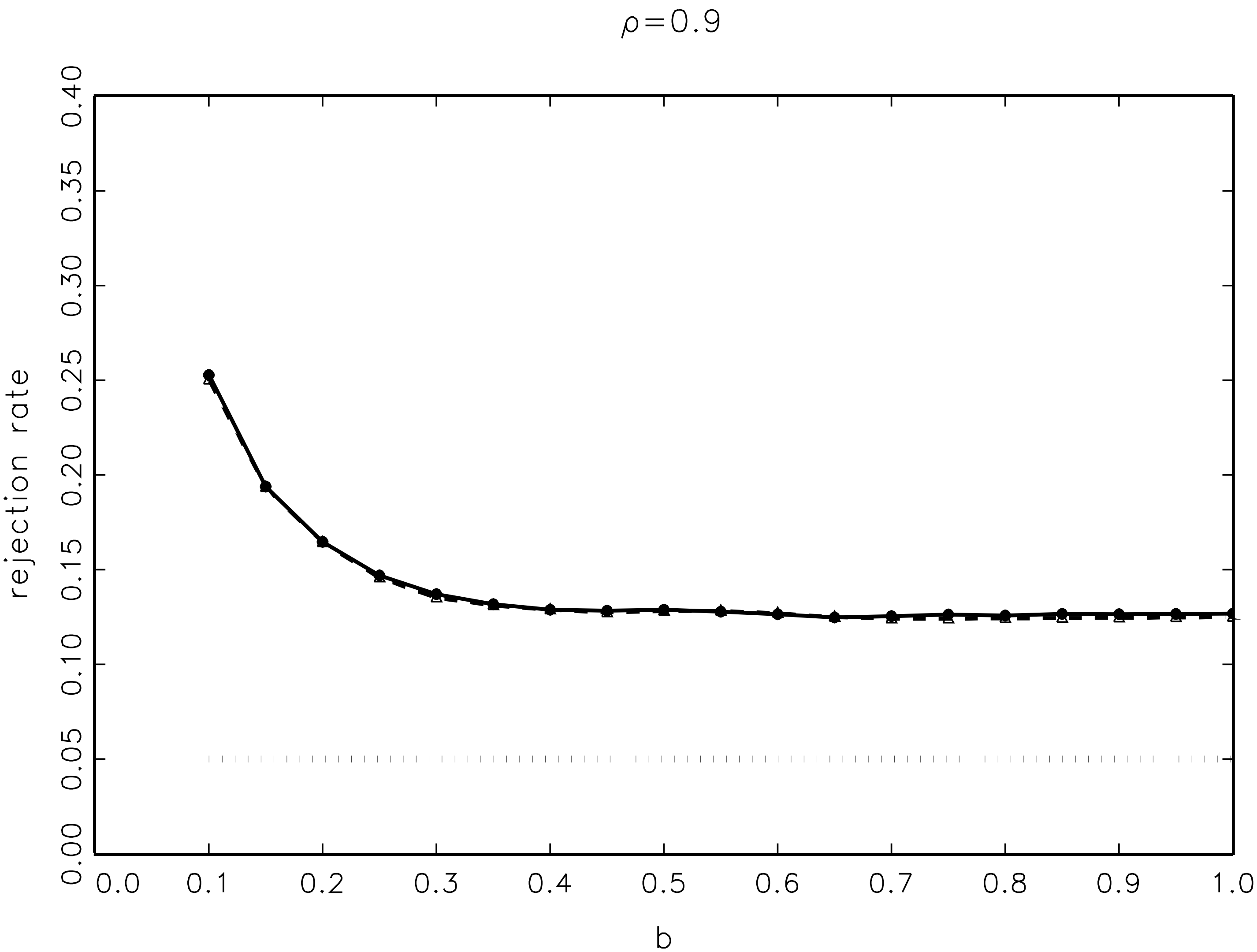


Figure 2.62: ES - Random Bernoulli ($p = 0.3$), Bartlett, $T = 200$

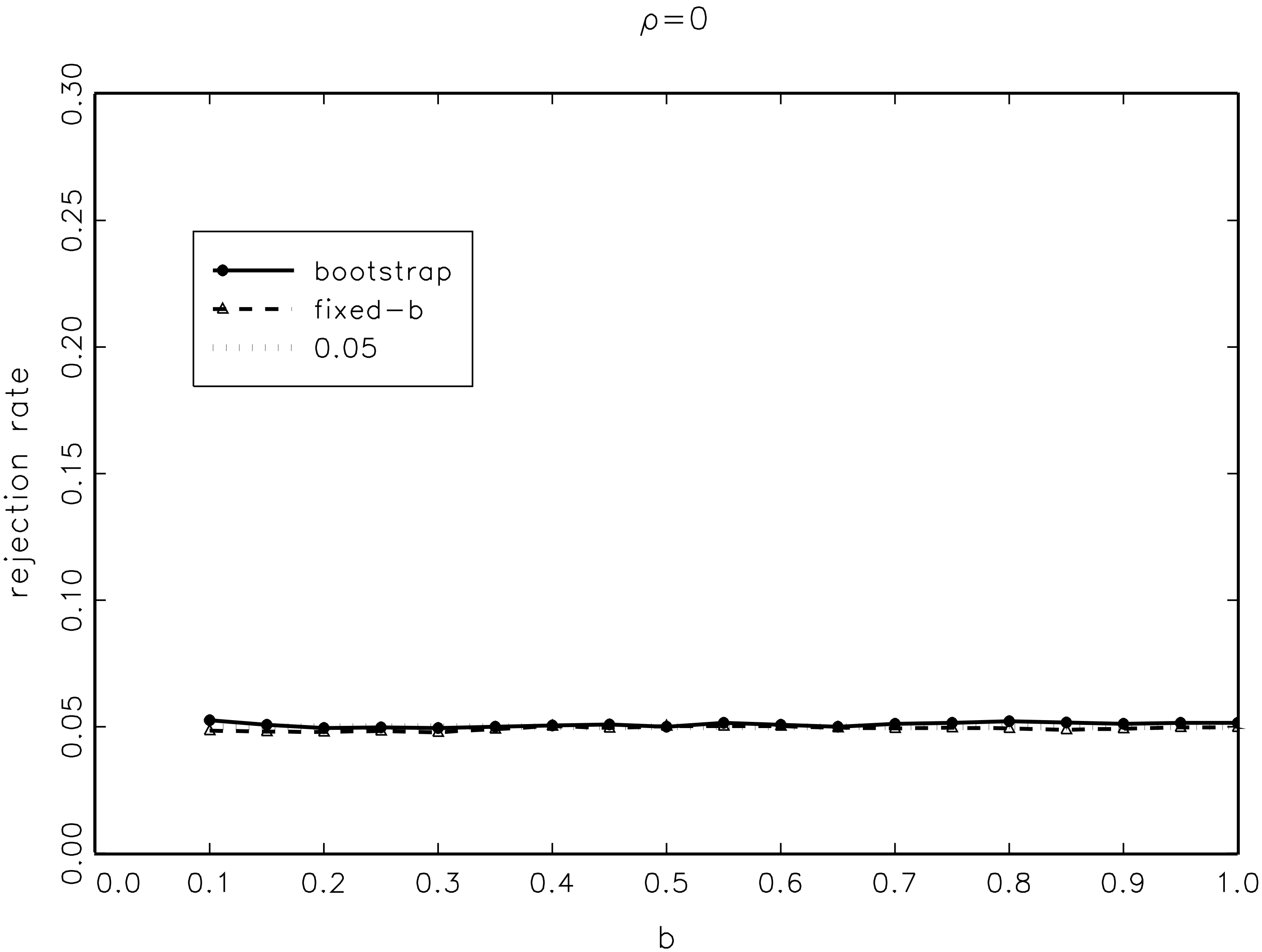


Figure 2.62: (cont'd)

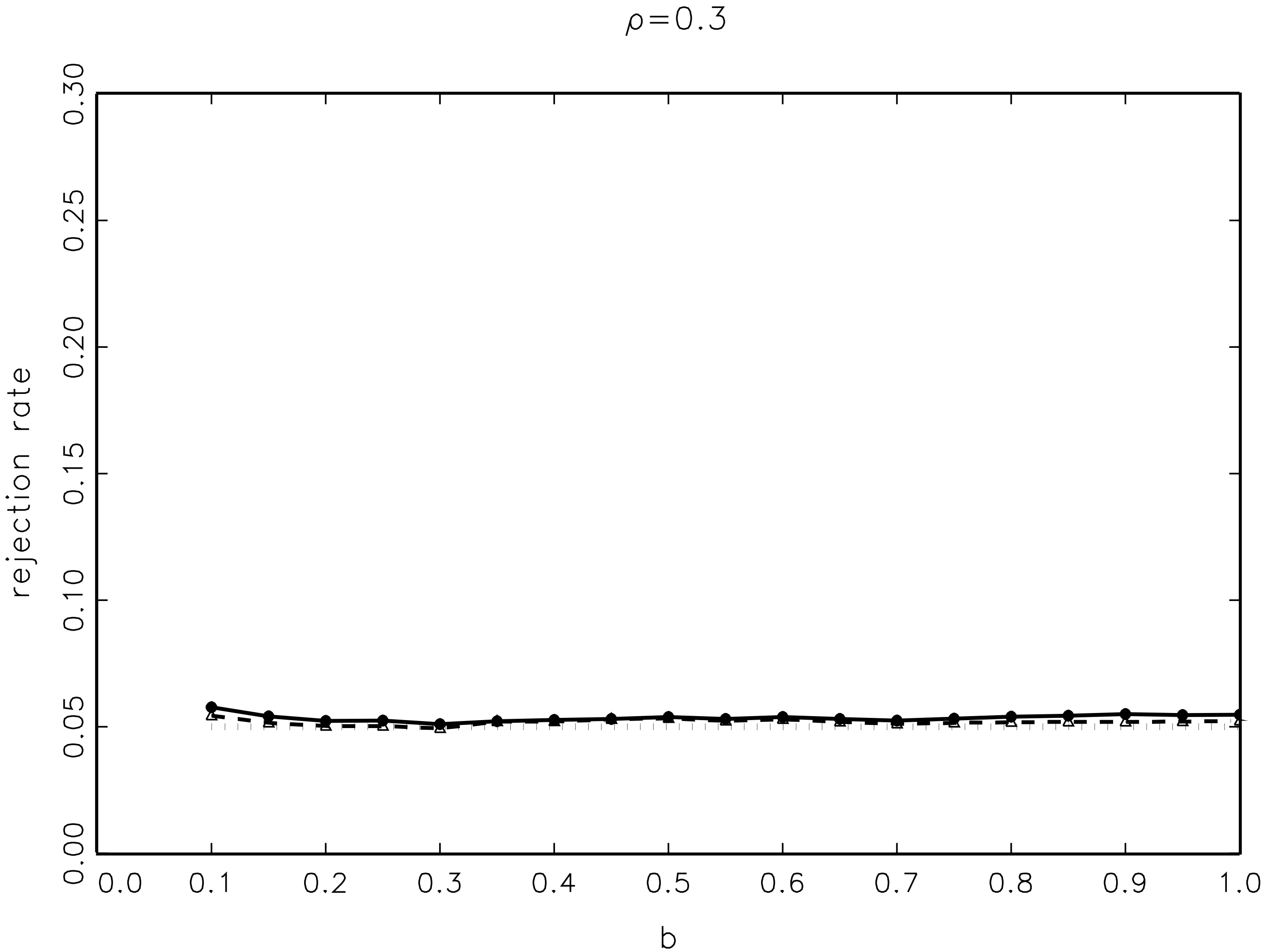


Figure 2.62: (cont'd)

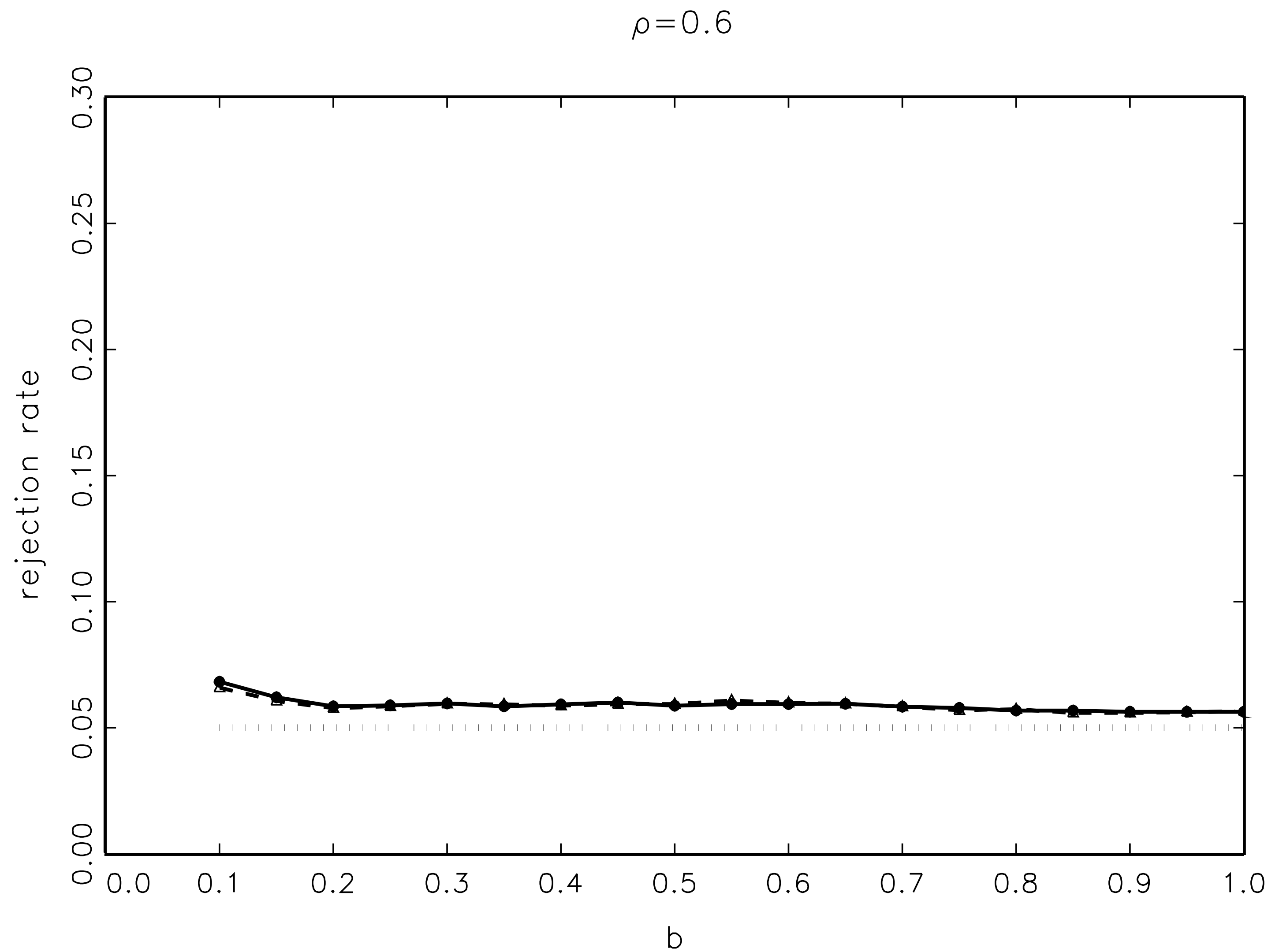


Figure 2.62: (cont'd)

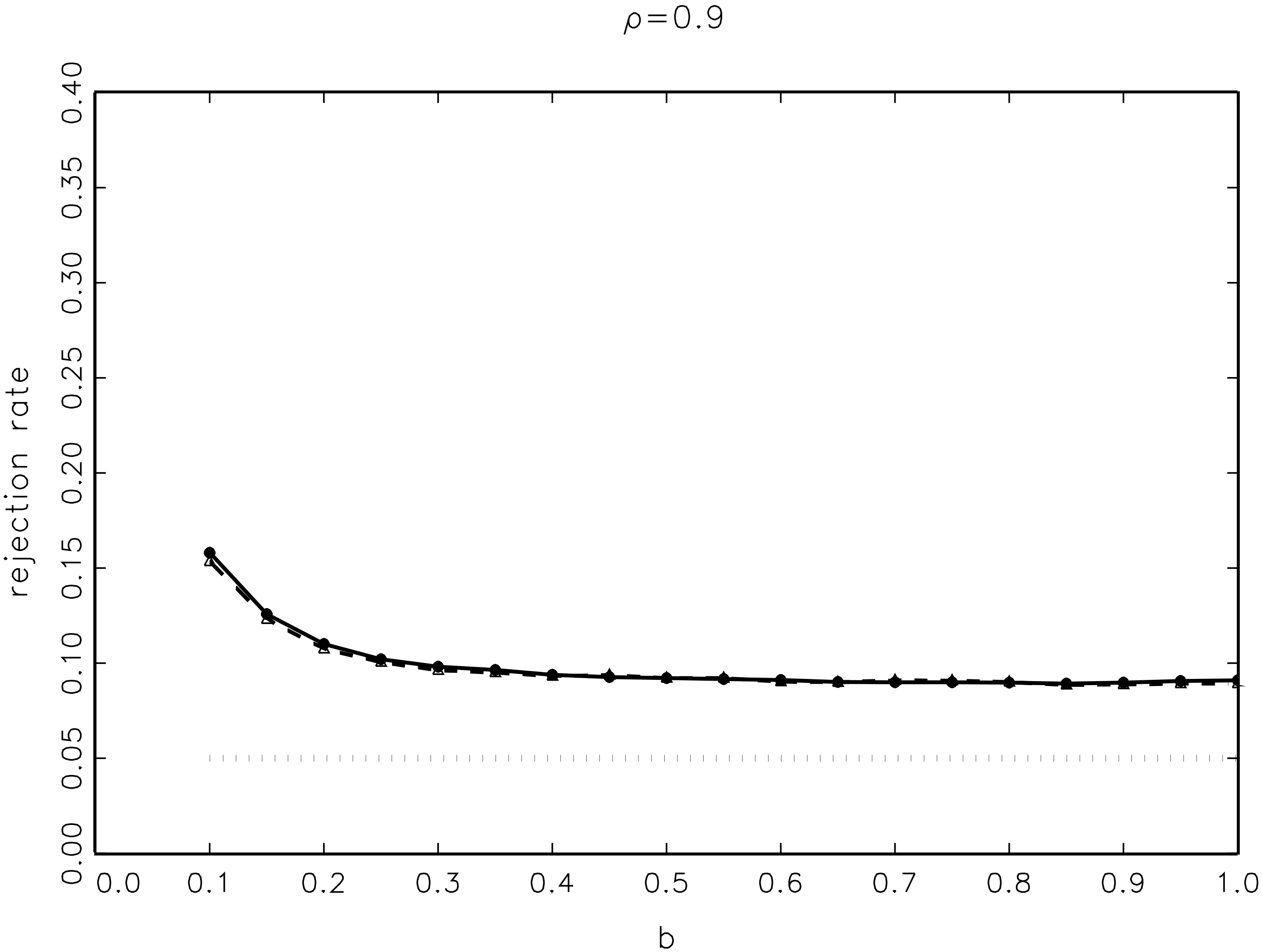


Figure 2.63: ES - Random Bernoulli ($p = 0.5$), Bartlett, $T = 200$

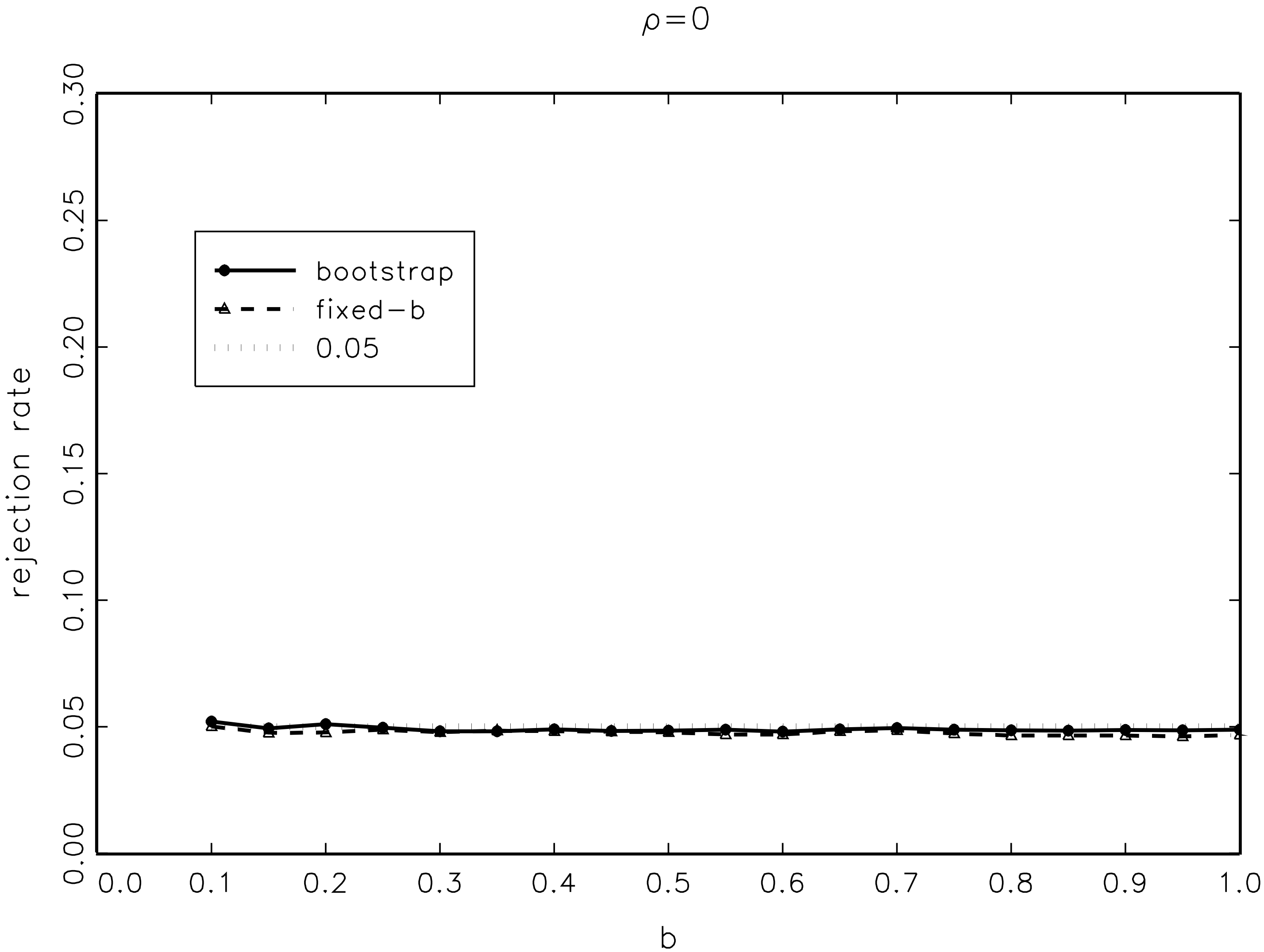


Figure 2.63: (cont'd)

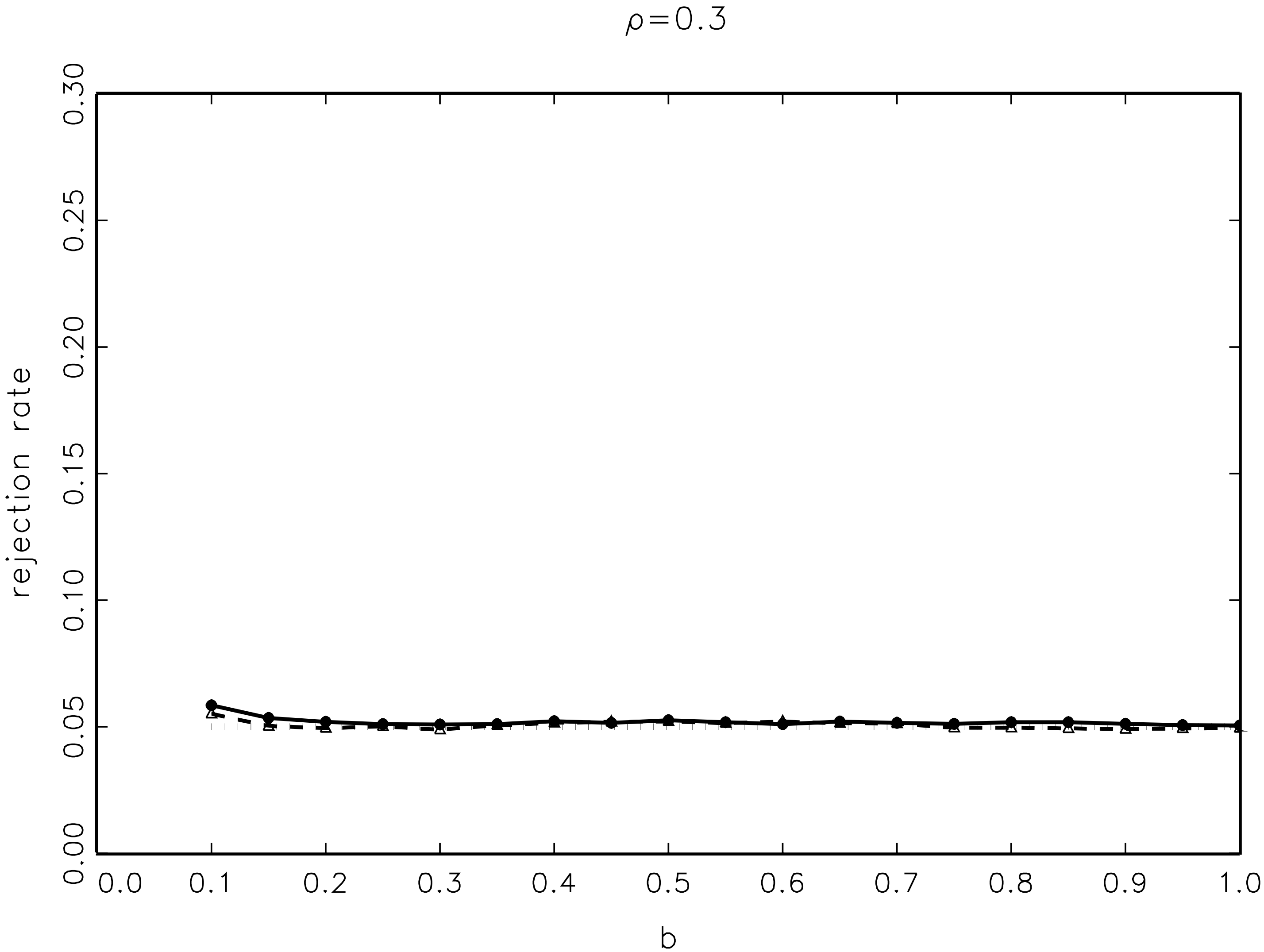


Figure 2.63: (cont'd)

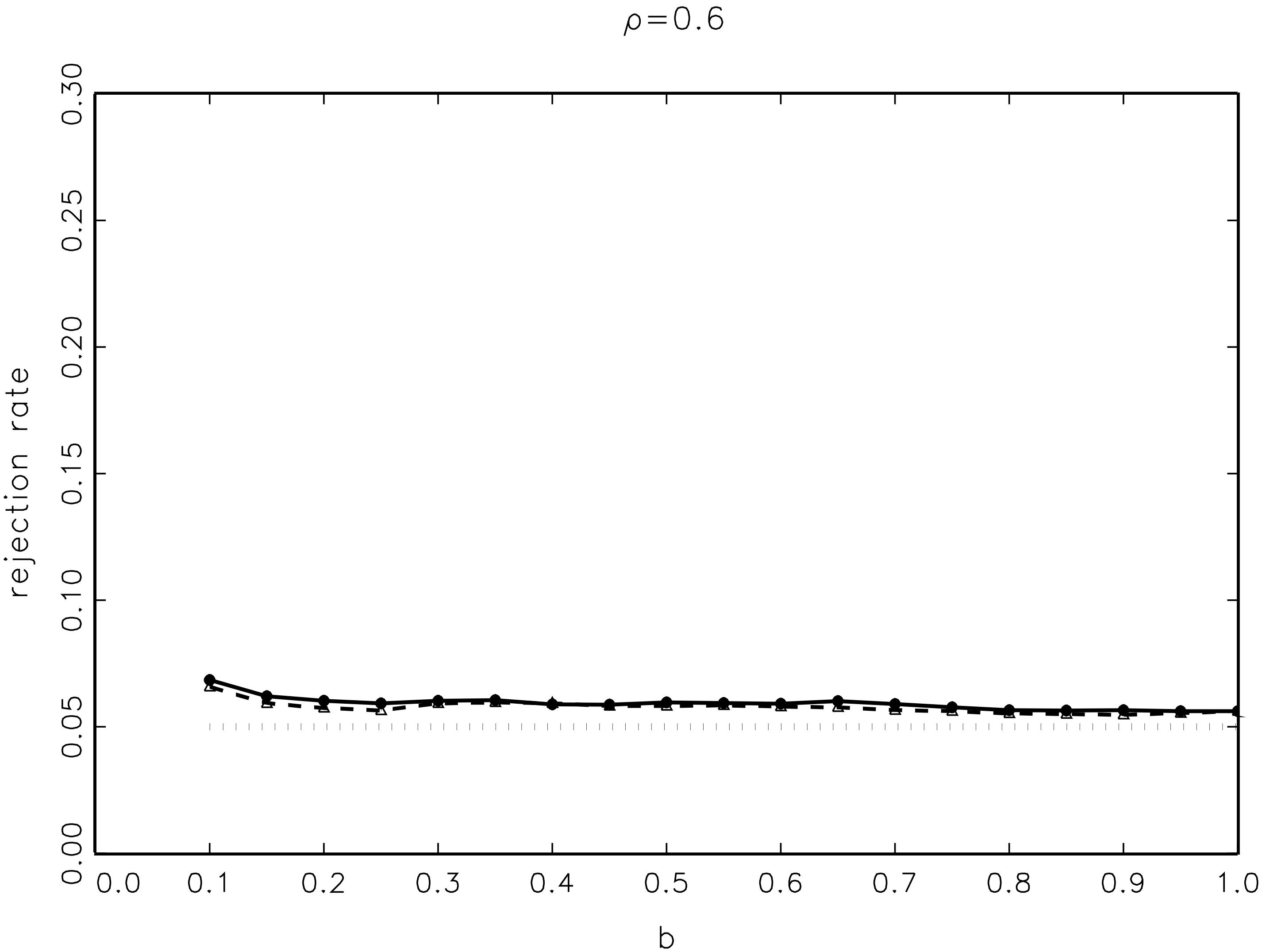


Figure 2.63: (cont'd)

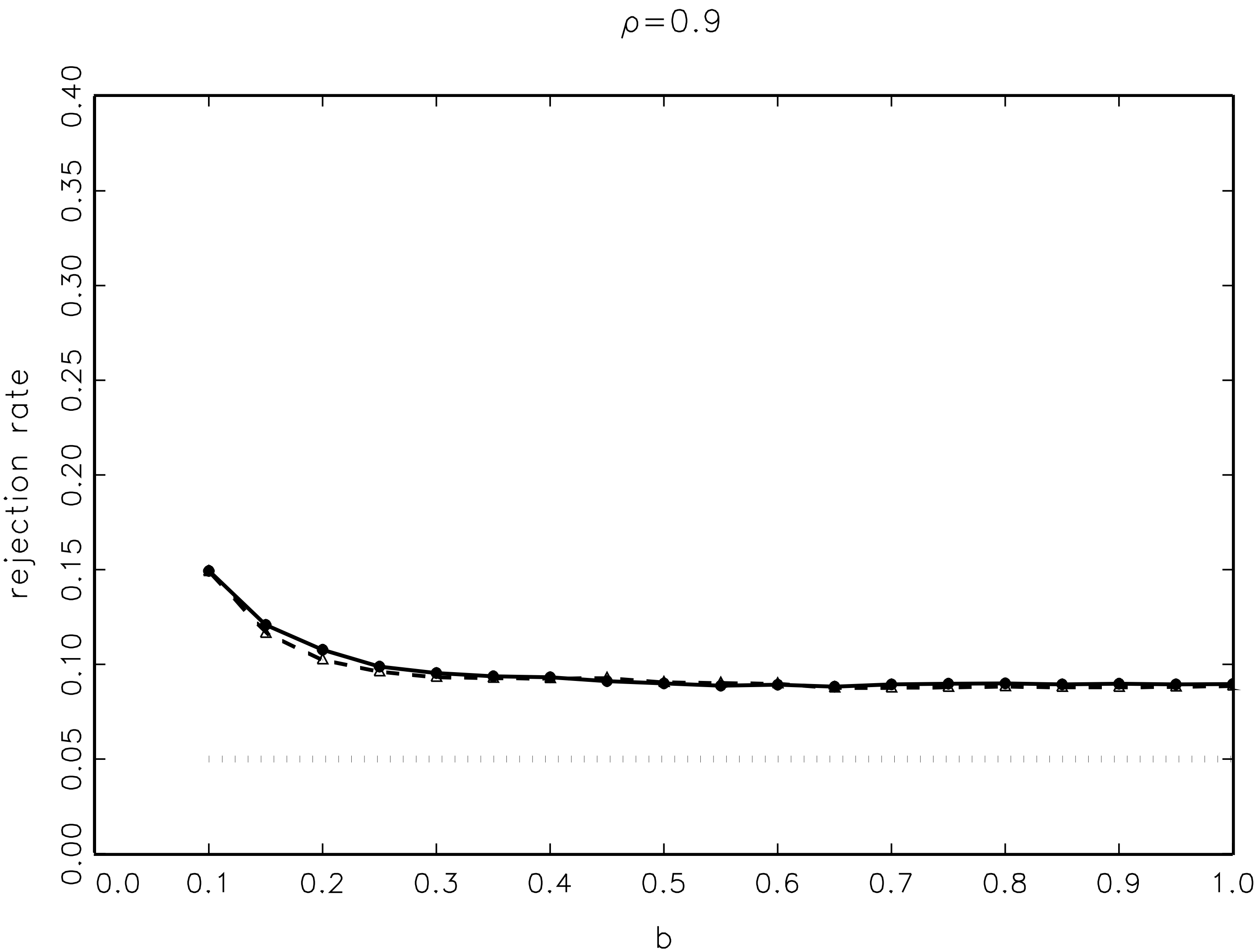


Figure 2.64: ES - Random Bernoulli ($p = 0.7$), Bartlett, $T = 200$

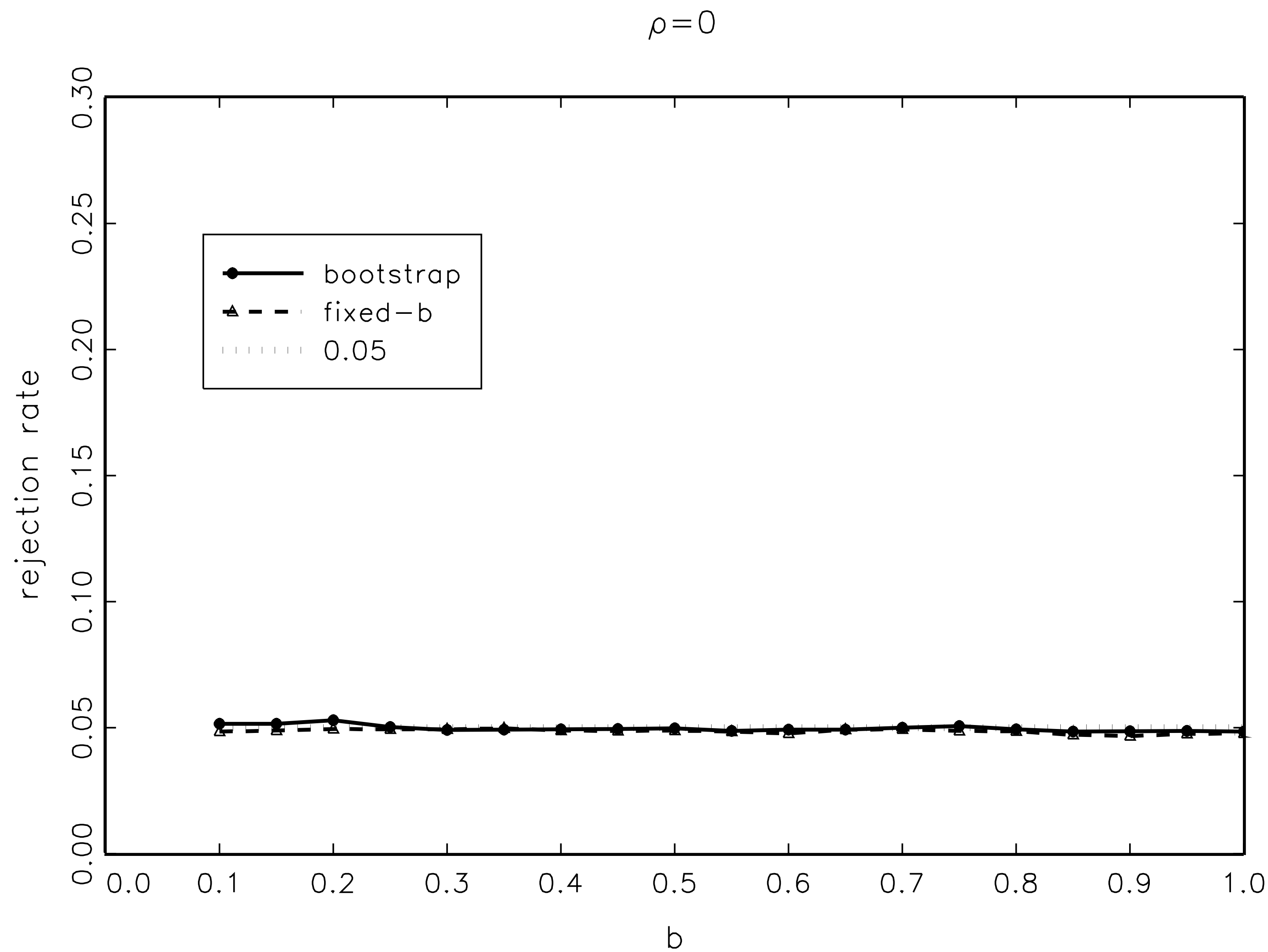


Figure 2.64: (cont'd)

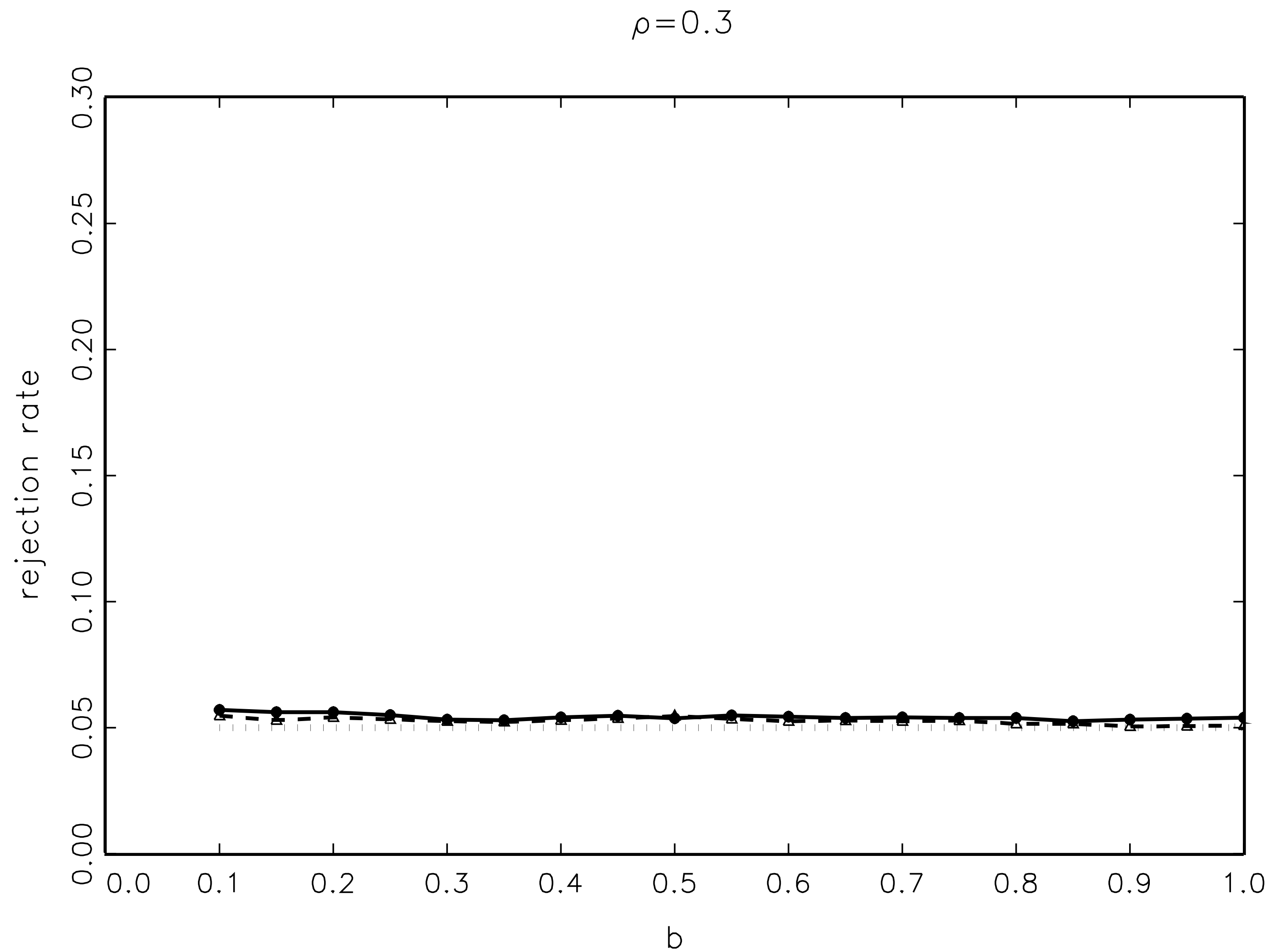


Figure 2.64: (cont'd)

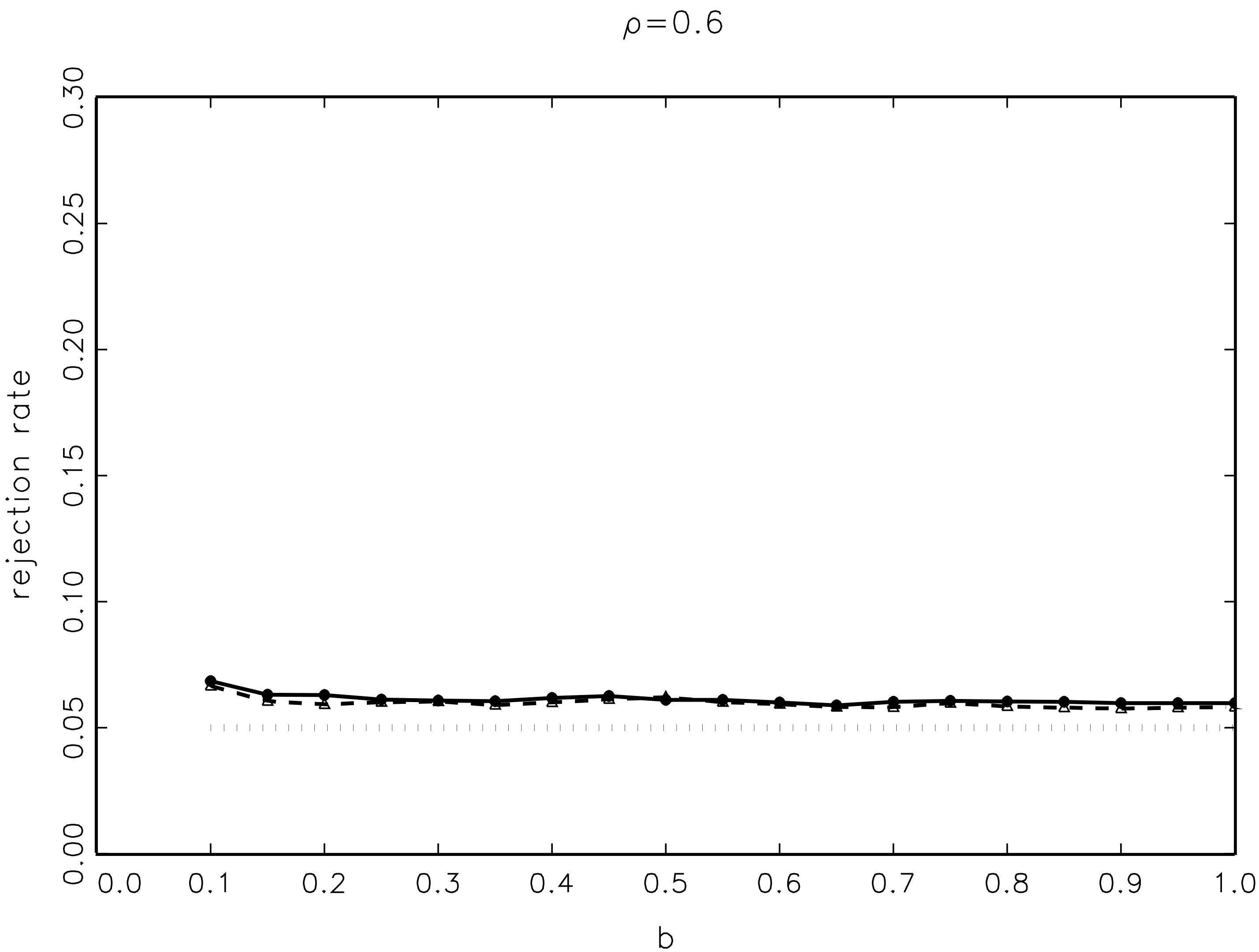


Figure 2.64: (cont'd)

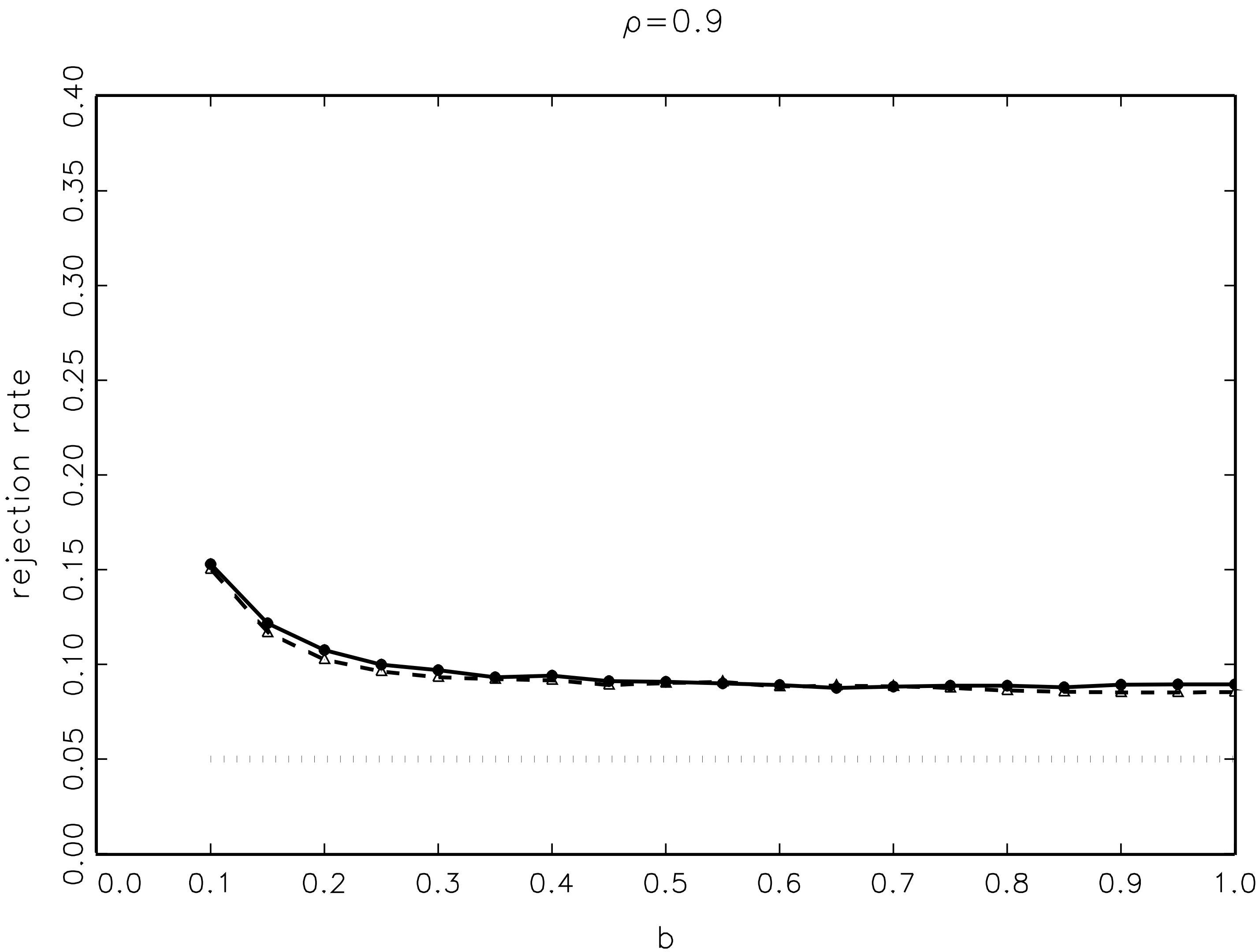


Figure 2.65: AM and ES - World War (quarterly), Bartlett, $T = 36$

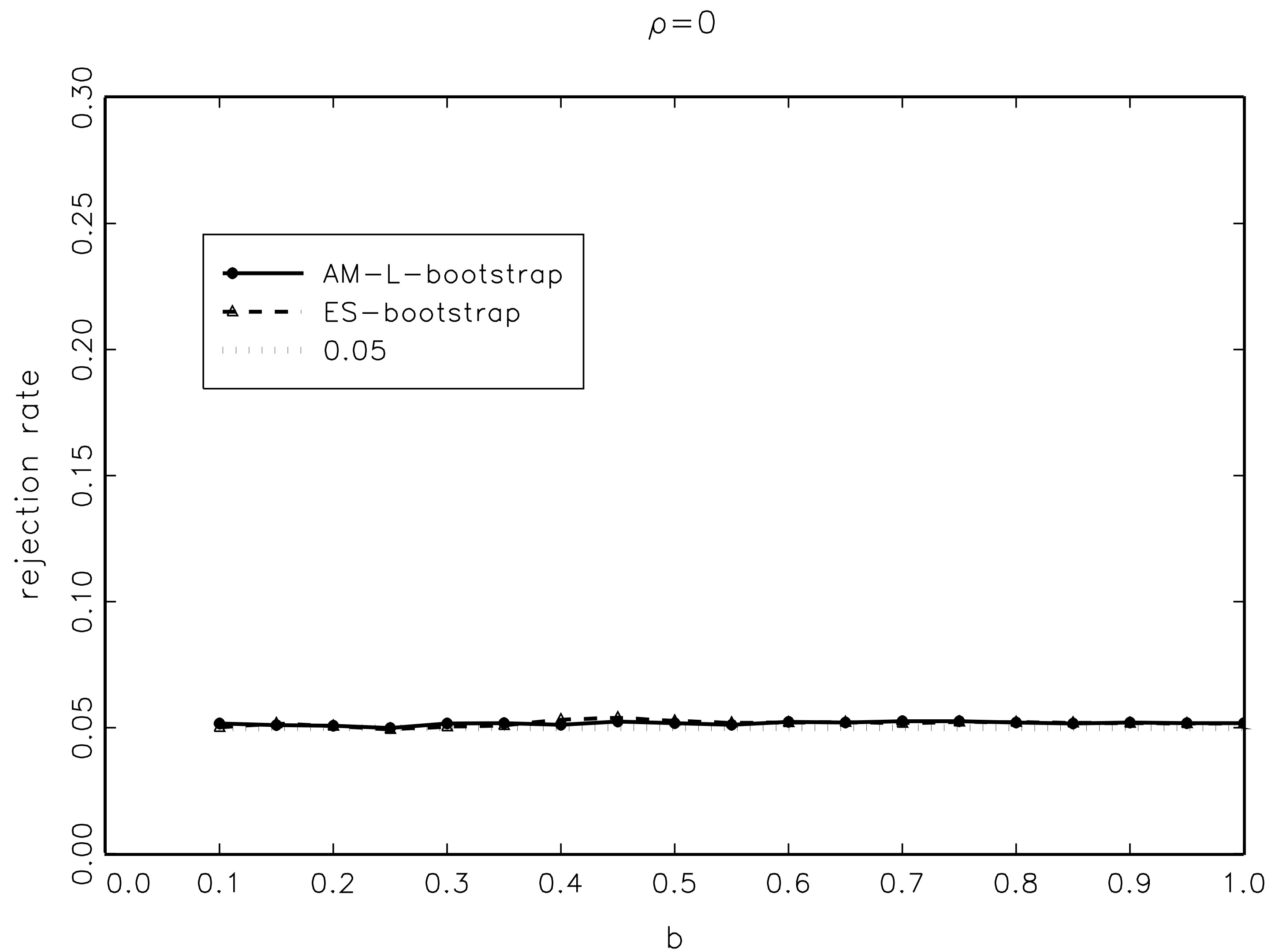


Figure 2.65: (cont'd)

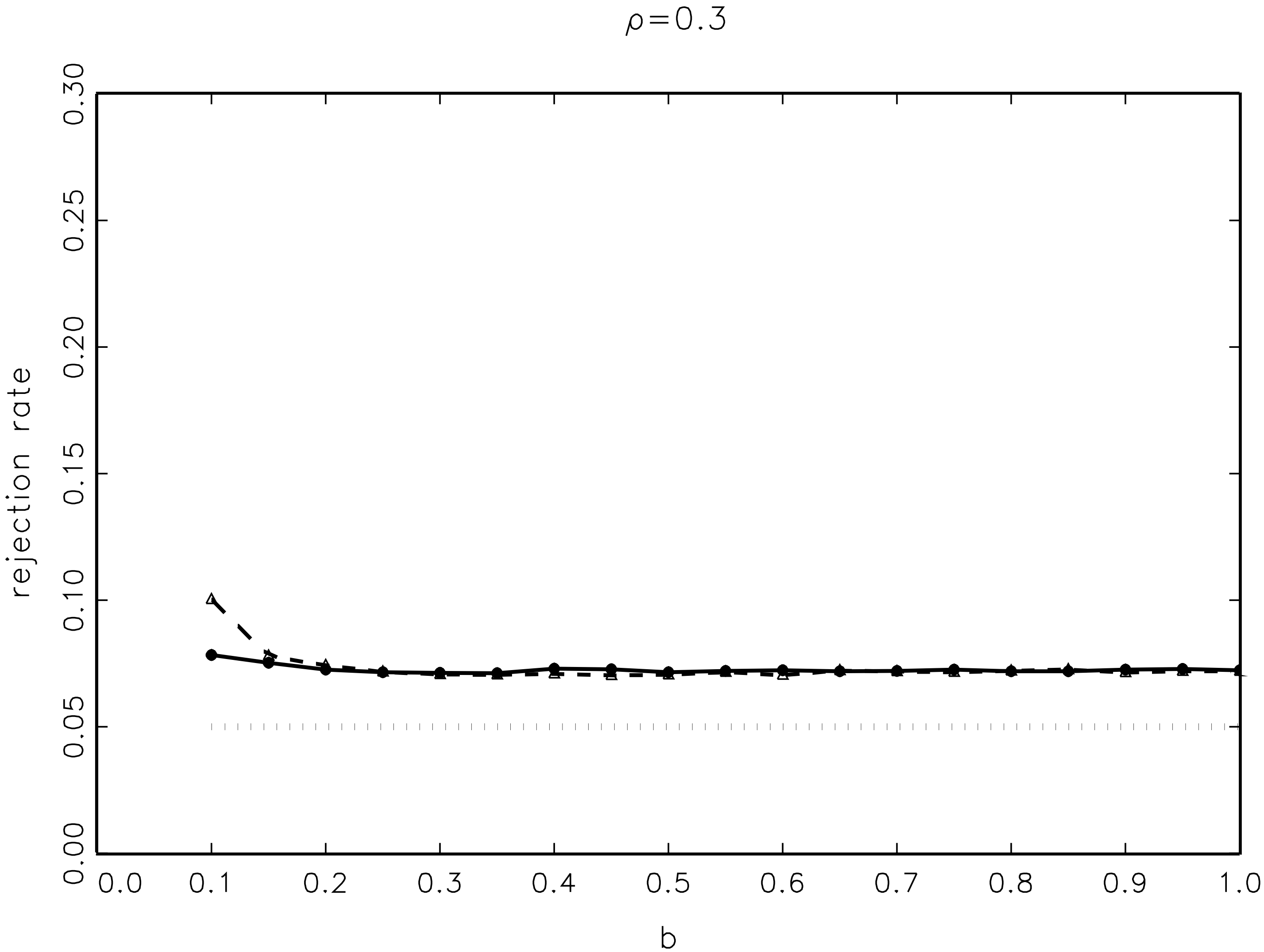


Figure 2.65: (cont'd)

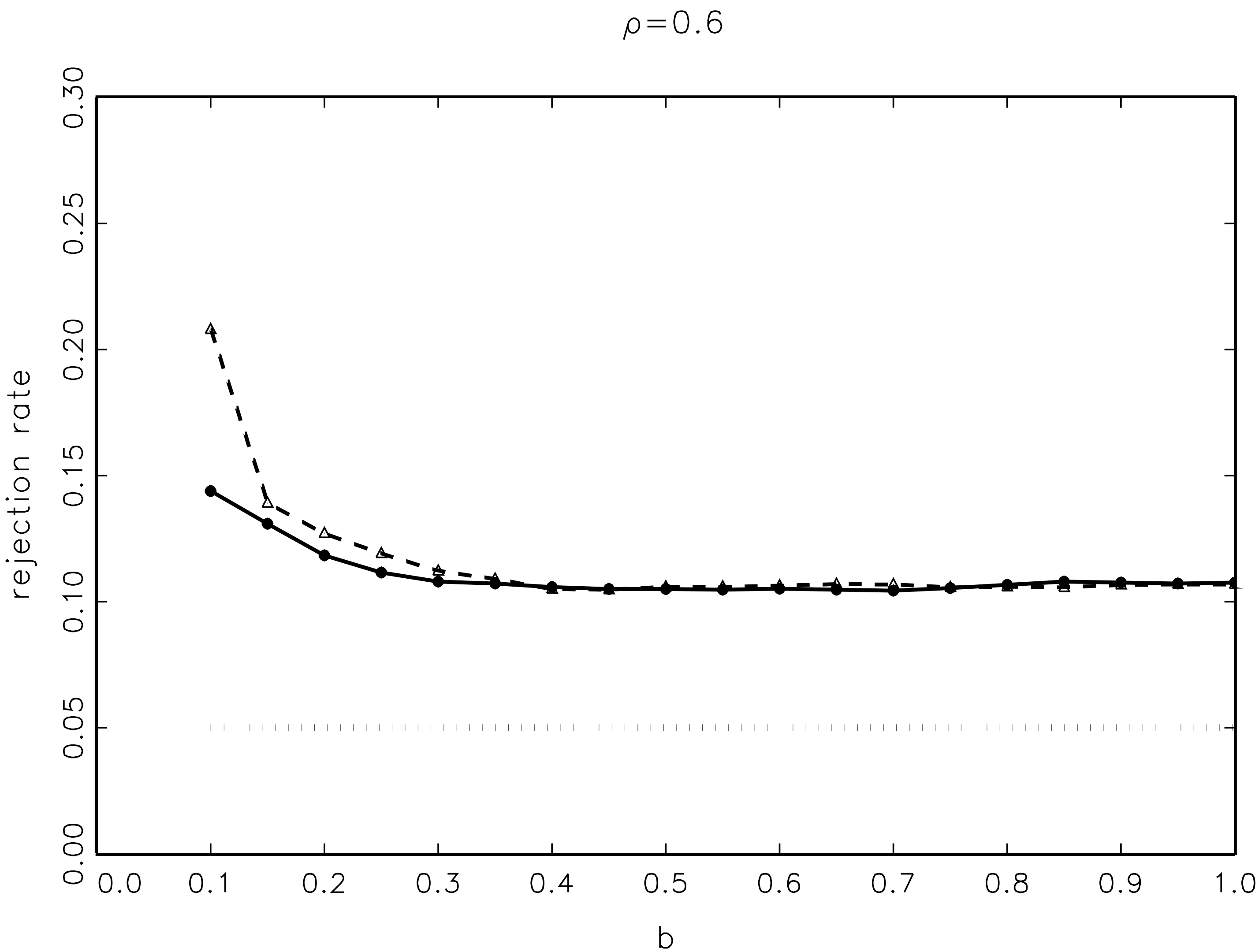


Figure 2.65: (cont'd)

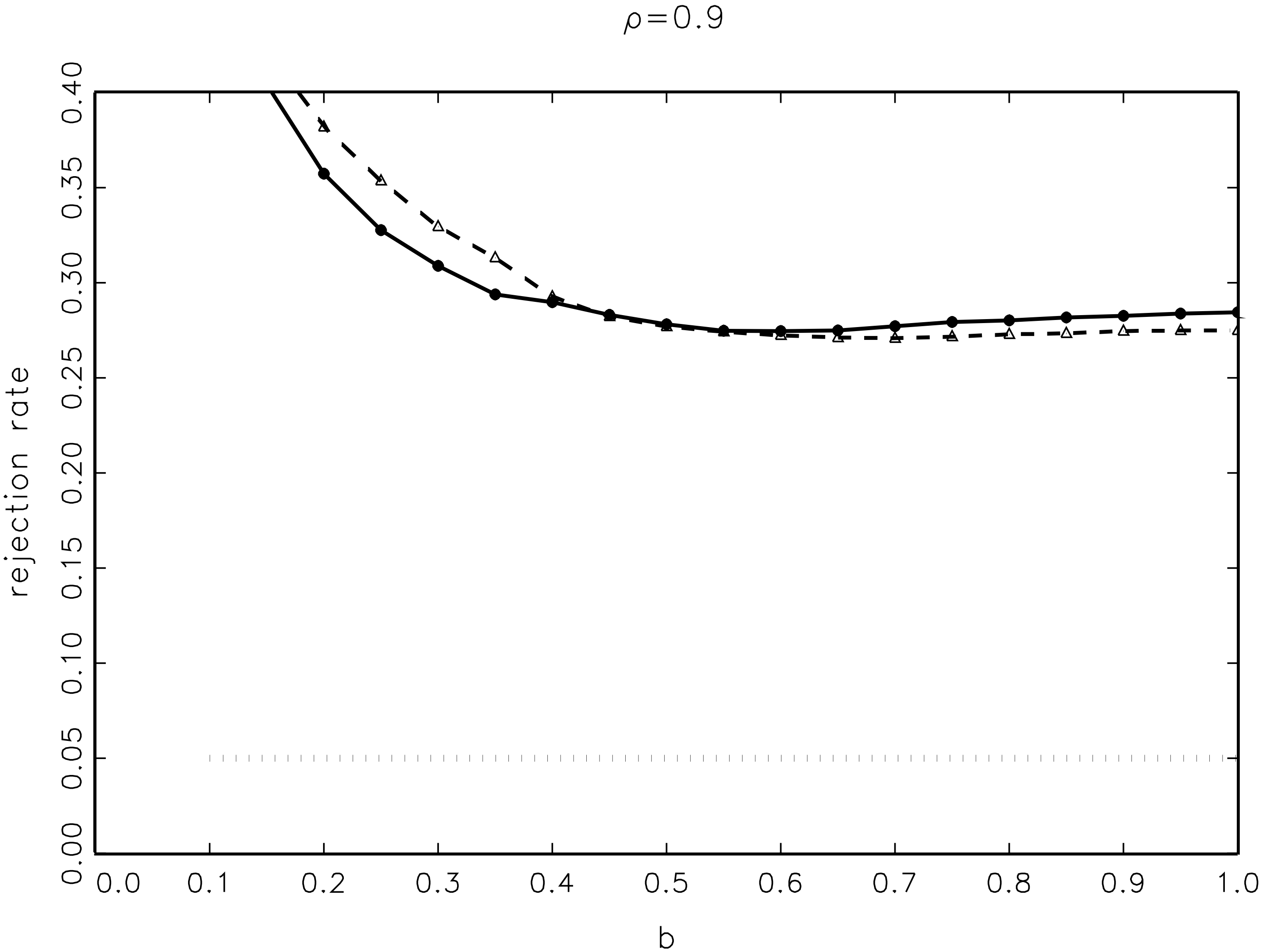


Figure 2.66: AM and ES - World War (quarterly), Bartlett, $T = 48$

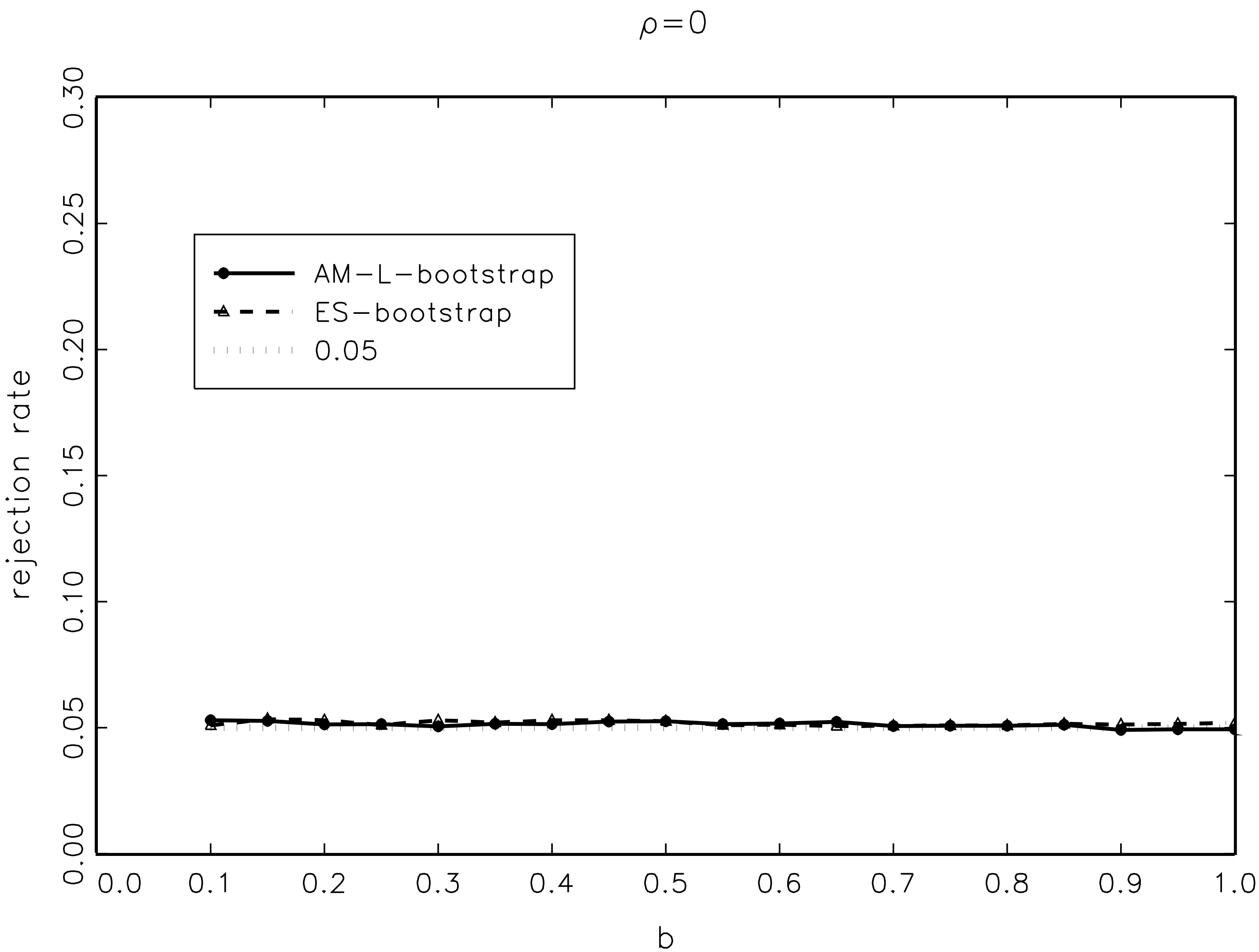


Figure 2.66: (cont'd)

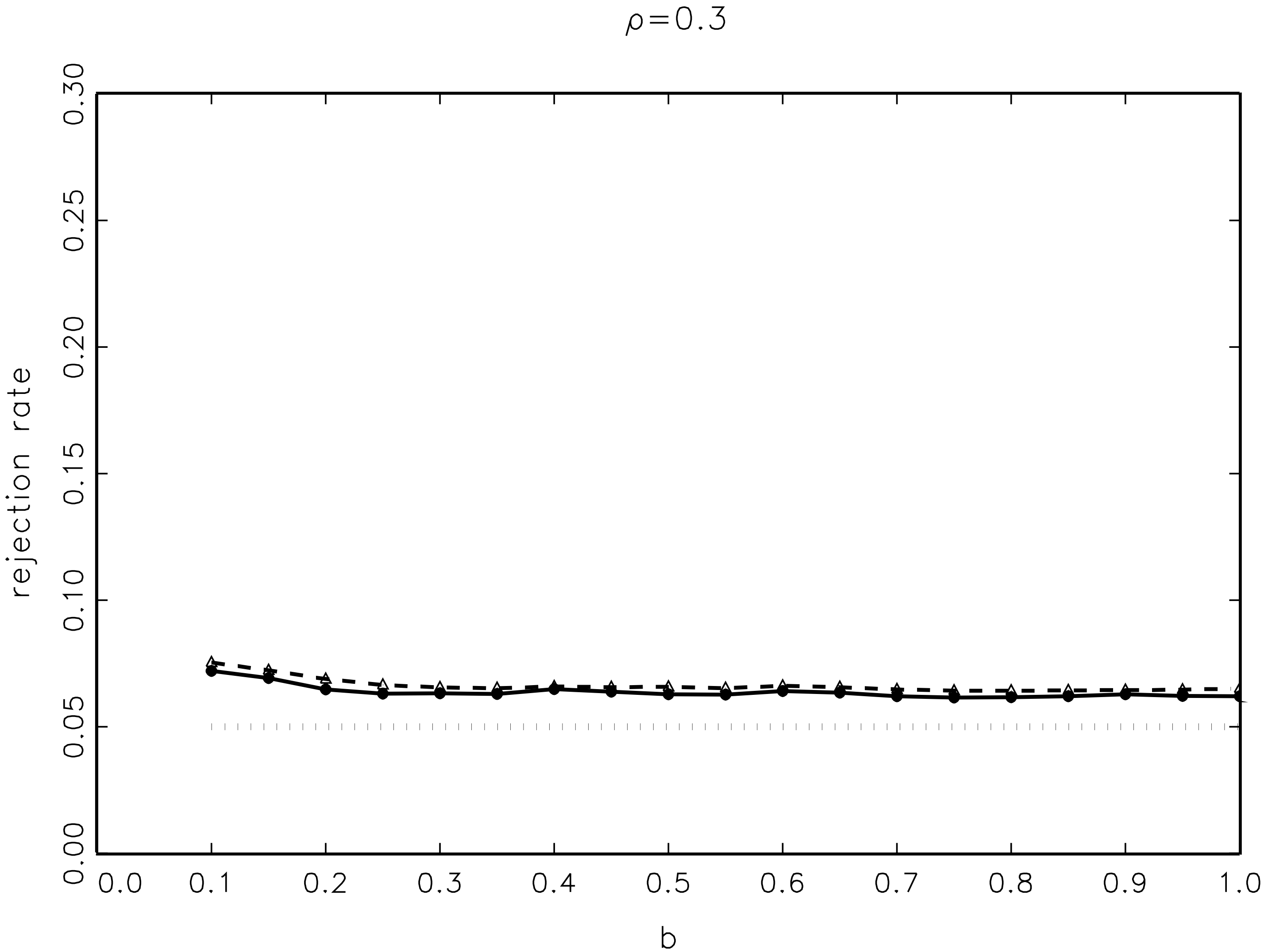


Figure 2.66: (cont'd)

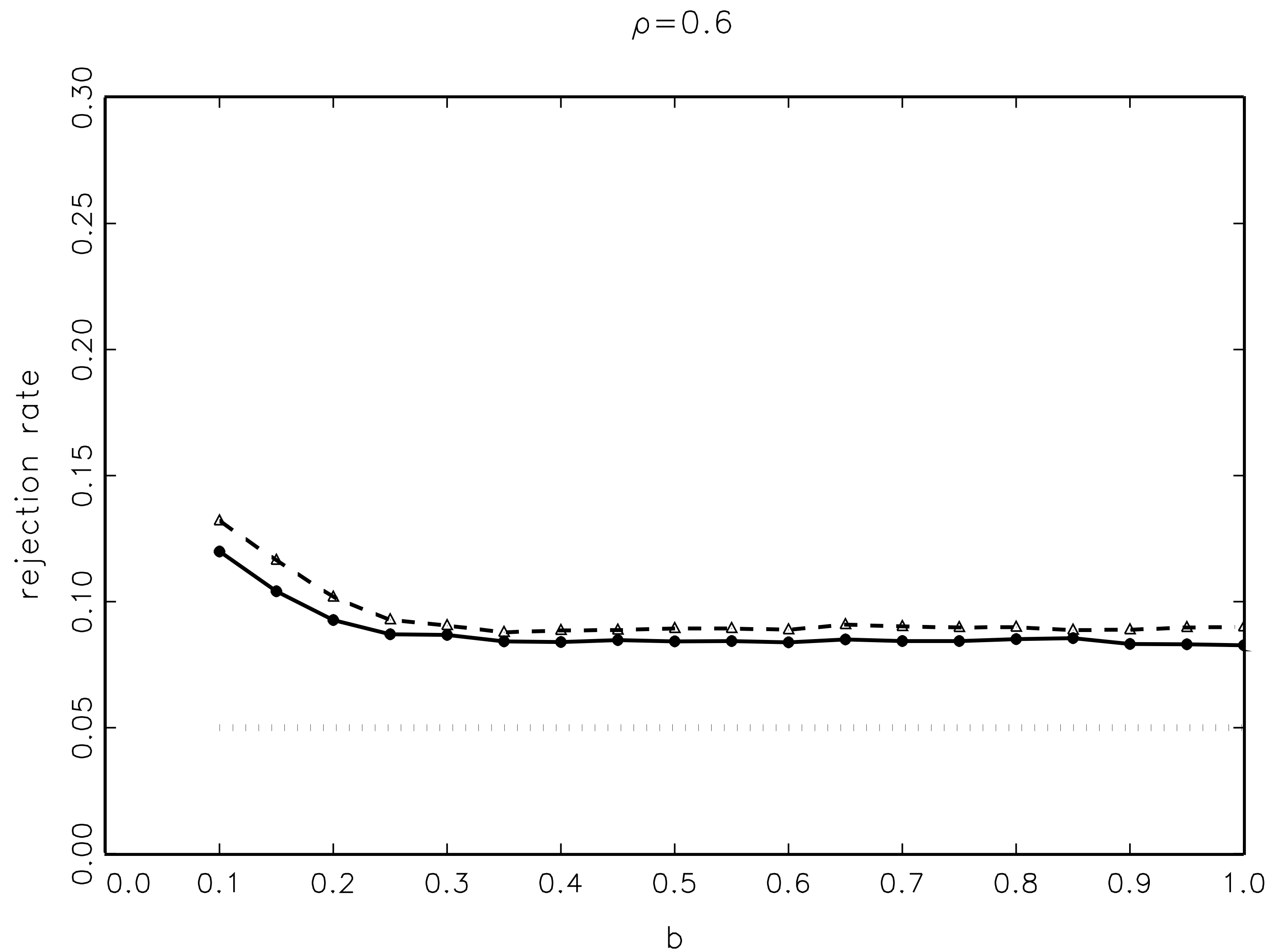


Figure 2.66: (cont'd)

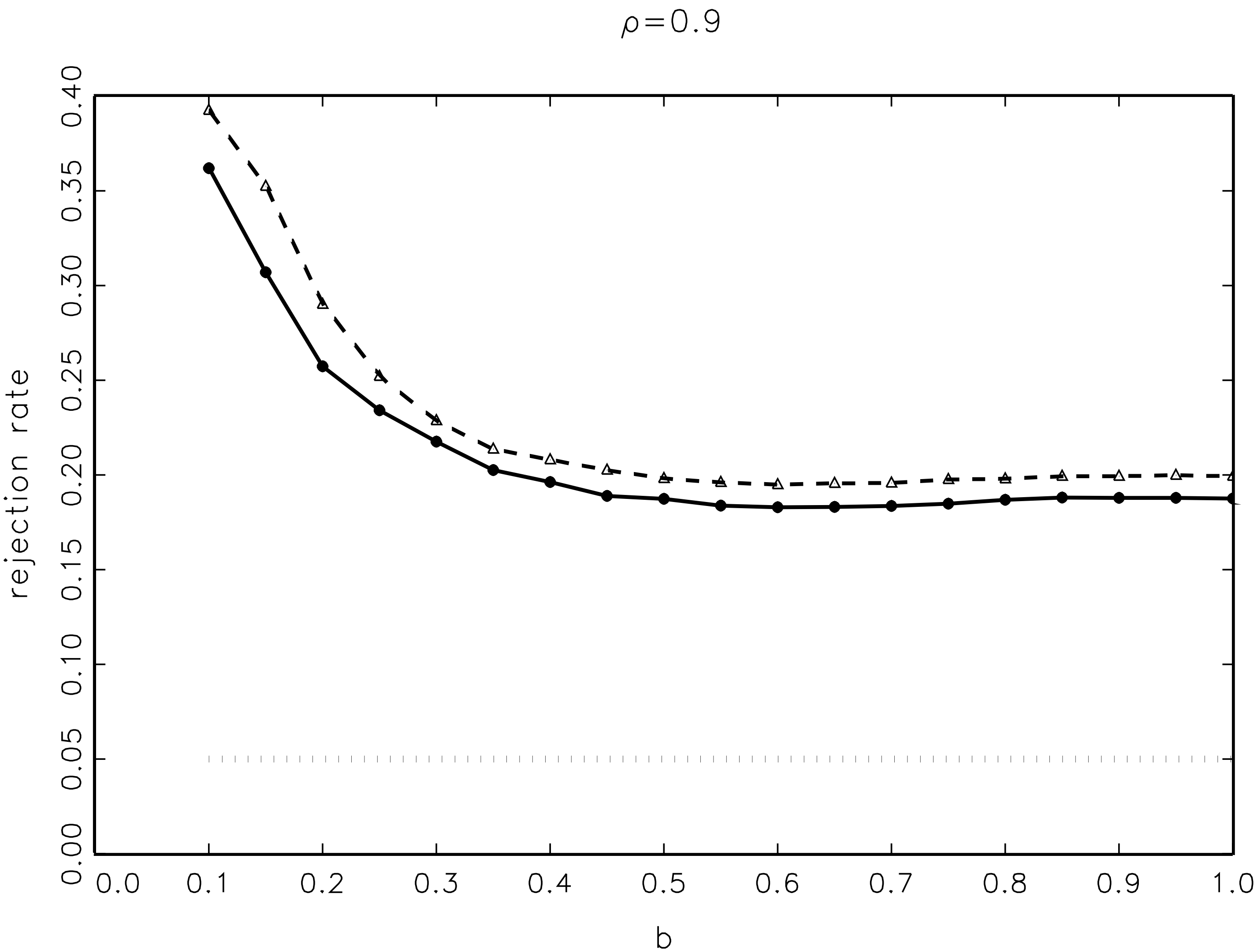


Figure 2.67: AM and ES - World War (quarterly), Bartlett, $T = 60$

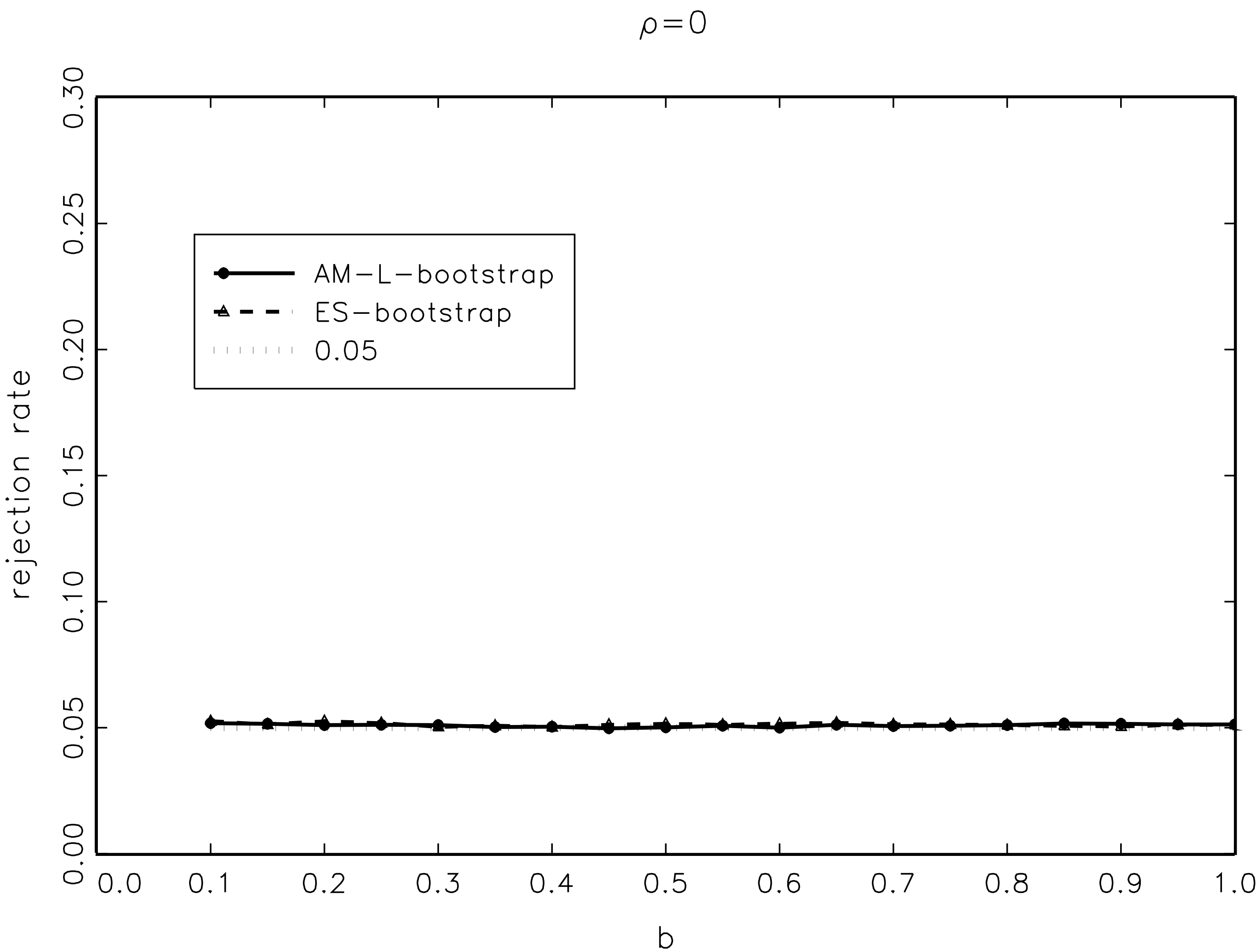


Figure 2.67: (cont'd)

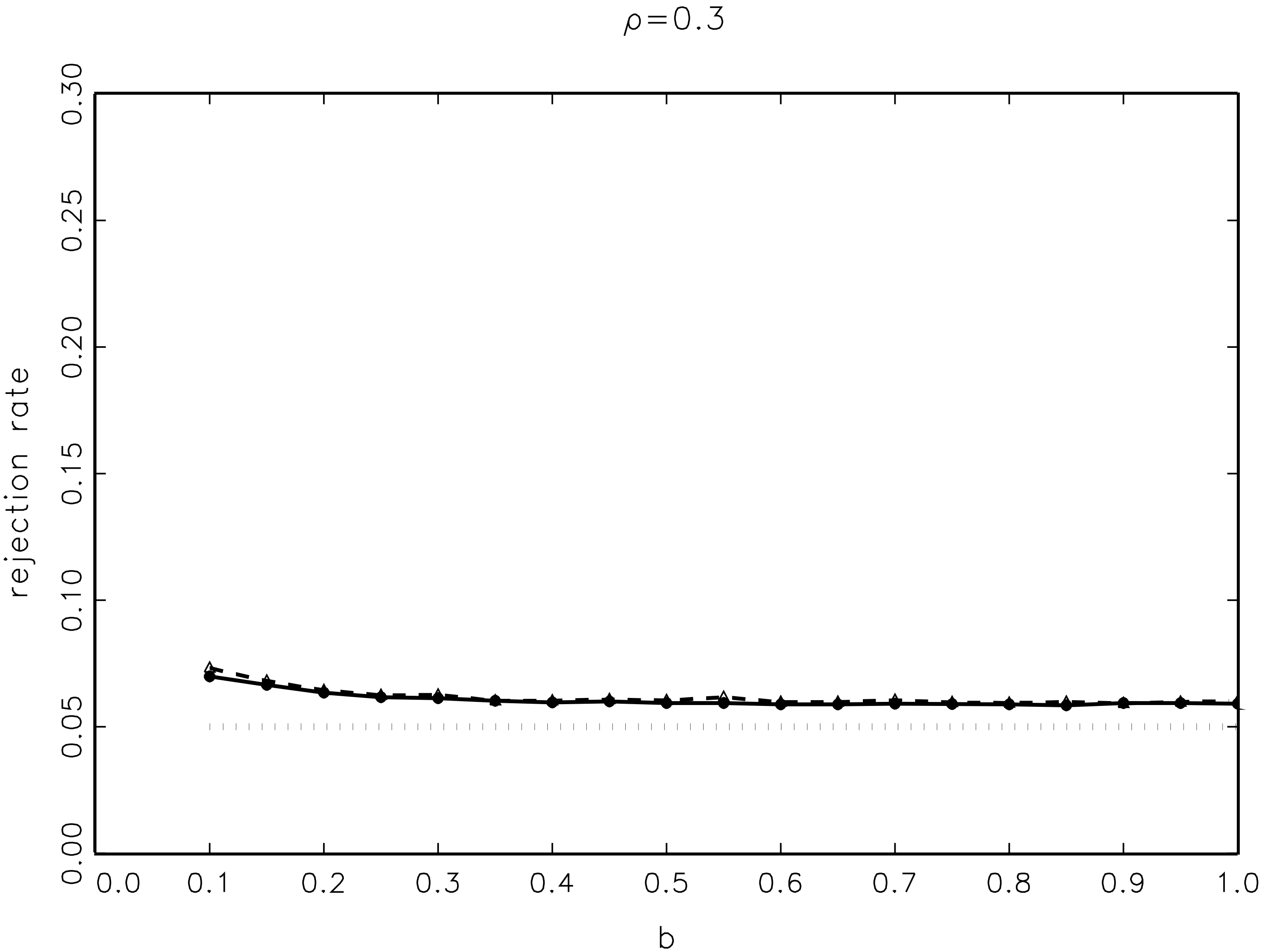


Figure 2.67: (cont'd)

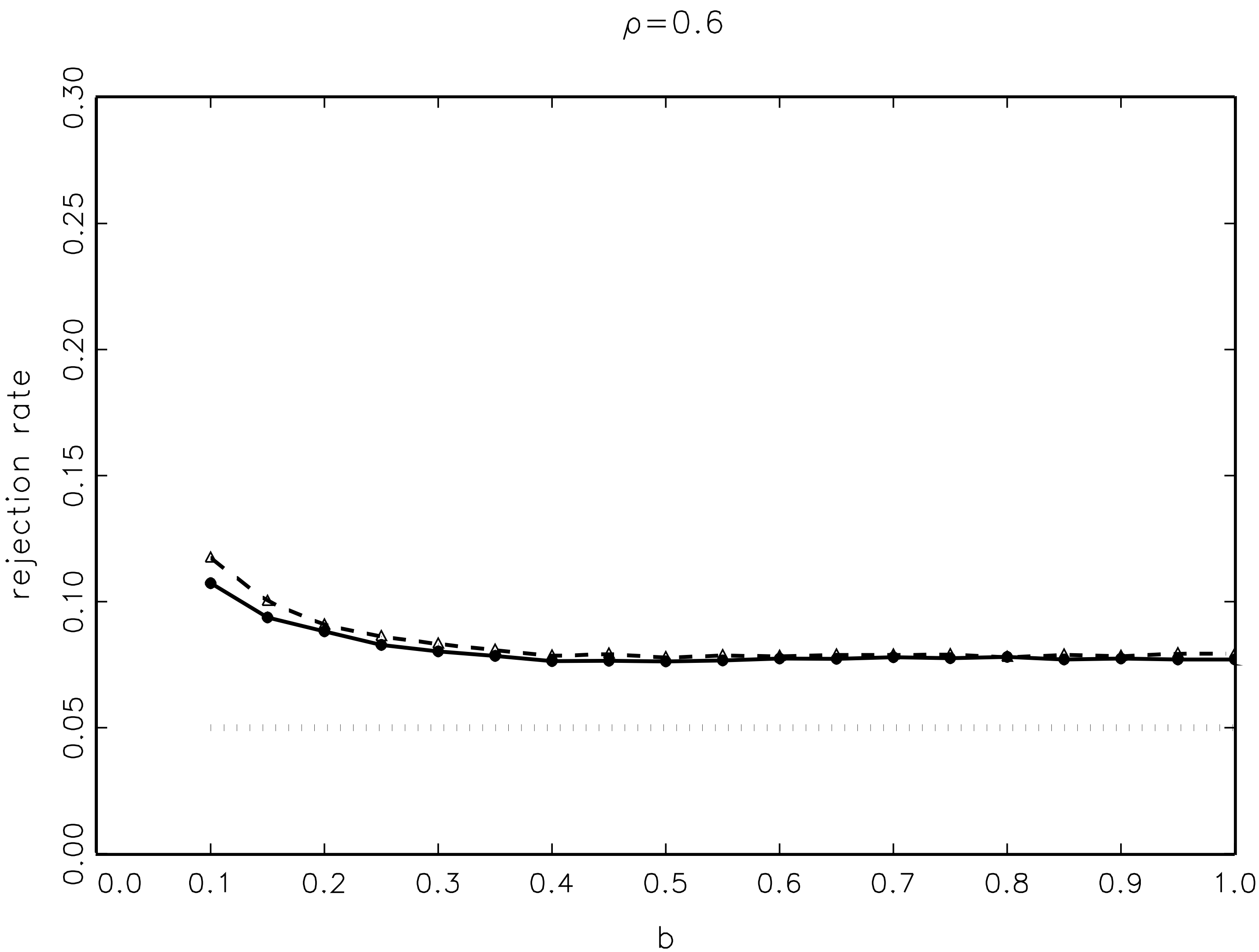


Figure 2.67: (cont'd)

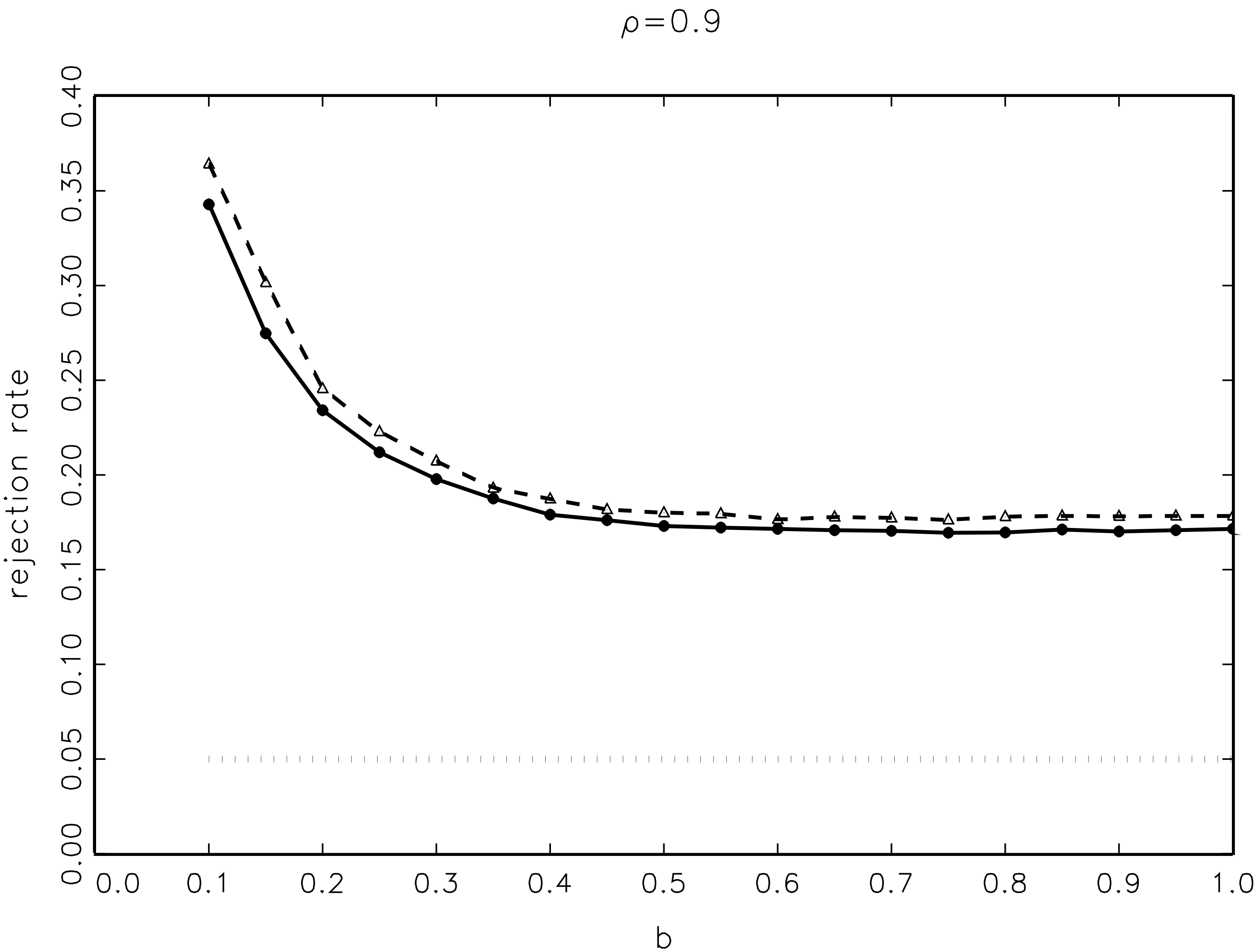


Figure 2.68: AM and ES - World War (quarterly), Bartlett, $T = 144$

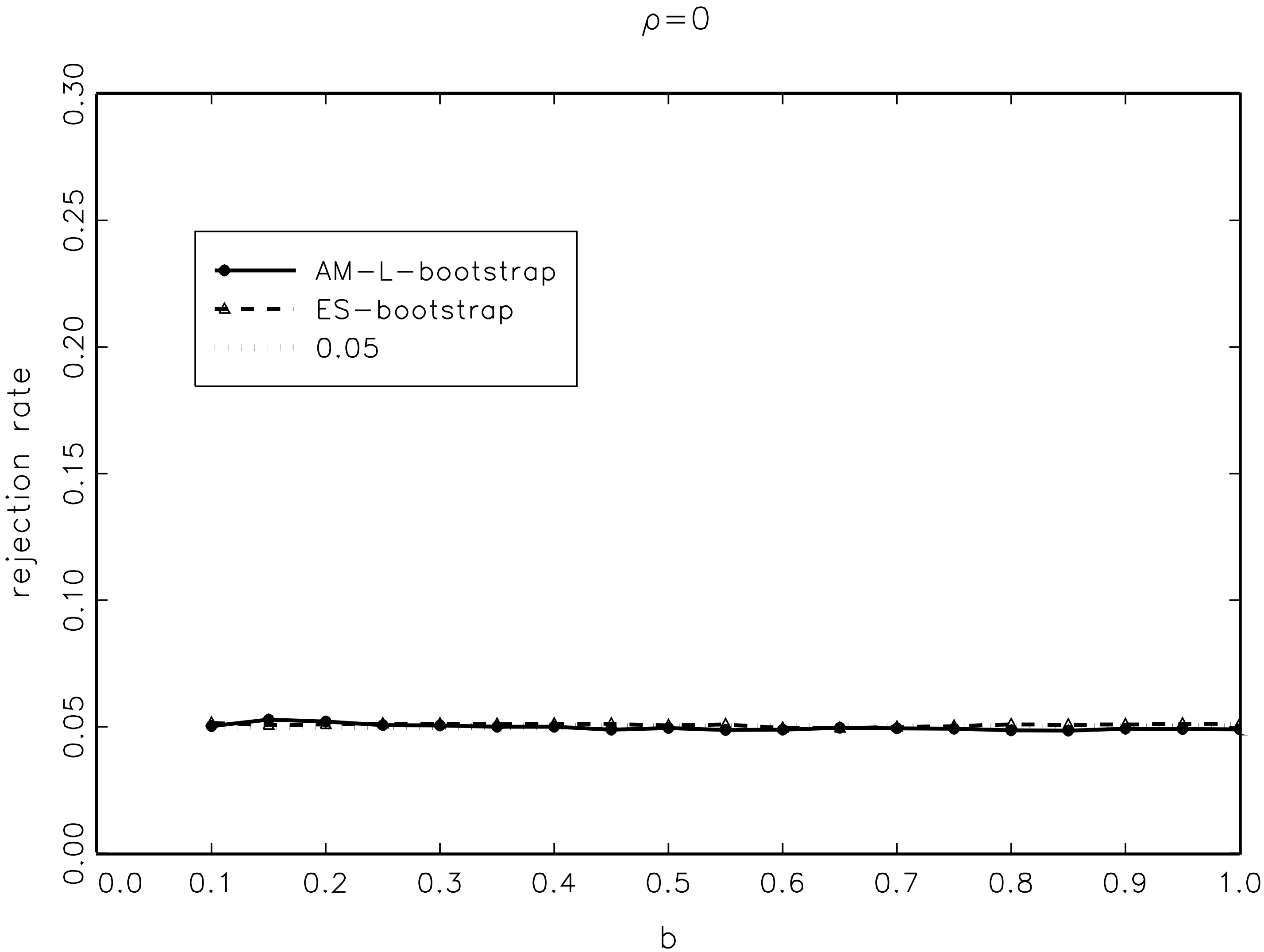


Figure 2.68: (cont'd)

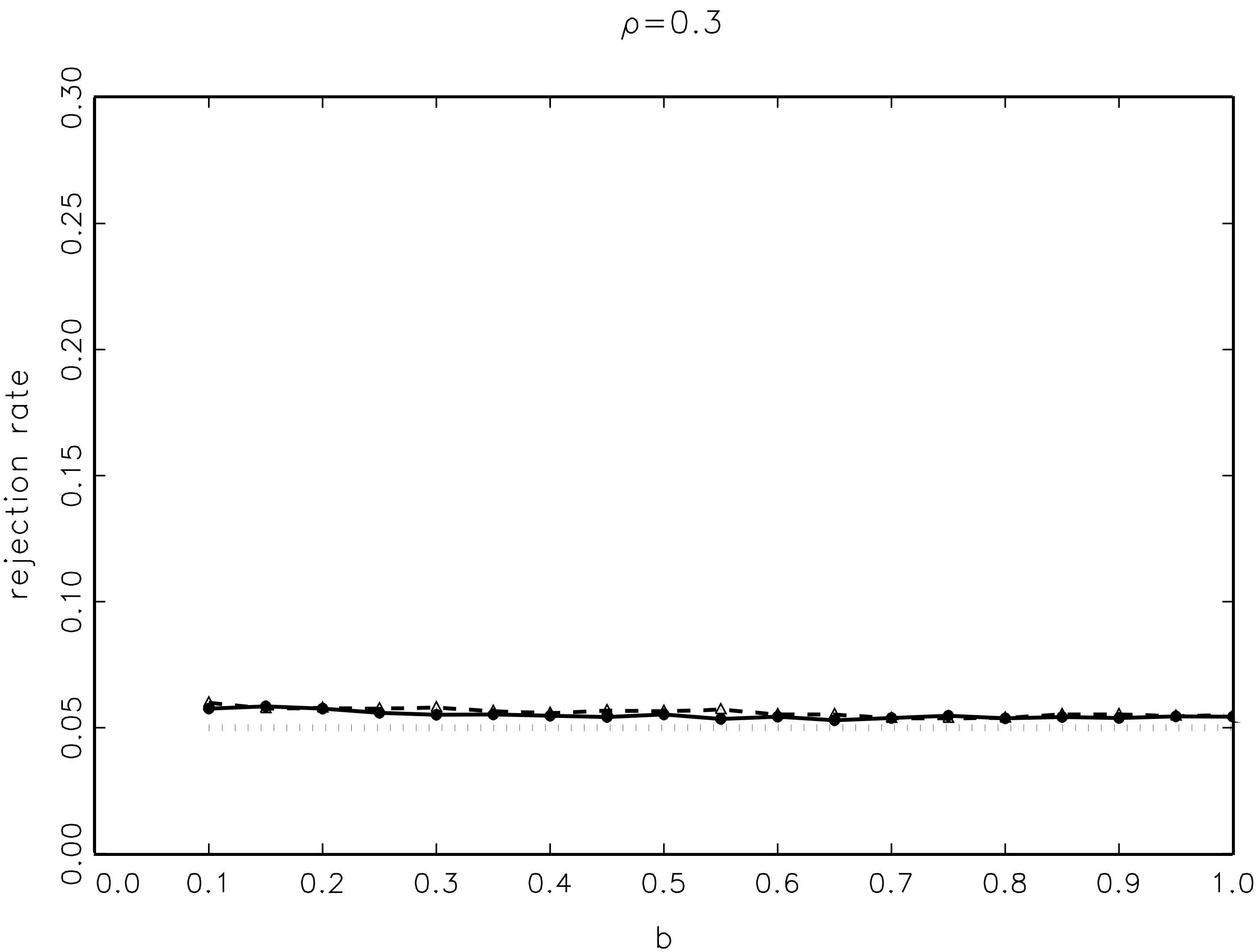


Figure 2.68: (cont'd)

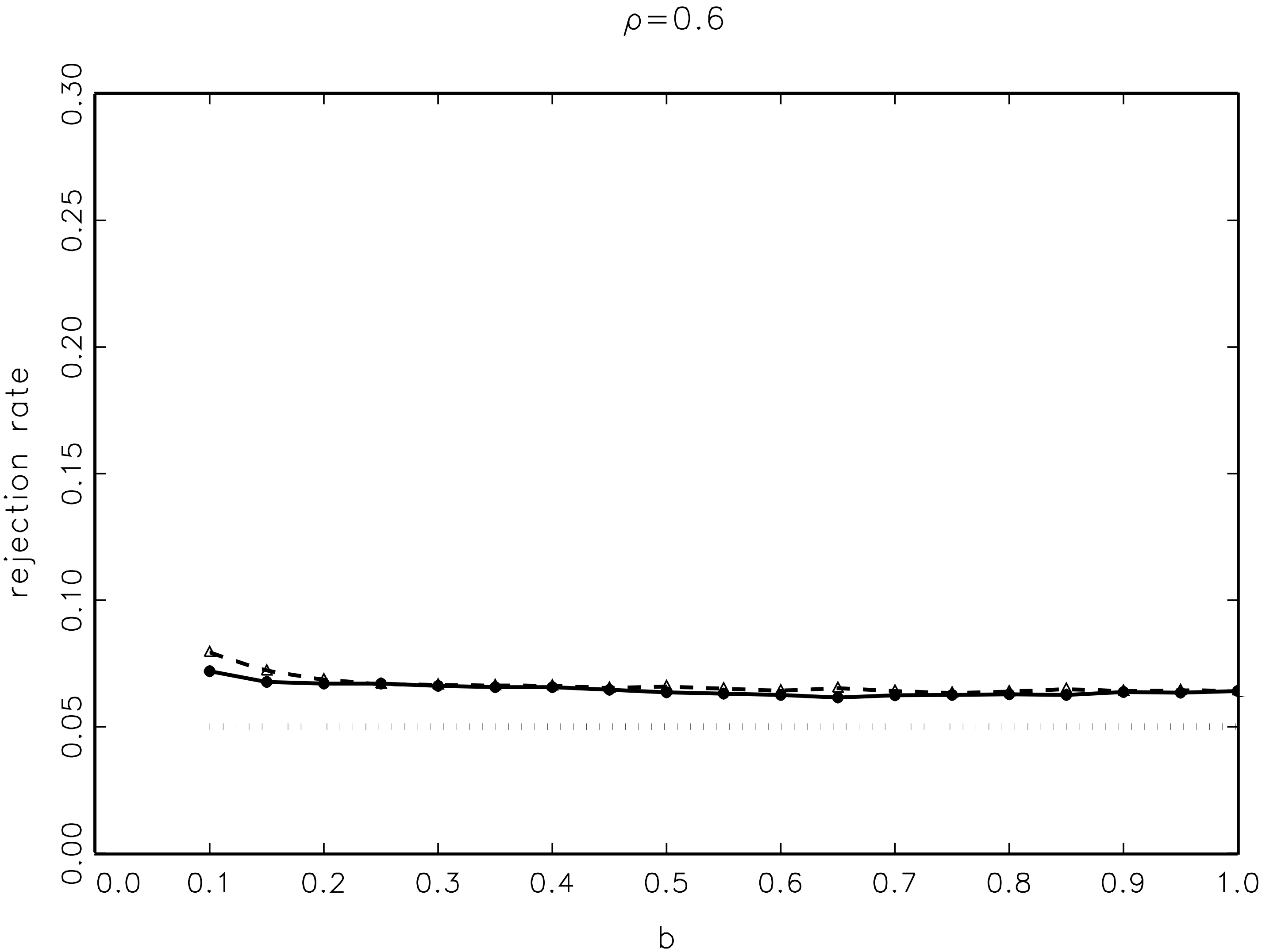


Figure 2.68: (cont'd)

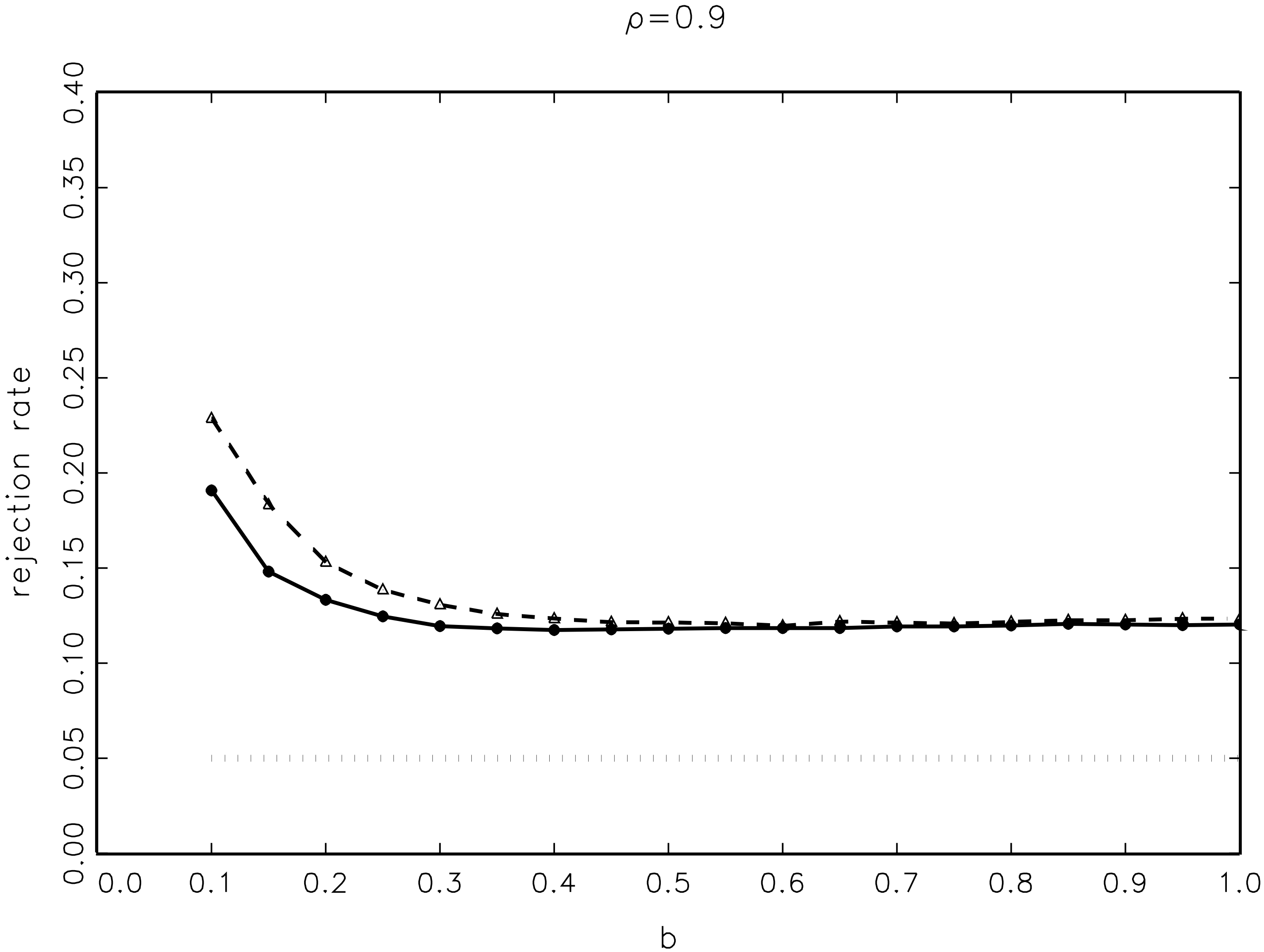


Figure 2.69: AM and ES - World War (quarterly), Bartlett, $T = 192$

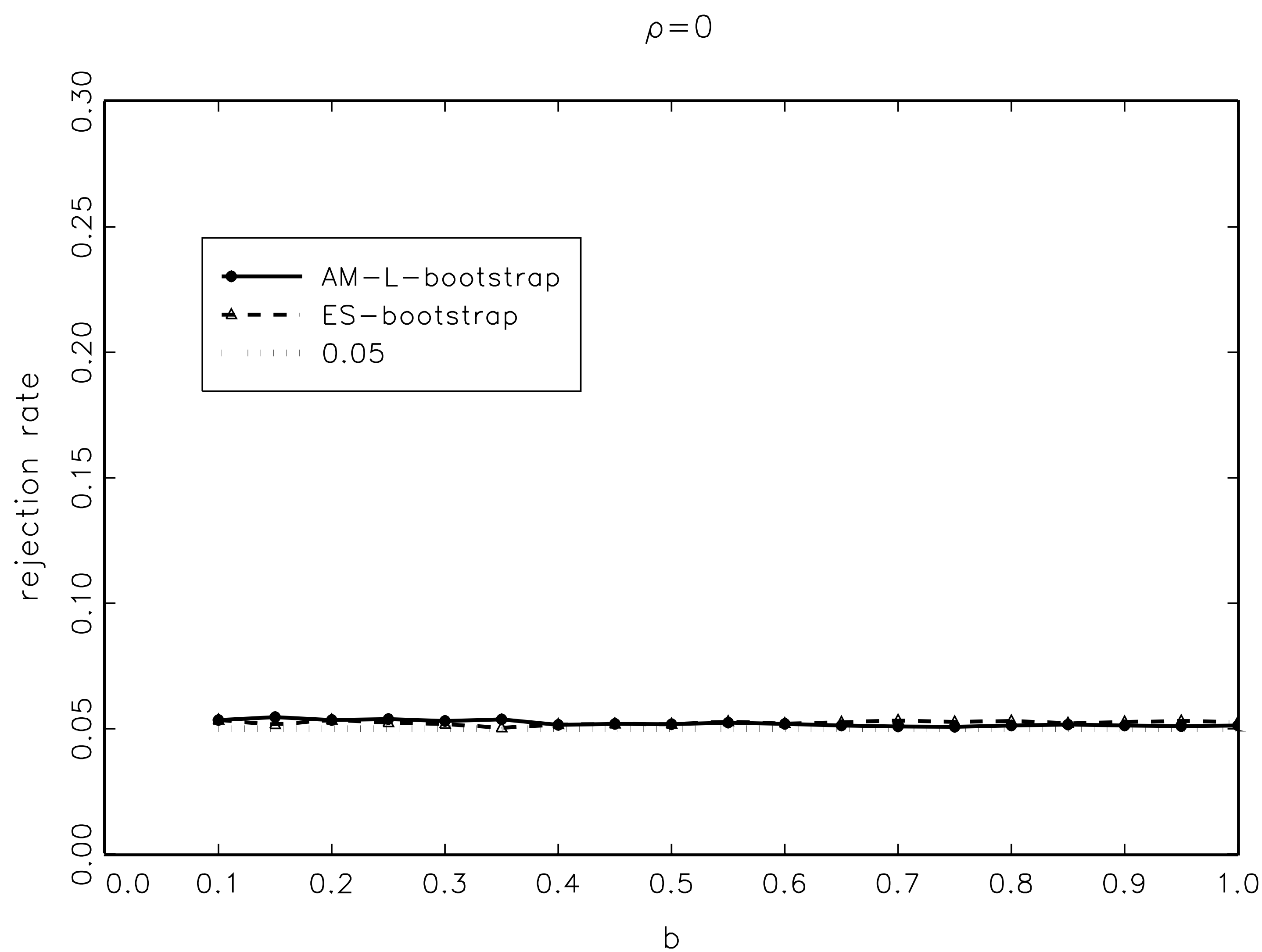


Figure 2.69: (cont'd)

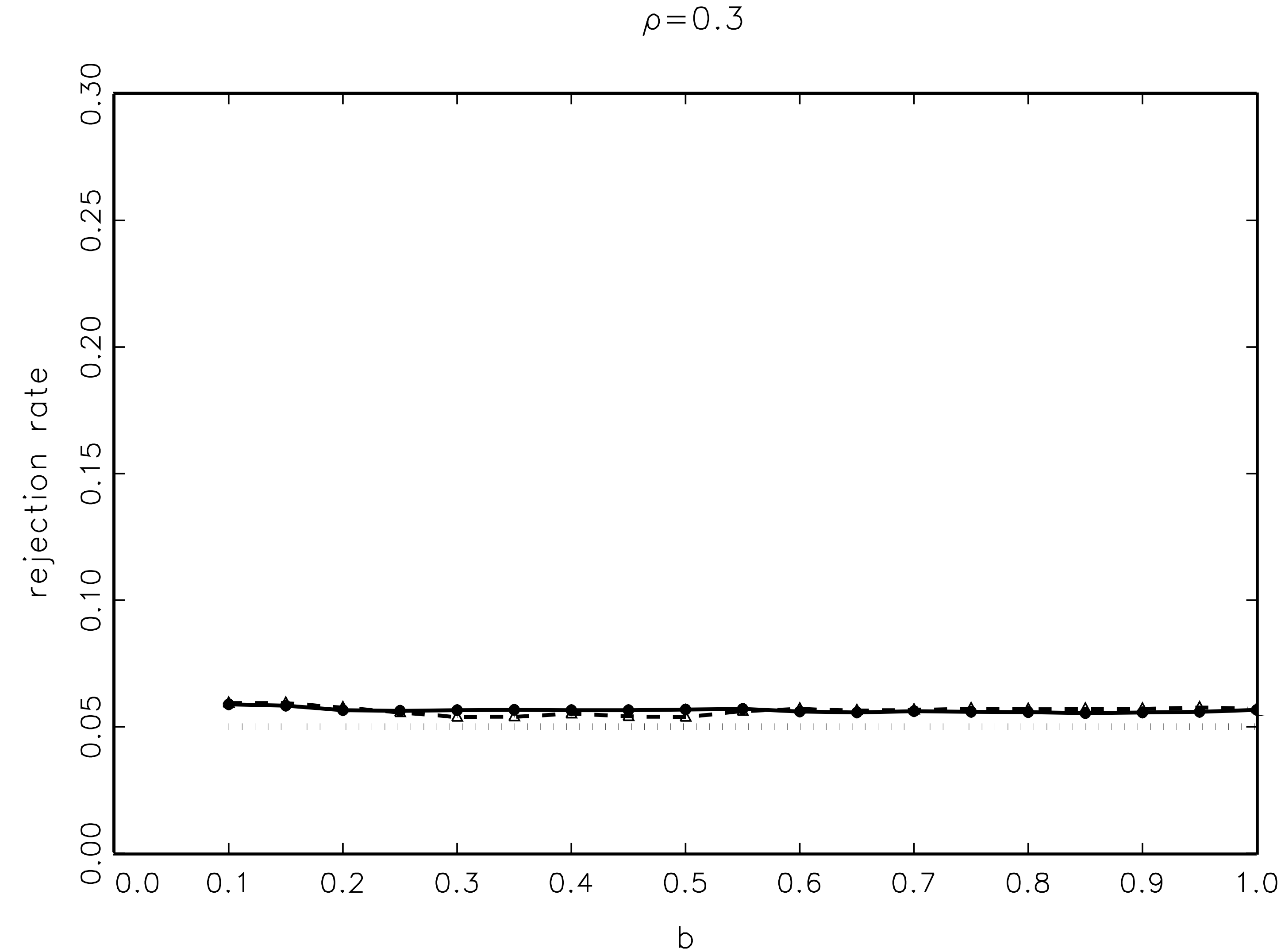


Figure 2.69: (cont'd)

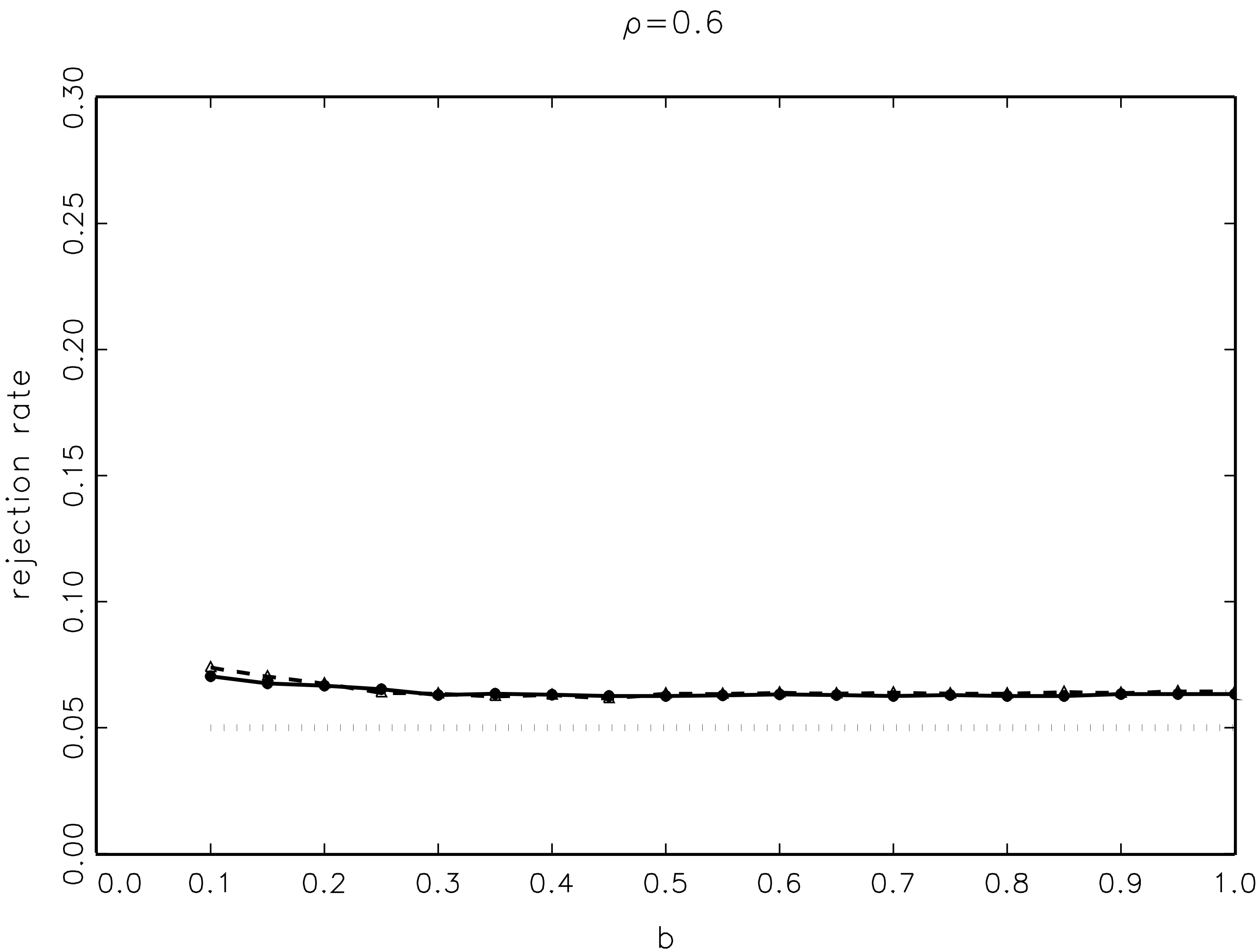


Figure 2.69: (cont'd)

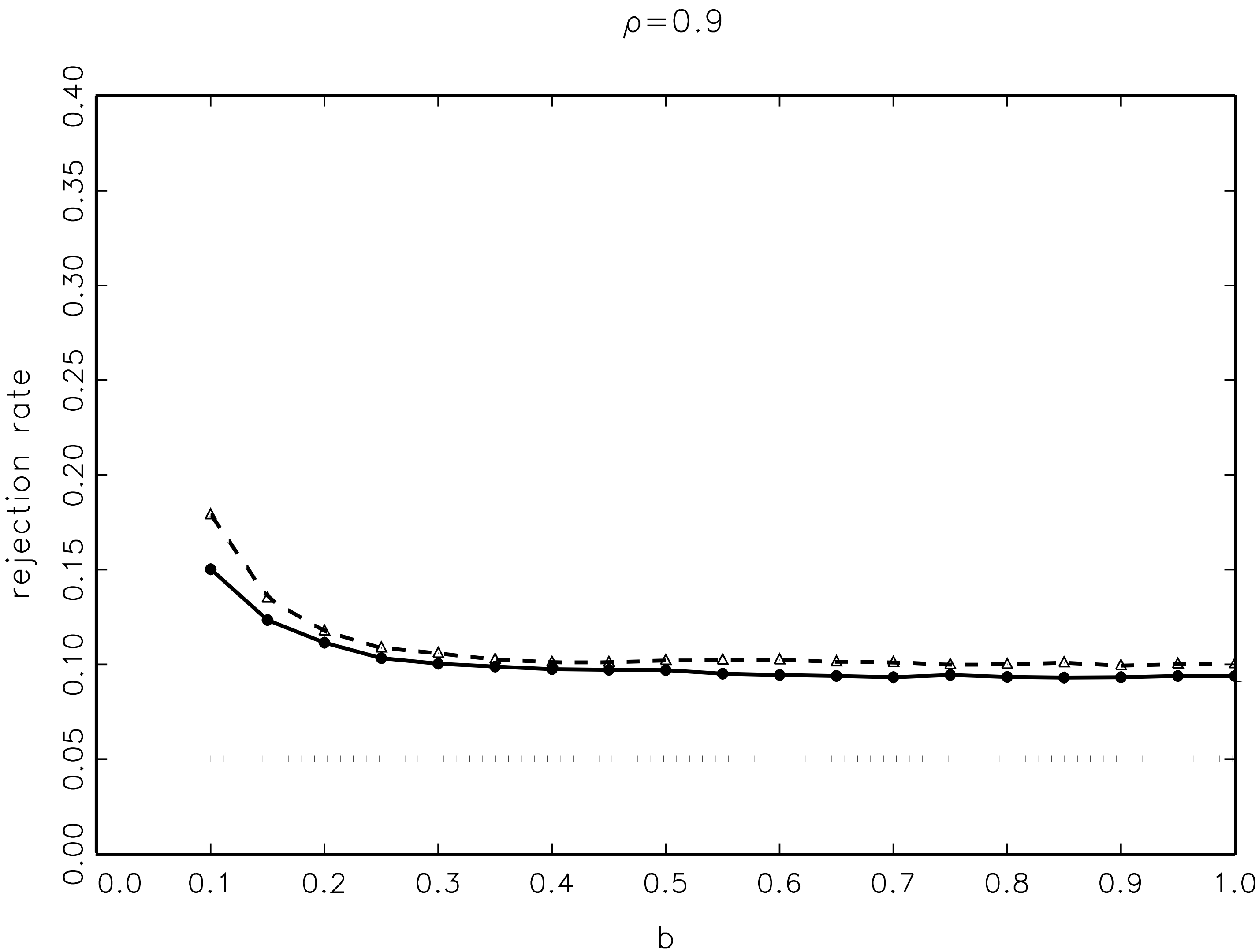


Figure 2.70: AM and ES - World War (quarterly), Bartlett, $T = 240$

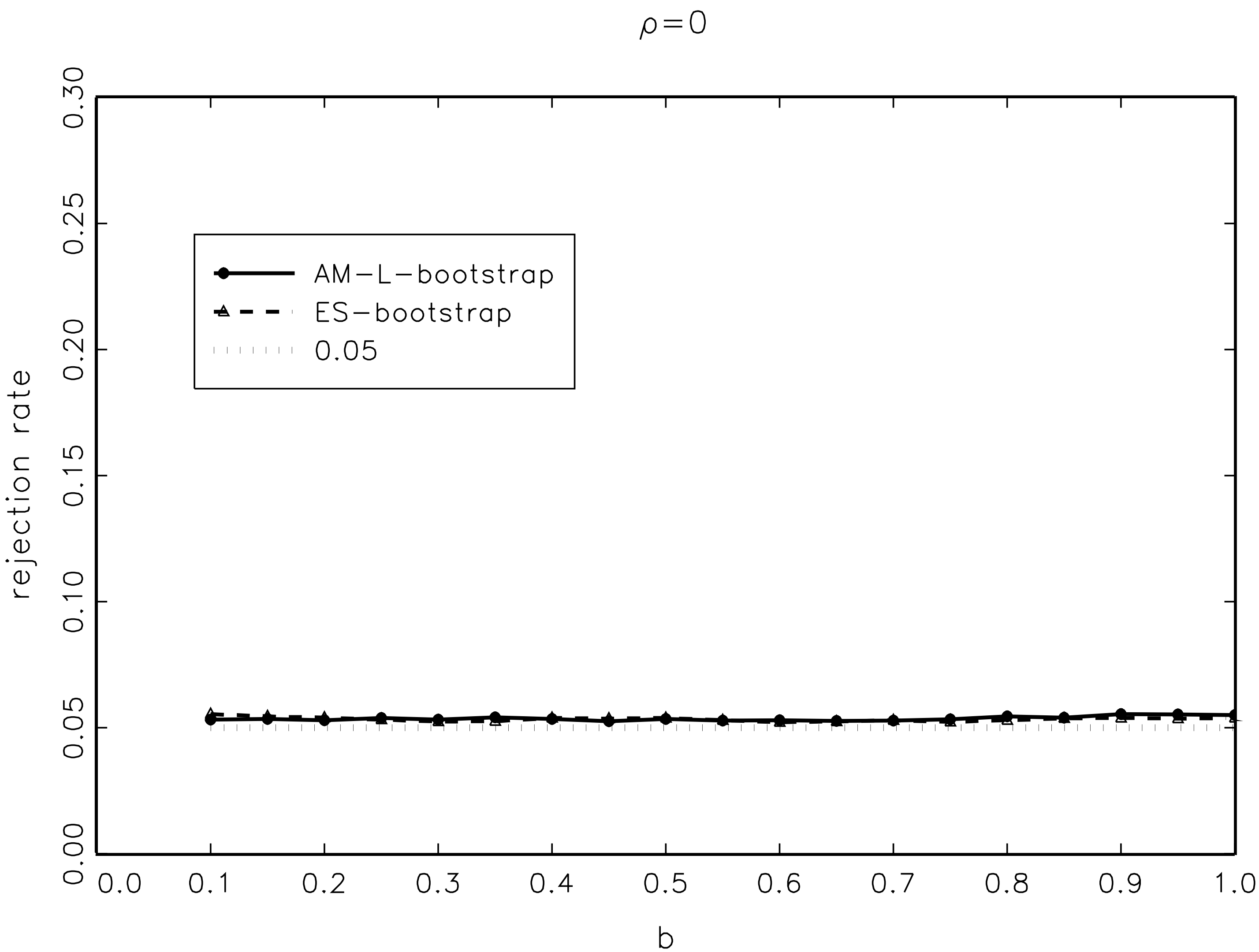


Figure 2.70: (cont'd)

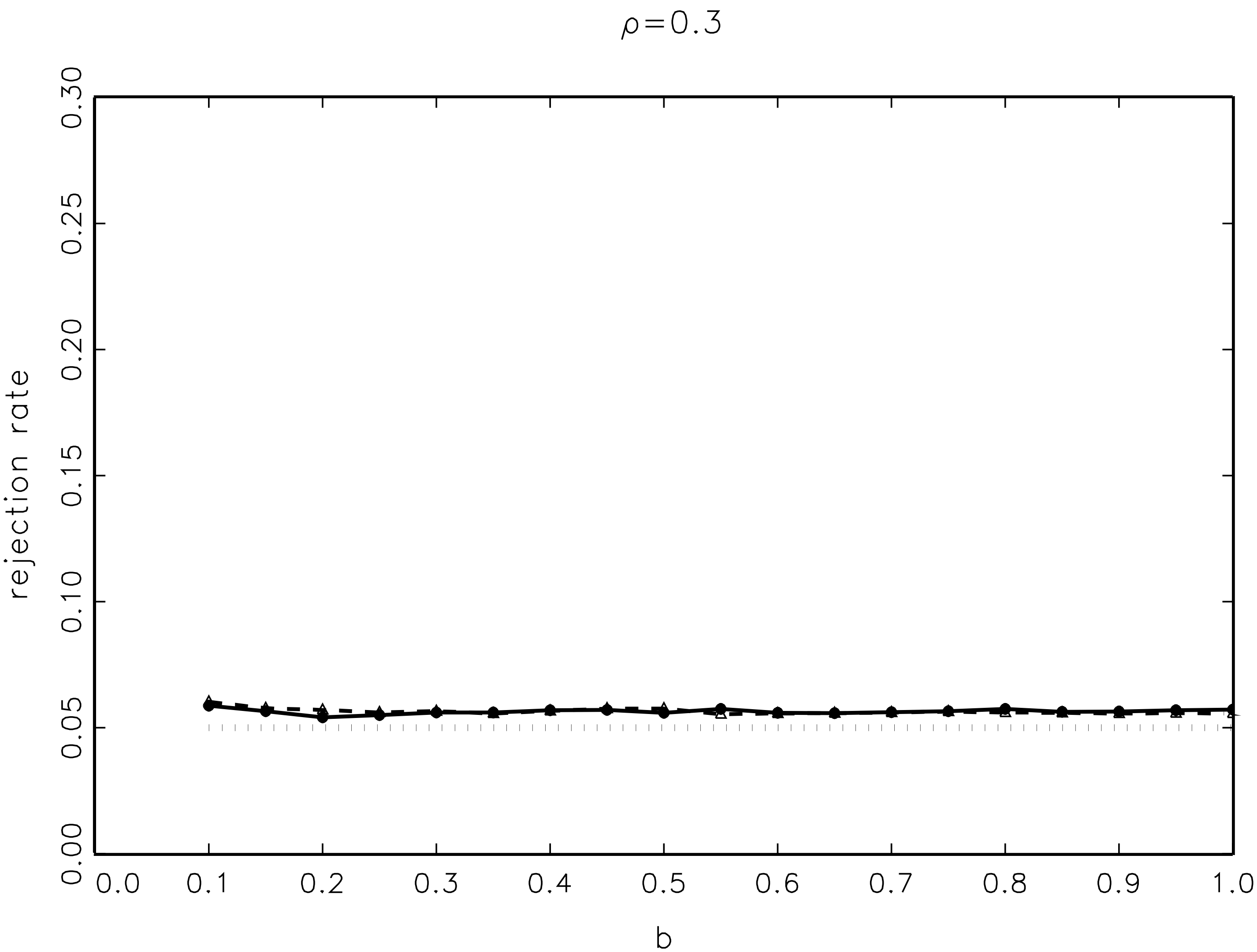


Figure 2.70: (cont'd)

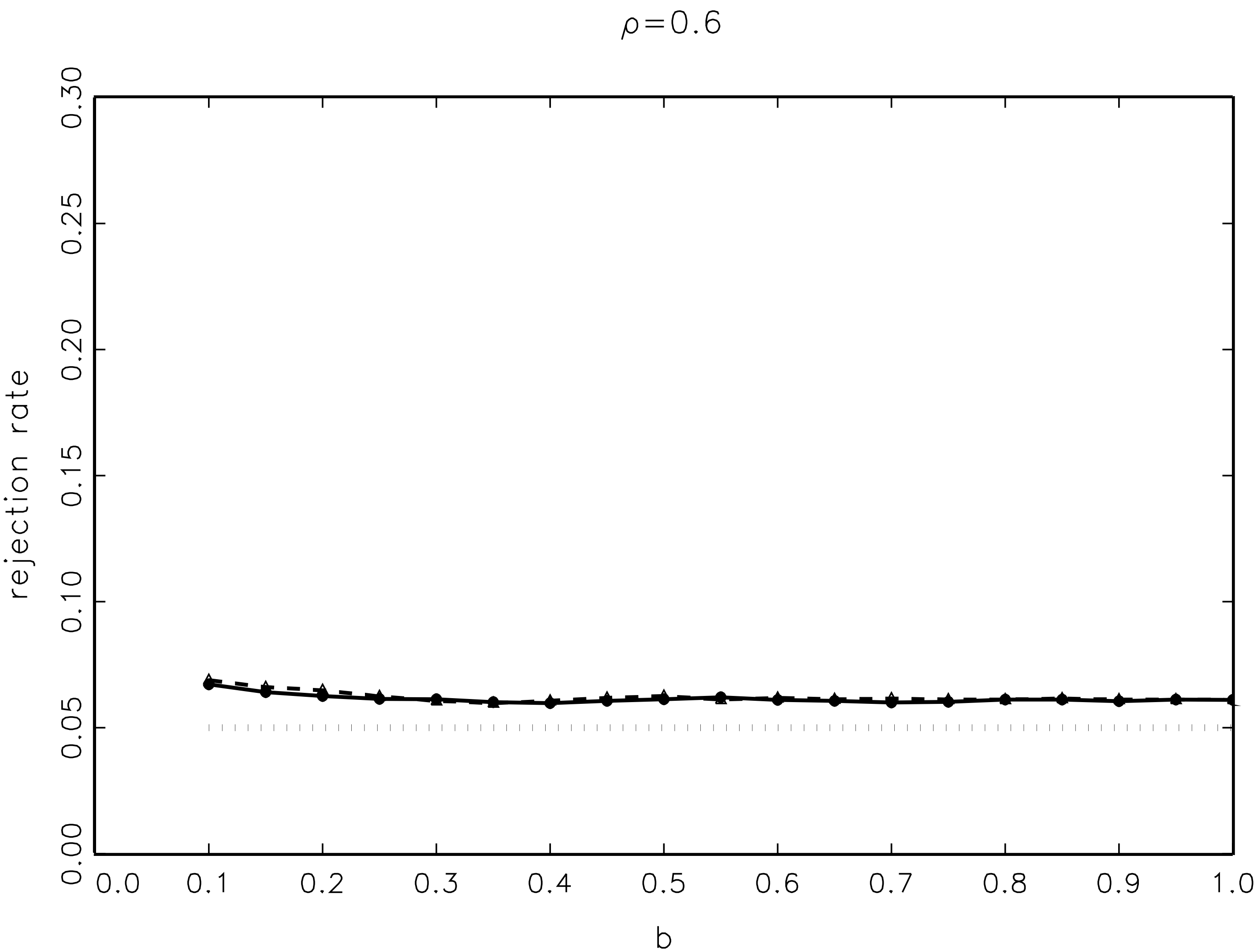


Figure 2.70: (cont'd)

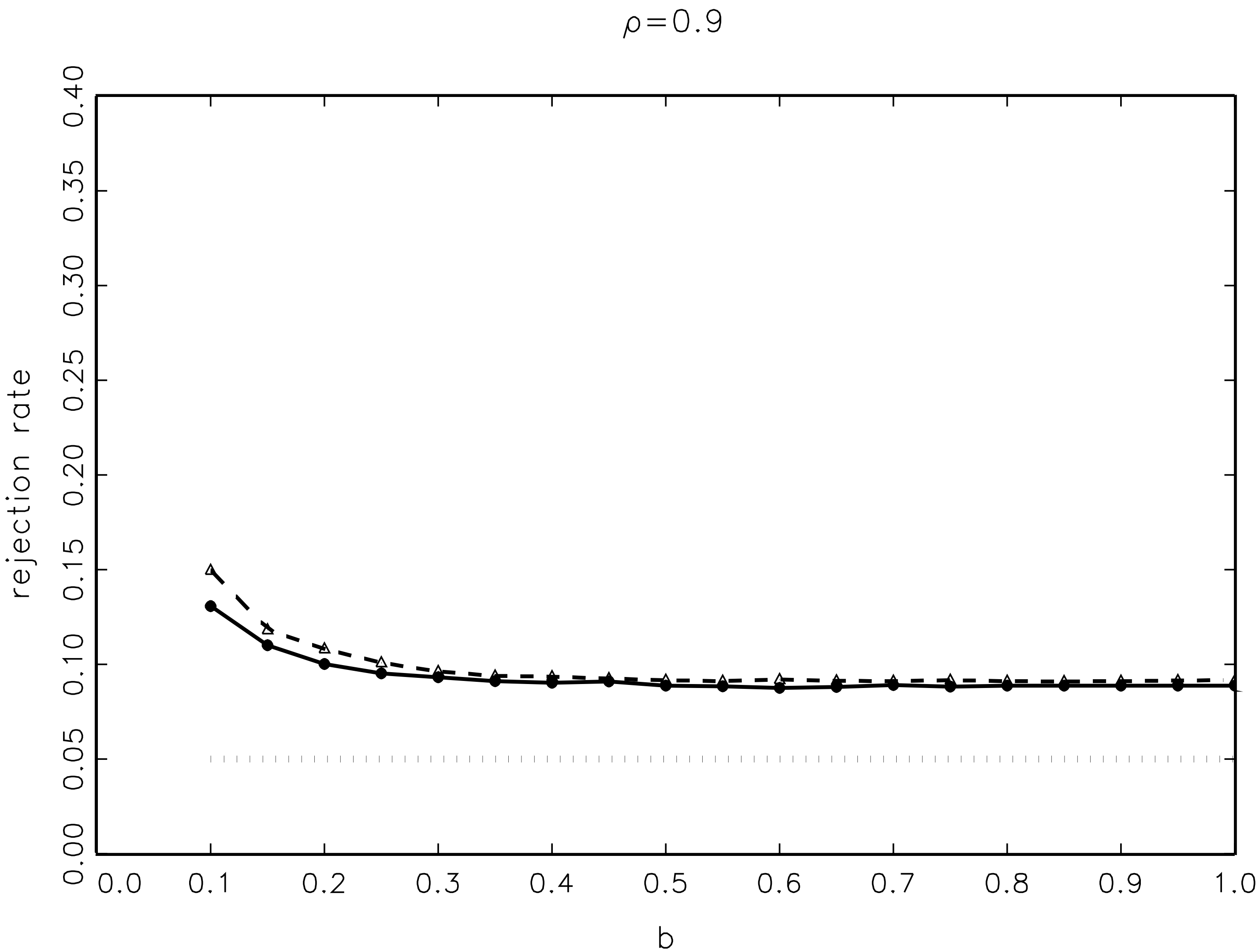


Figure 2.71: AM and ES - Initially Scarce Data, Bartlett, $N_Q = 12$ $N_M = 12$

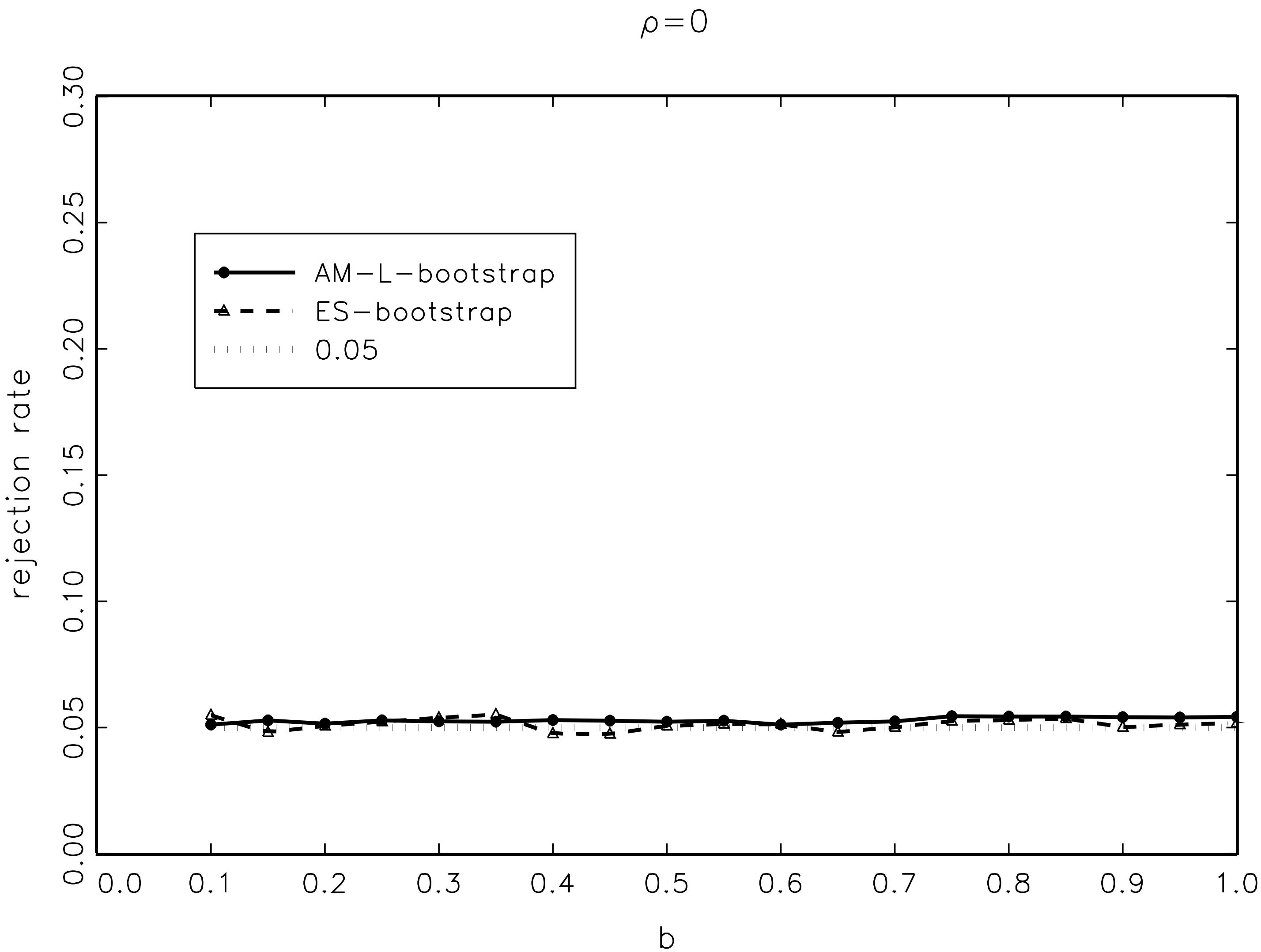


Figure 2.71: (cont'd)

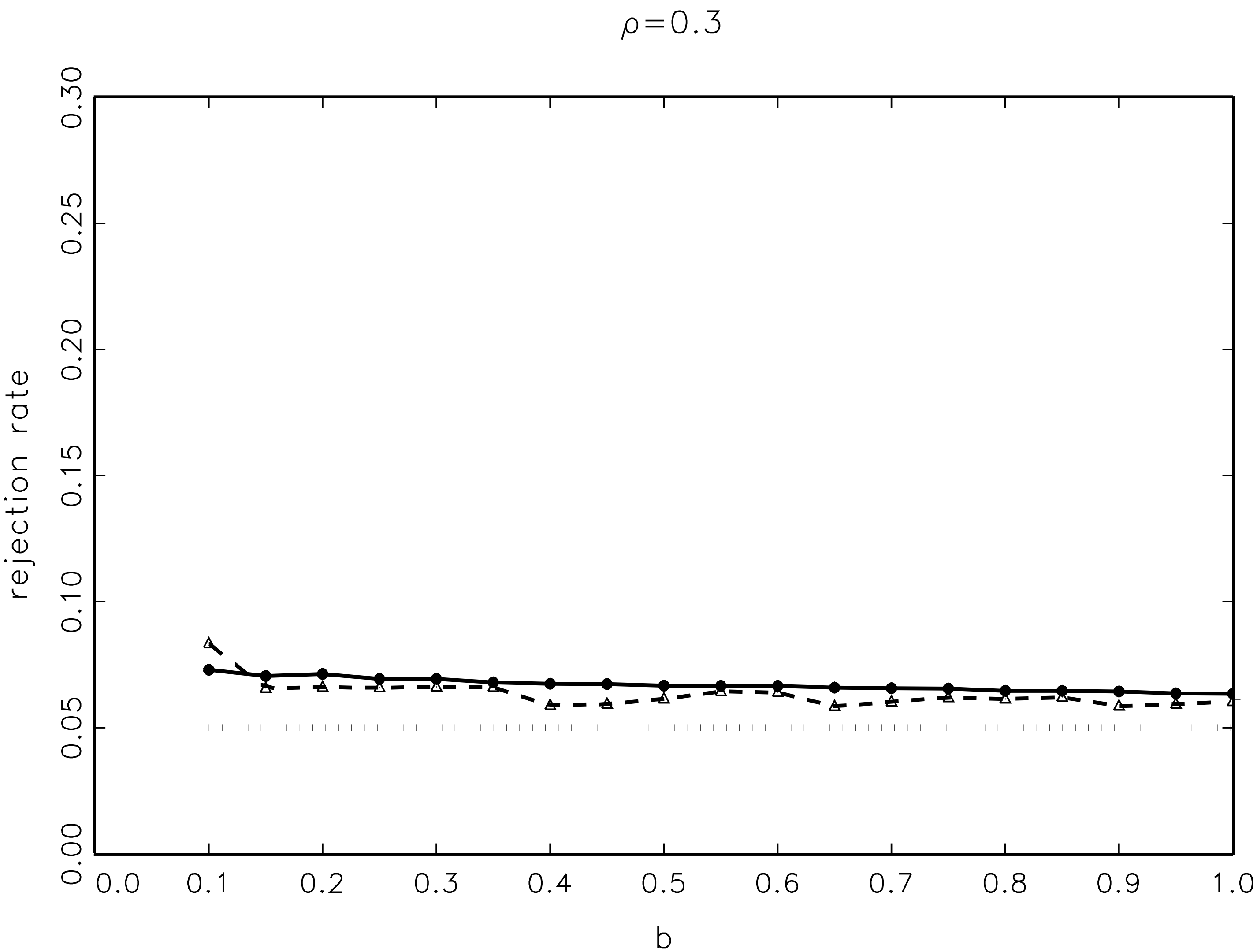


Figure 2.71: (cont'd)

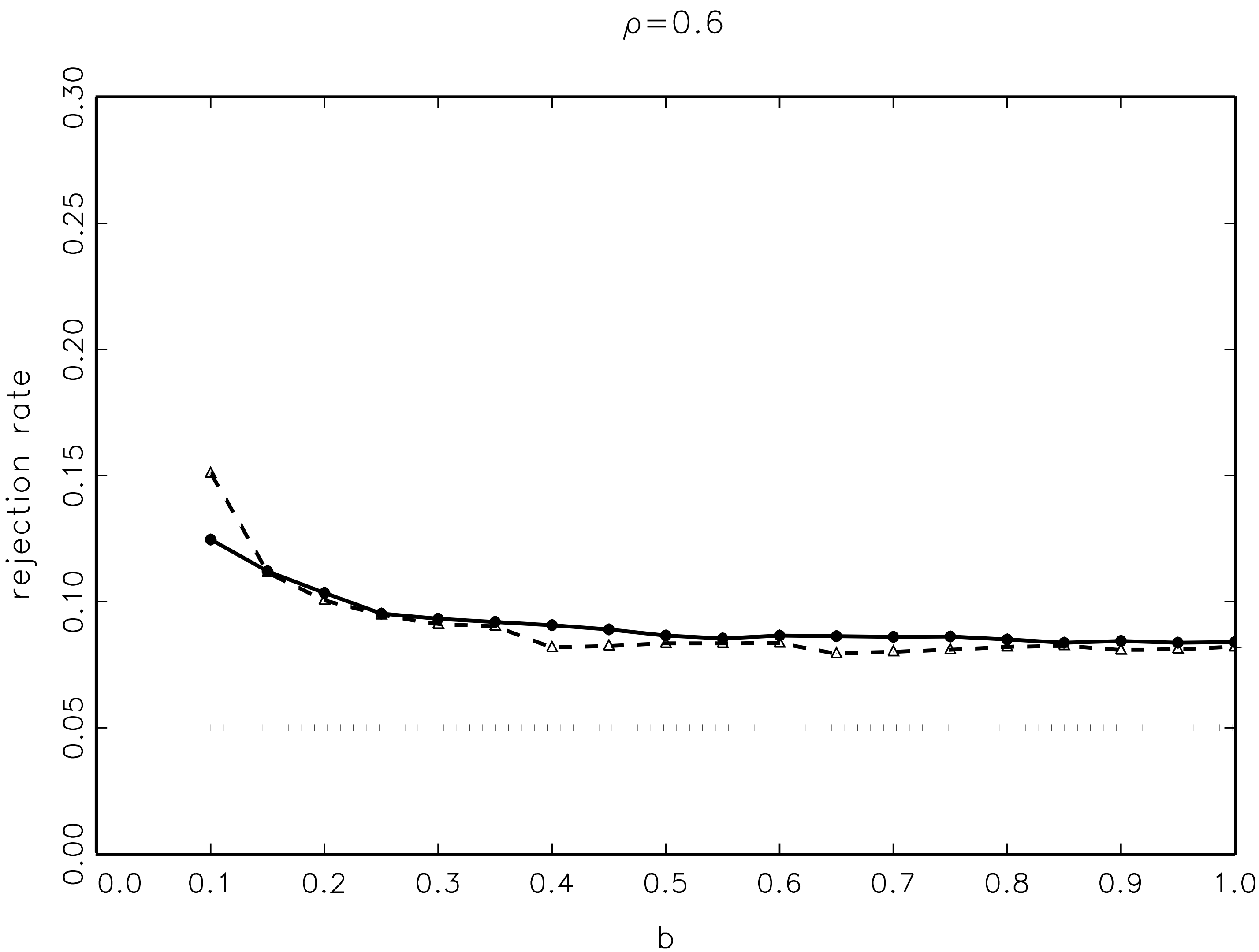


Figure 2.71: (cont'd)

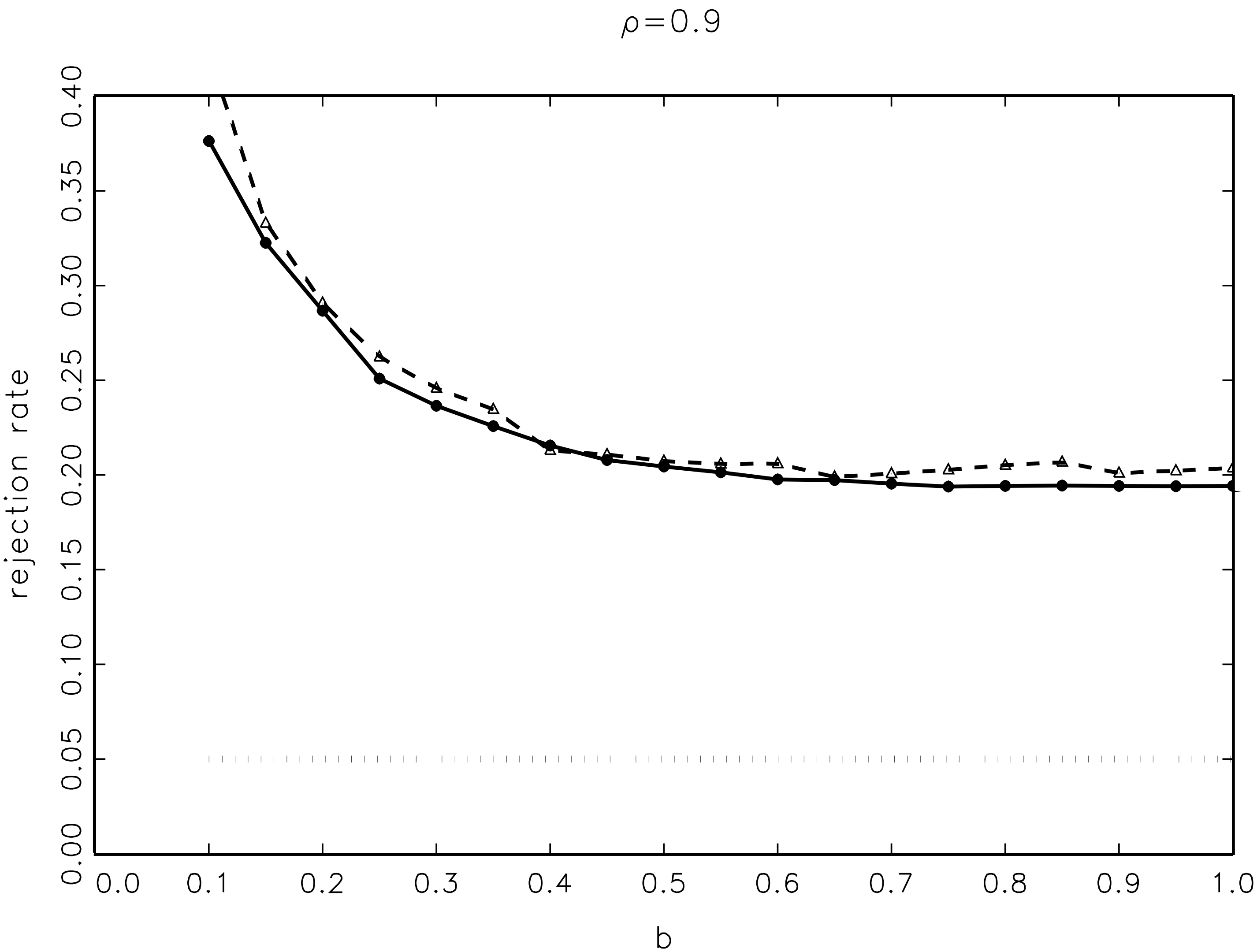


Figure 2.72: AM and ES - Initially Scarce Data, Bartlett, $N_Q = 12$ $N_M = 24$

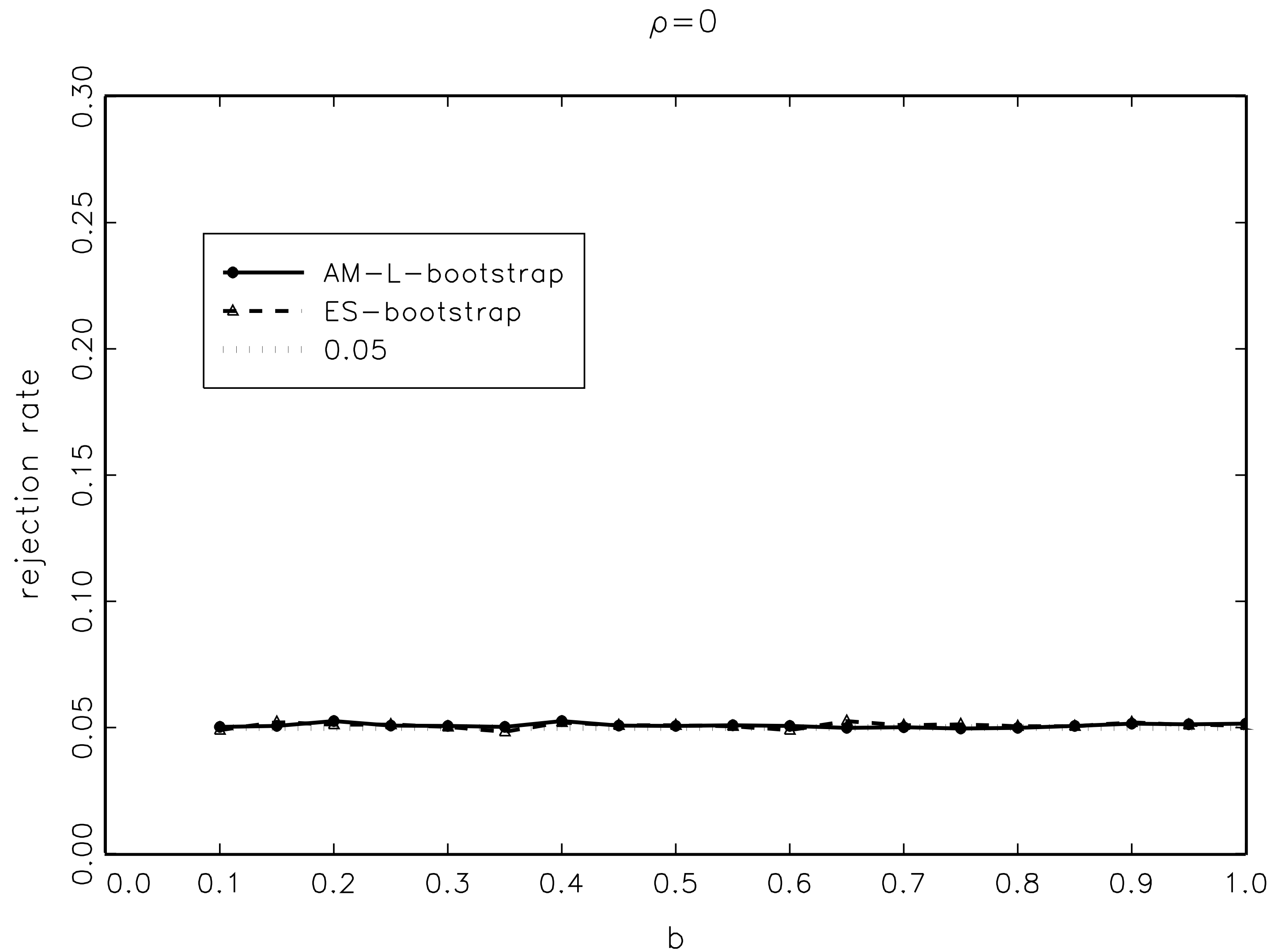


Figure 2.72: (cont'd)

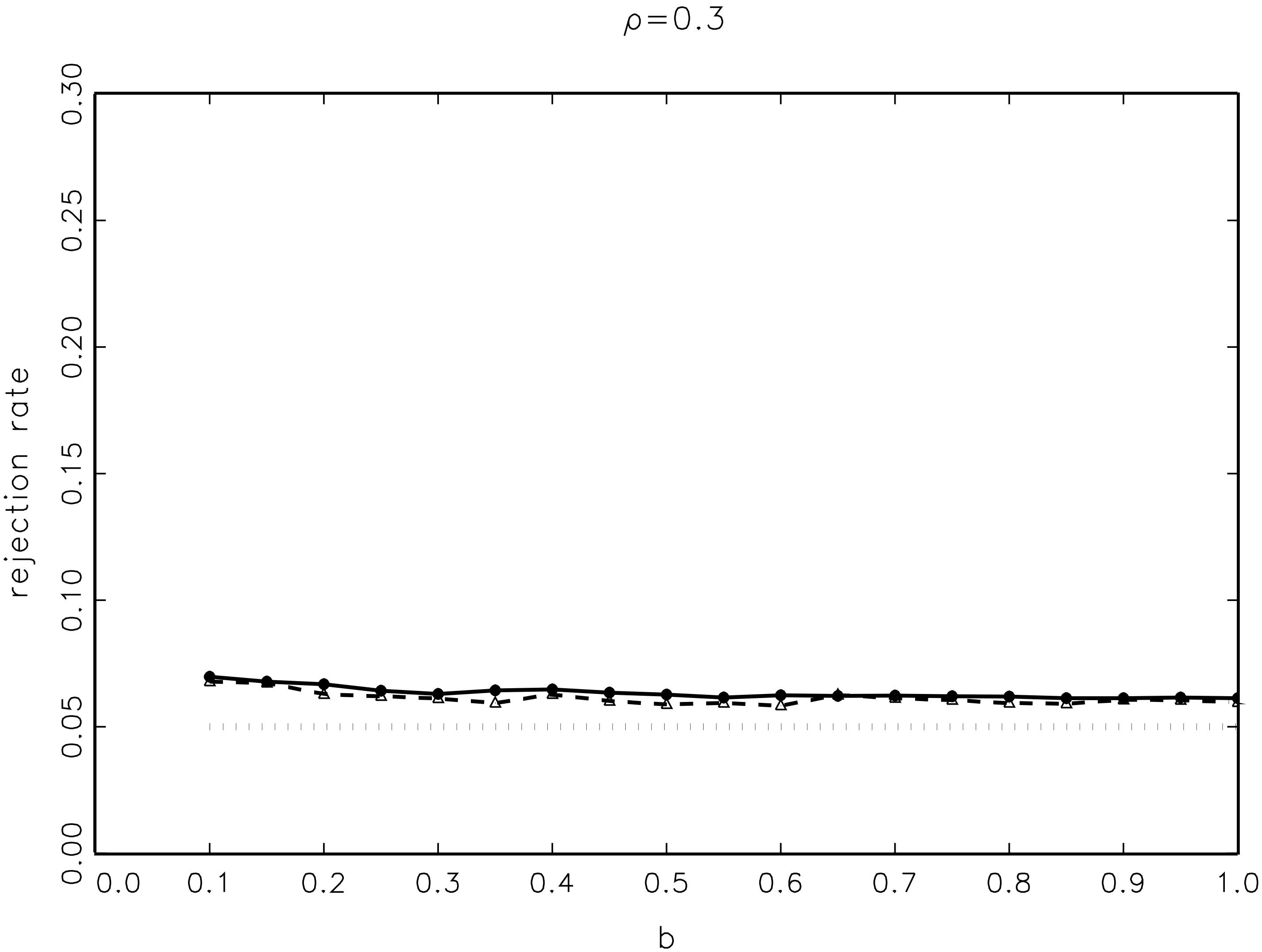


Figure 2.72: (cont'd)

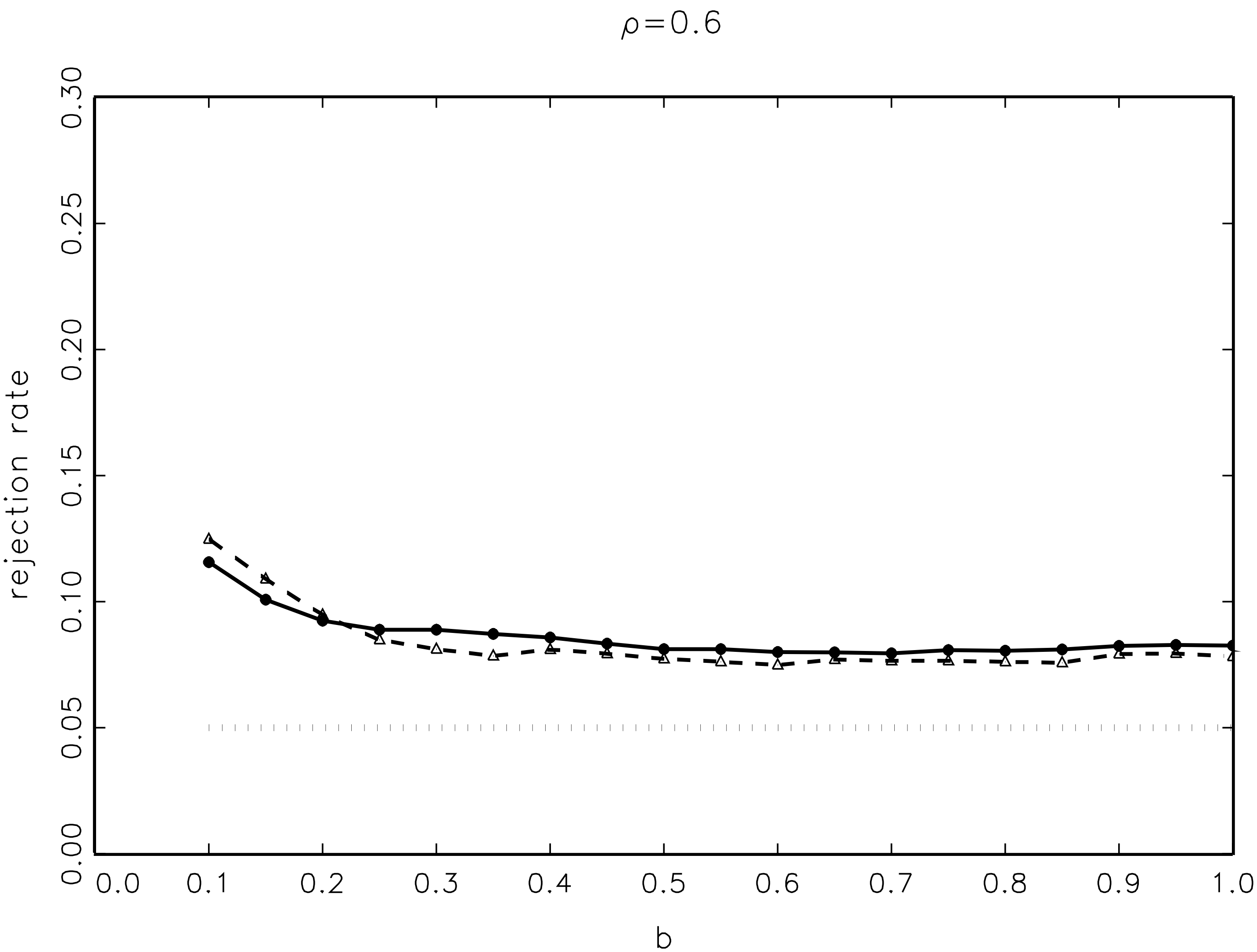


Figure 2.72: (cont'd)

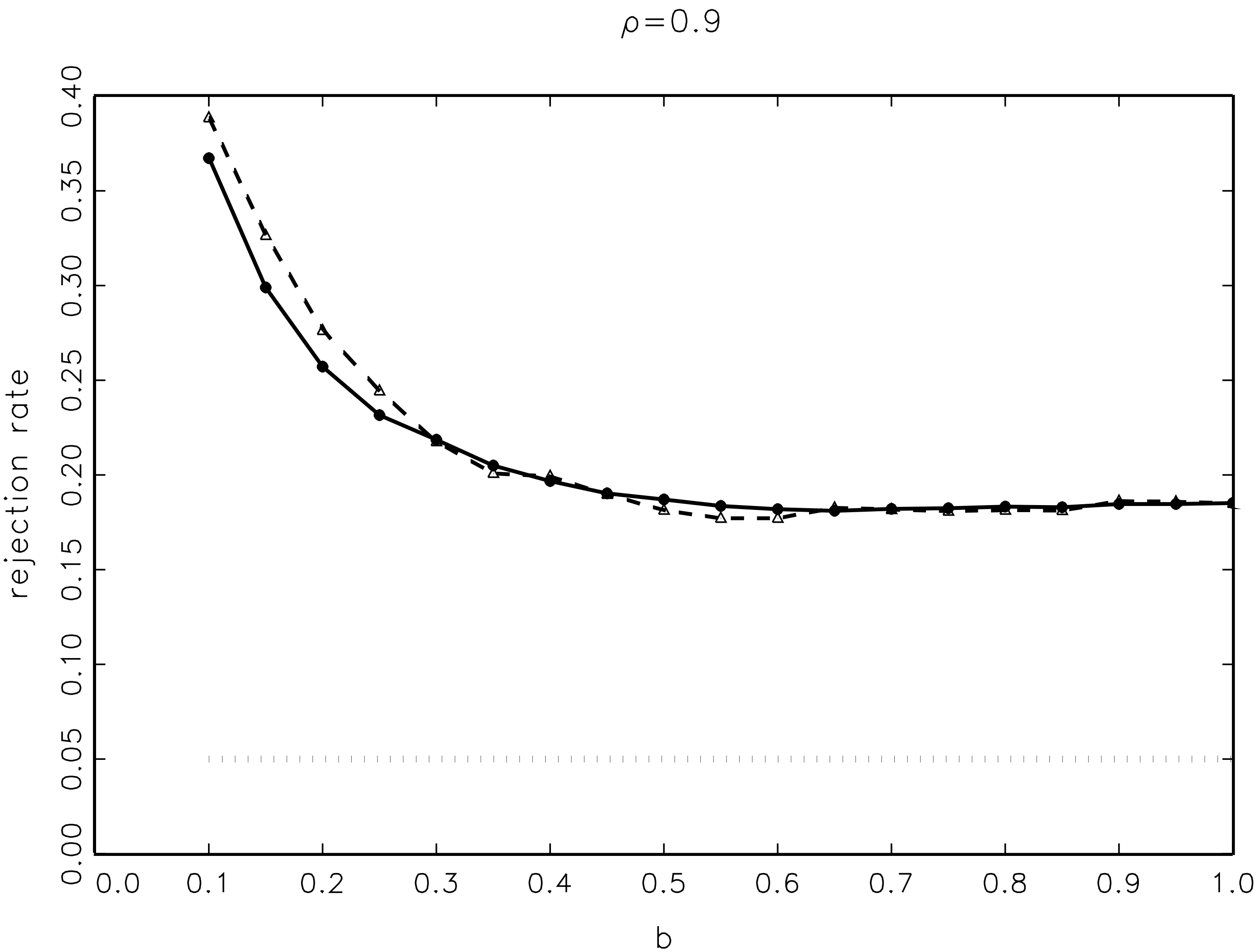


Figure 2.73: AM and ES - Initially Scarce Data, Bartlett, $N_Q = 12$ $N_M = 48$

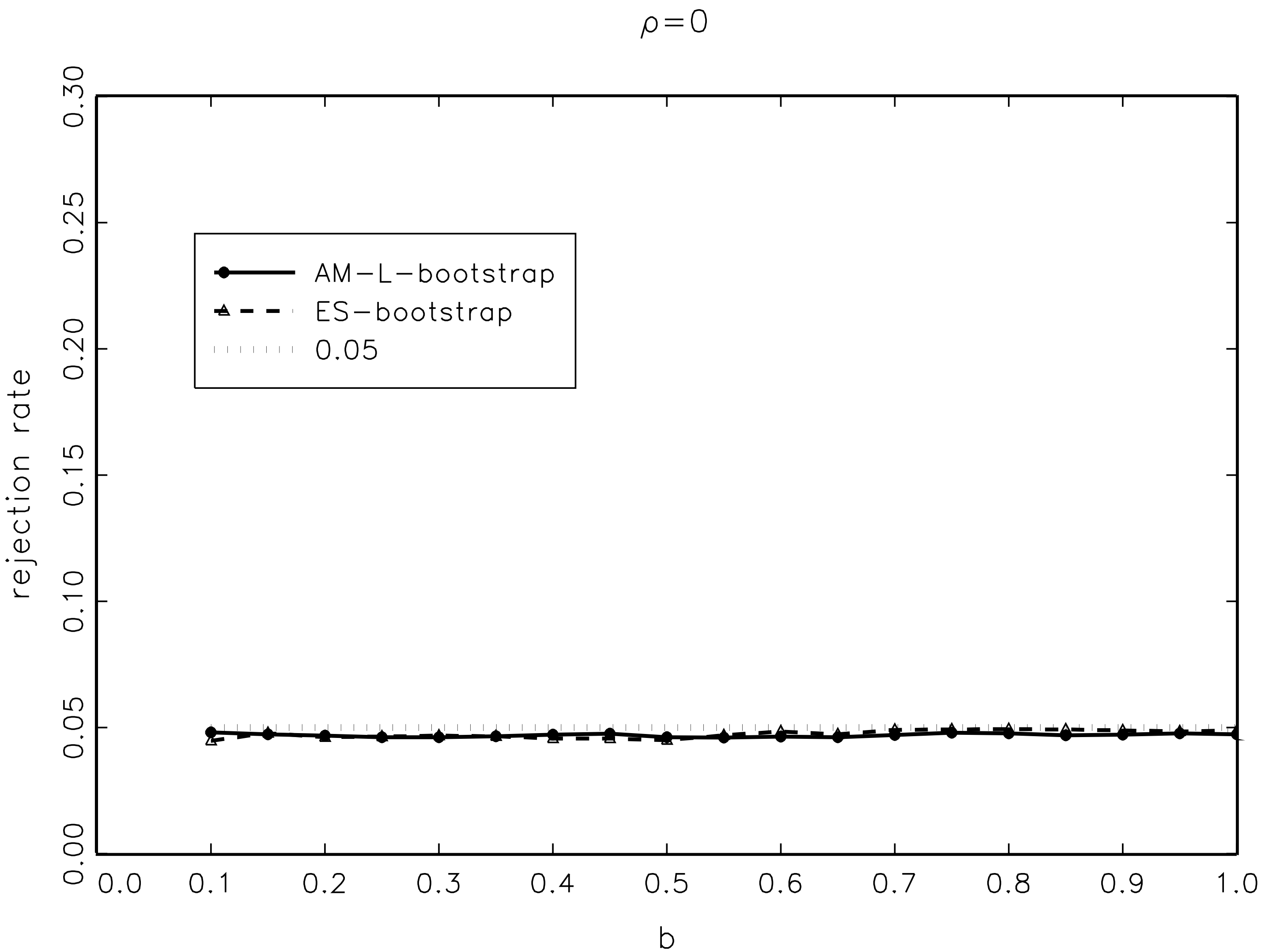


Figure 2.73: (cont'd)

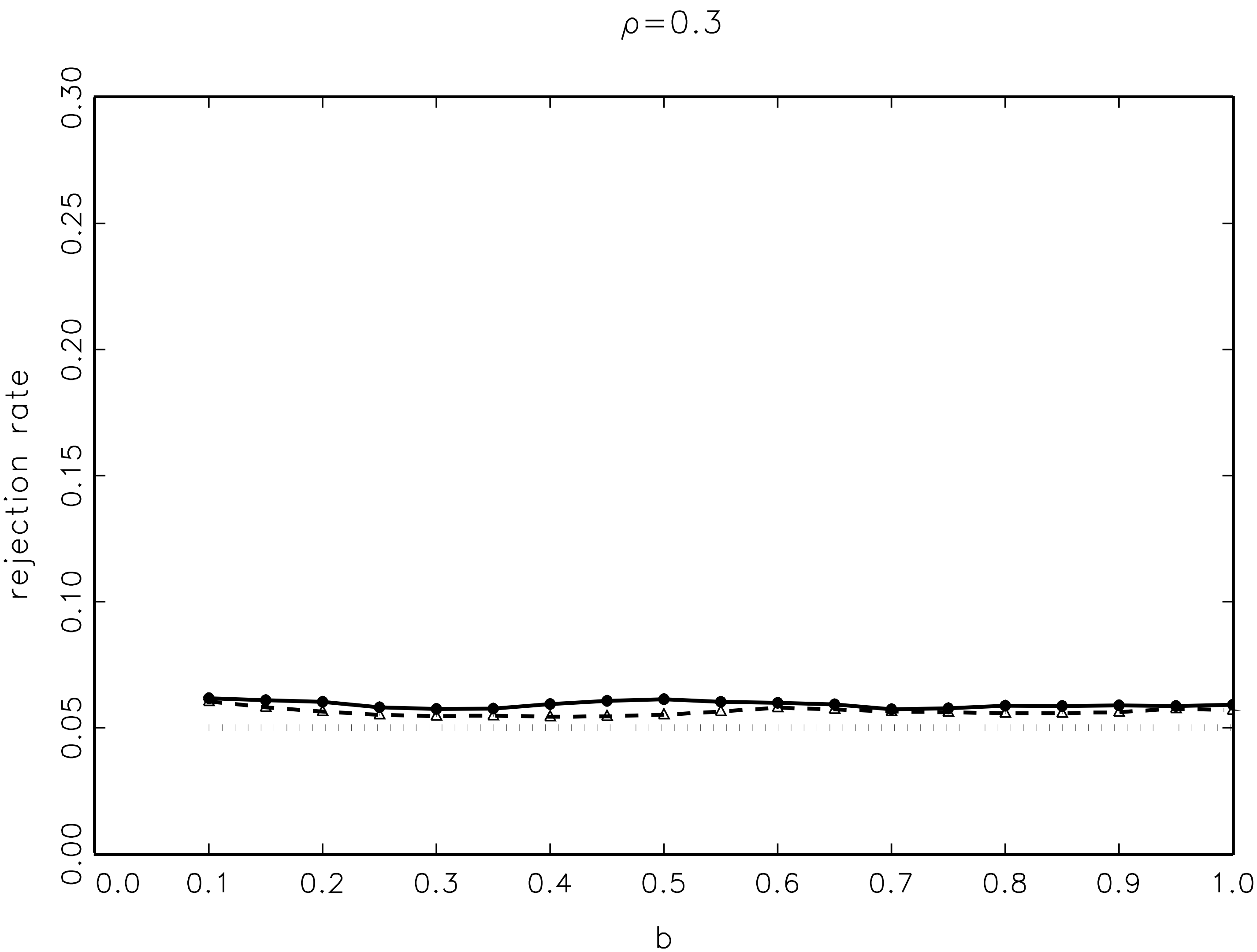


Figure 2.73: (cont'd)

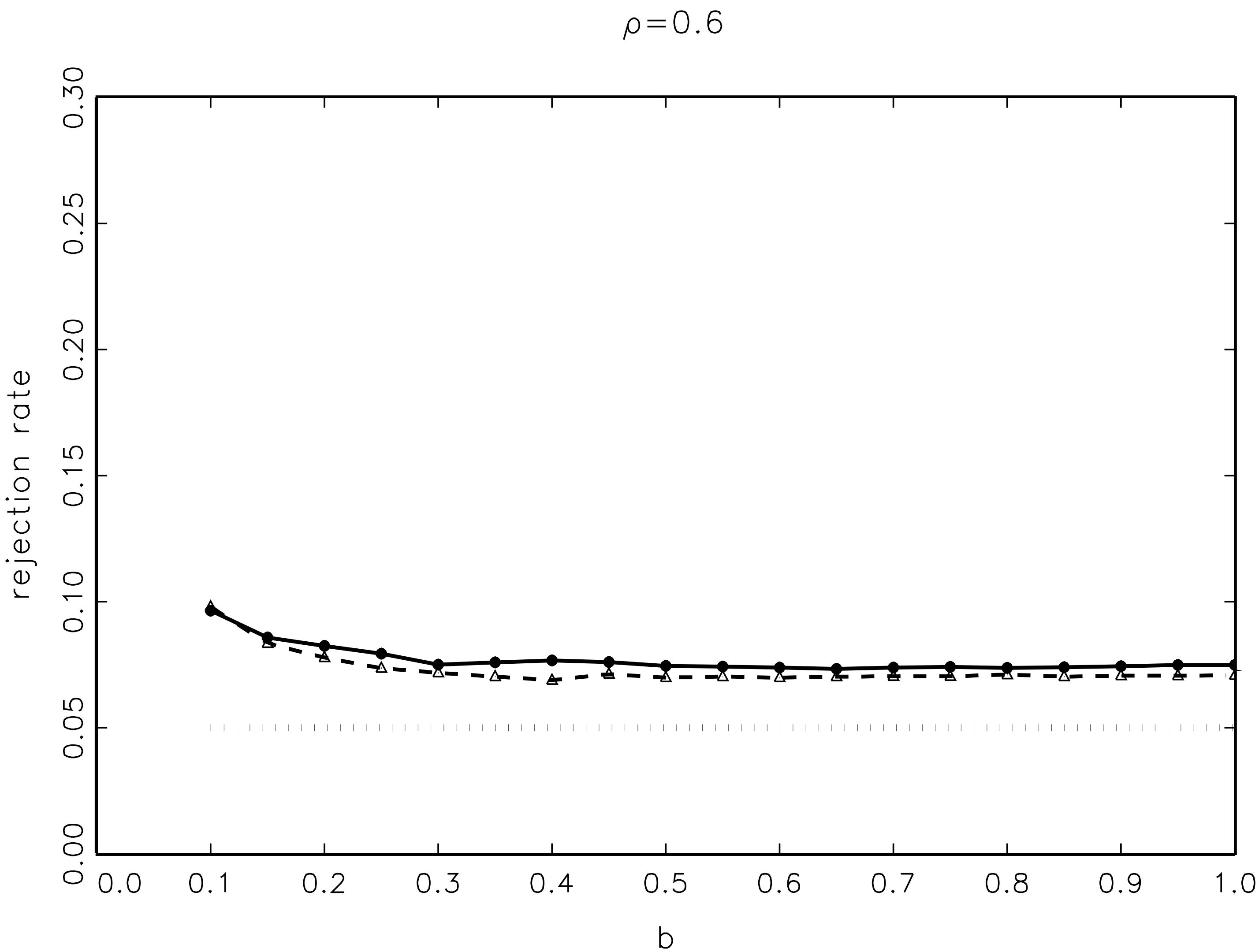


Figure 2.73: (cont'd)

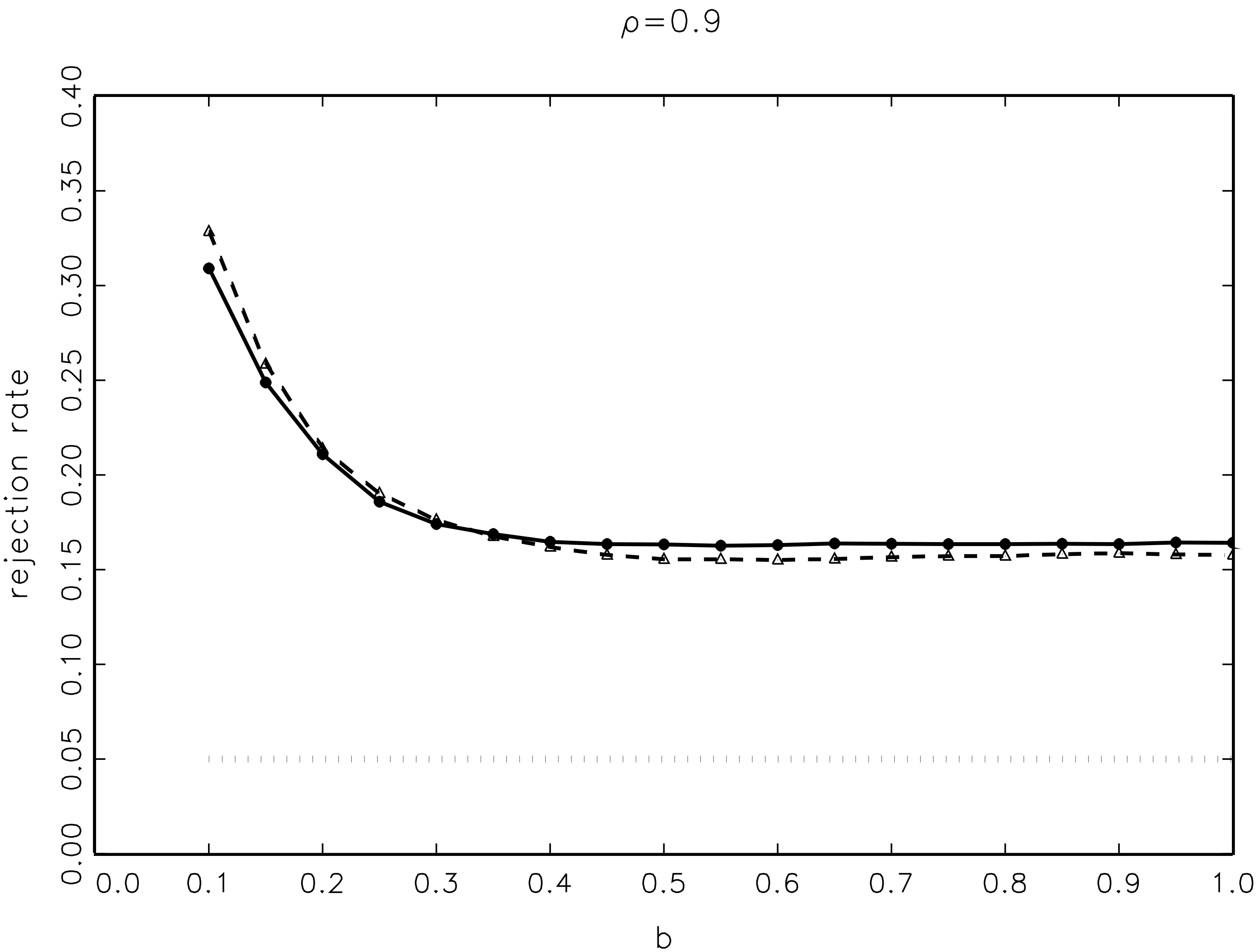


Figure 2.74: AM and ES - Initially Scarce Data, Bartlett, $N_Q = 24$ $N_M = 12$

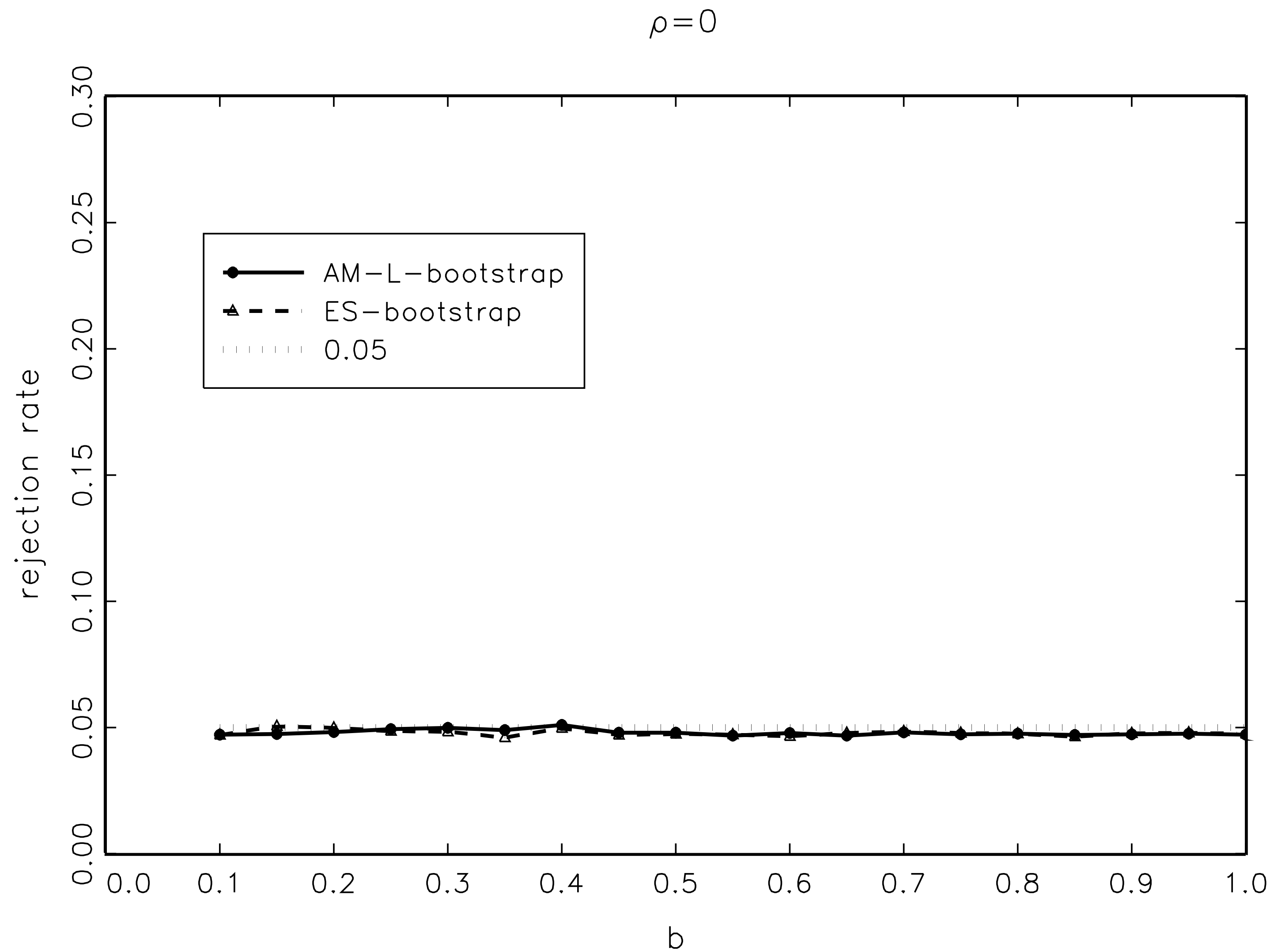


Figure 2.74: (cont'd)

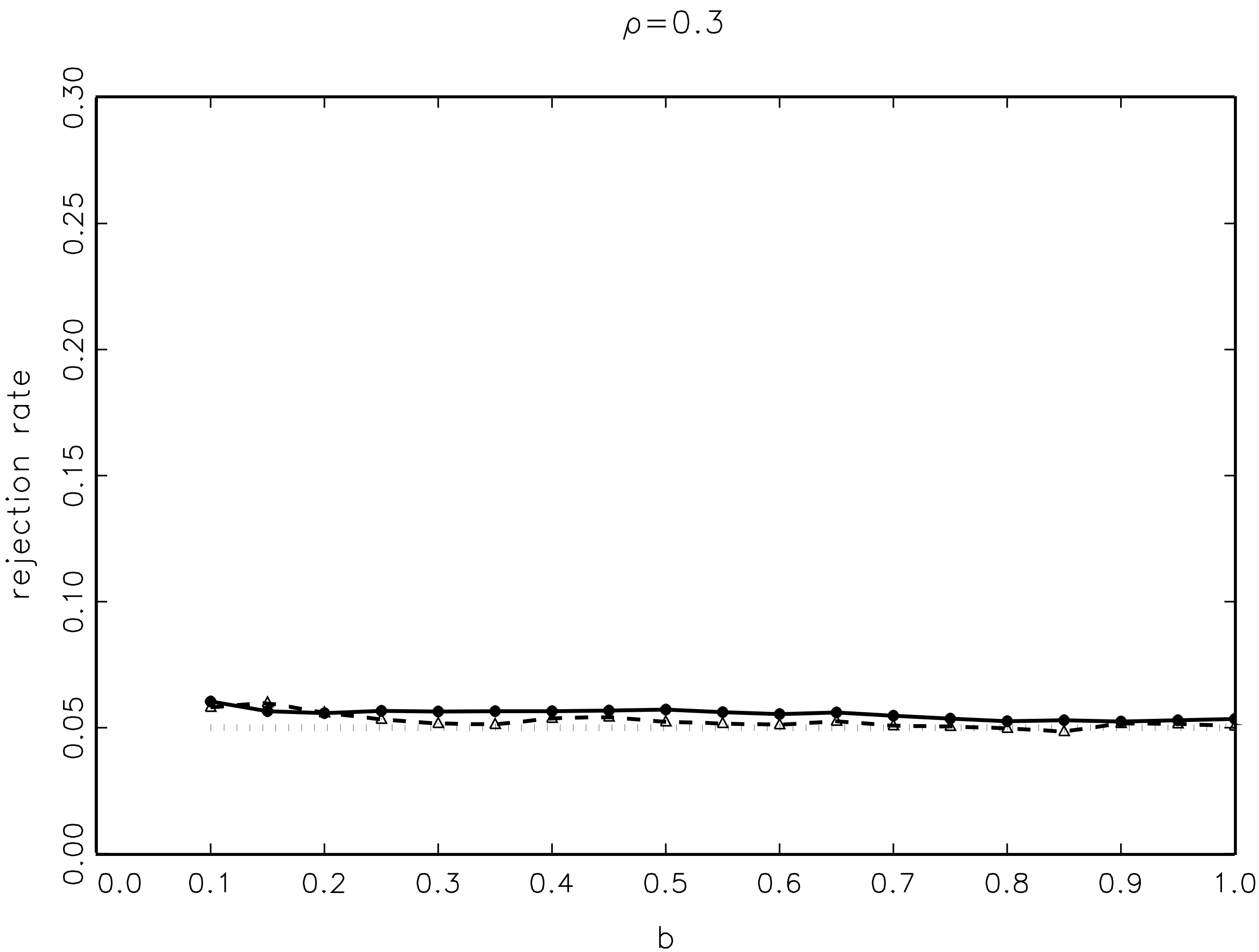


Figure 2.74: (cont'd)

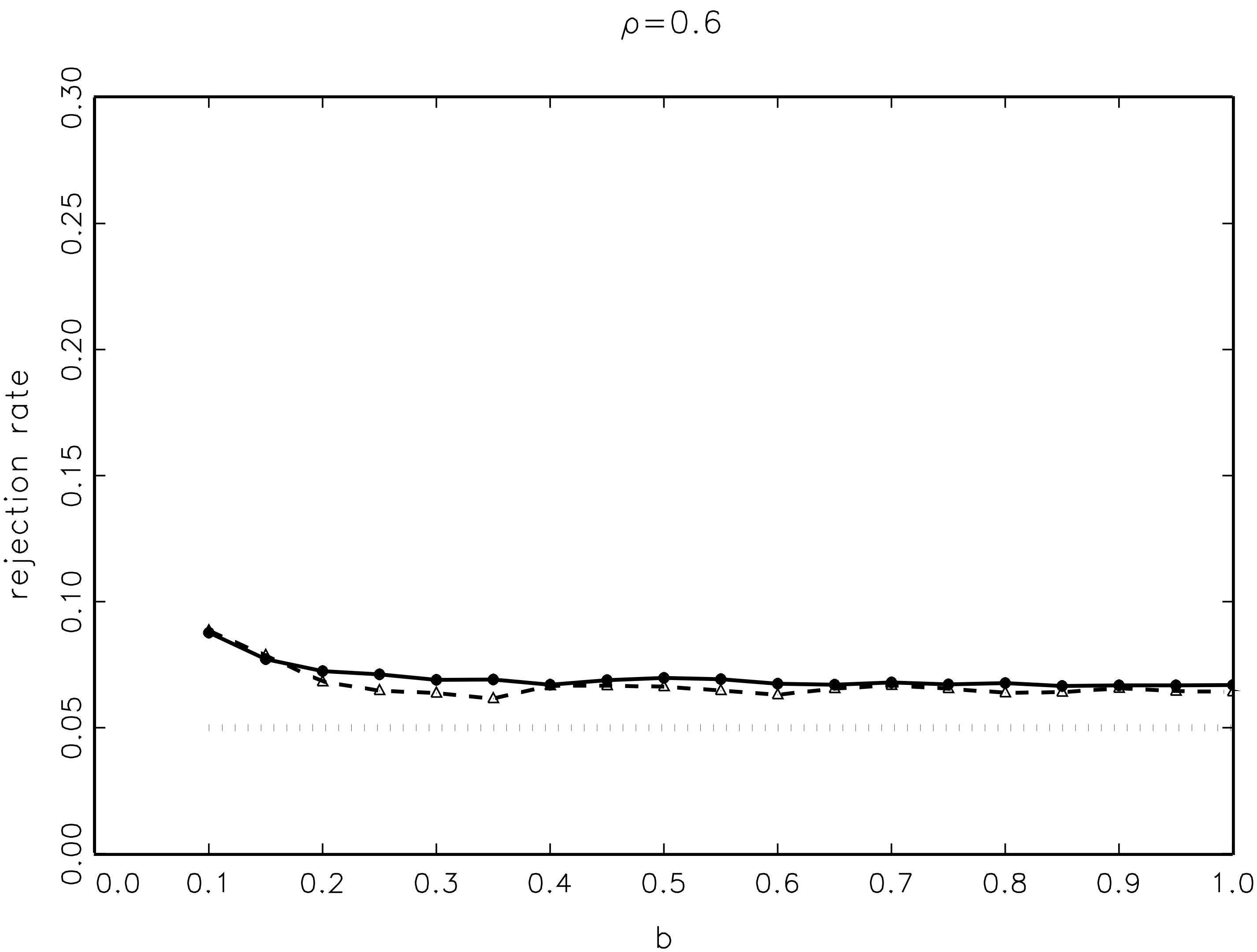


Figure 2.74: (cont'd)

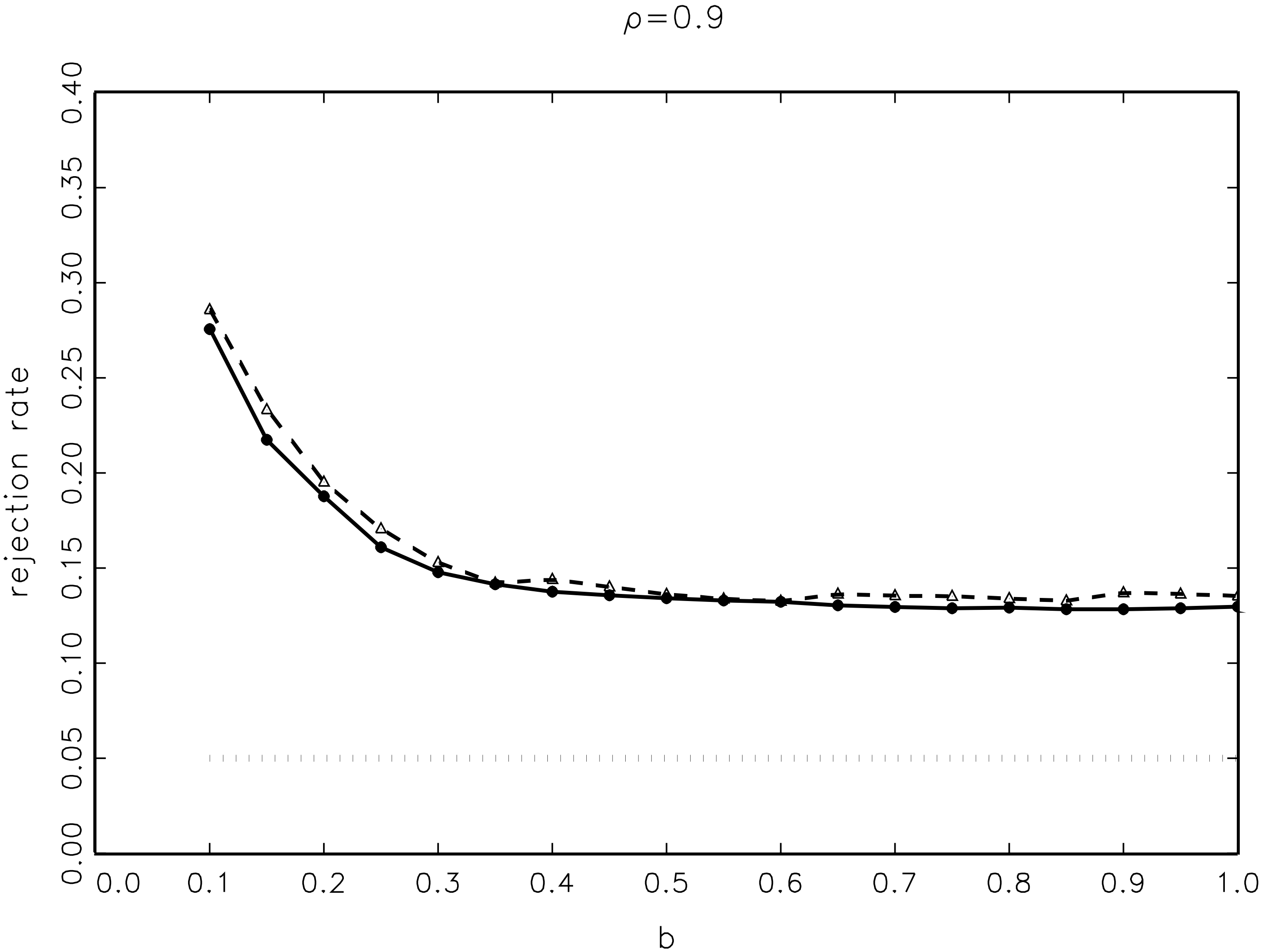


Figure 2.75: AM and ES - Initially Scarce Data, Bartlett, $N_Q = 24$ $N_M = 24$

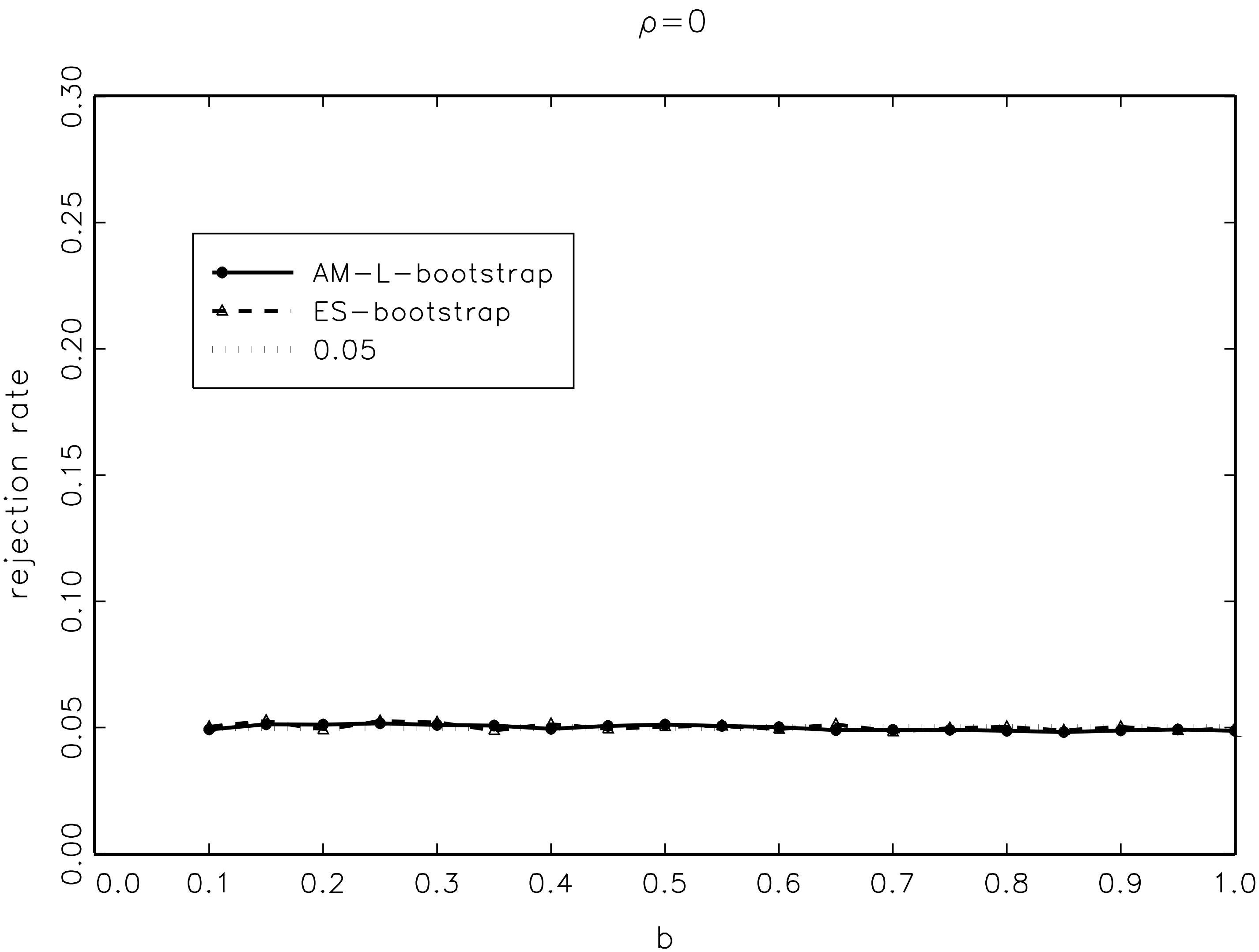


Figure 2.75: (cont'd)

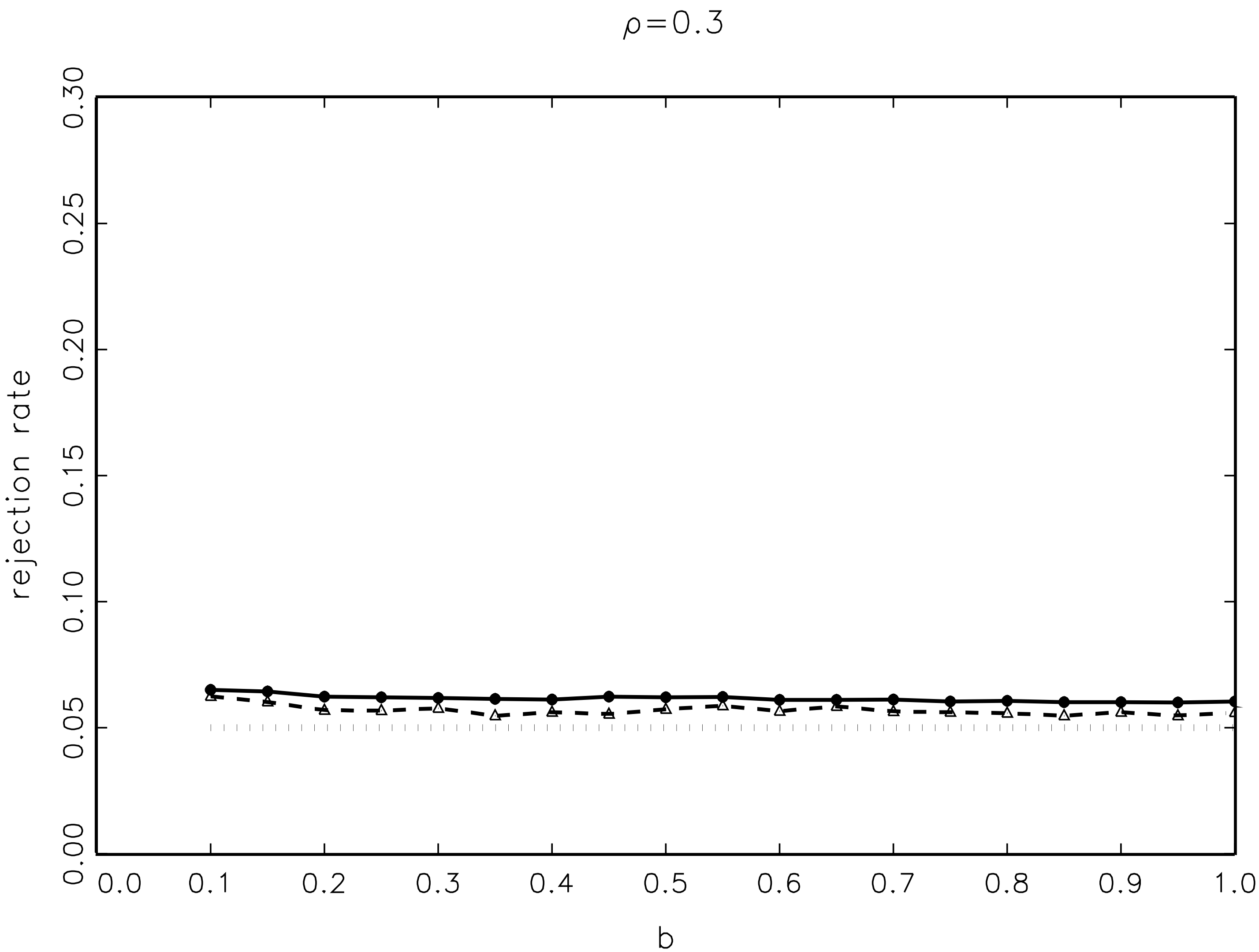


Figure 2.75: (cont'd)

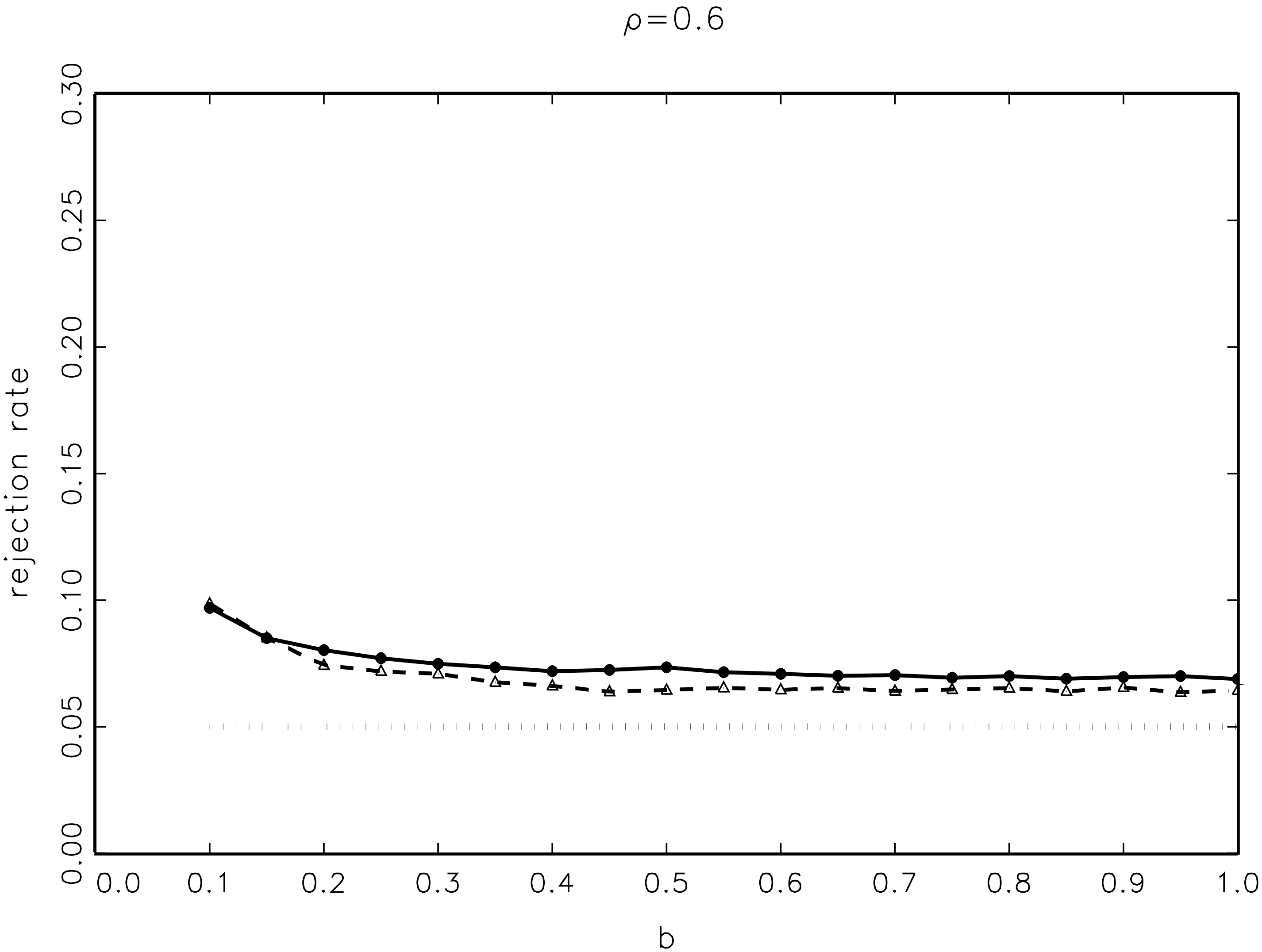


Figure 2.75: (cont'd)

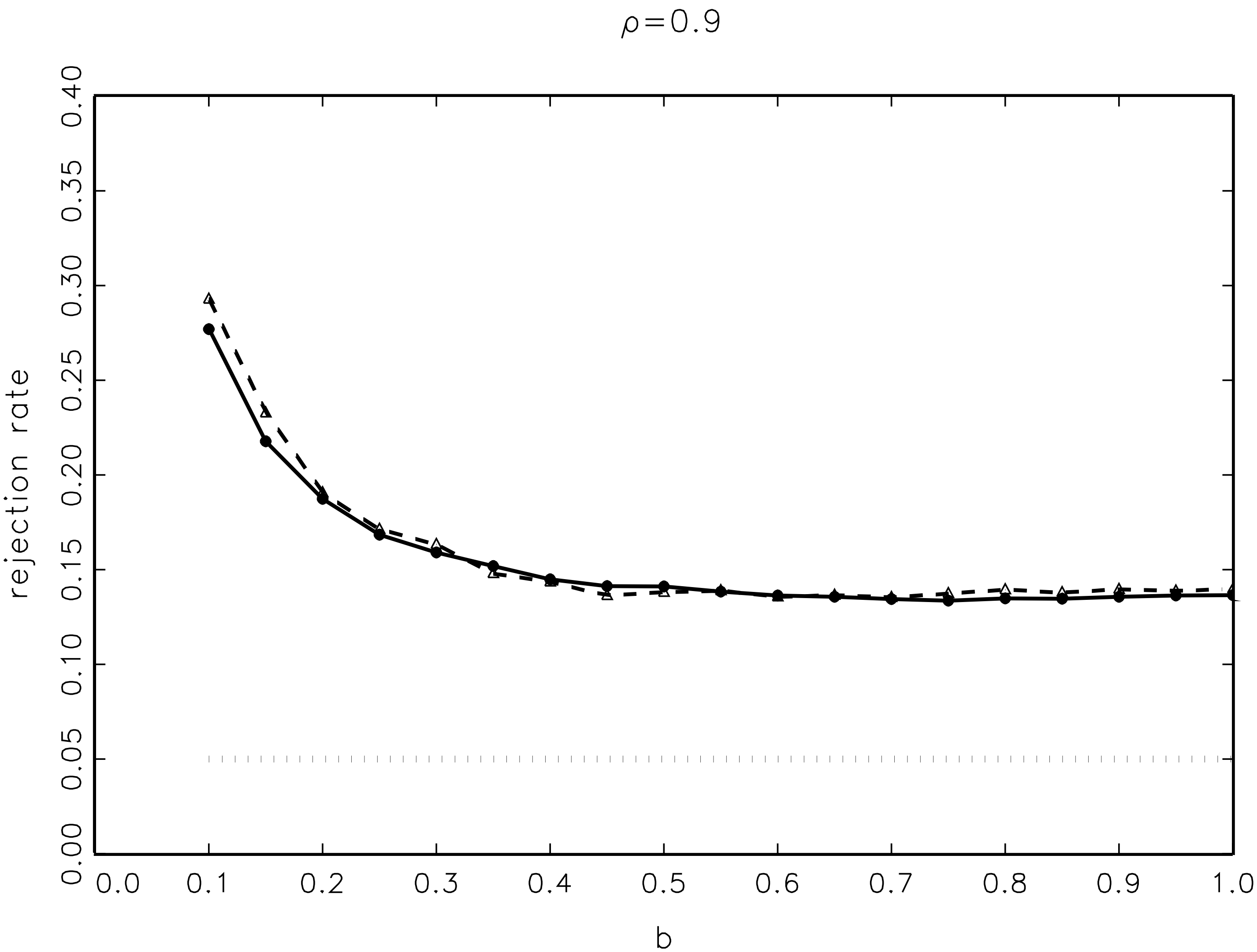


Figure 2.76: AM and ES - Initially Scarce Data, Bartlett, $N_Q = 24$ $N_M = 48$

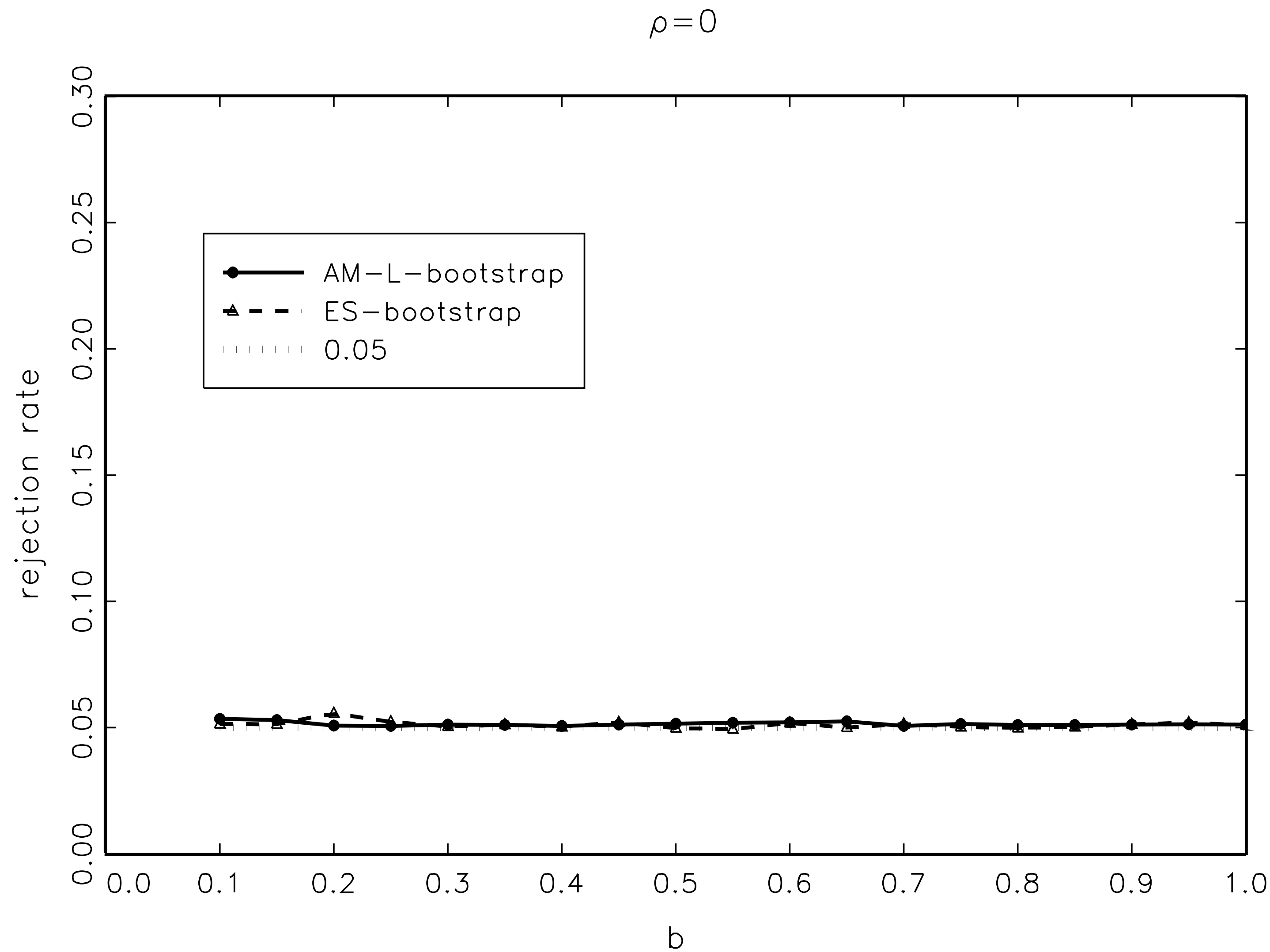


Figure 2.76: (cont'd)

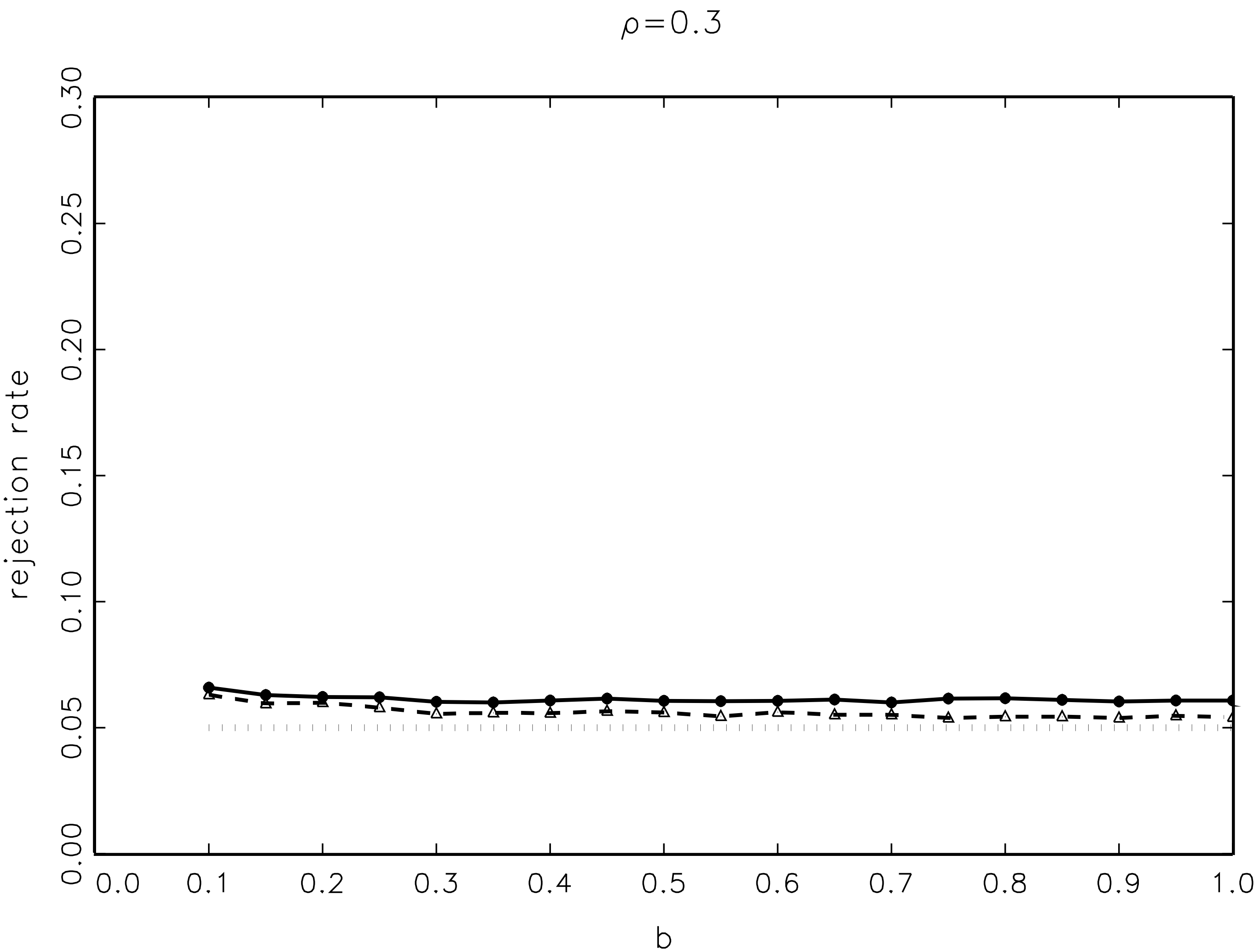


Figure 2.76: (cont'd)

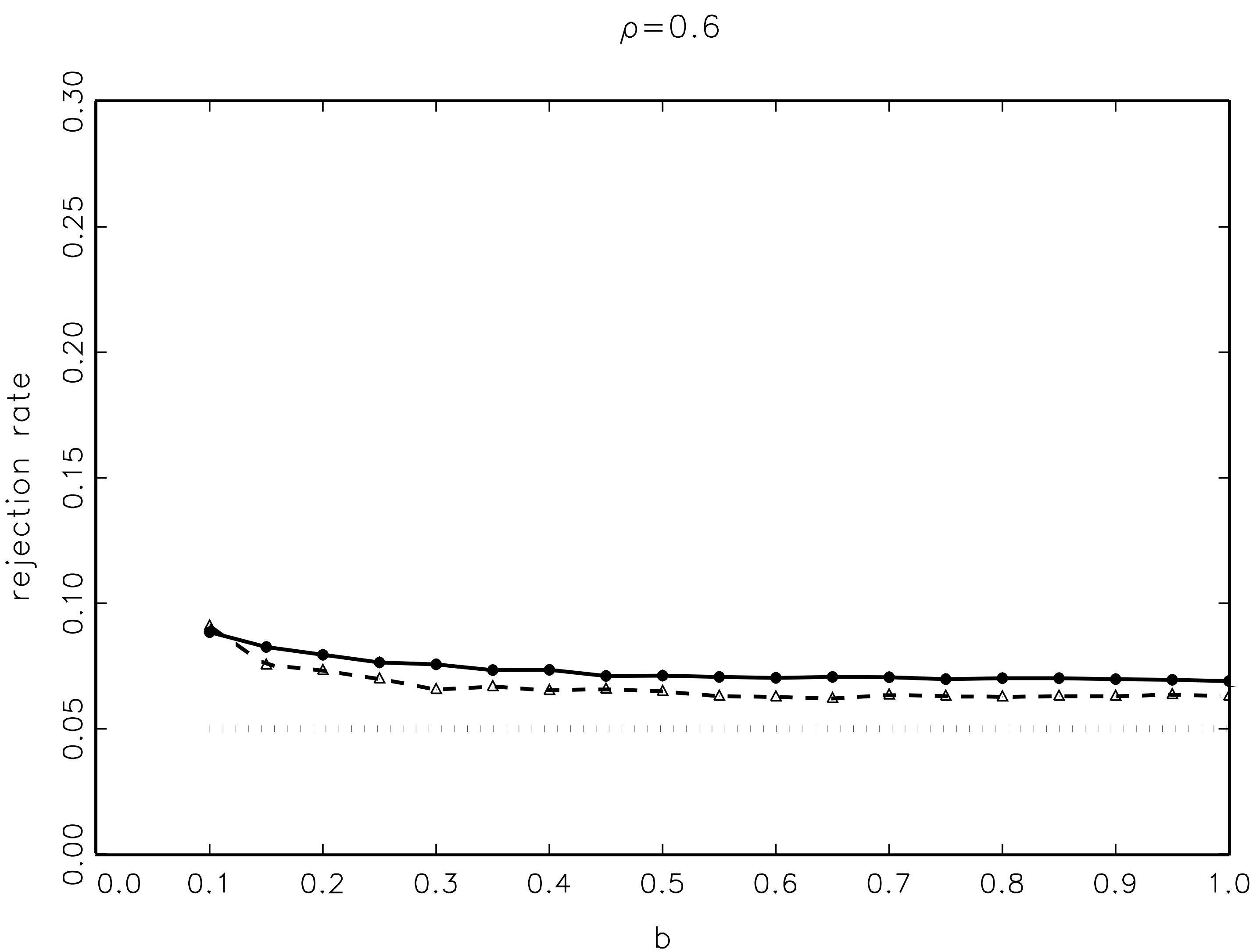


Figure 2.76: (cont'd)

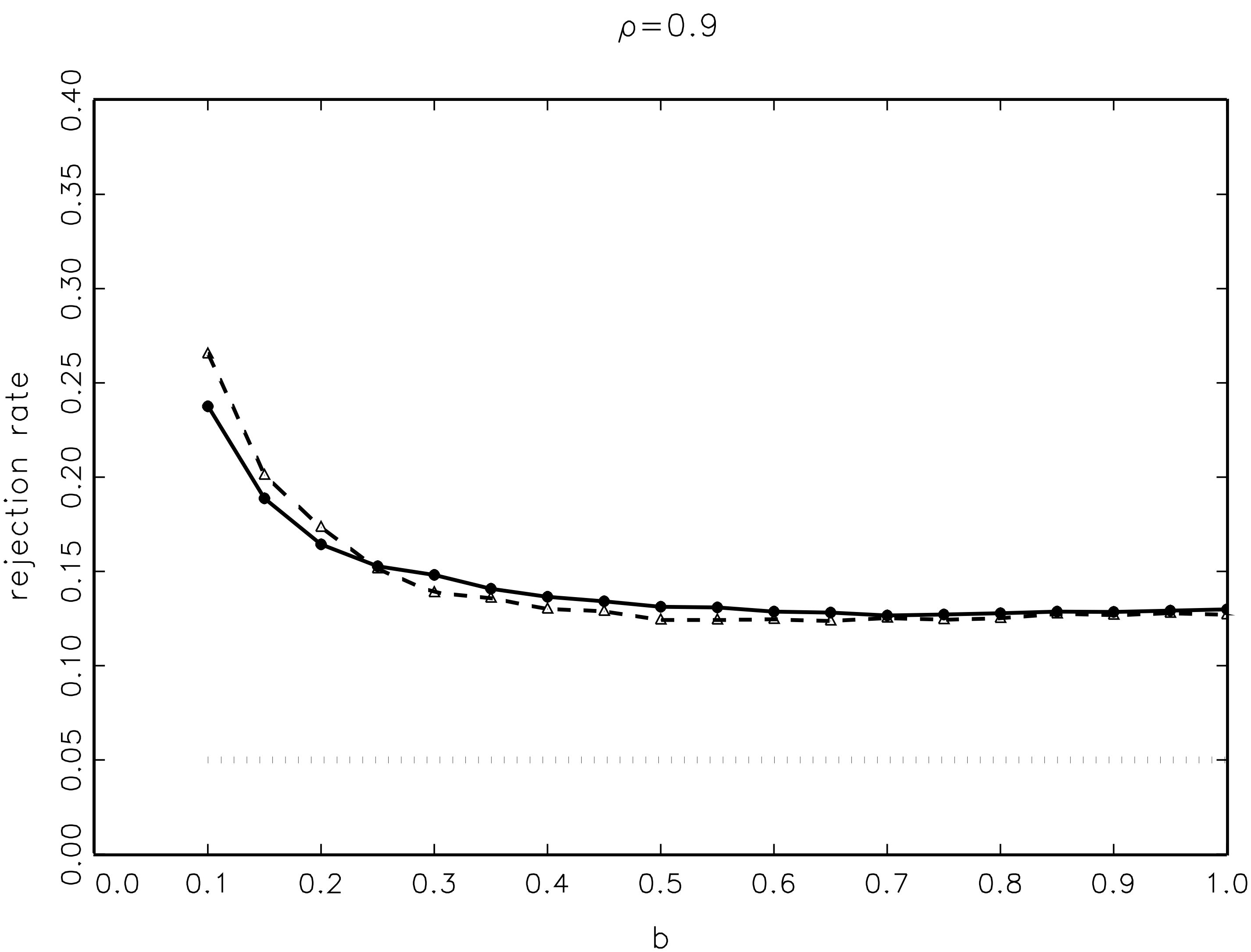


Figure 2.77: AM and ES - Conditional Bernoulli ($p = 0.3$), Bartlett, $T = 50$

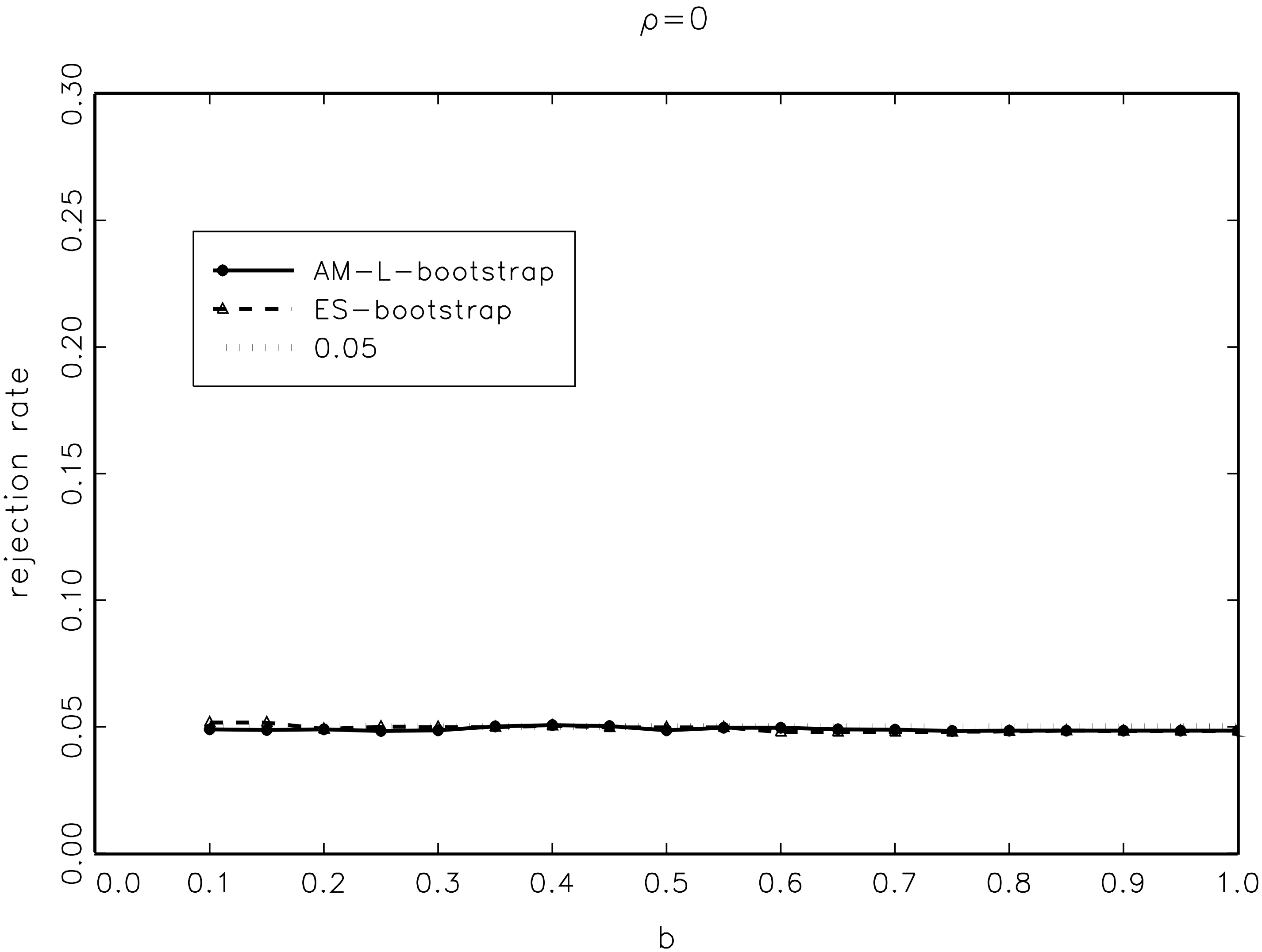


Figure 2.77: (cont'd)

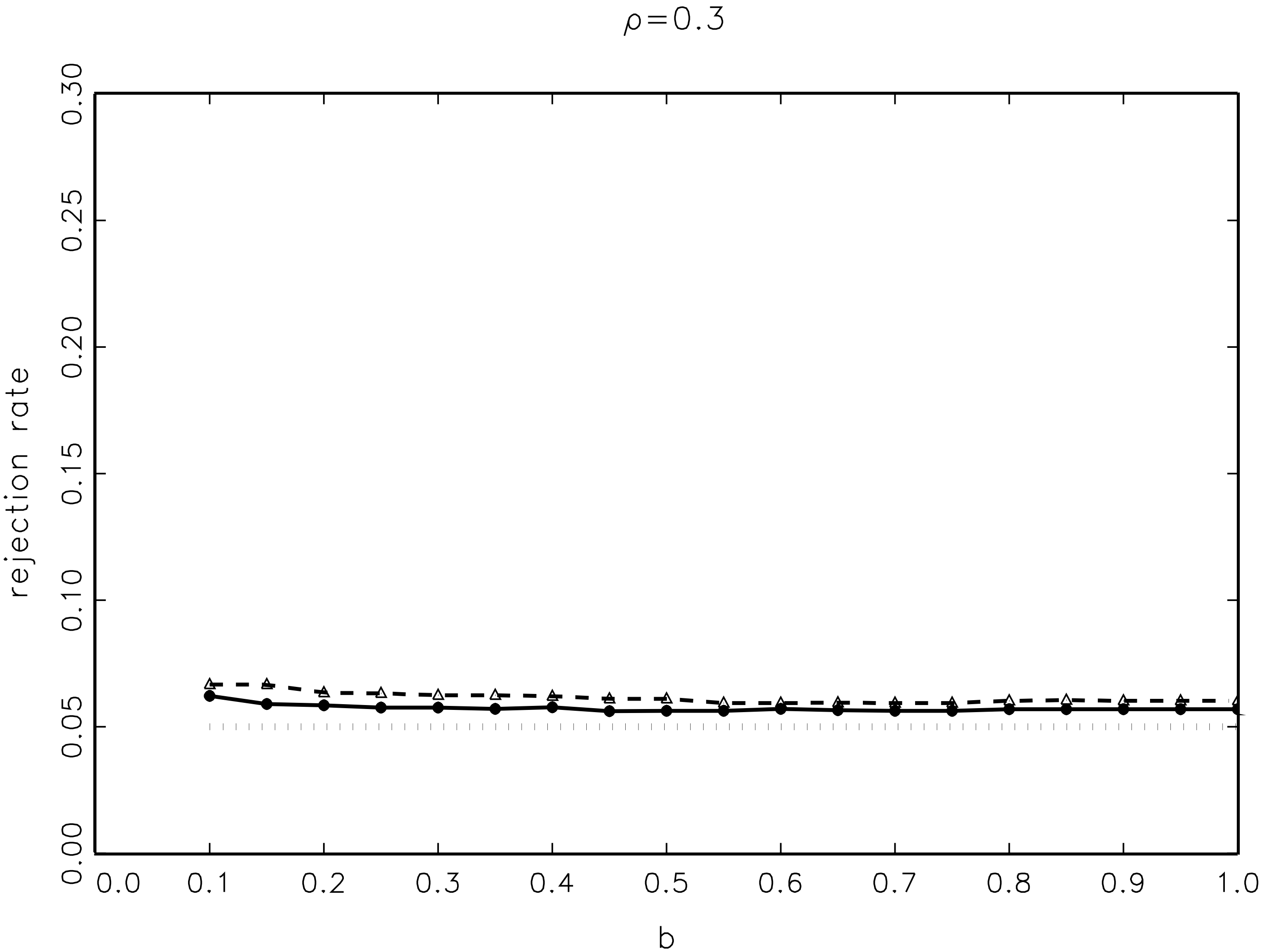


Figure 2.77: (cont'd)

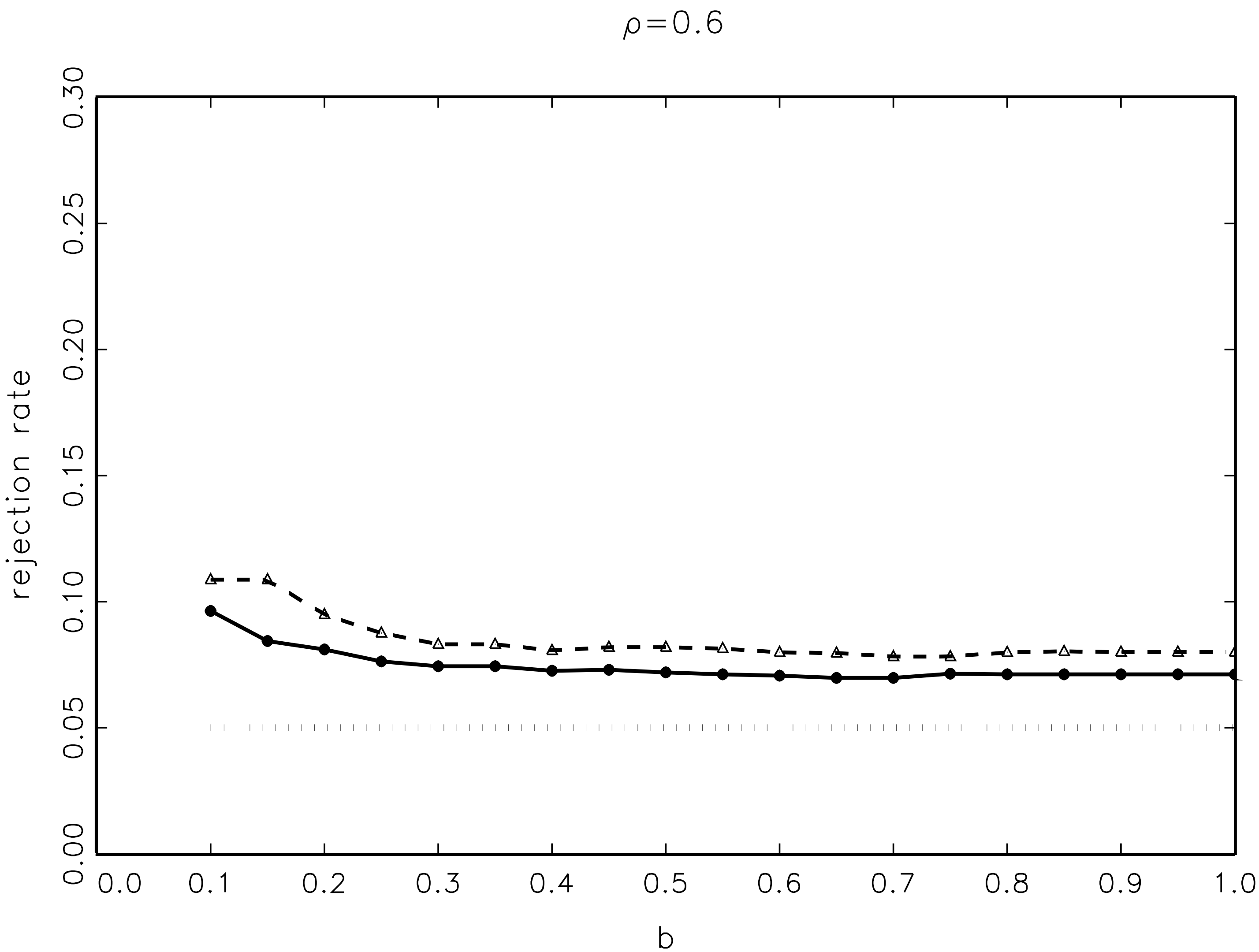


Figure 2.77: (cont'd)

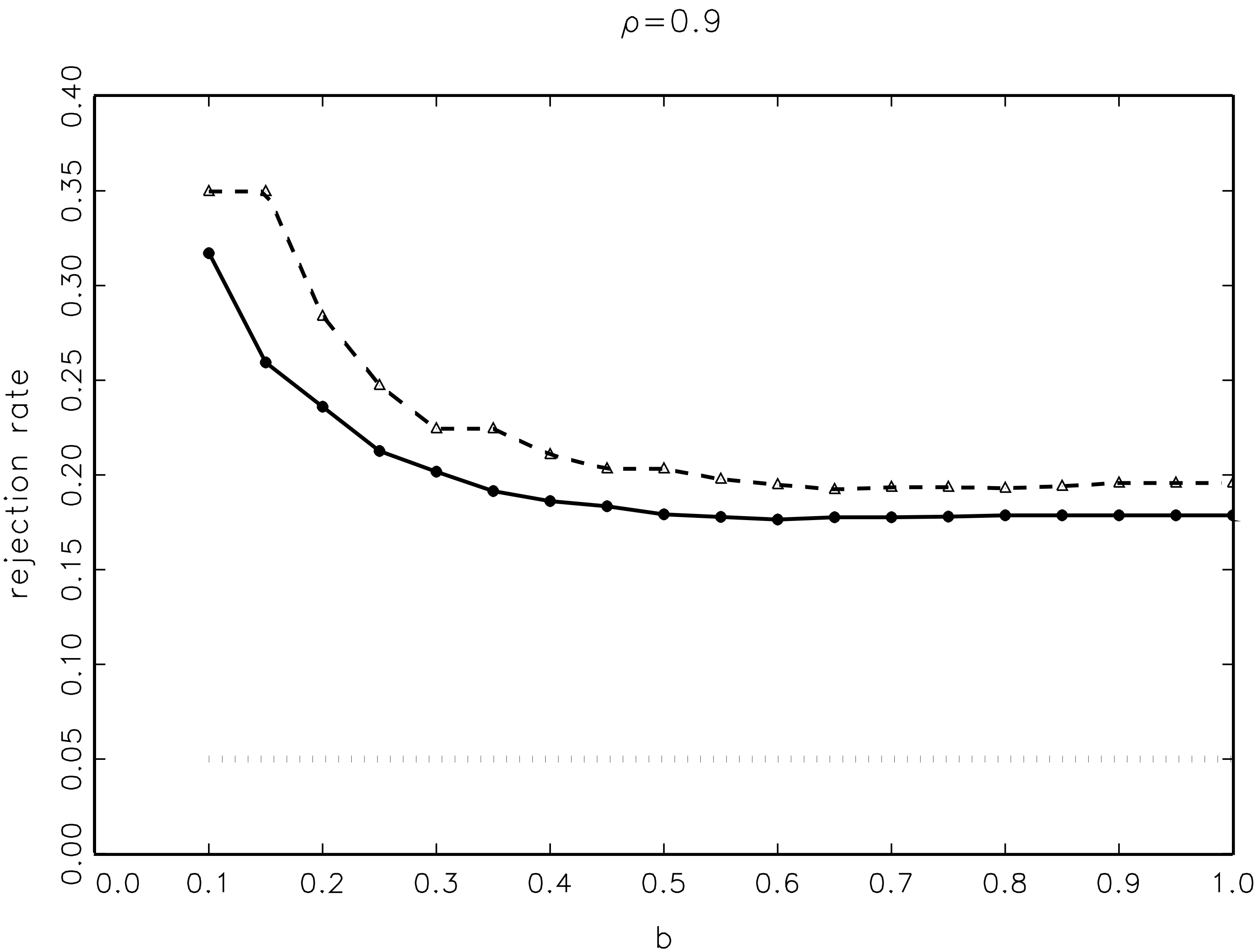


Figure 2.78: AM and ES - Conditional Bernoulli ($p = 0.5$), Bartlett, $T = 50$

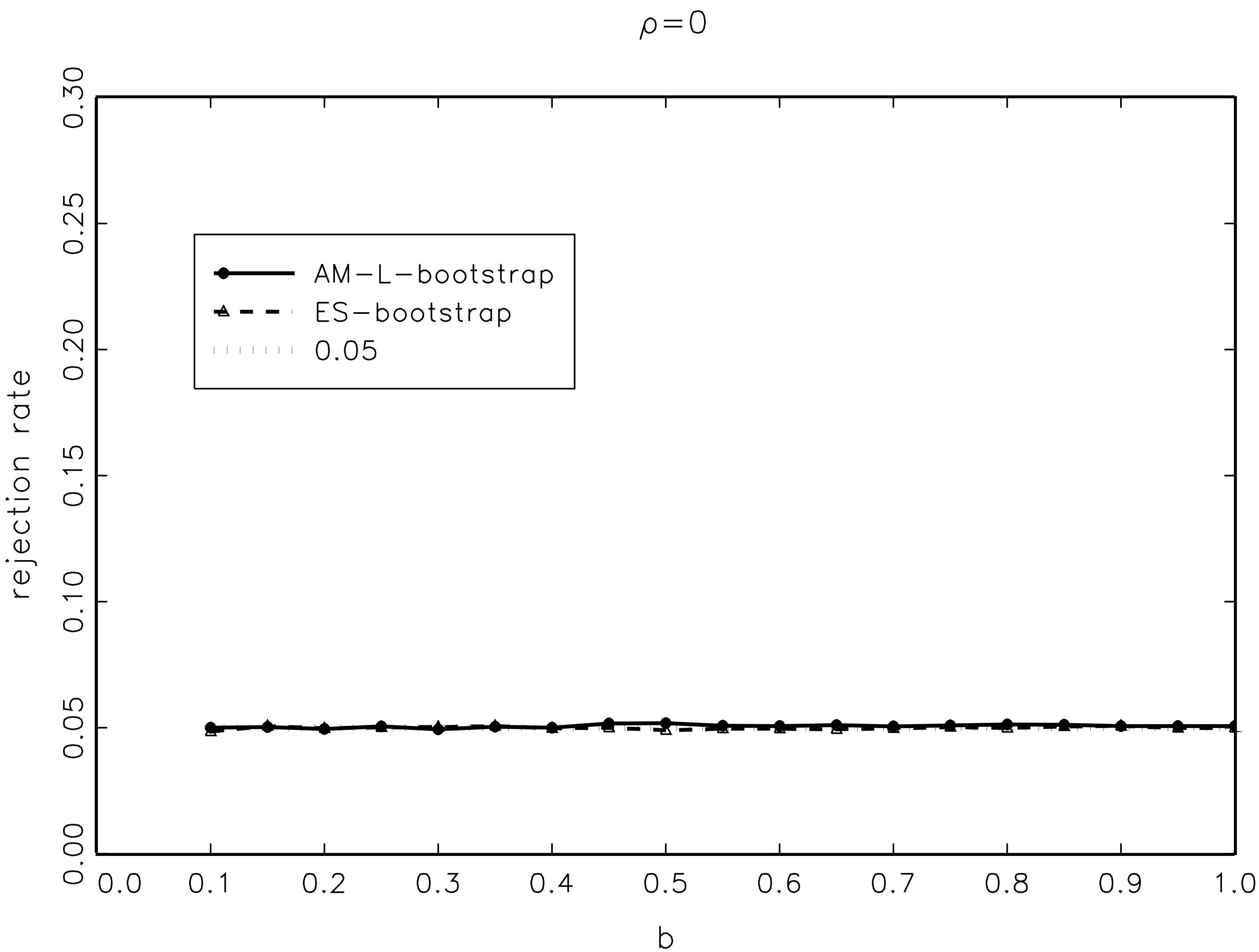


Figure 2.78: (cont'd)

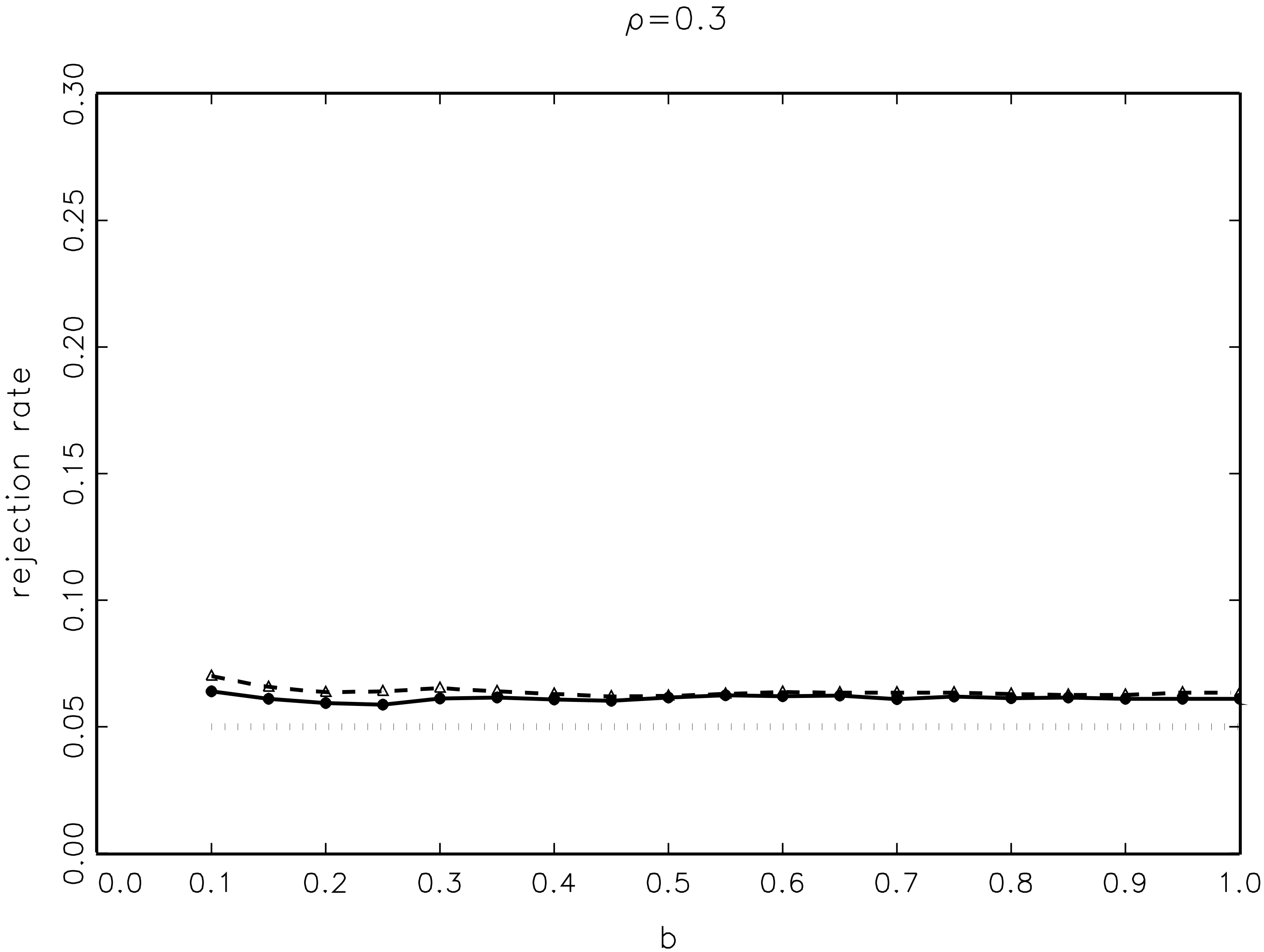


Figure 2.78: (cont'd)

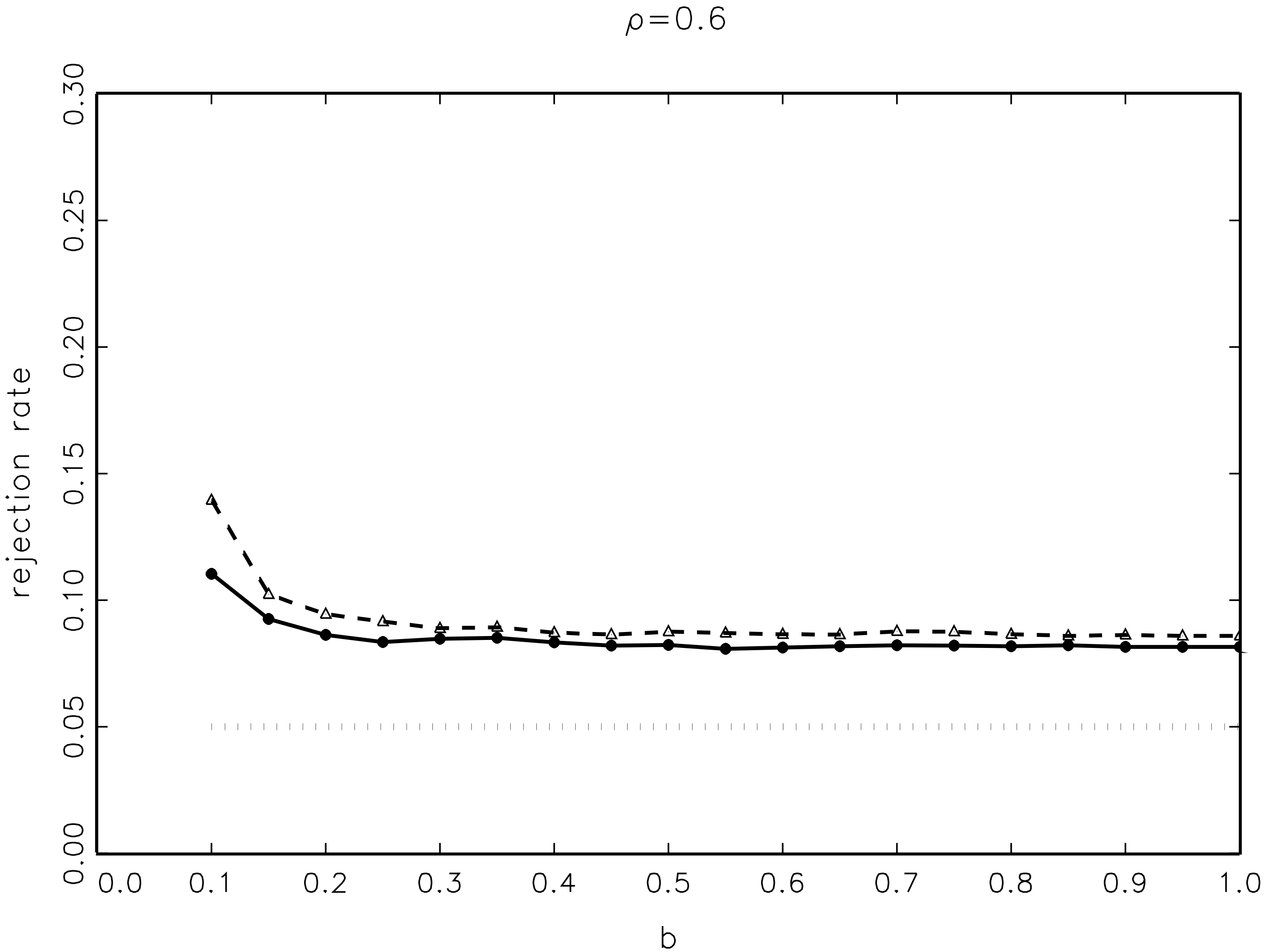


Figure 2.78: (cont'd)

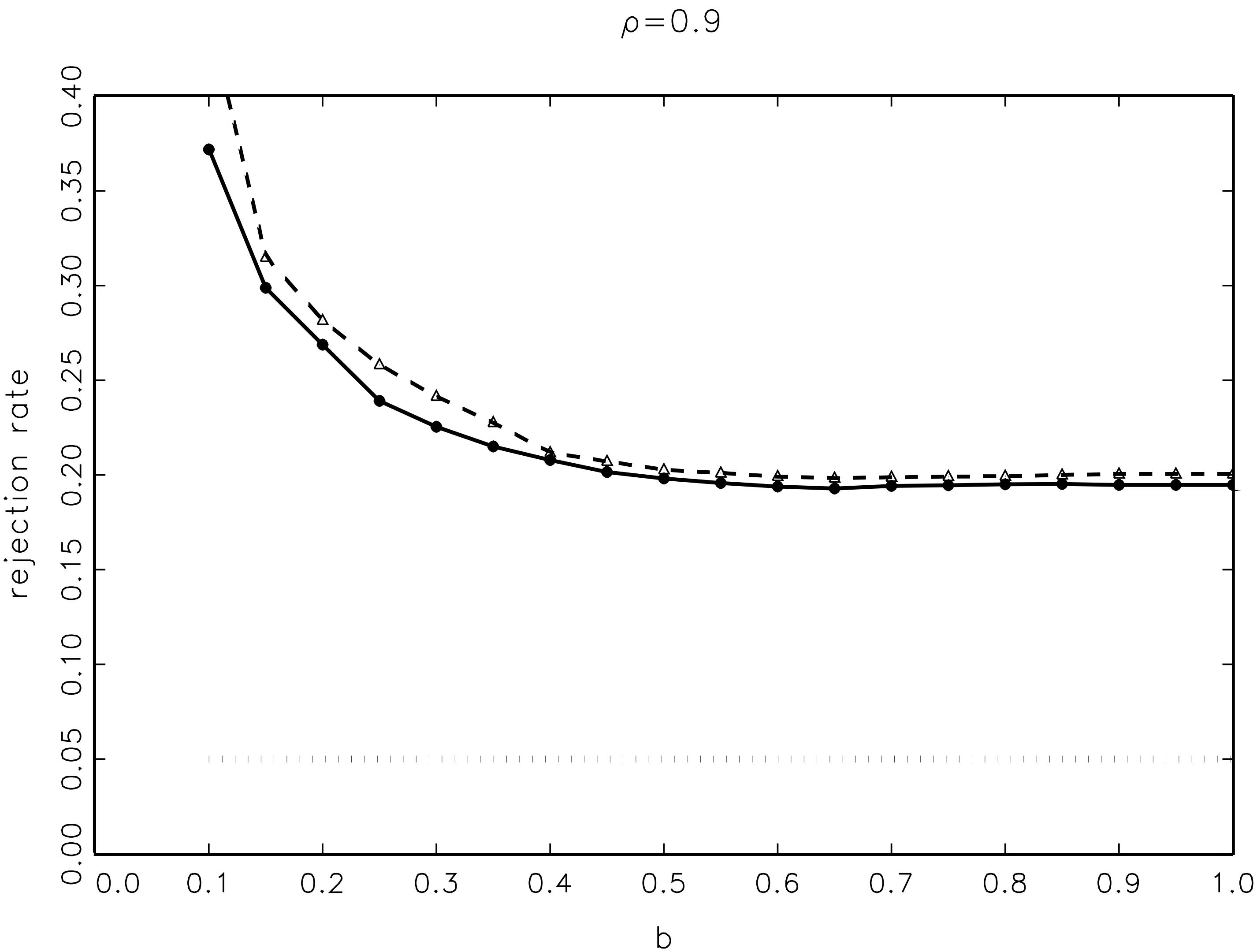


Figure 2.79: AM and ES - Conditional Bernoulli ($p = 0.7$), Bartlett, $T = 50$

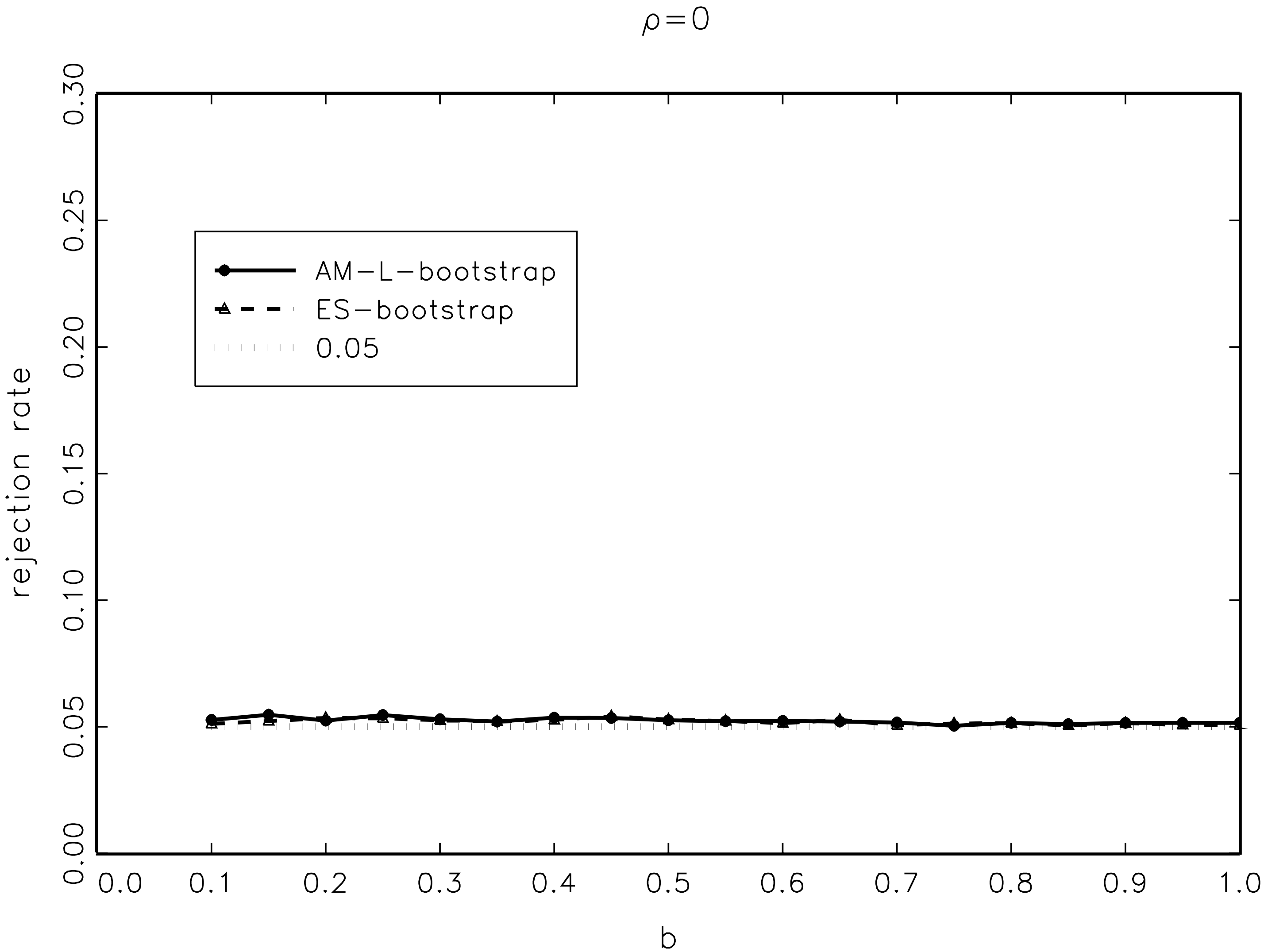


Figure 2.79: (cont'd)

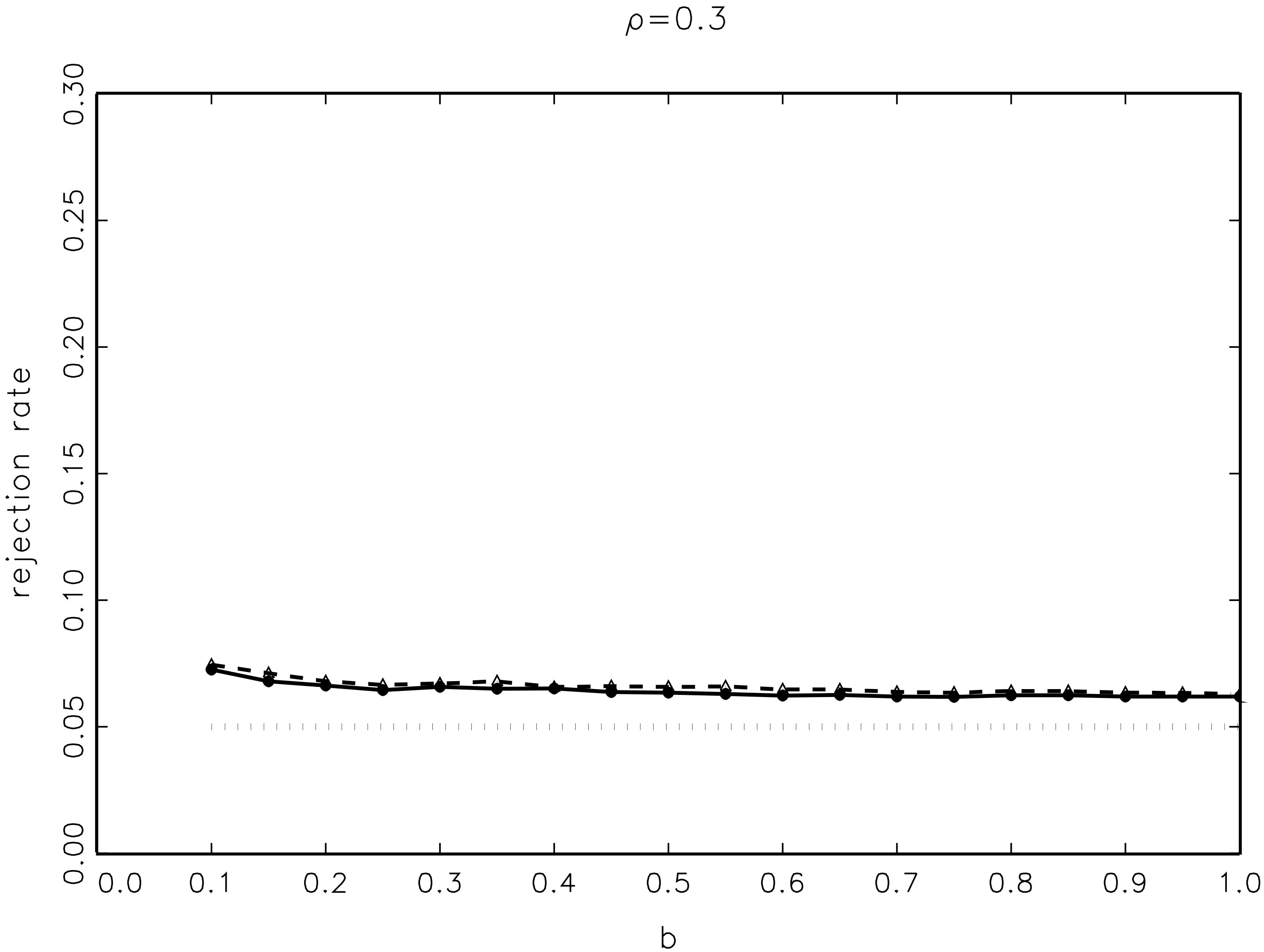


Figure 2.79: (cont'd)

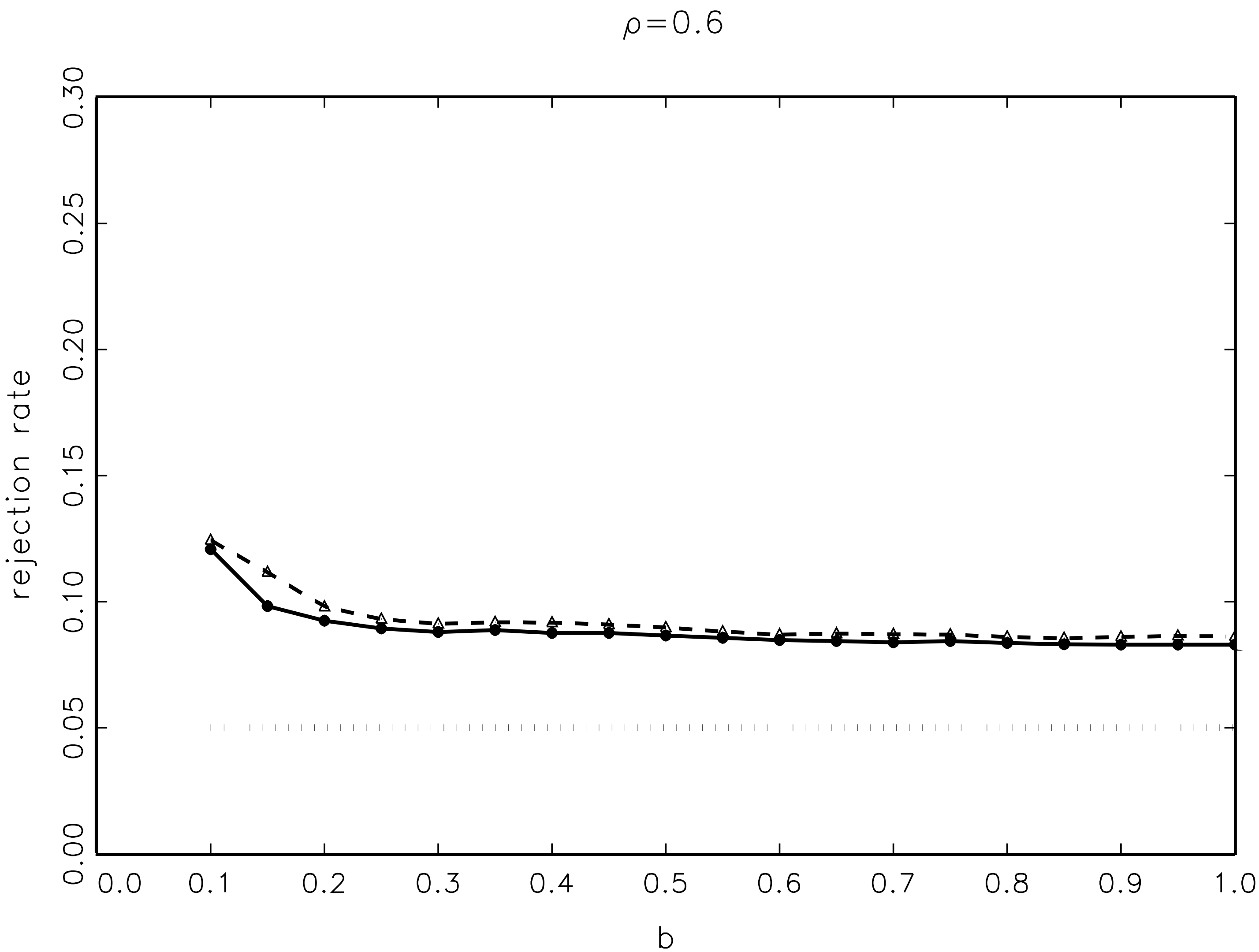


Figure 2.79: (cont'd)

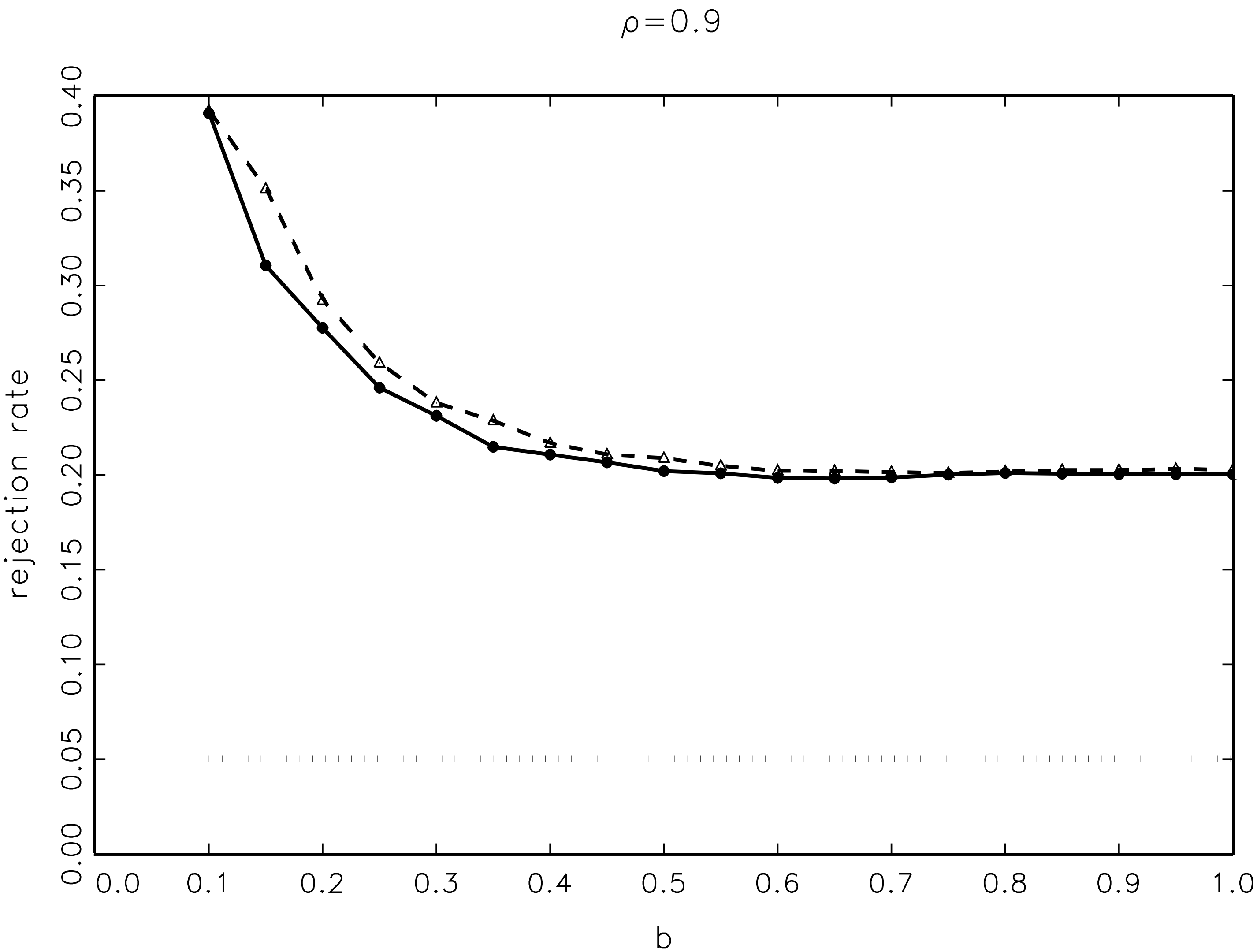


Figure 2.80: AM and ES - Conditional Bernoulli ($p = 0.3$), Bartlett, $T = 100$

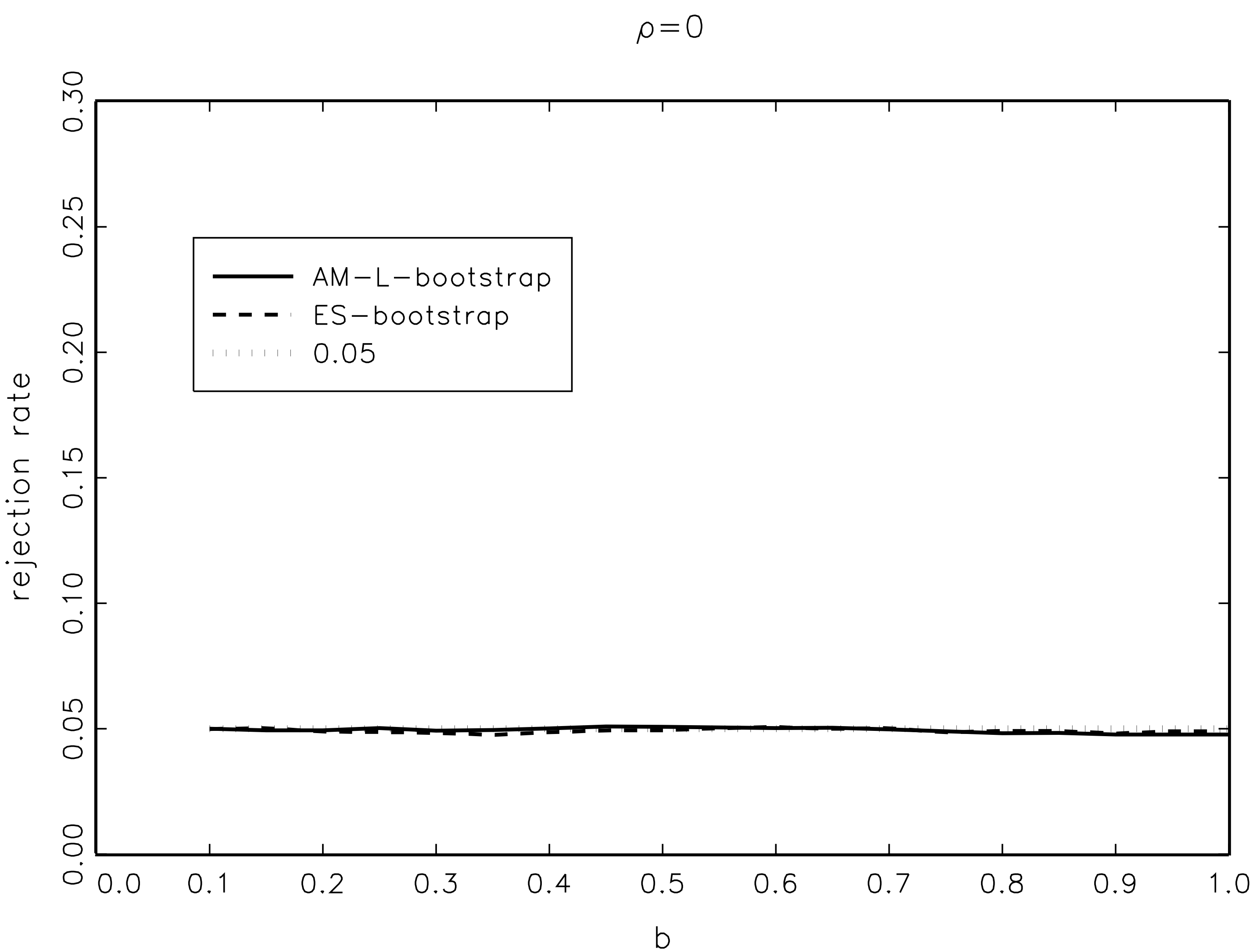


Figure 2.80: (cont'd)

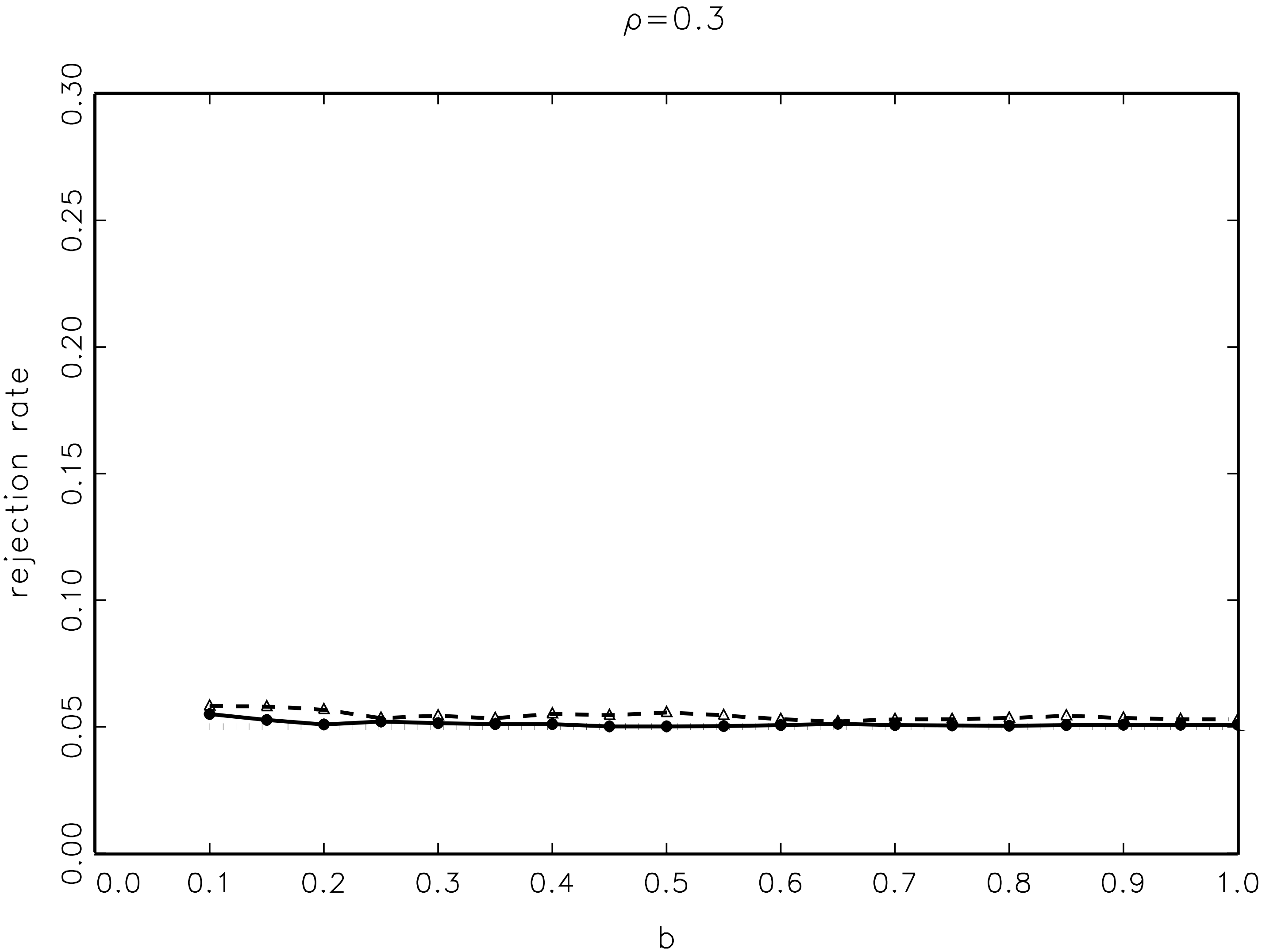


Figure 2.80: (cont'd)

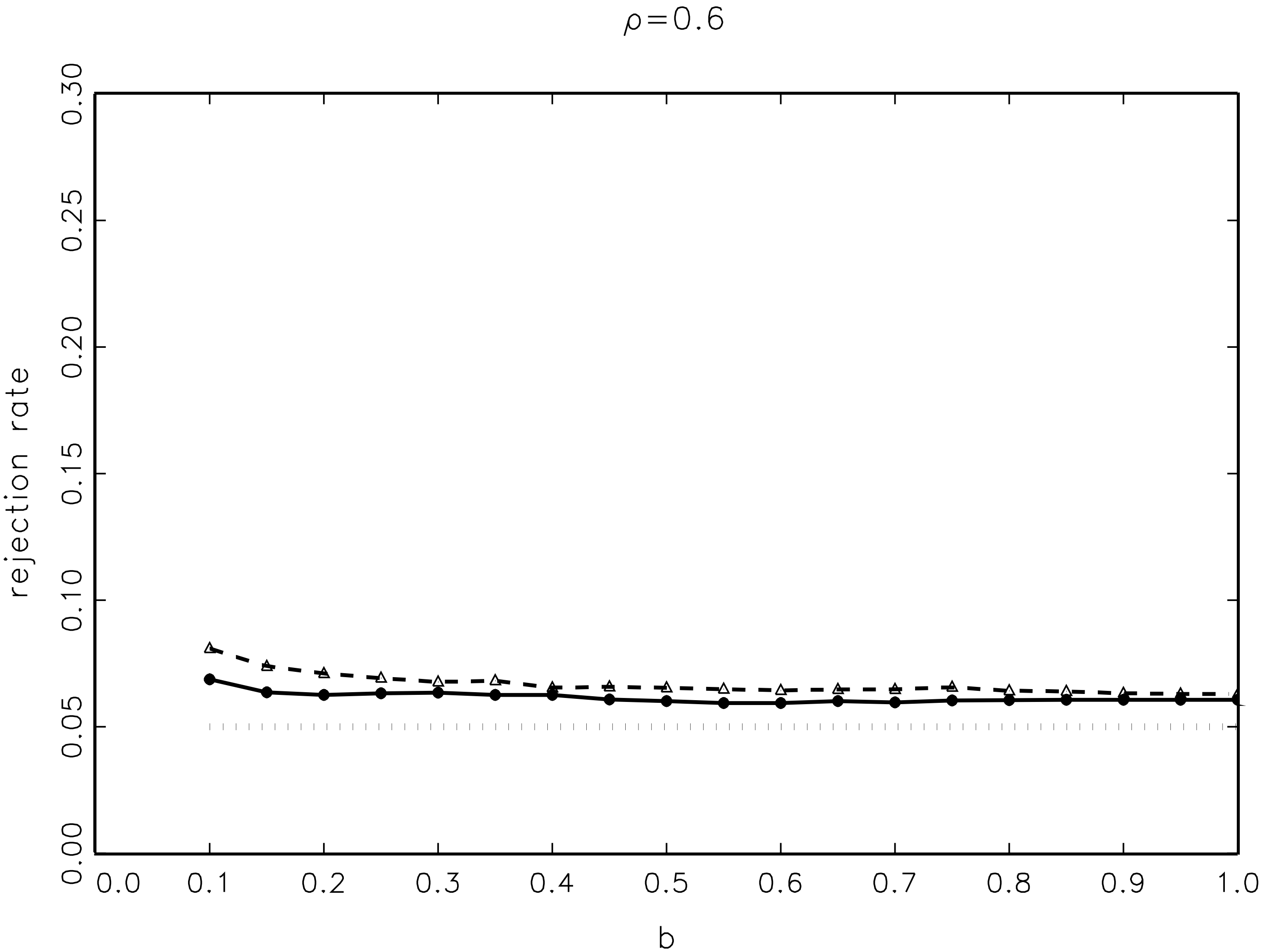


Figure 2.80: (cont'd)

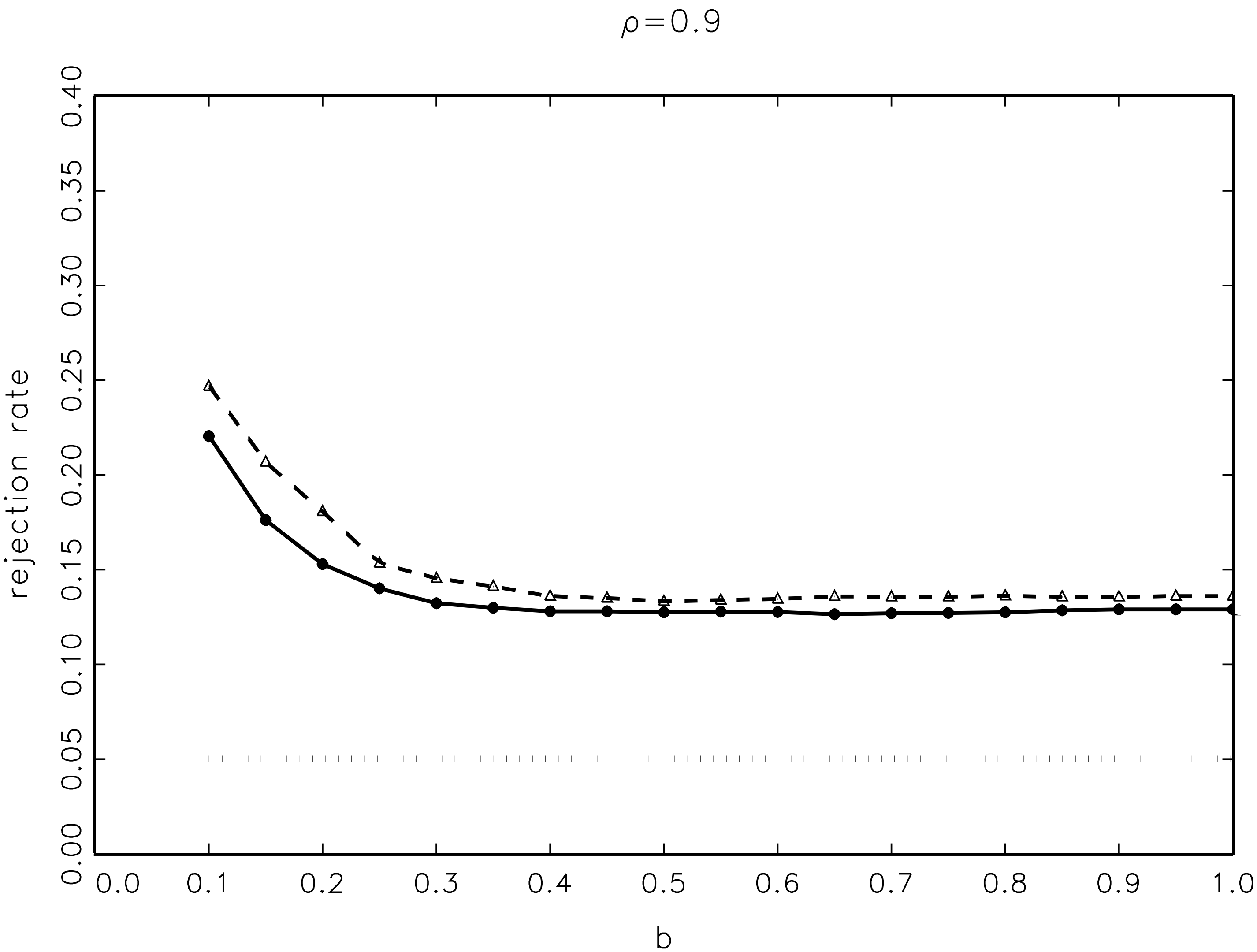


Figure 2.81: AM and ES - Conditional Bernoulli ($p = 0.5$), Bartlett, $T = 100$

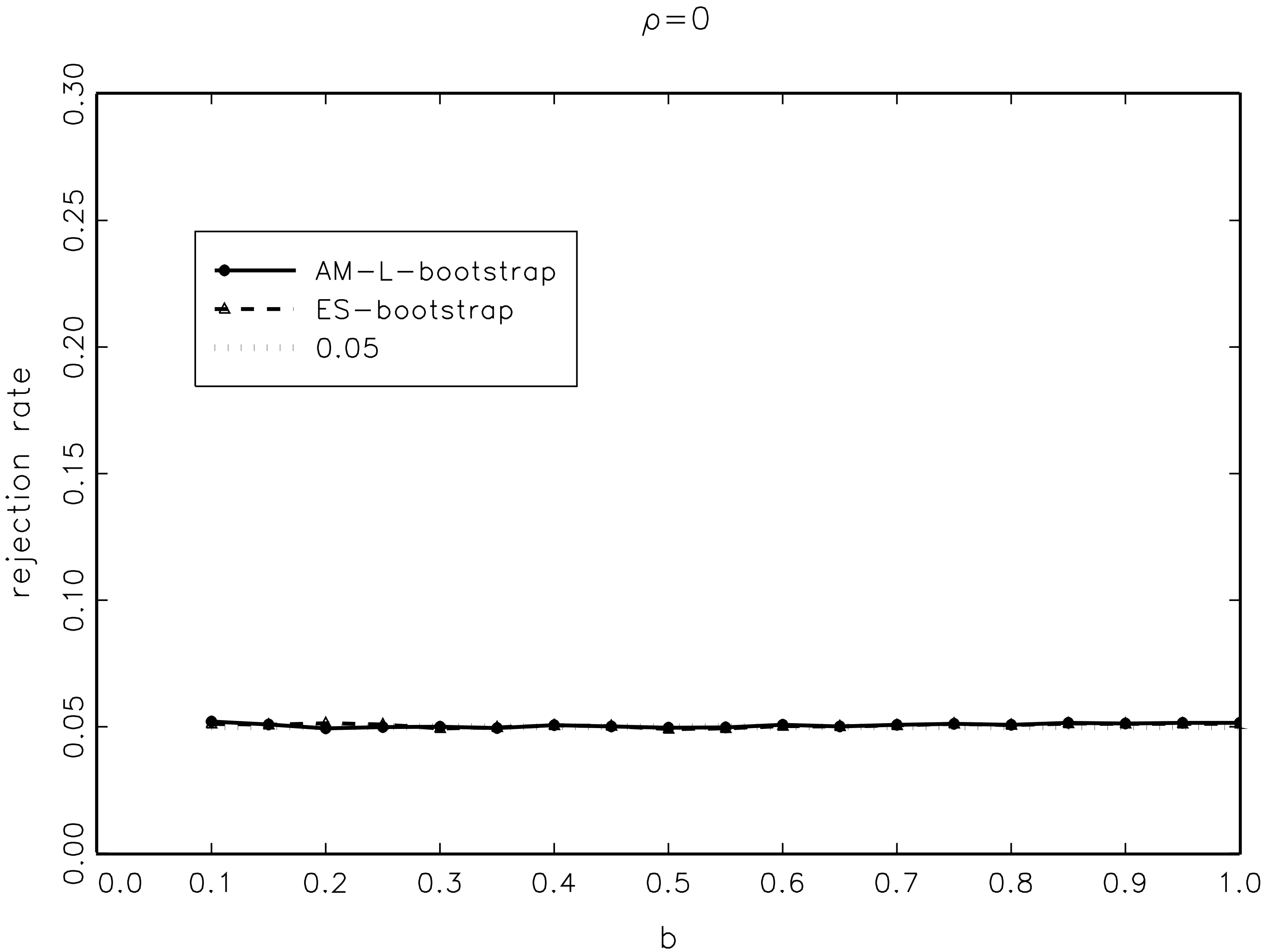


Figure 2.81: (cont'd)

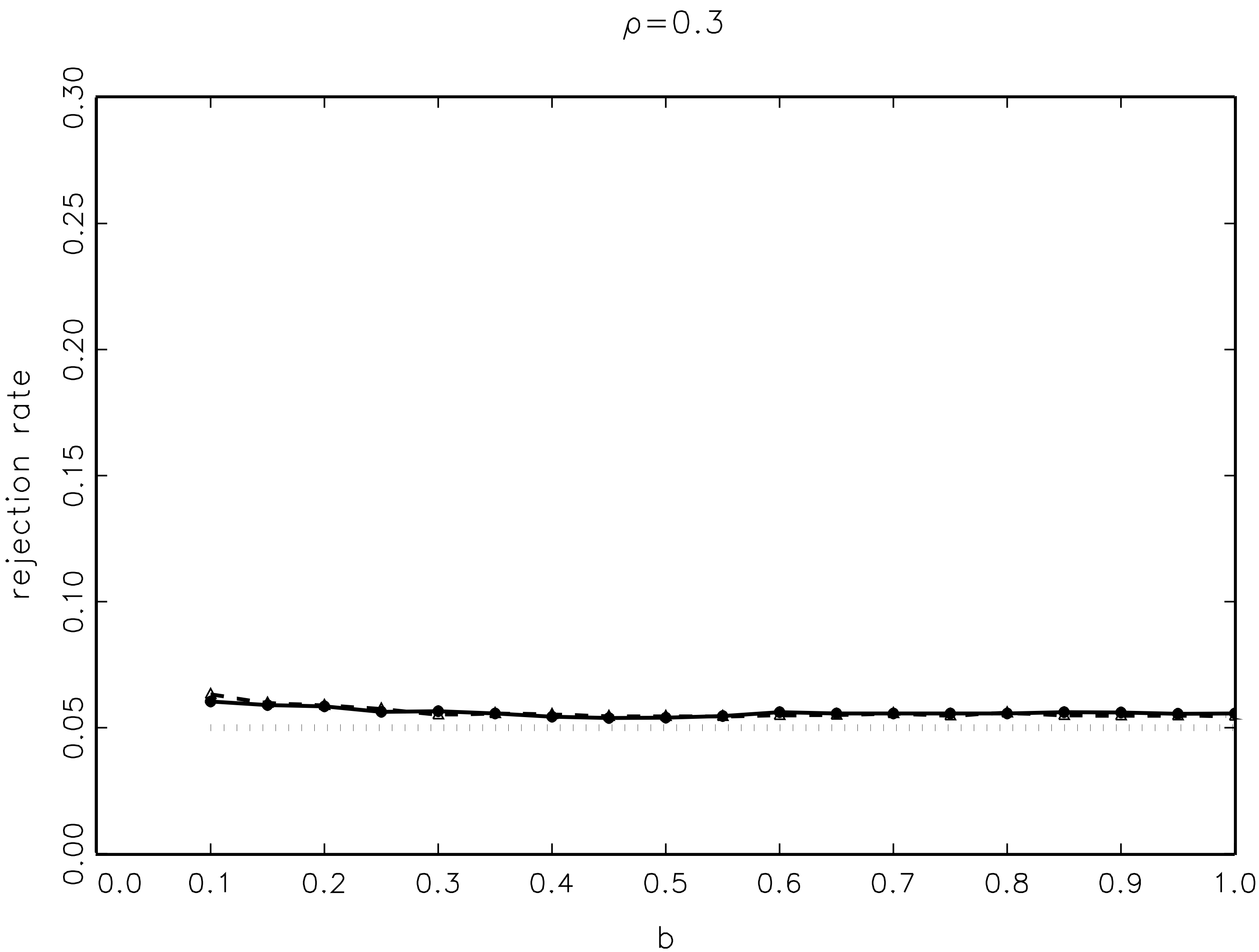


Figure 2.81: (cont'd)

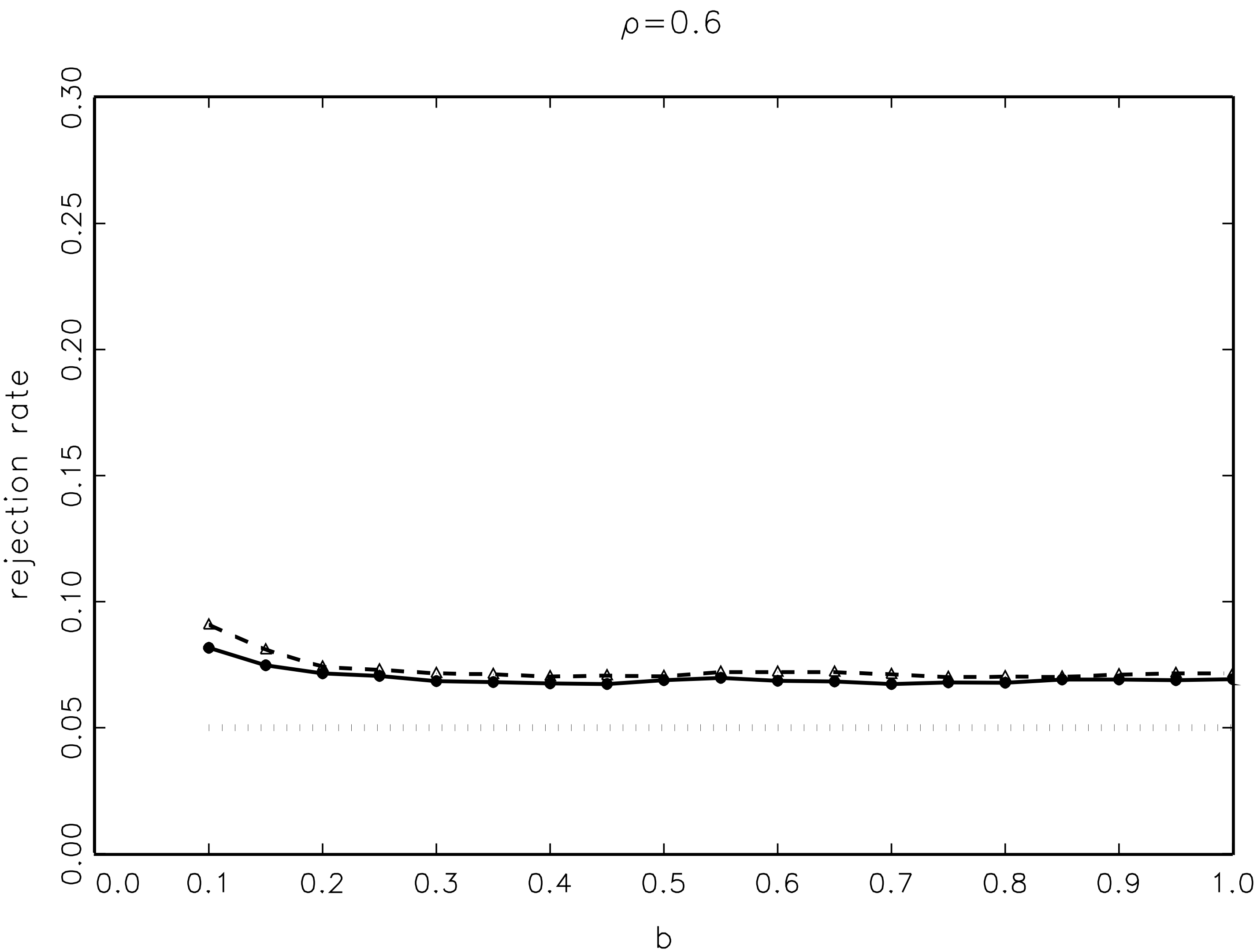


Figure 2.81: (cont'd)

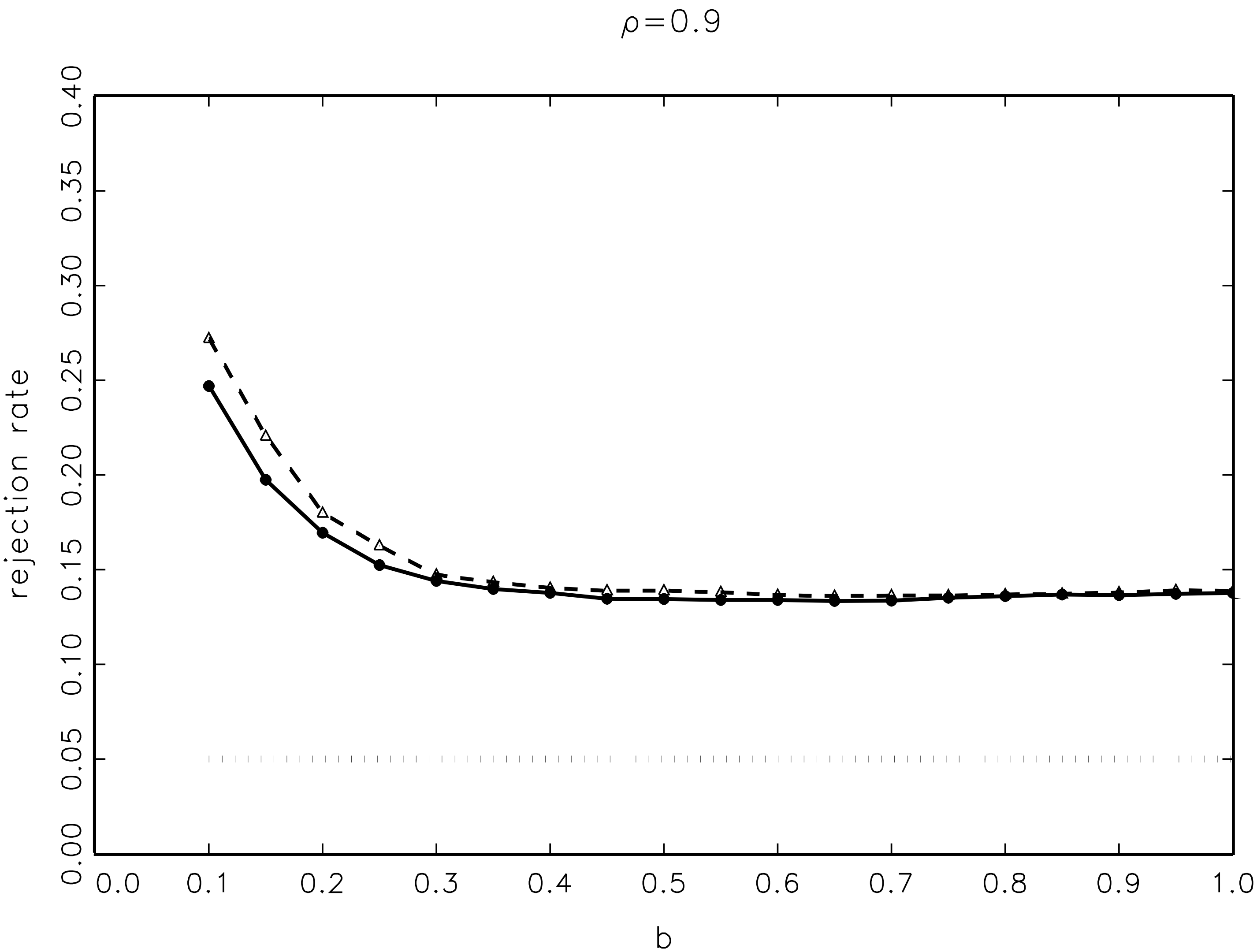


Figure 2.82: AM and ES - Conditional Bernoulli ($p = 0.7$), Bartlett, $T = 100$

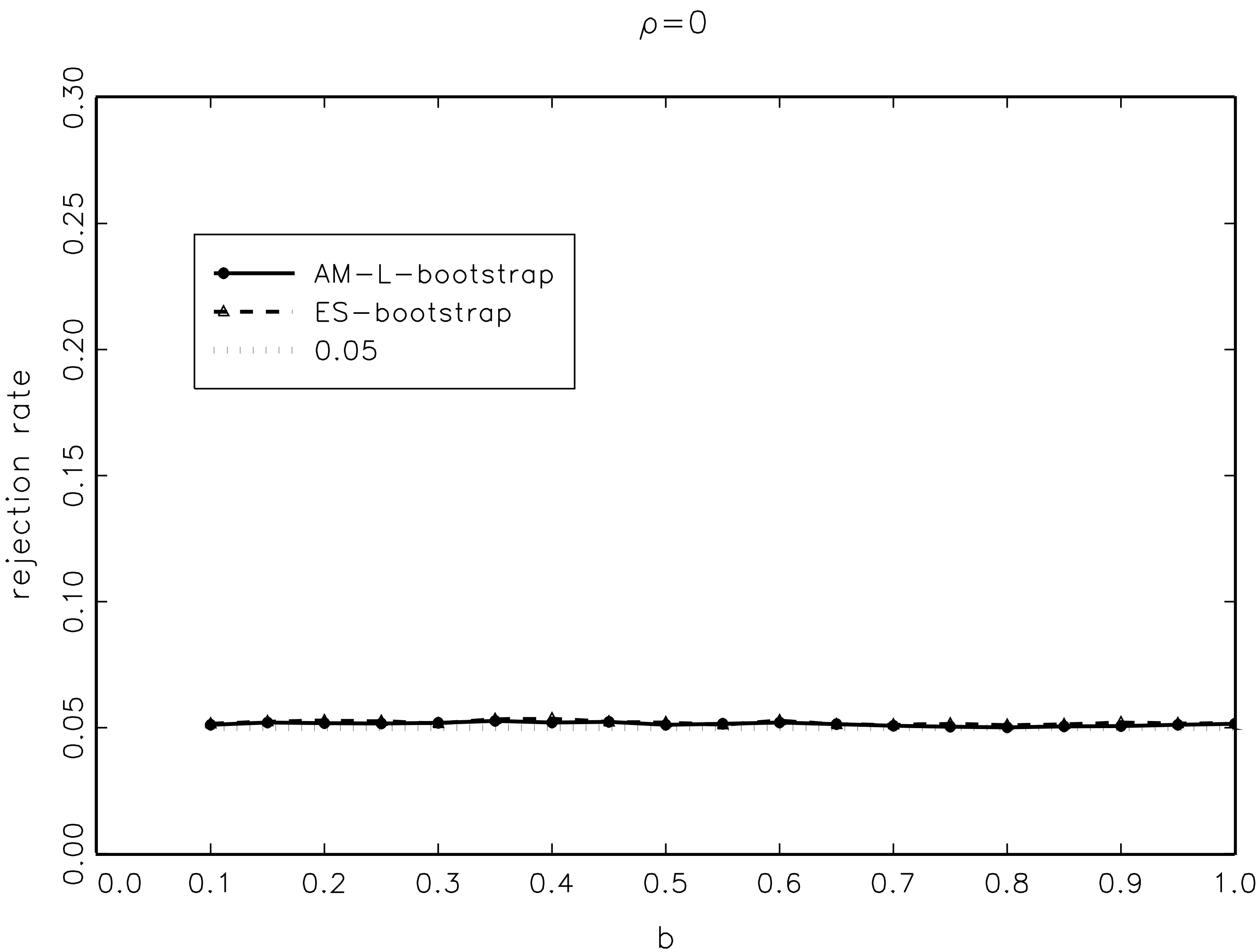


Figure 2.82: (cont'd)

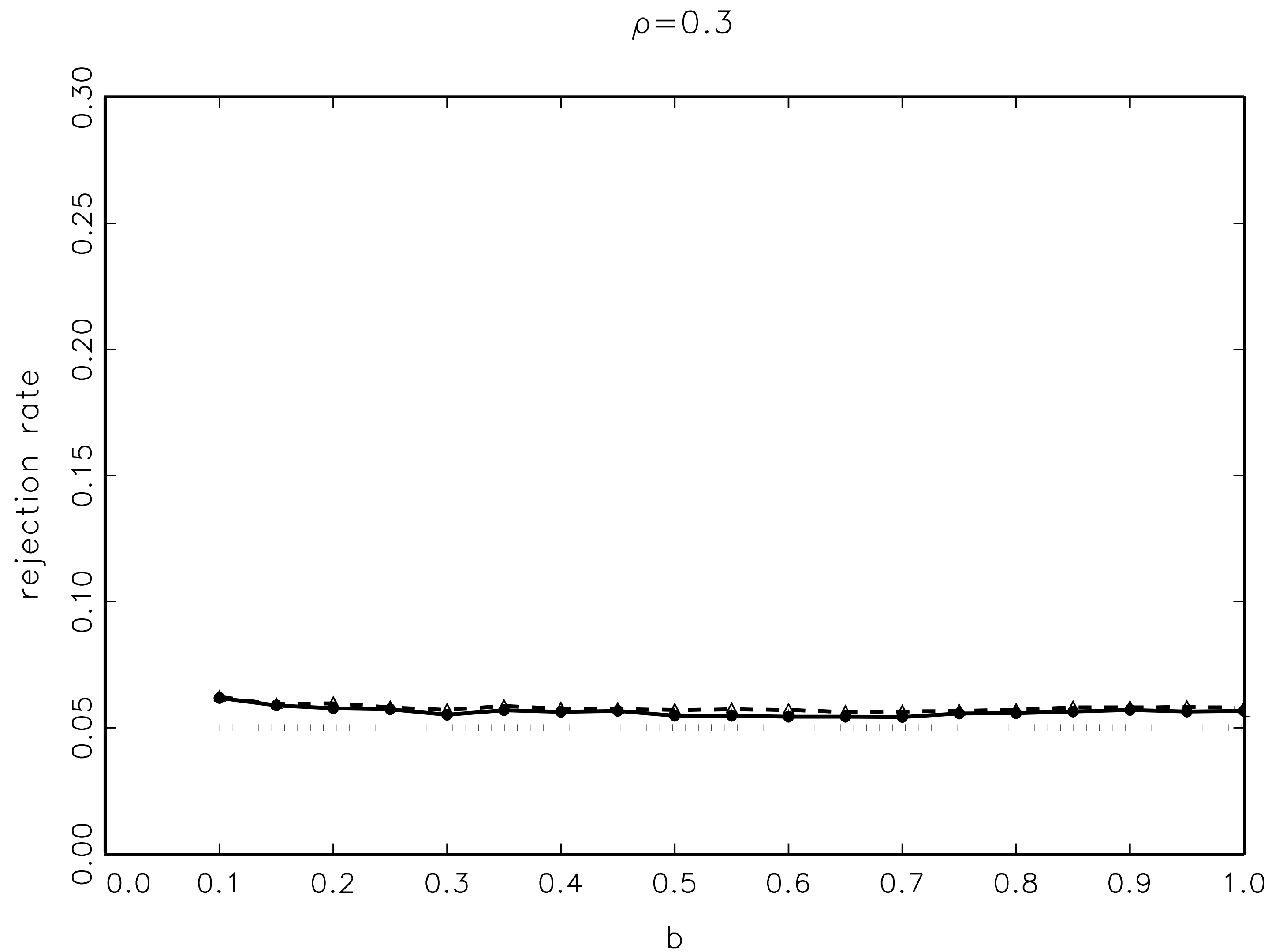


Figure 2.82: (cont'd)

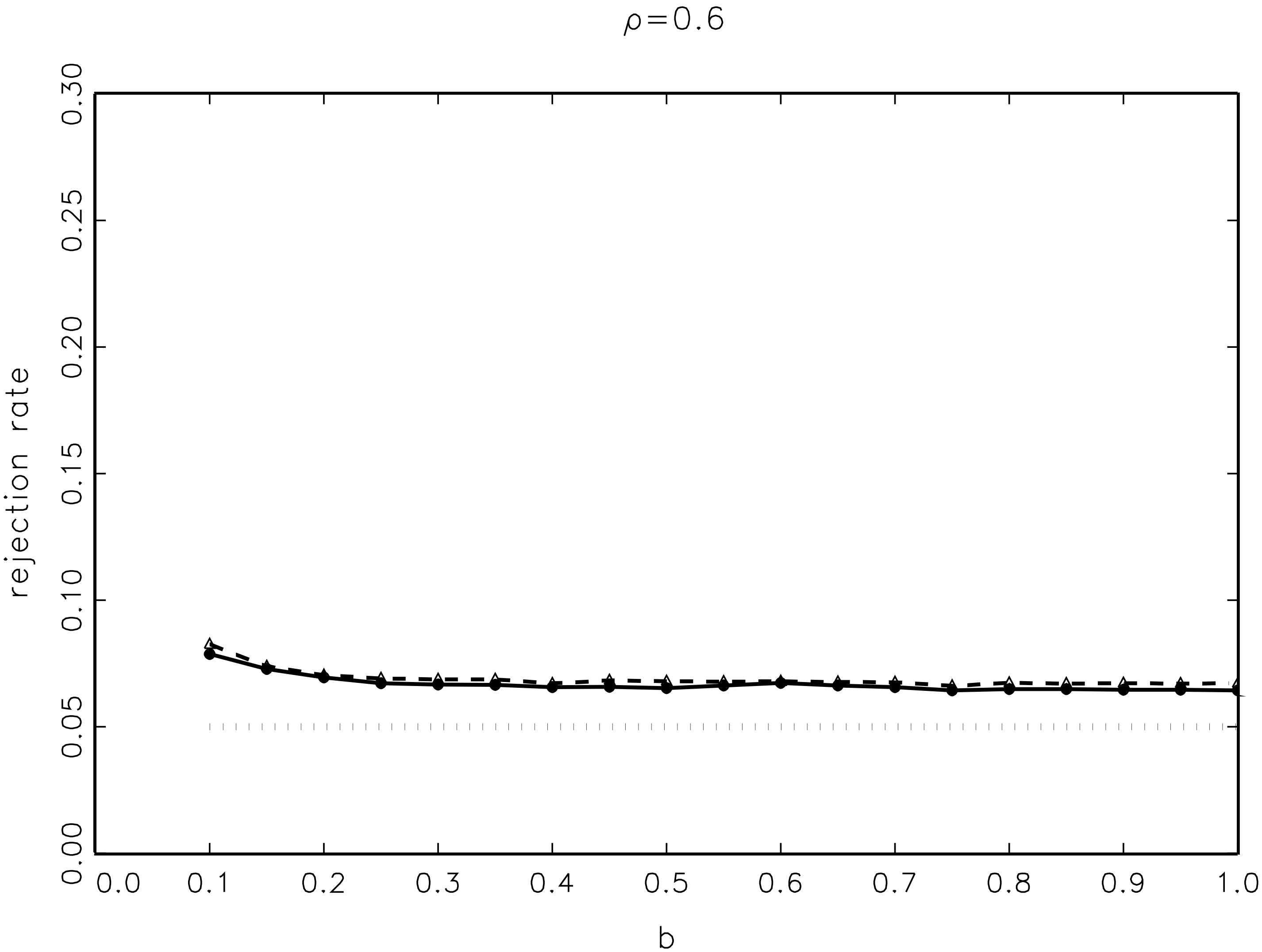


Figure 2.82: (cont'd)

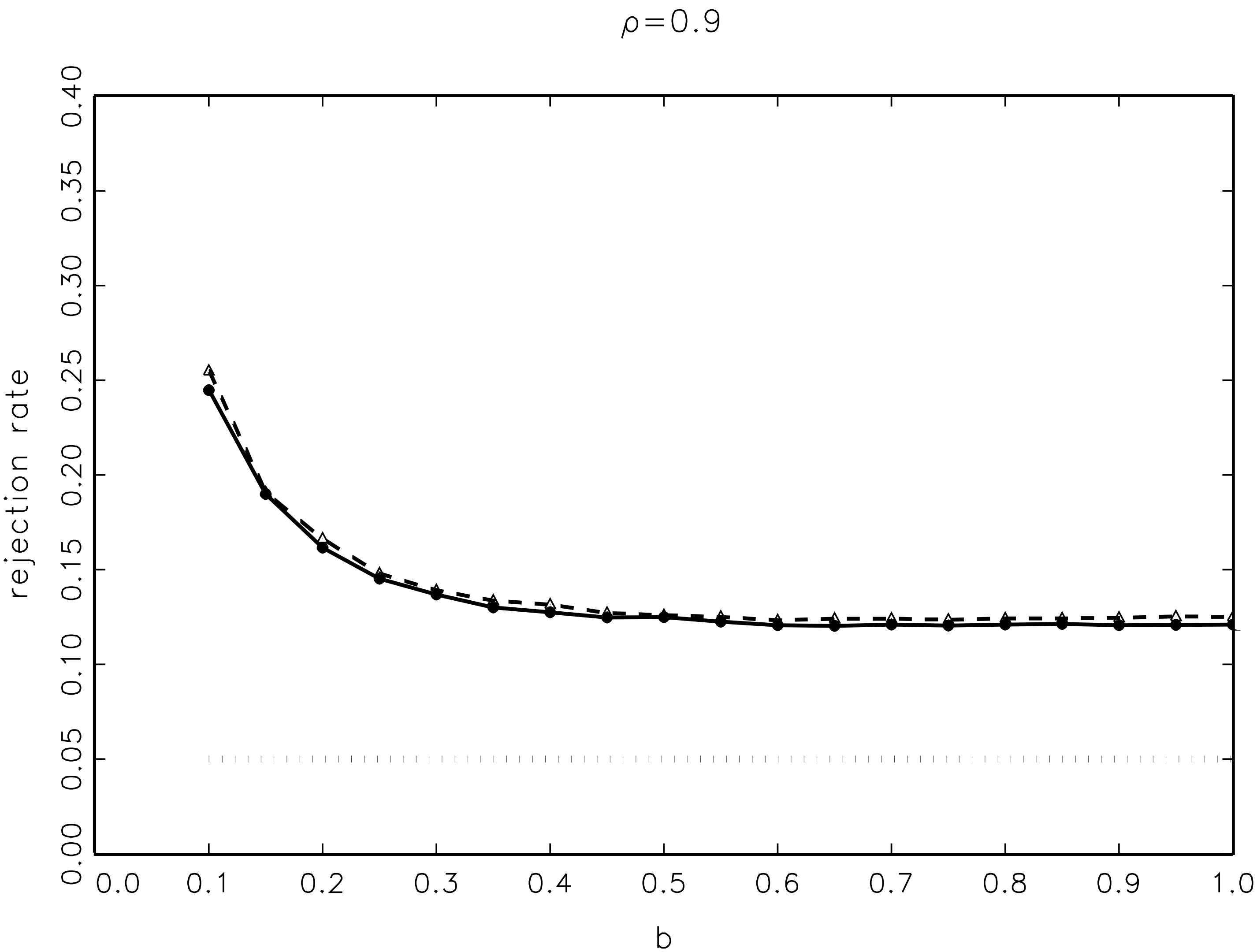


Figure 2.83: AM and ES - Conditional Bernoulli ($p = 0.3$), Bartlett, $T = 200$

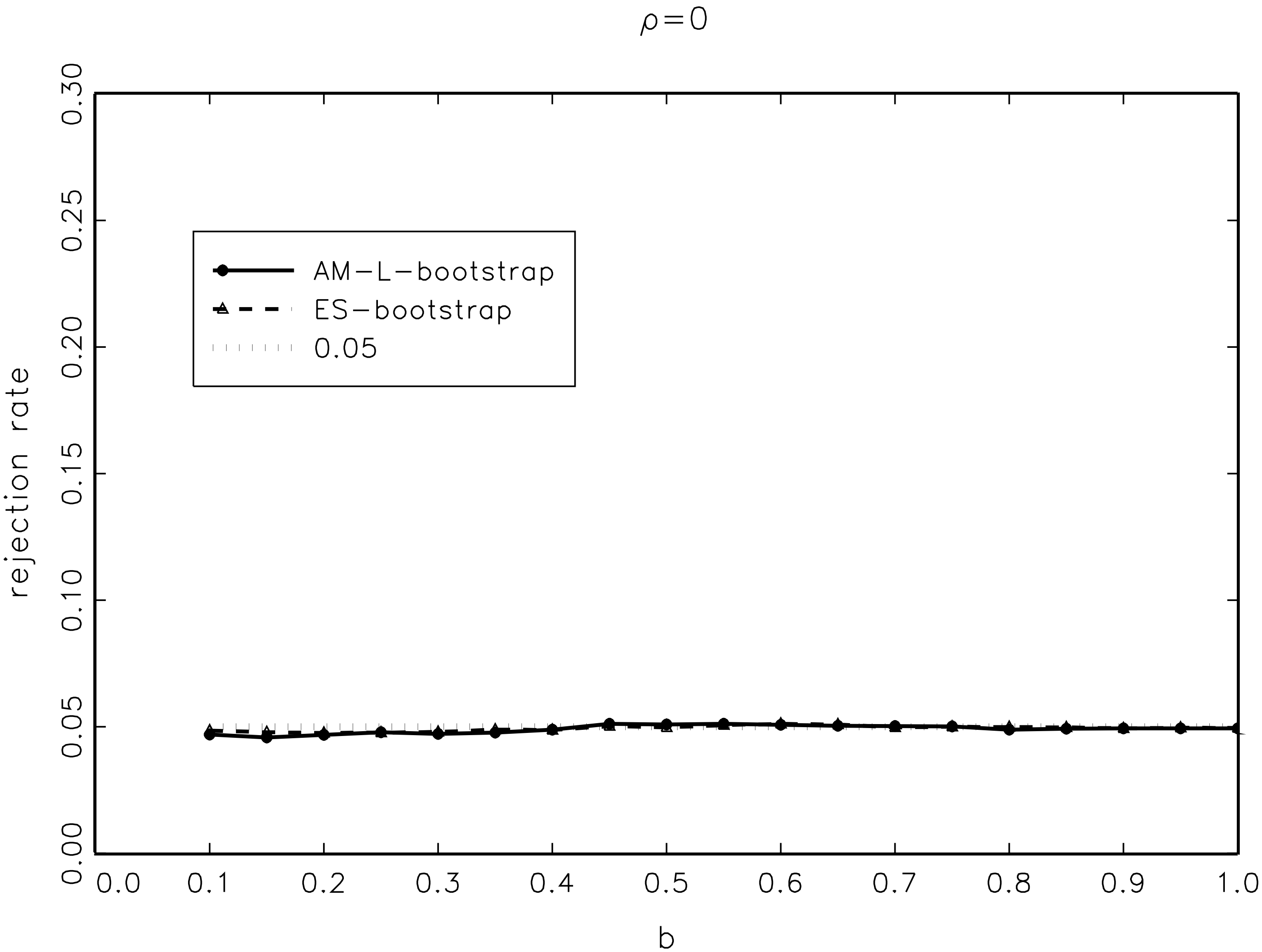


Figure 2.83: (cont'd)

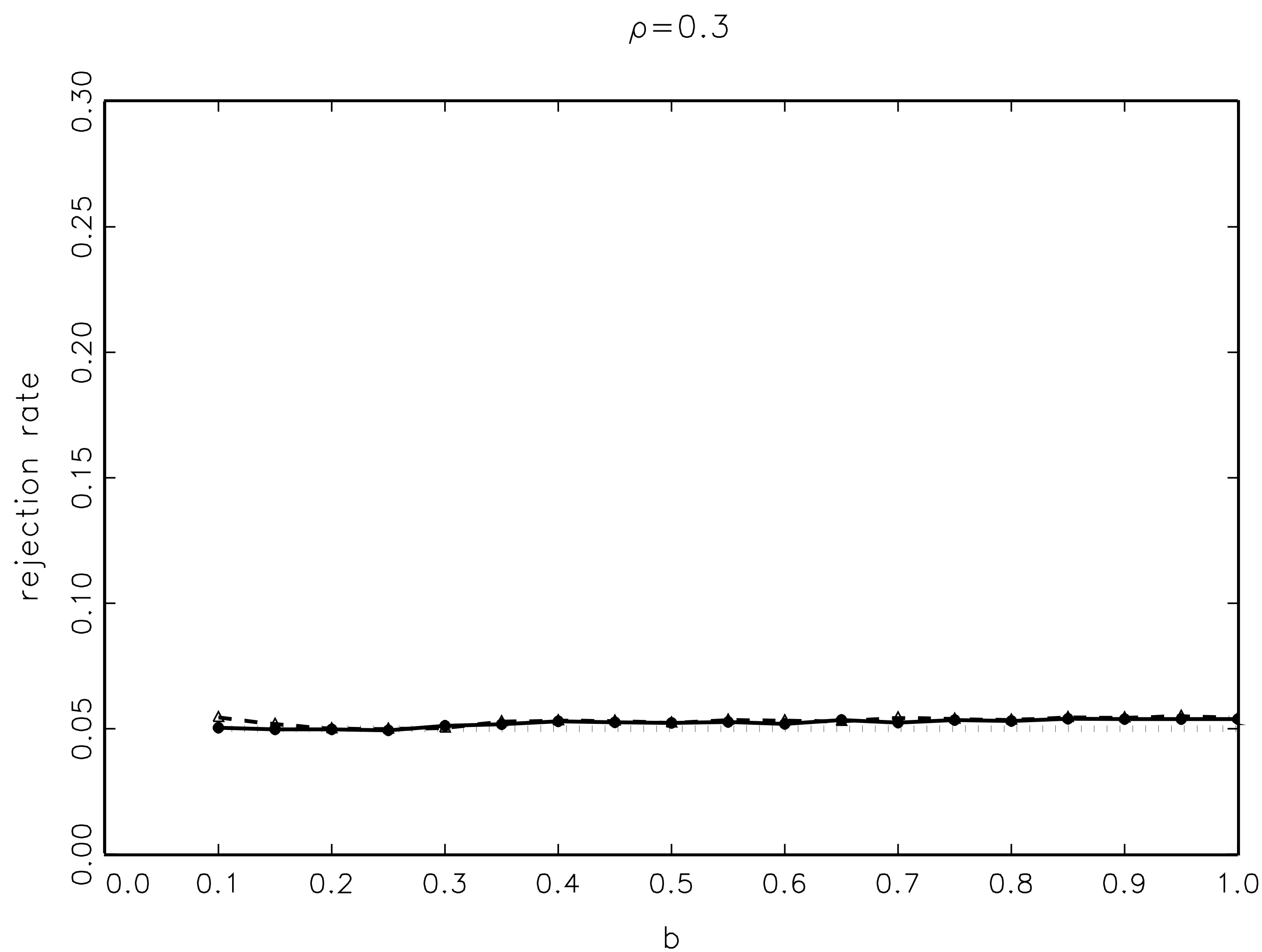


Figure 2.83: (cont'd)

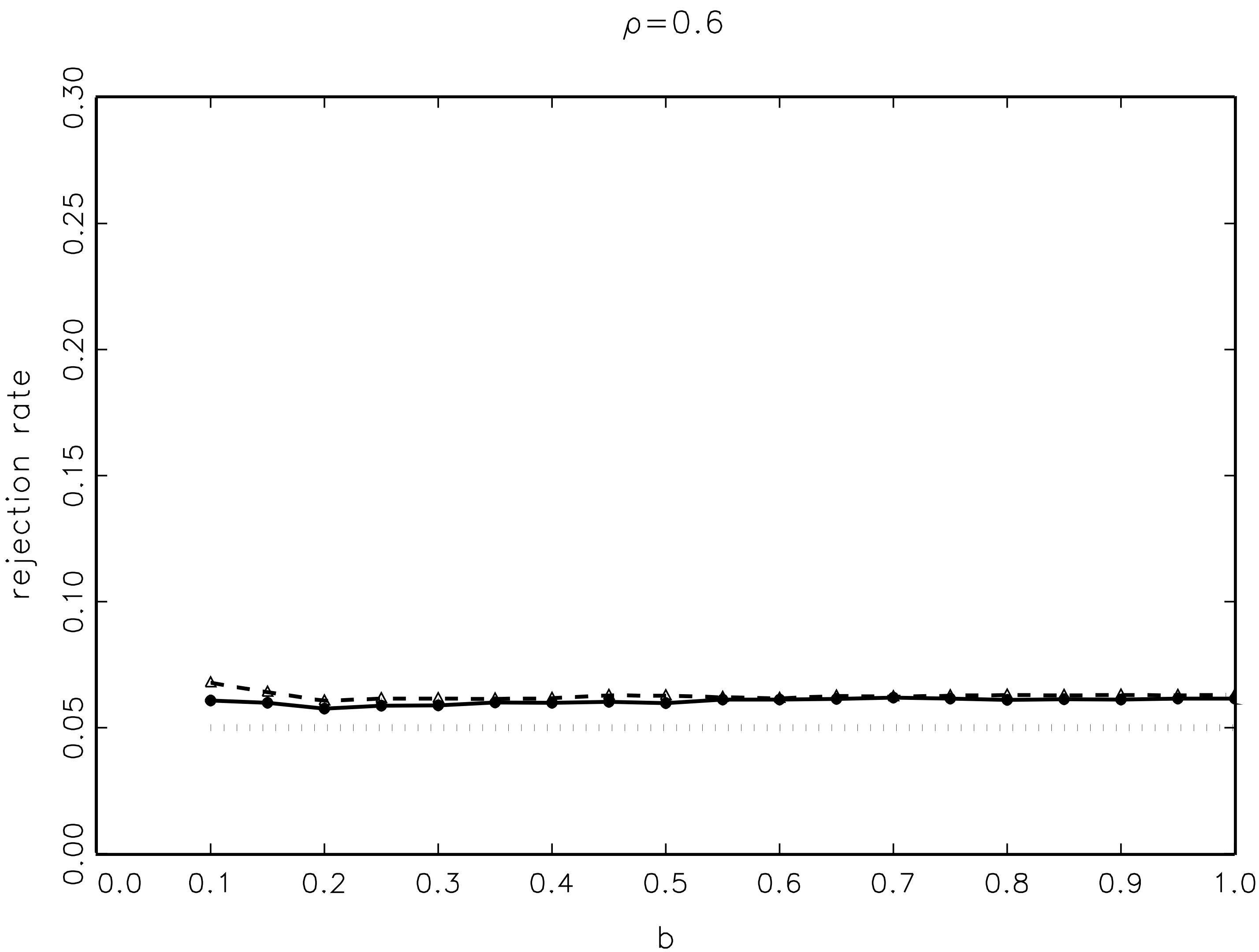


Figure 2.83: (cont'd)

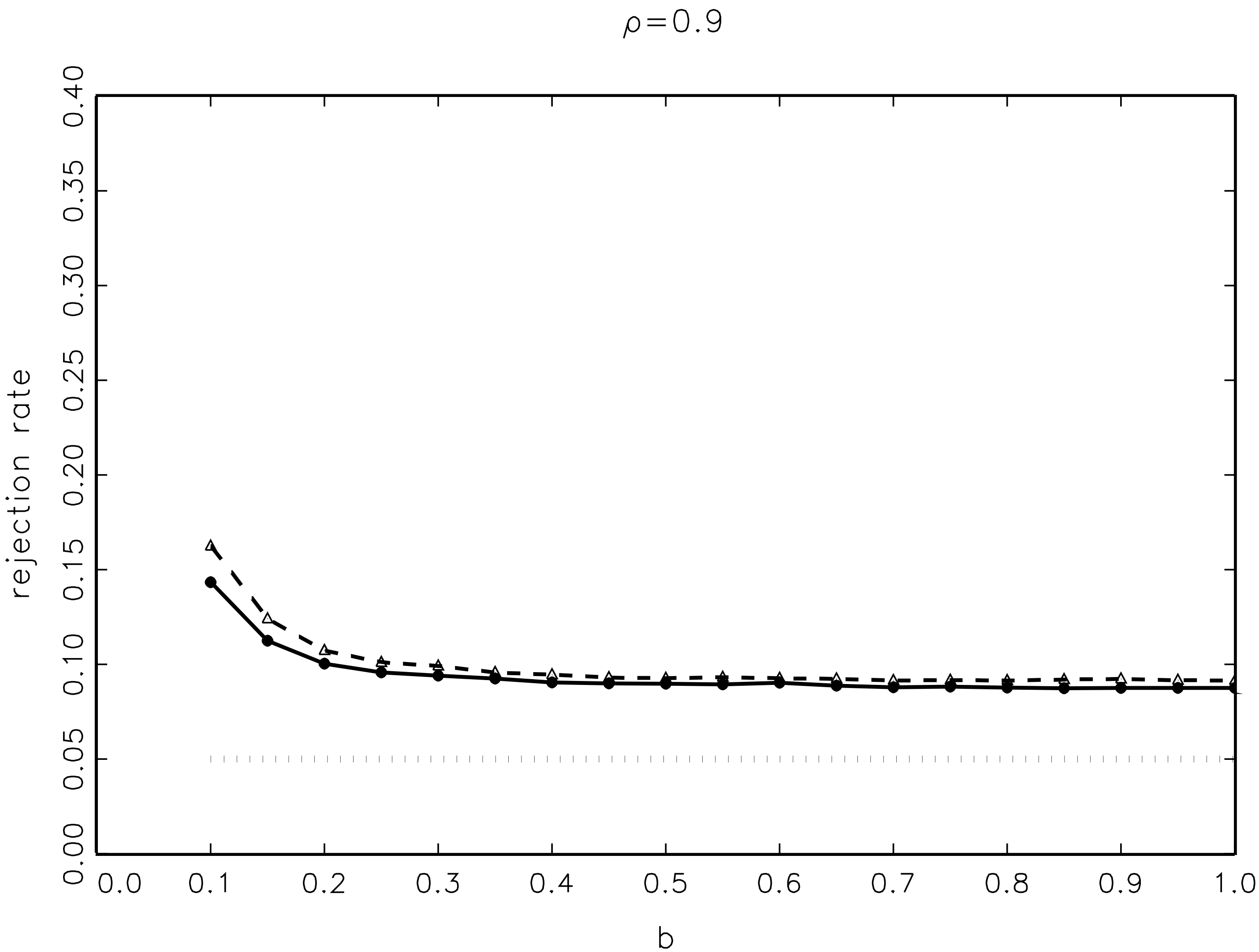


Figure 2.84: AM and ES - Conditional Bernoulli ($p = 0.5$), Bartlett, $T = 200$

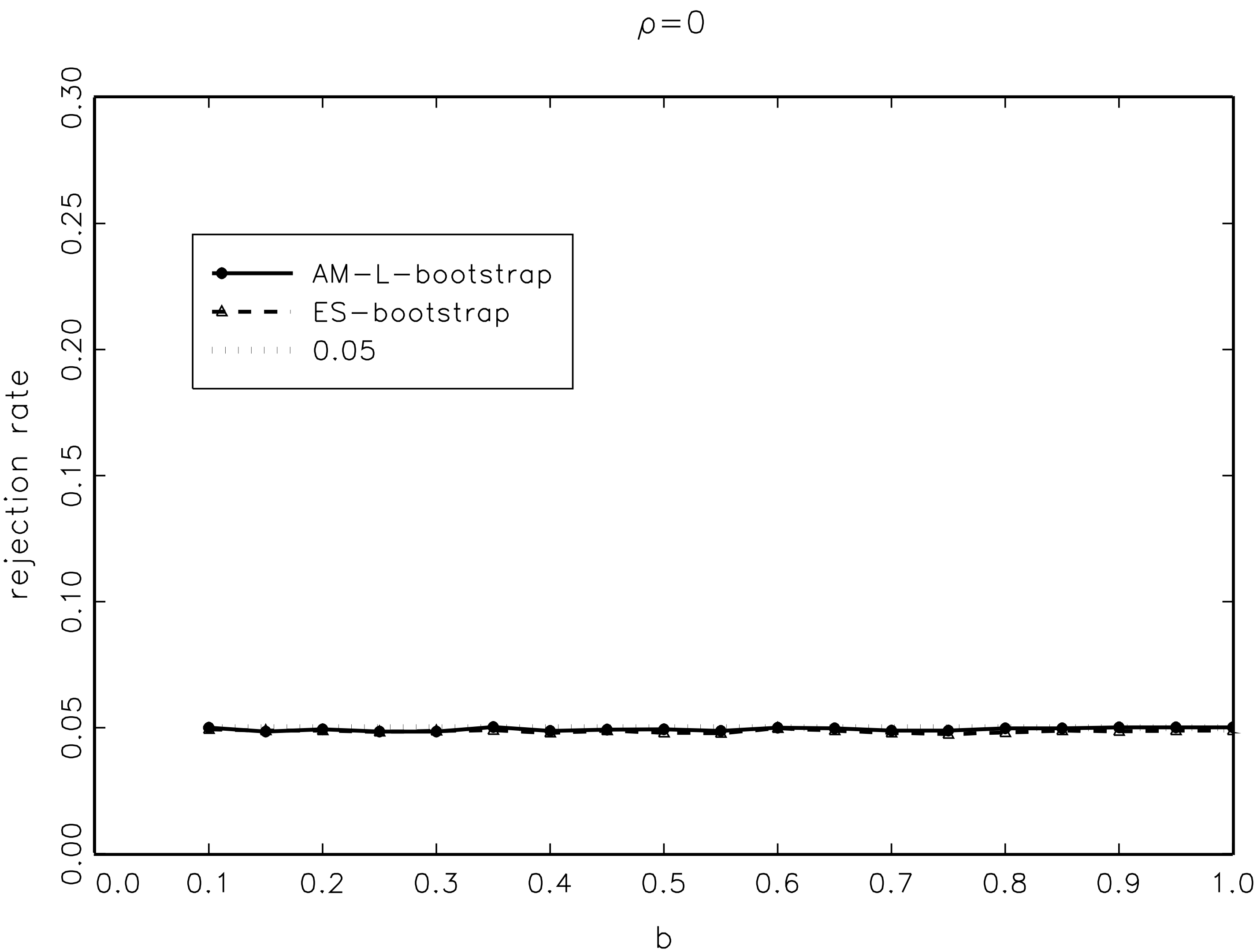


Figure 2.84: (cont'd)

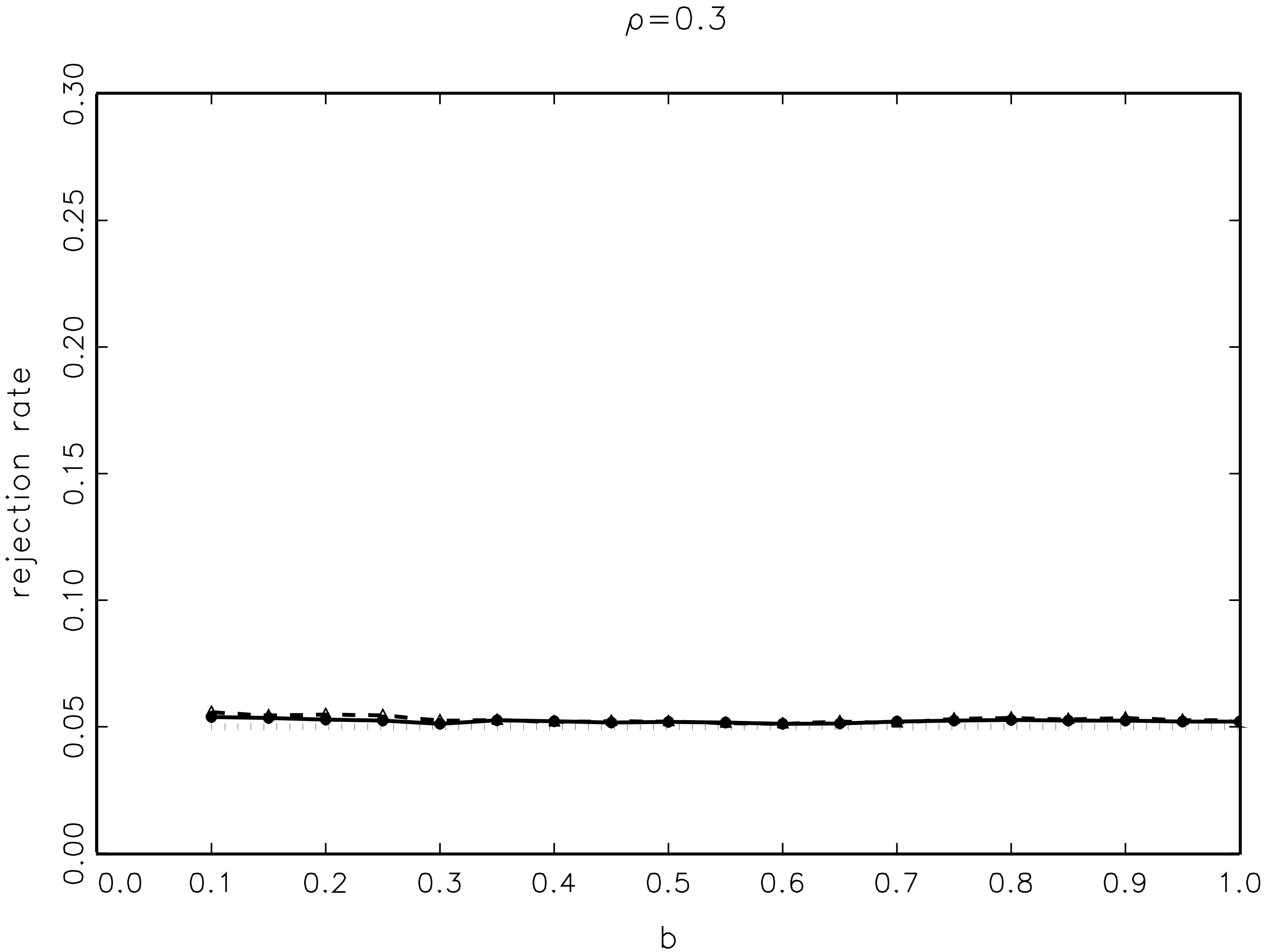


Figure 2.84: (cont'd)

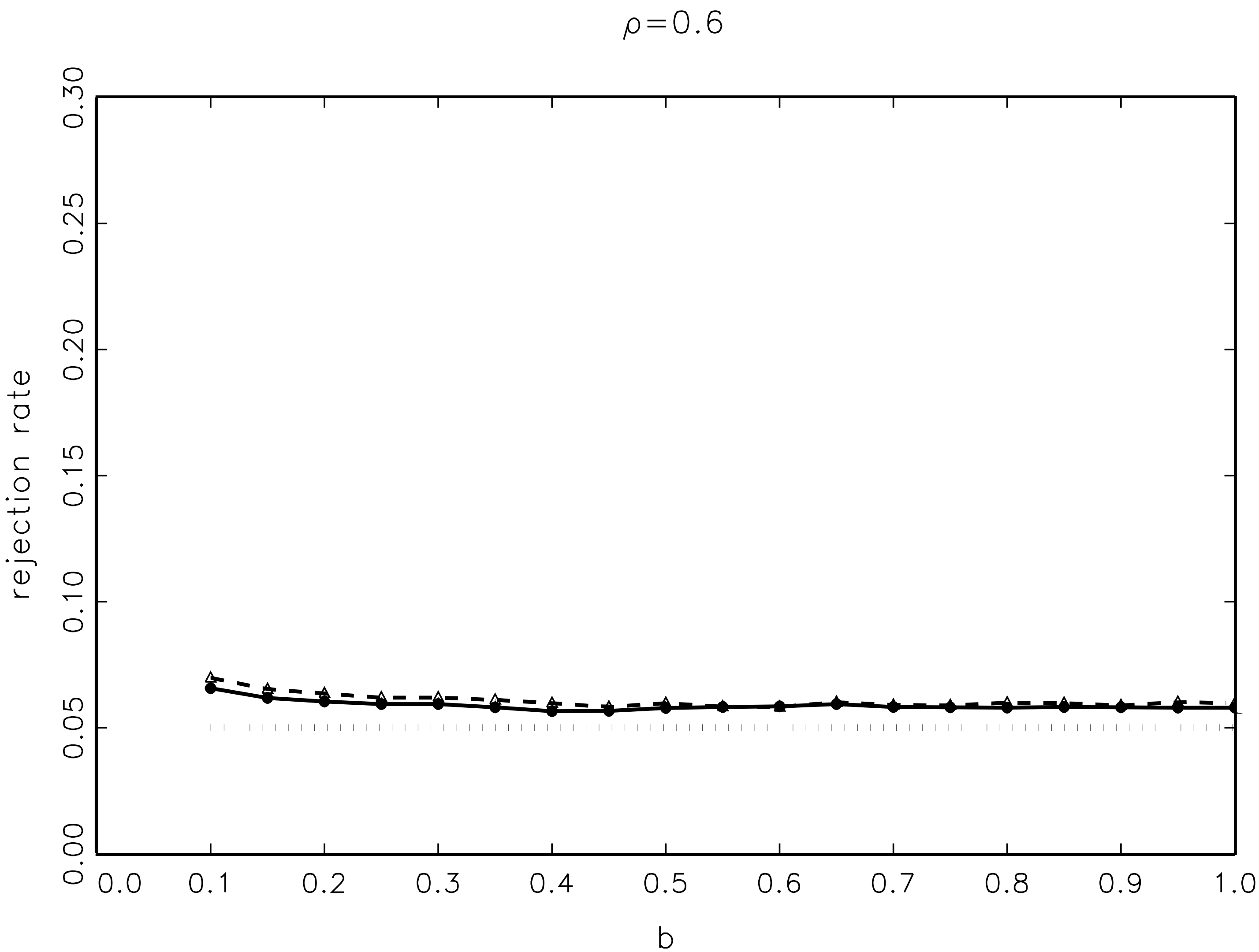


Figure 2.84: (cont'd)

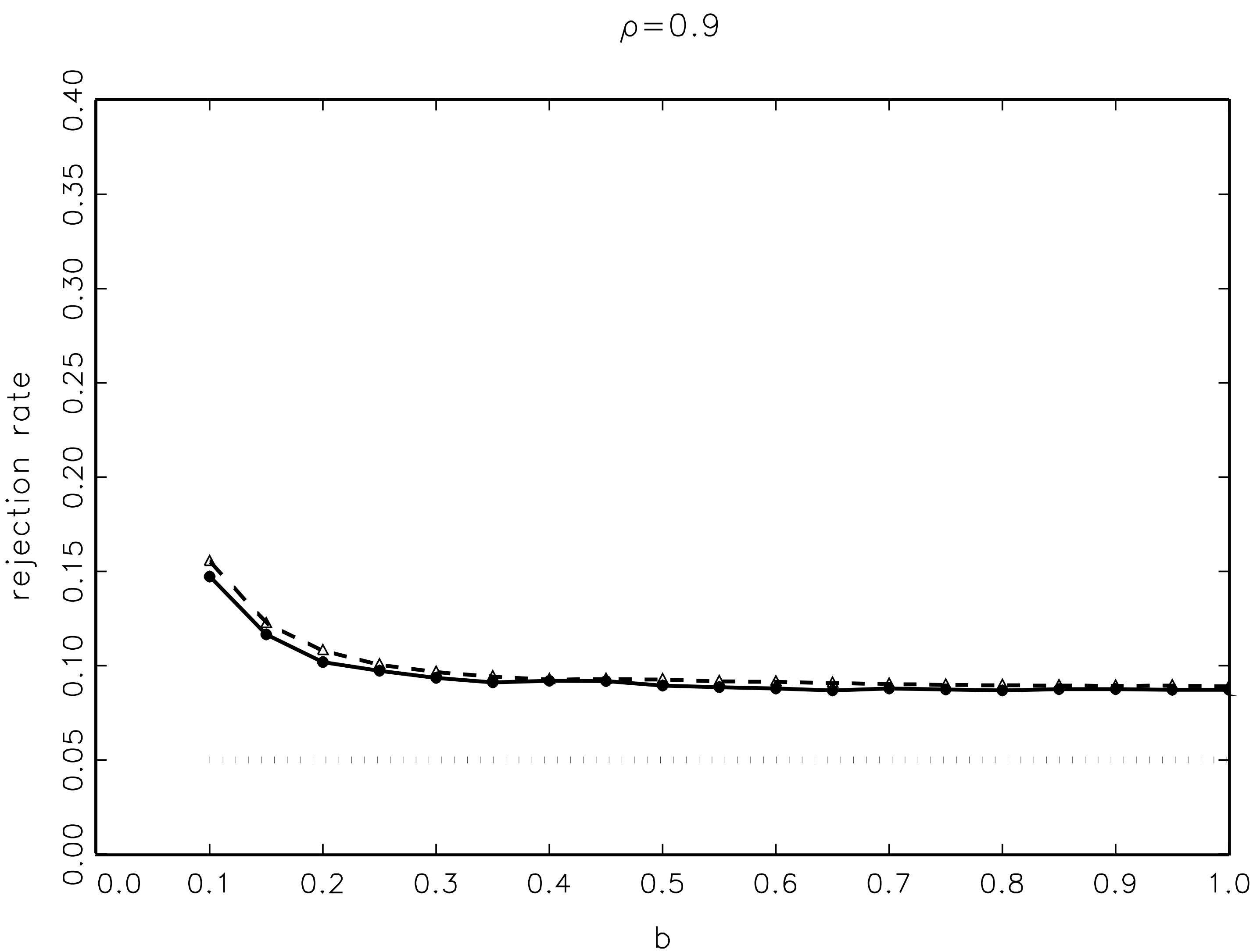


Figure 2.85: AM and ES - Conditional Bernoulli ($p = 0.7$), Bartlett, $T = 200$

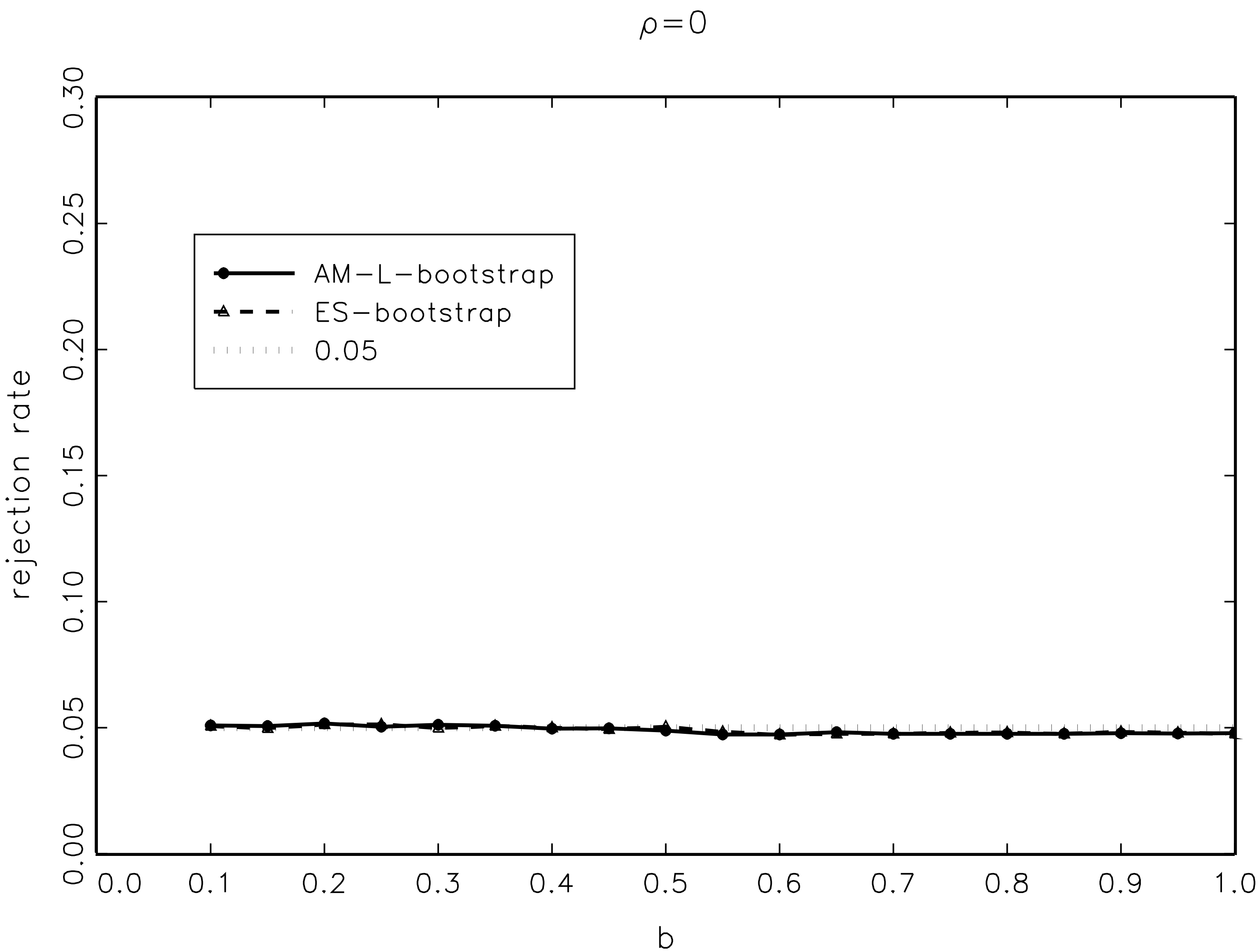


Figure 2.85: (cont'd)

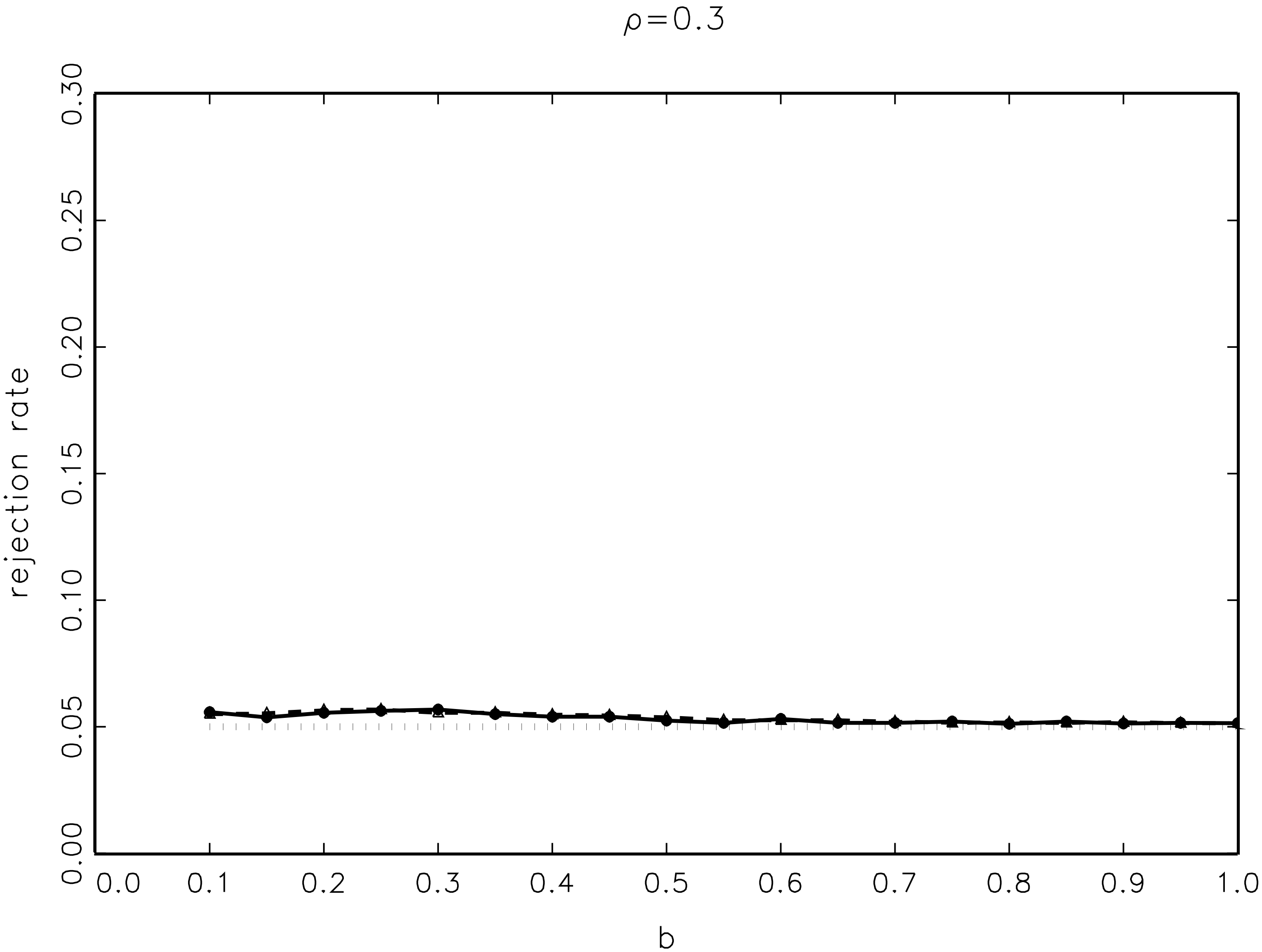


Figure 2.85: (cont'd)

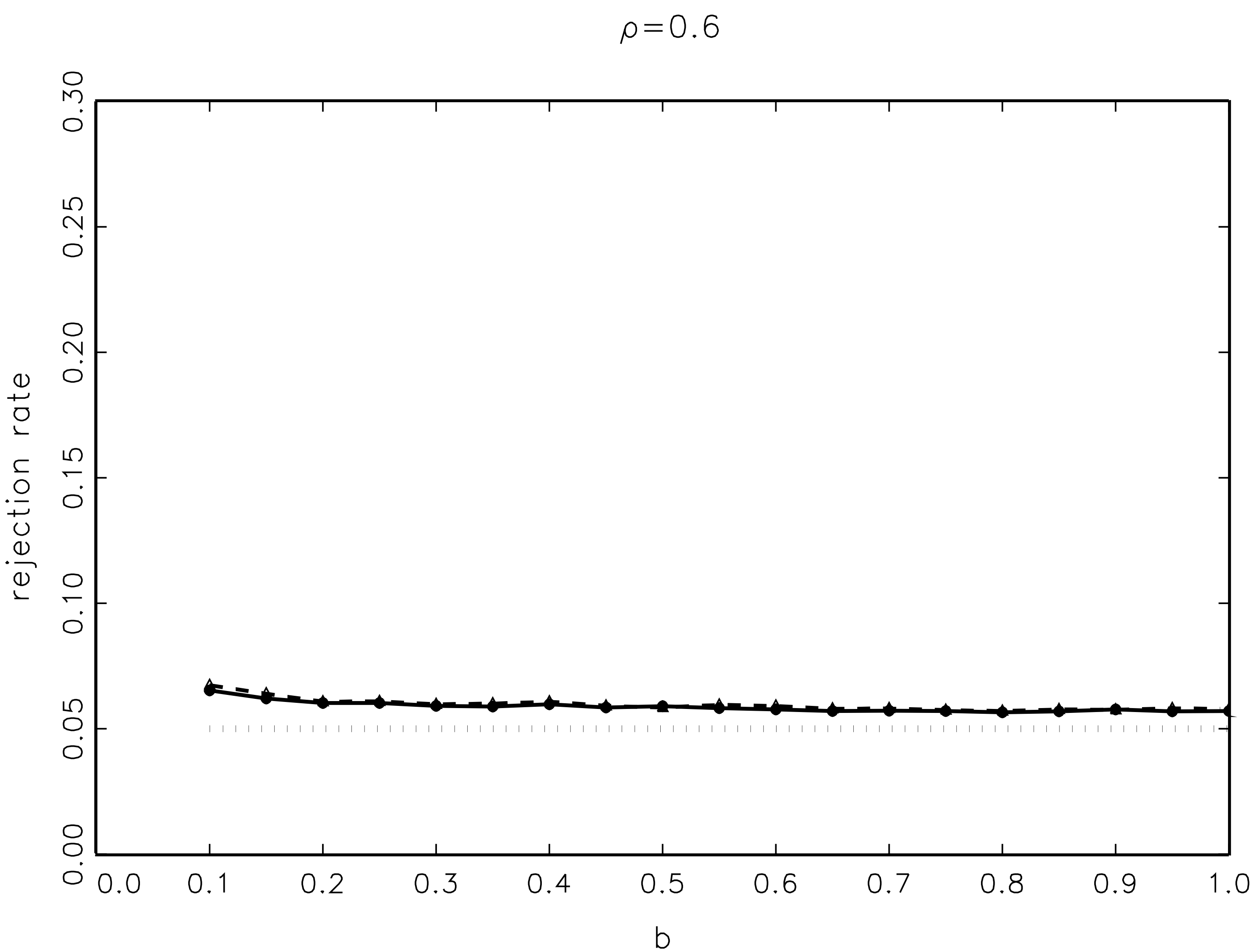


Figure 2.85: (cont'd)

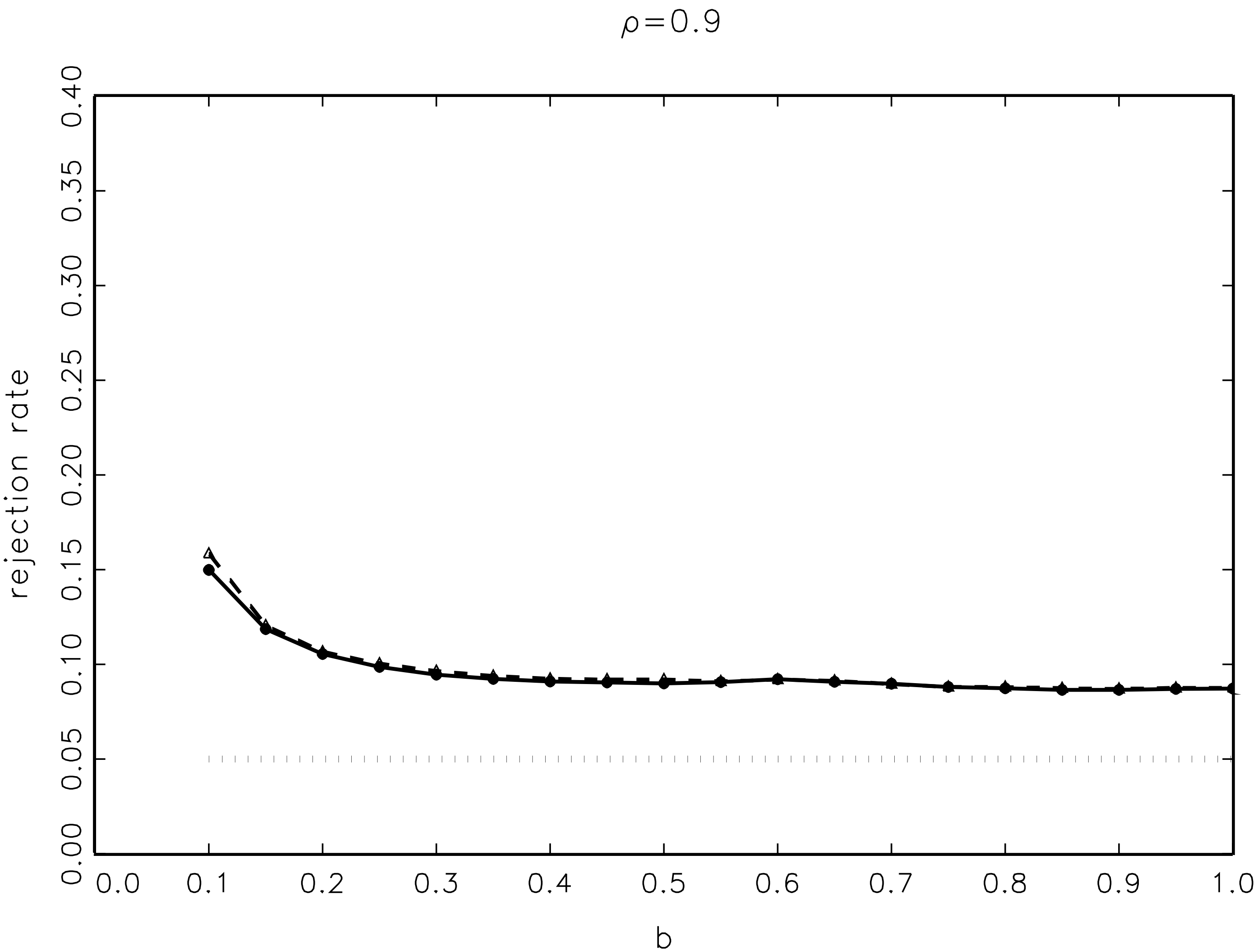


Figure 2.86: AM and ES - Random Bernoulli ($p = 0.3$), Bartlett, $T = 50$

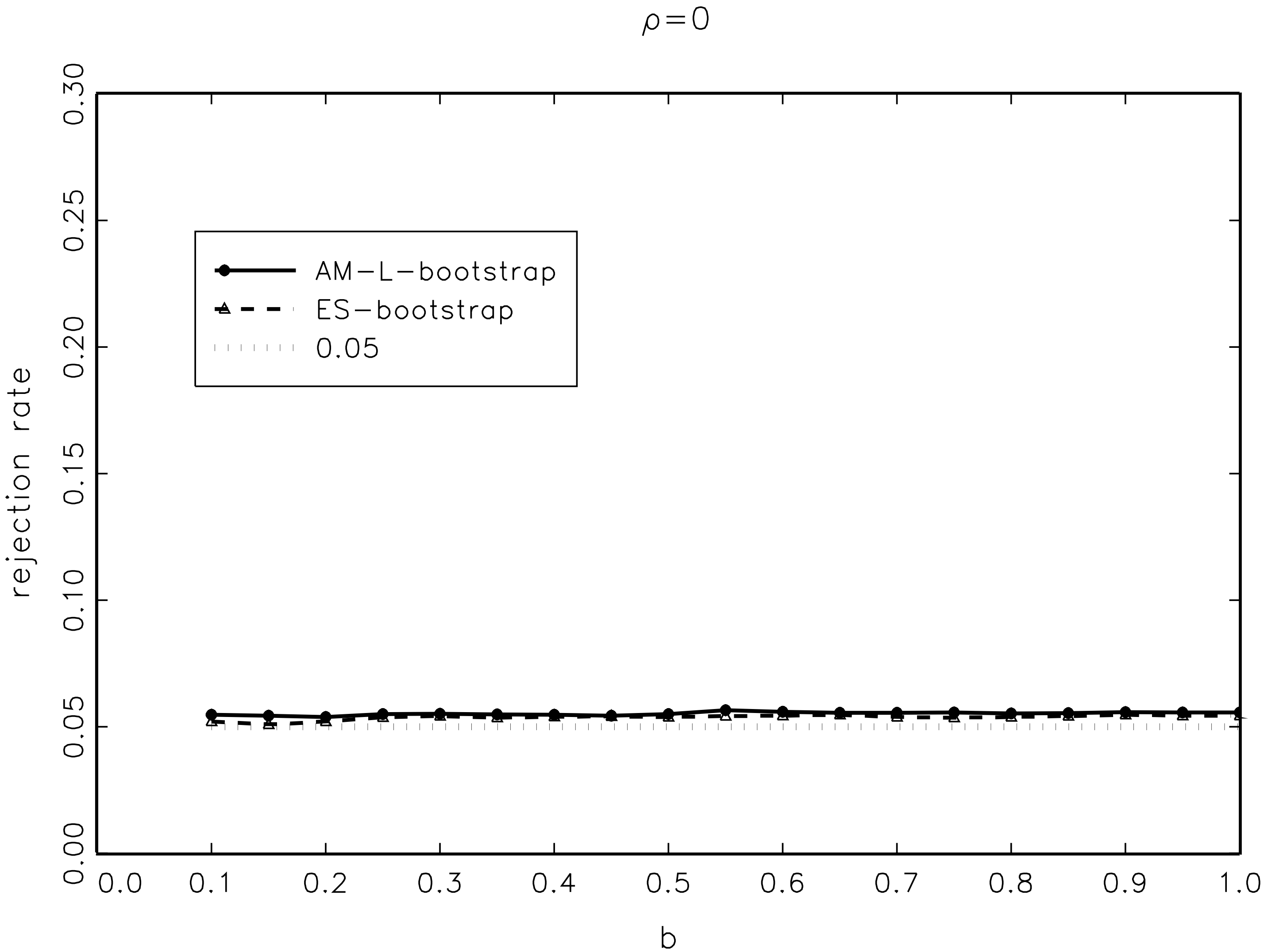


Figure 2.86: (cont'd)

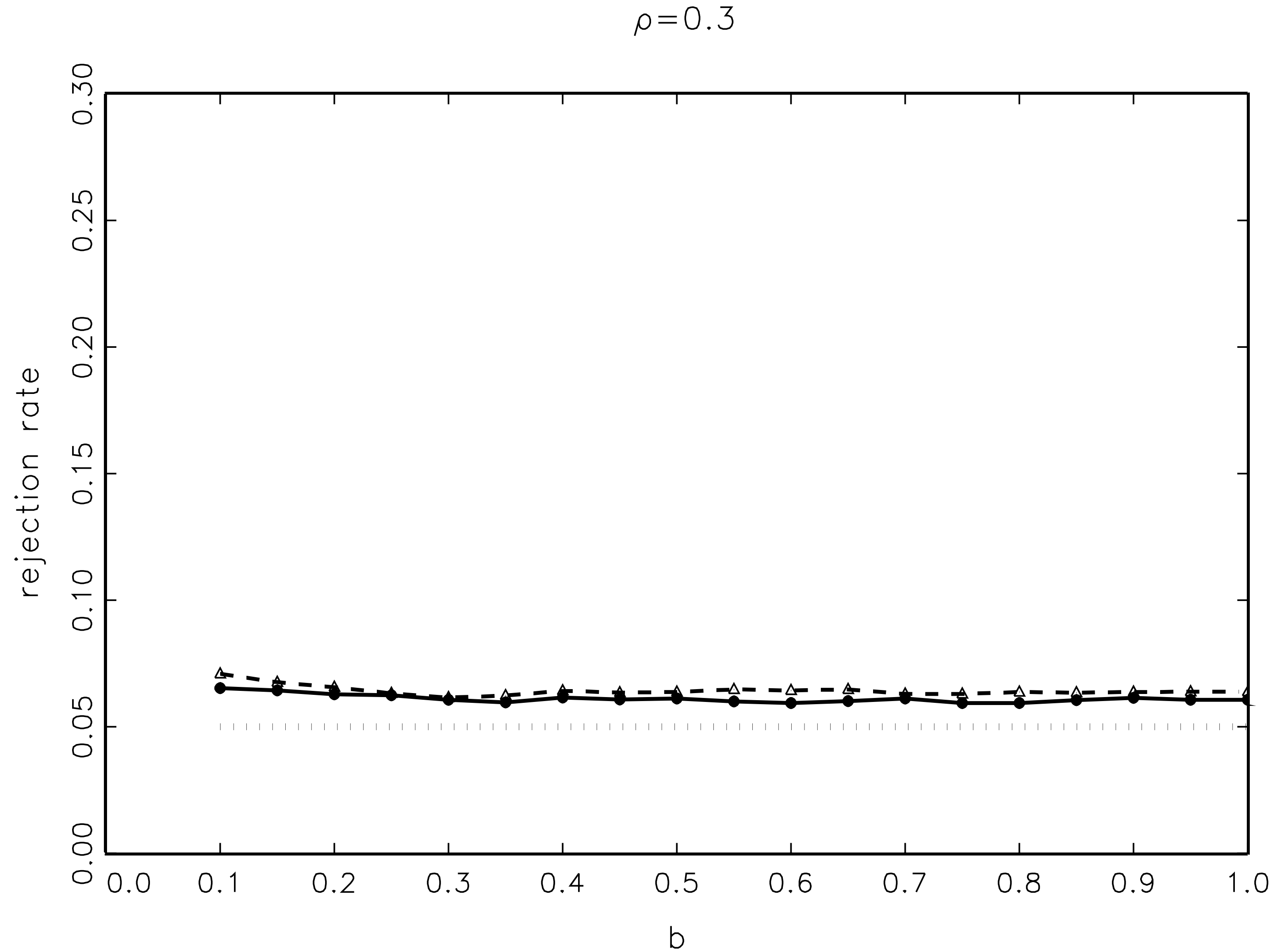


Figure 2.86: (cont'd)

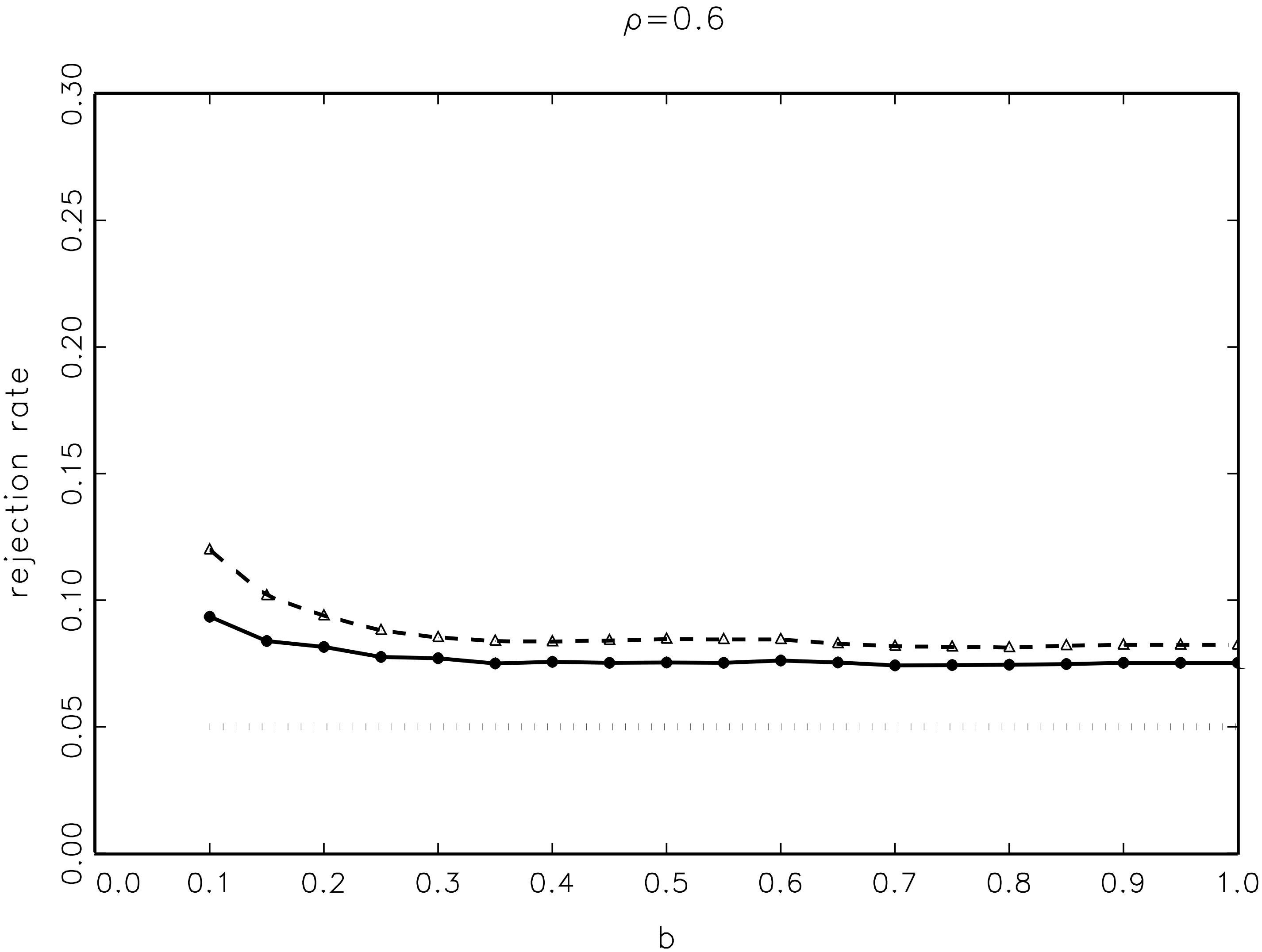


Figure 2.86: (cont'd)

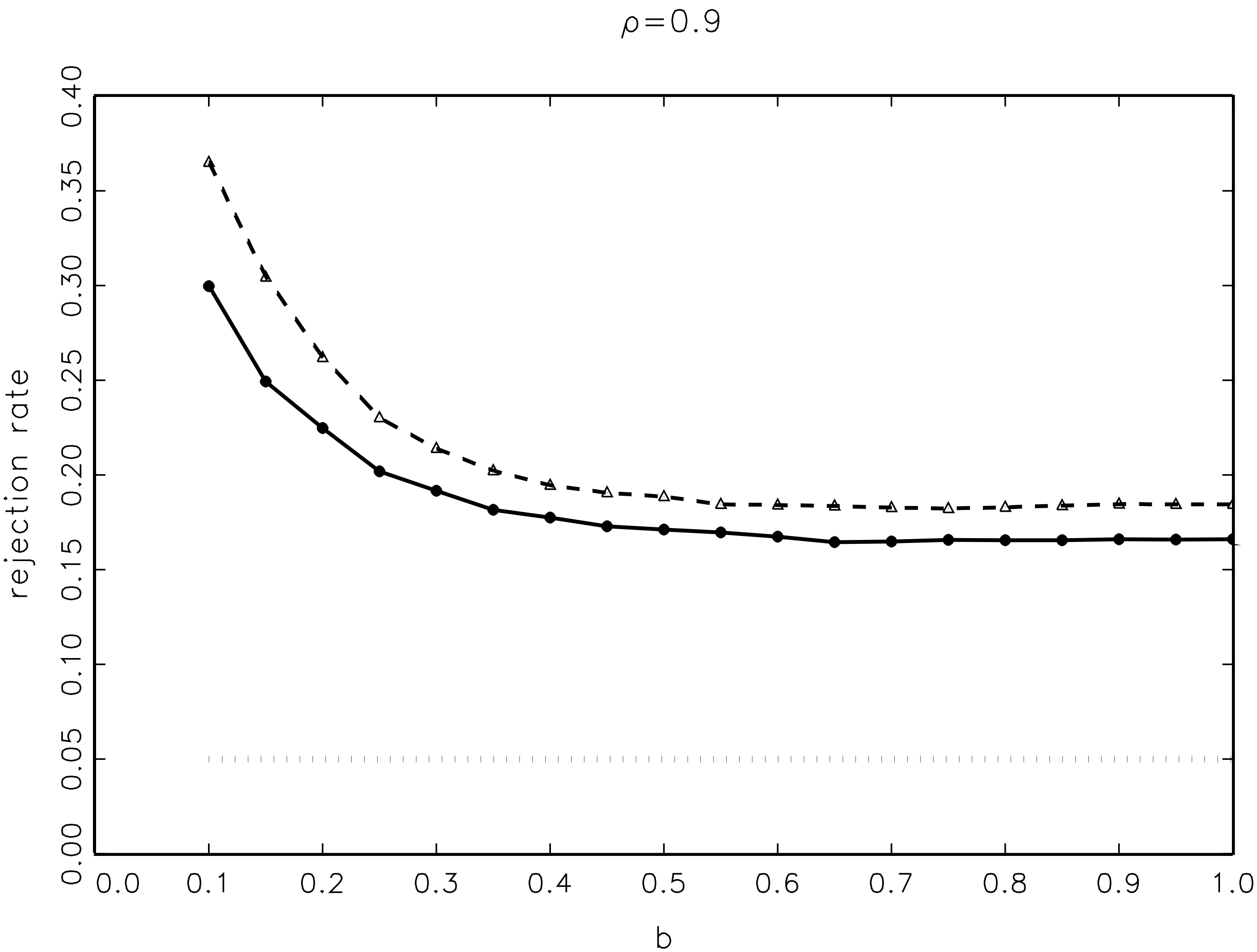


Figure 2.87: AM and ES - Random Bernoulli ($p = 0.5$), Bartlett, $T = 50$

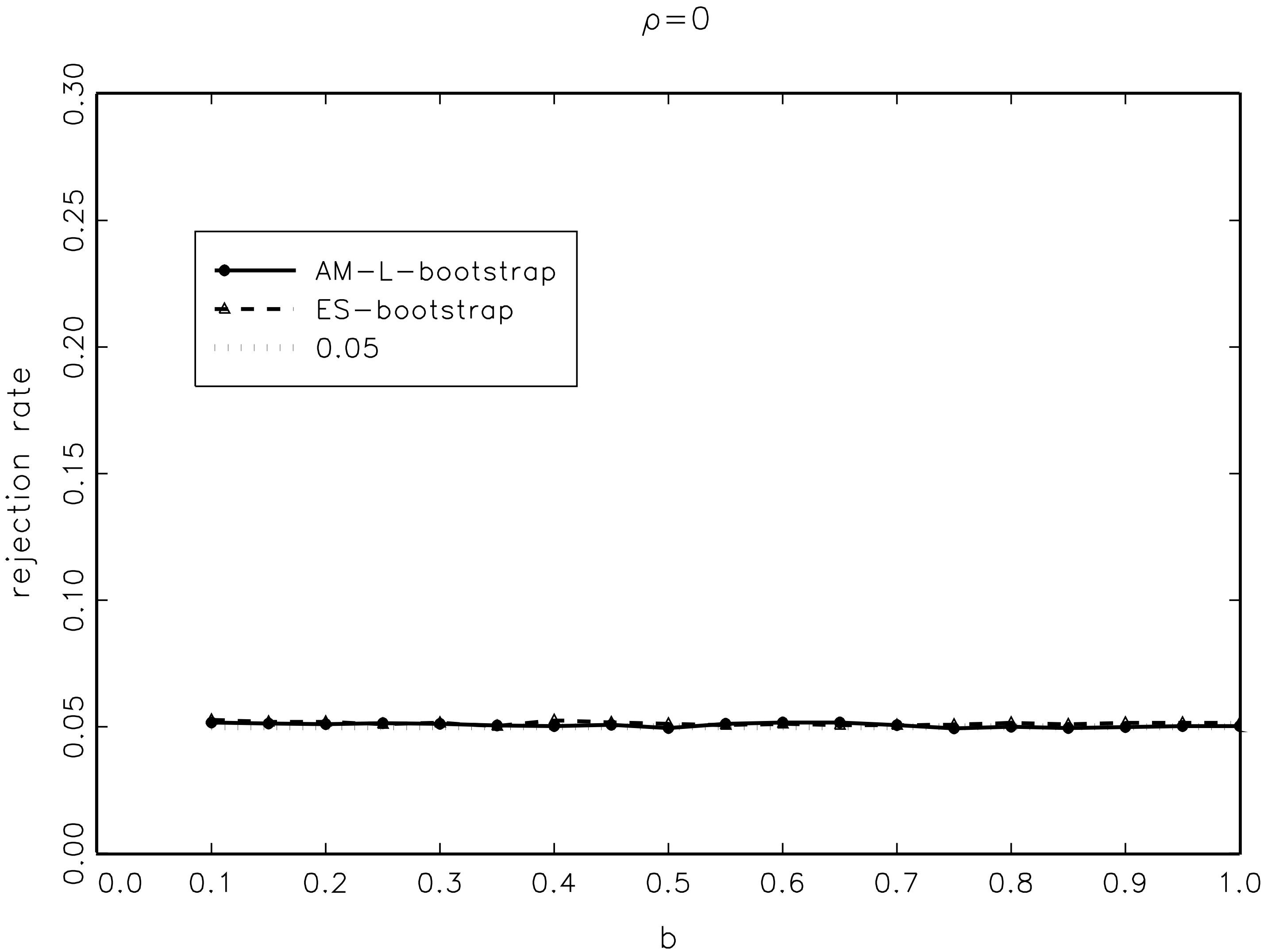


Figure 2.87: (cont'd)

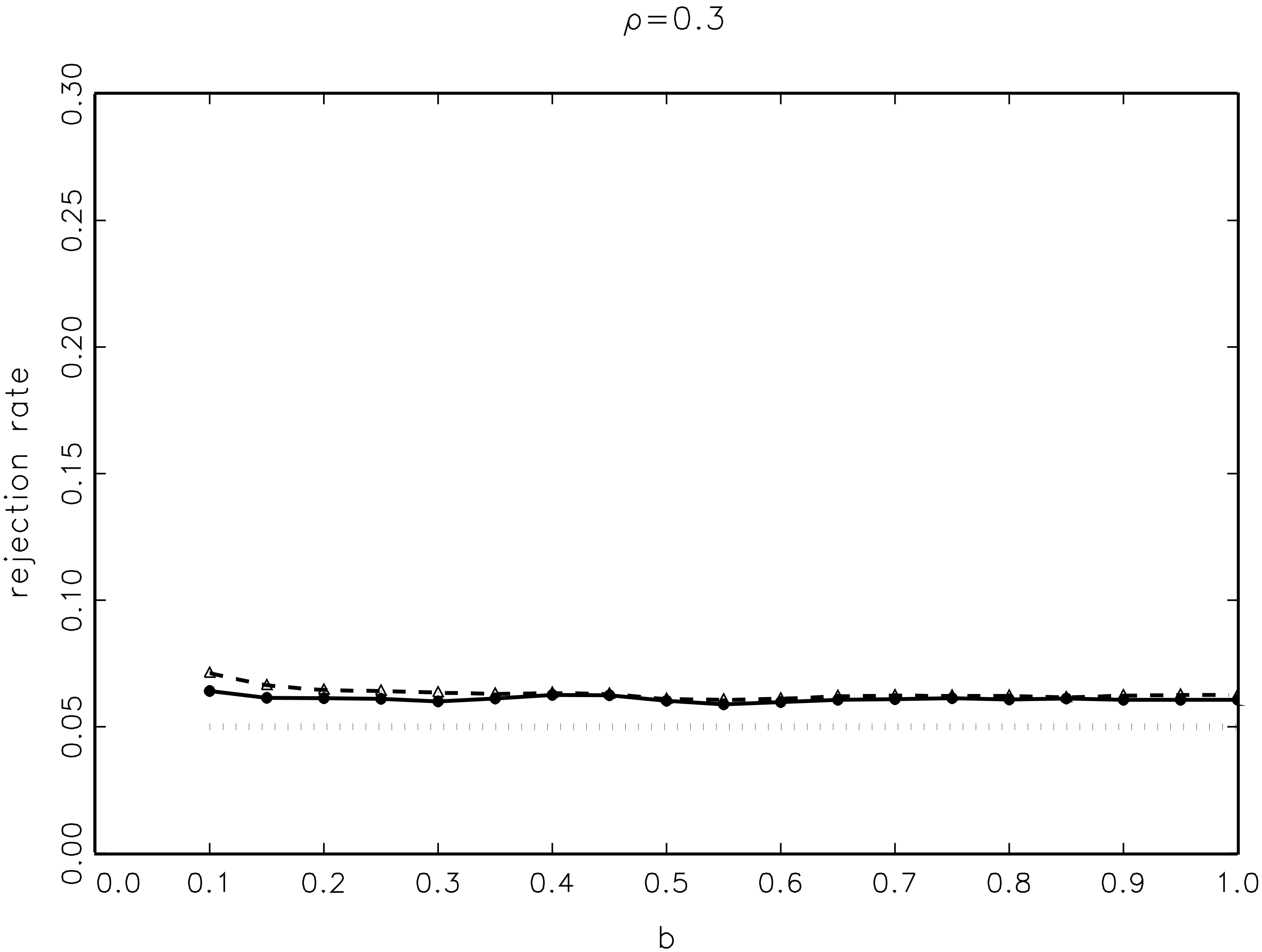


Figure 2.87: (cont'd)

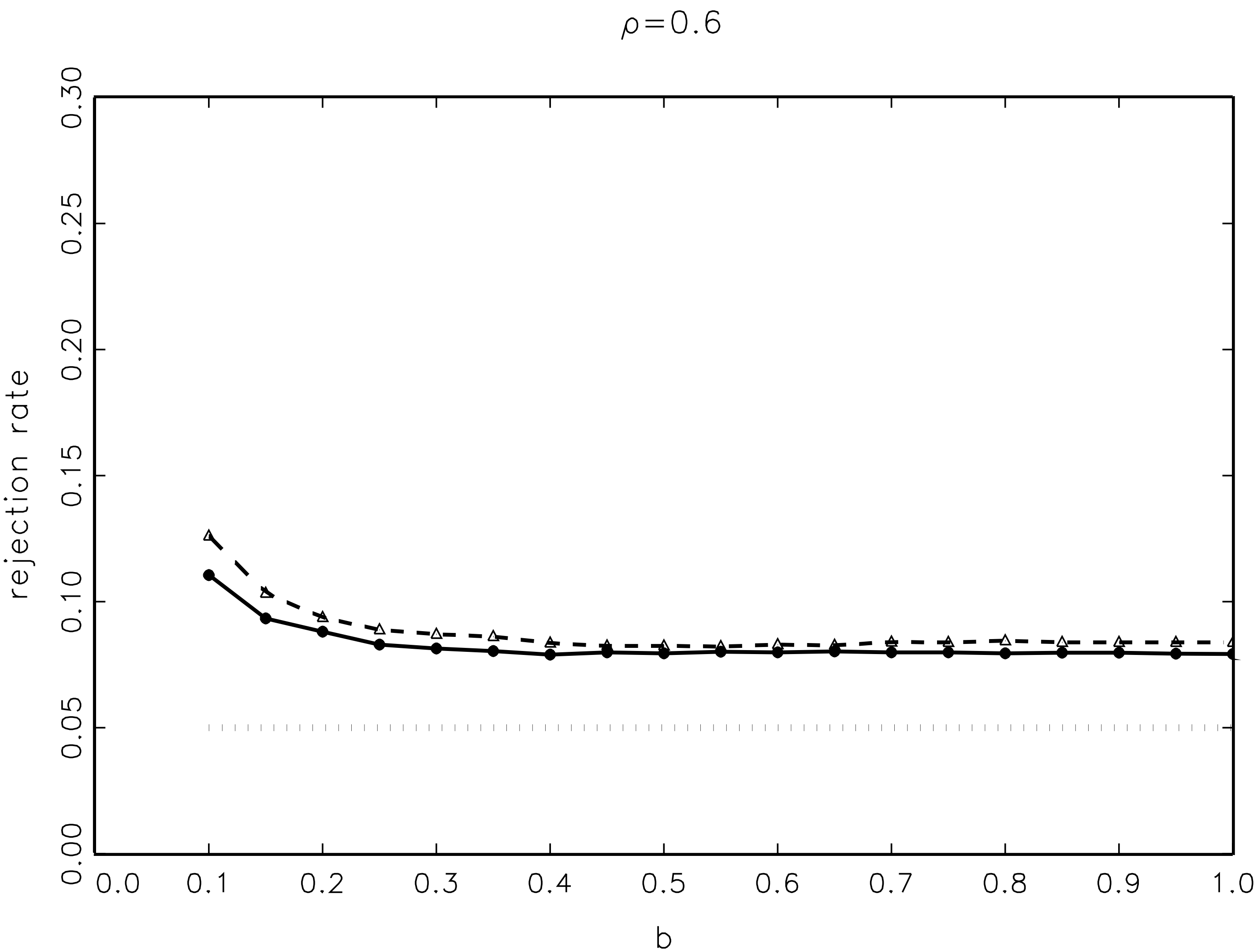


Figure 2.87: (cont'd)

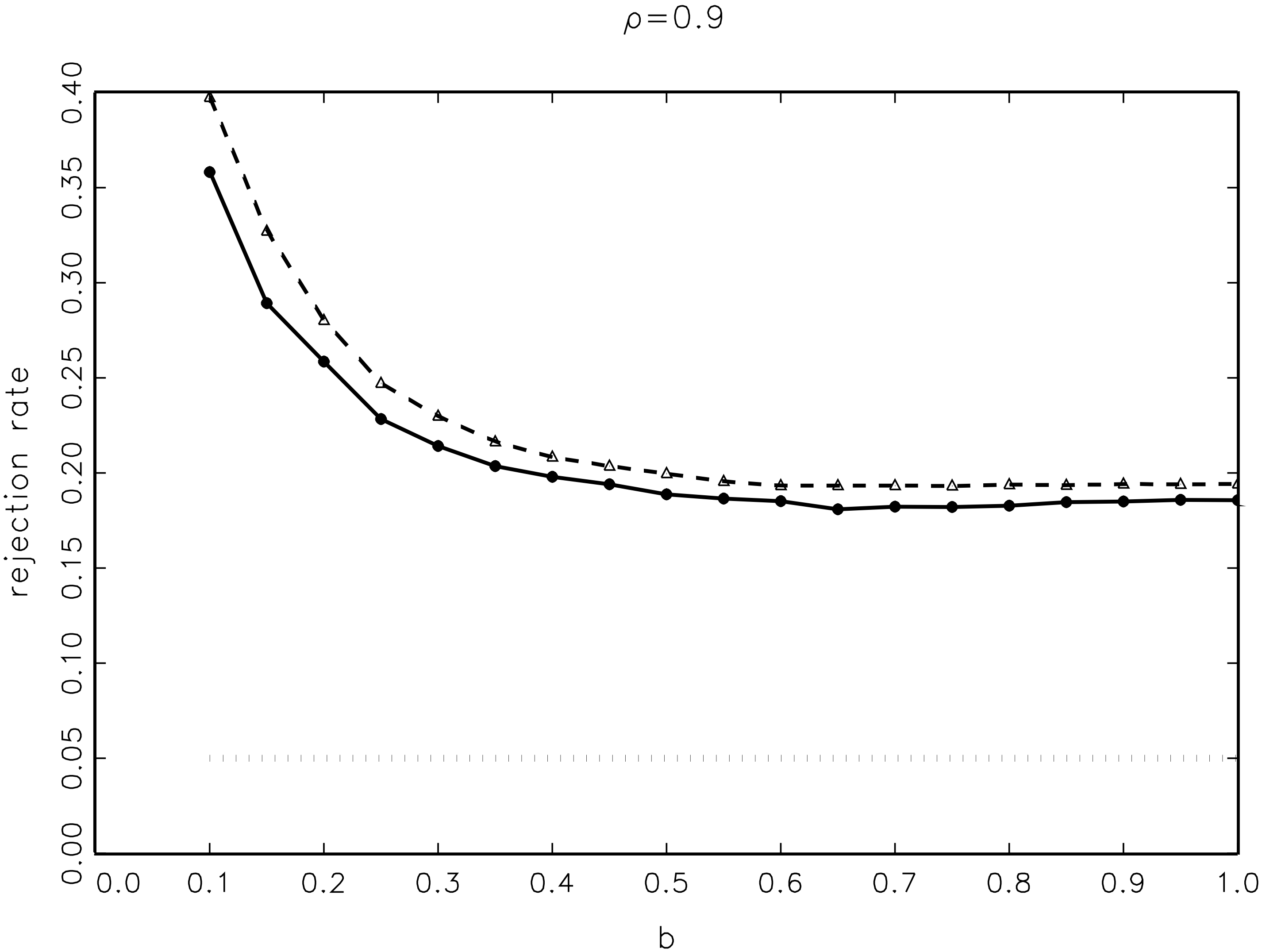


Figure 2.88: AM and ES - Random Bernoulli ($p = 0.7$), Bartlett, $T = 50$

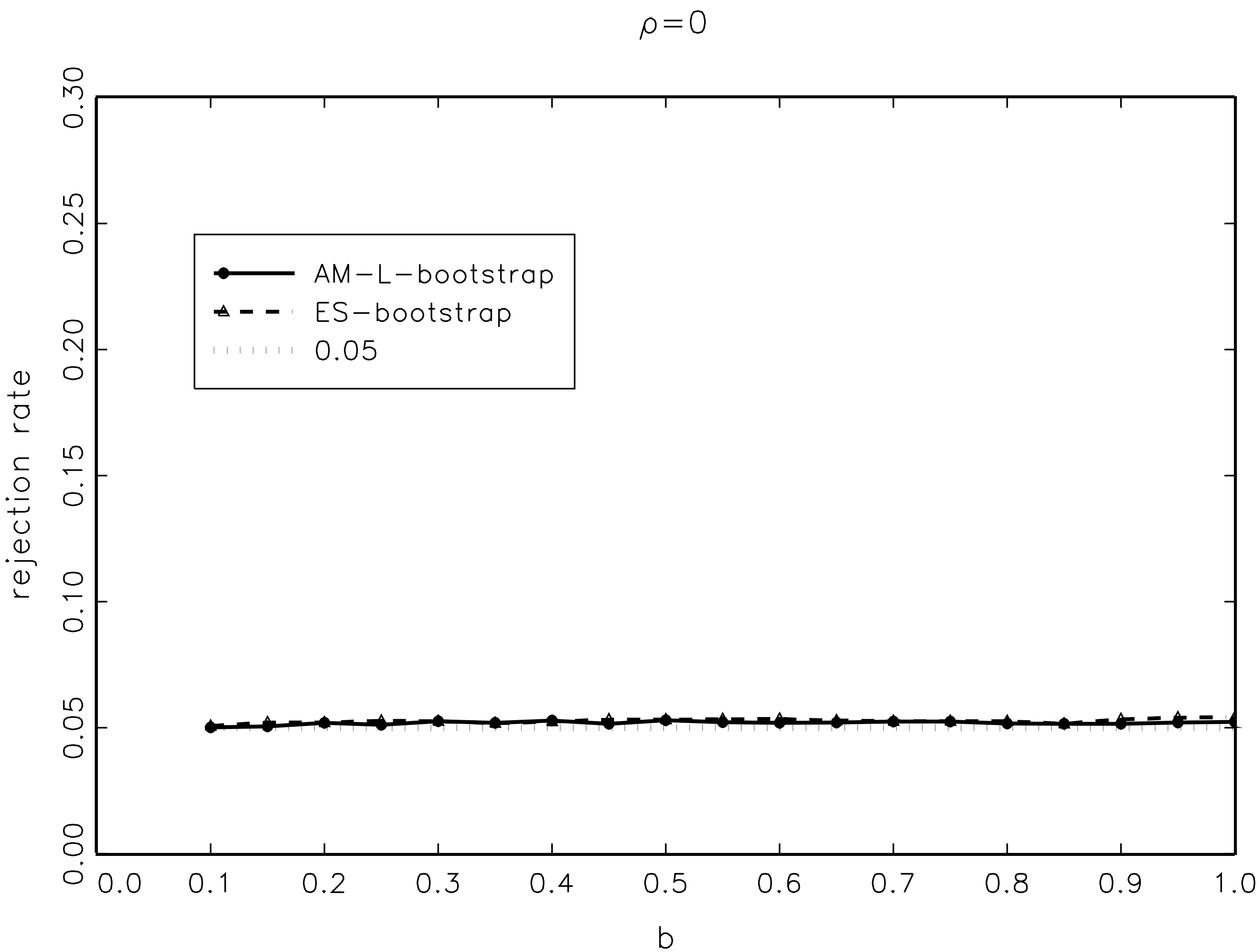


Figure 2.88: (cont'd)

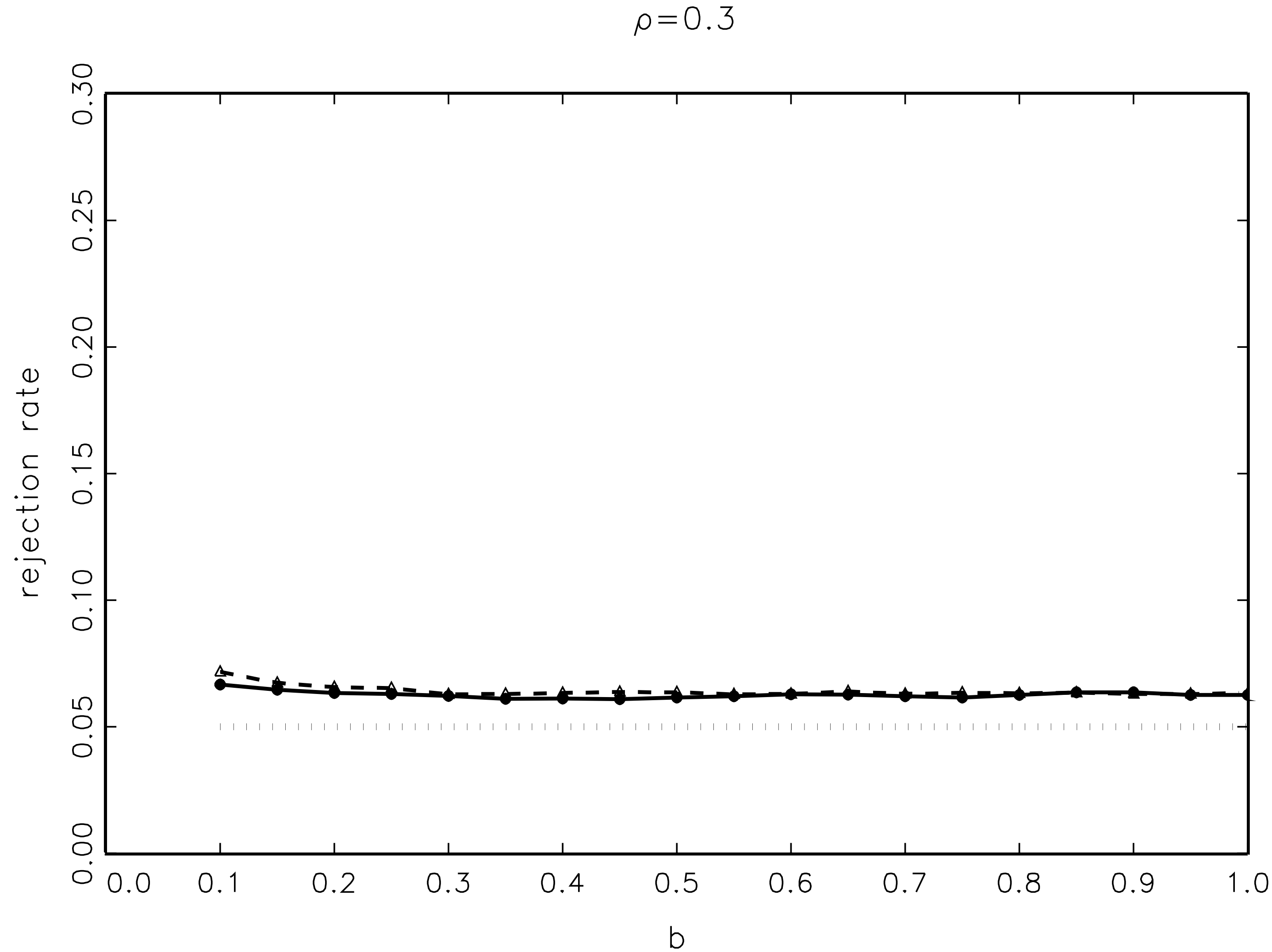


Figure 2.88: (cont'd)

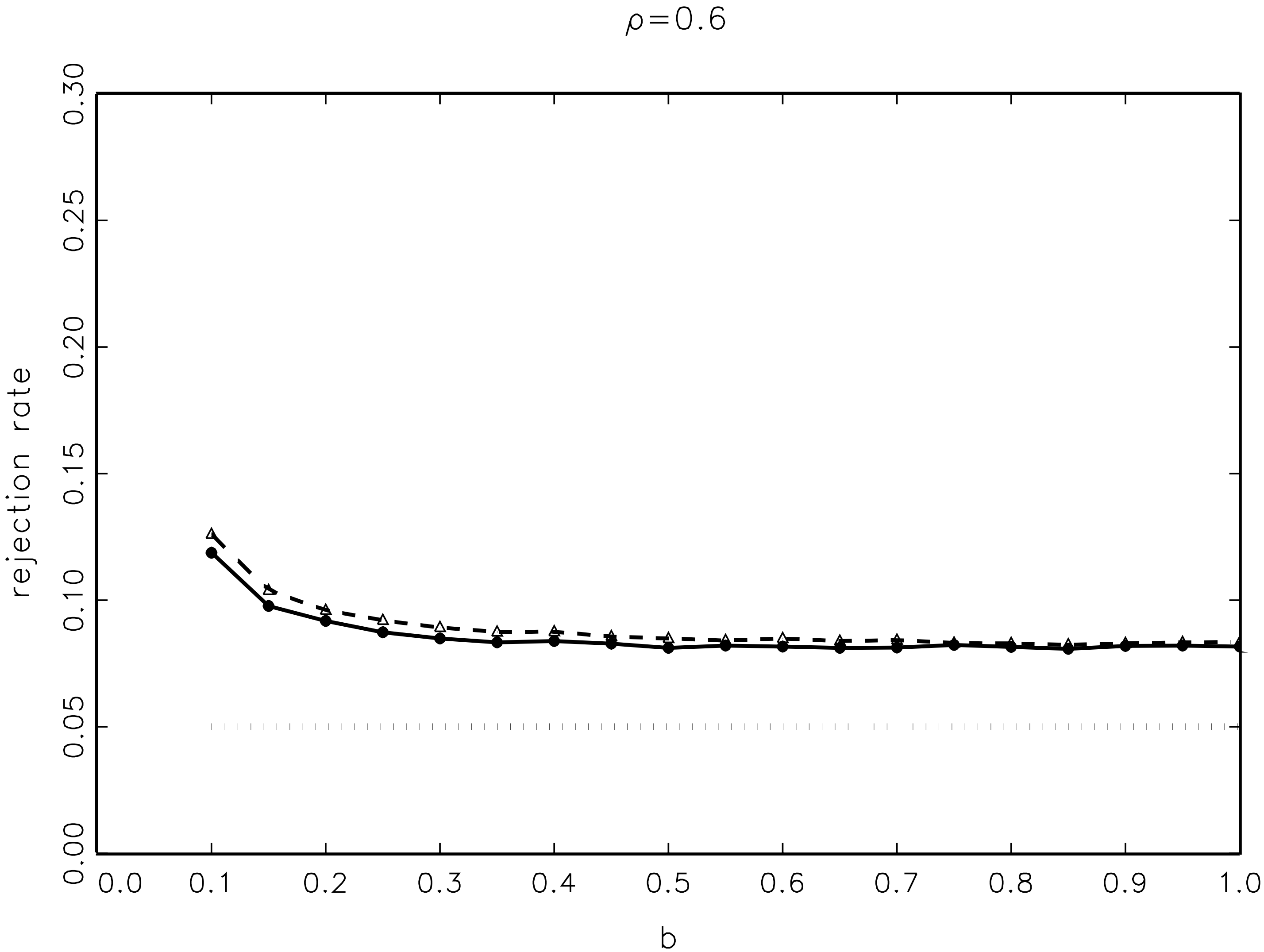


Figure 2.88: (cont'd)

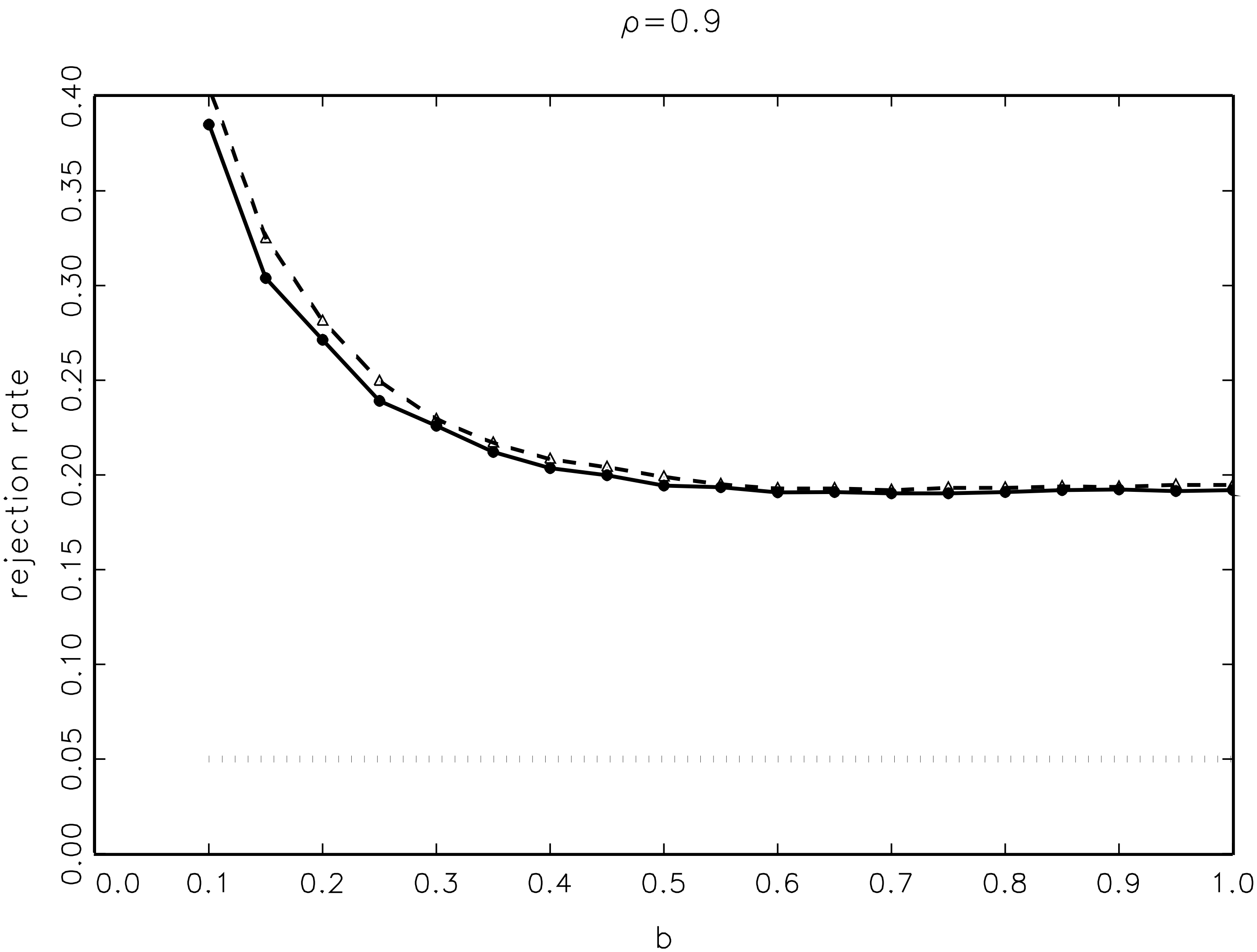


Figure 2.89: AM and ES - Random Bernoulli ($p = 0.3$), Bartlett, $T = 100$

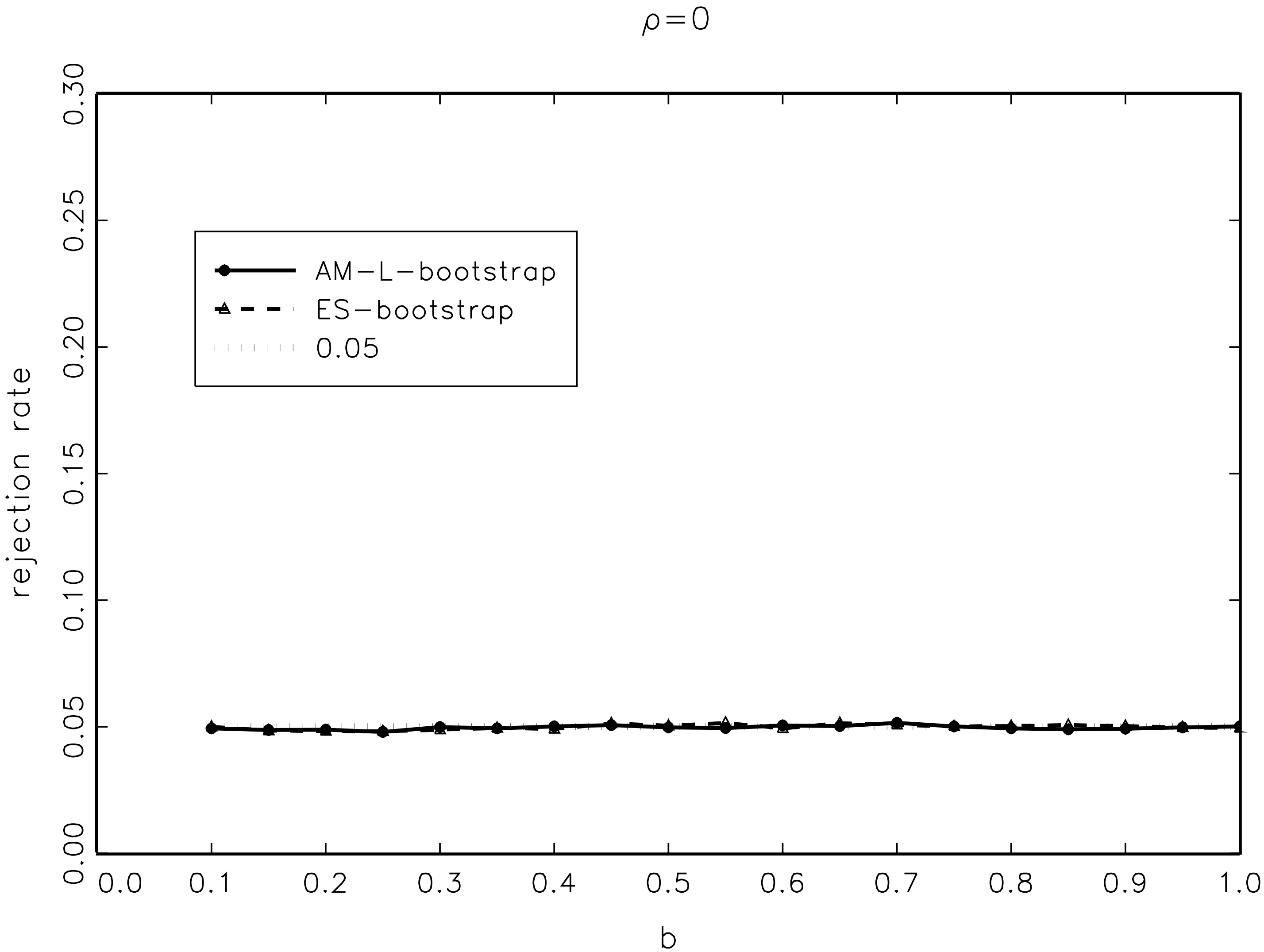


Figure 2.89: (cont'd)

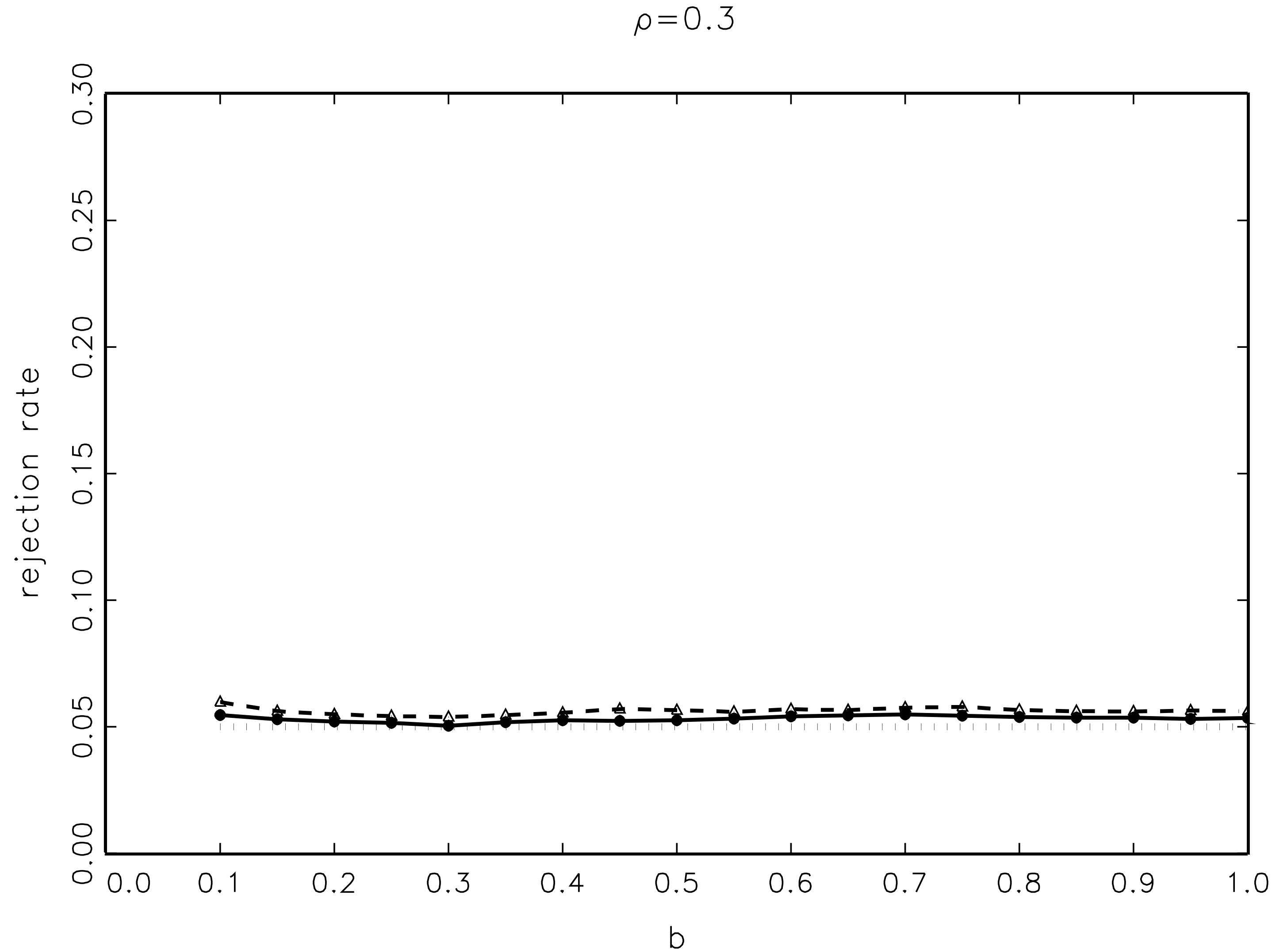


Figure 2.89: (cont'd)

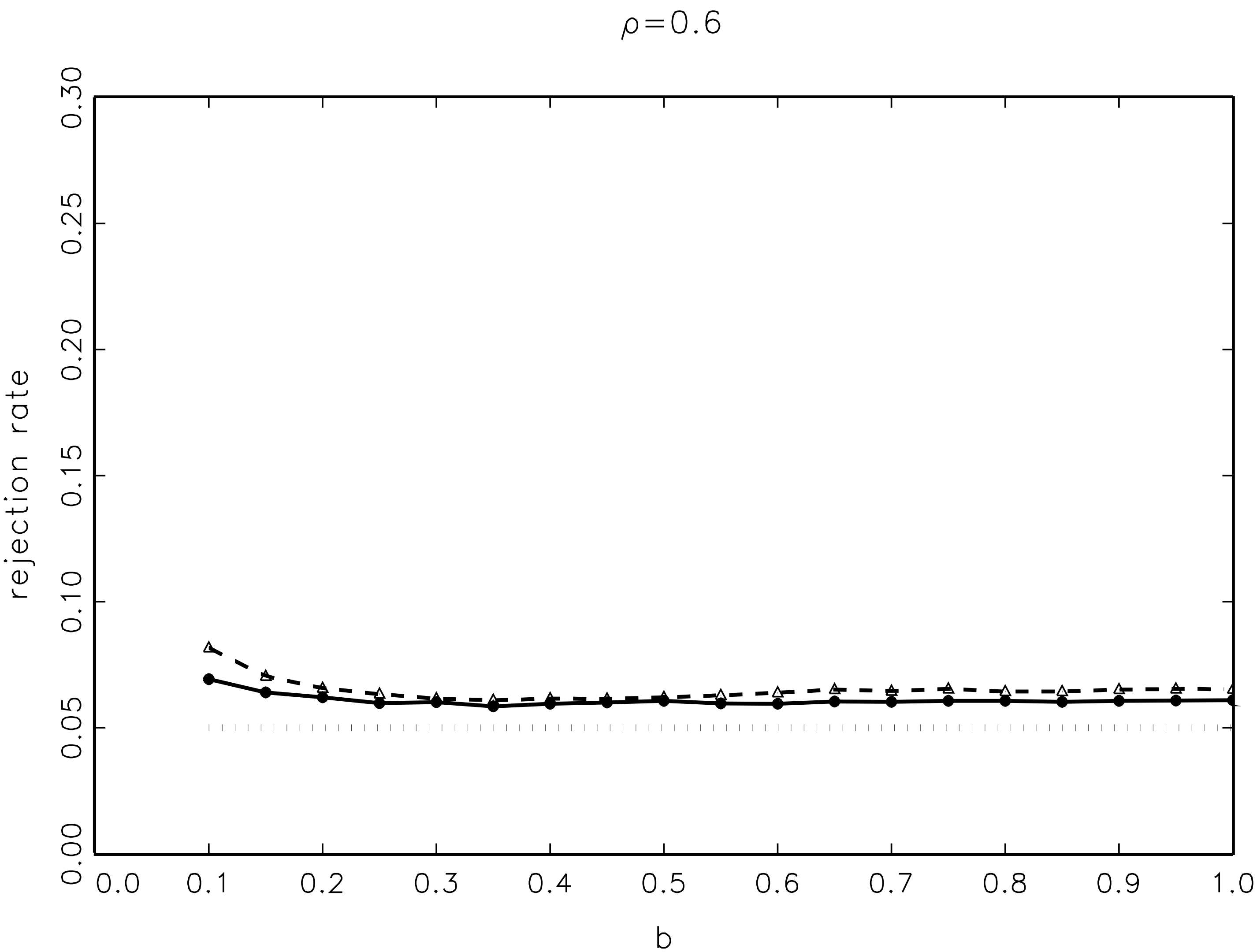


Figure 2.89: (cont'd)

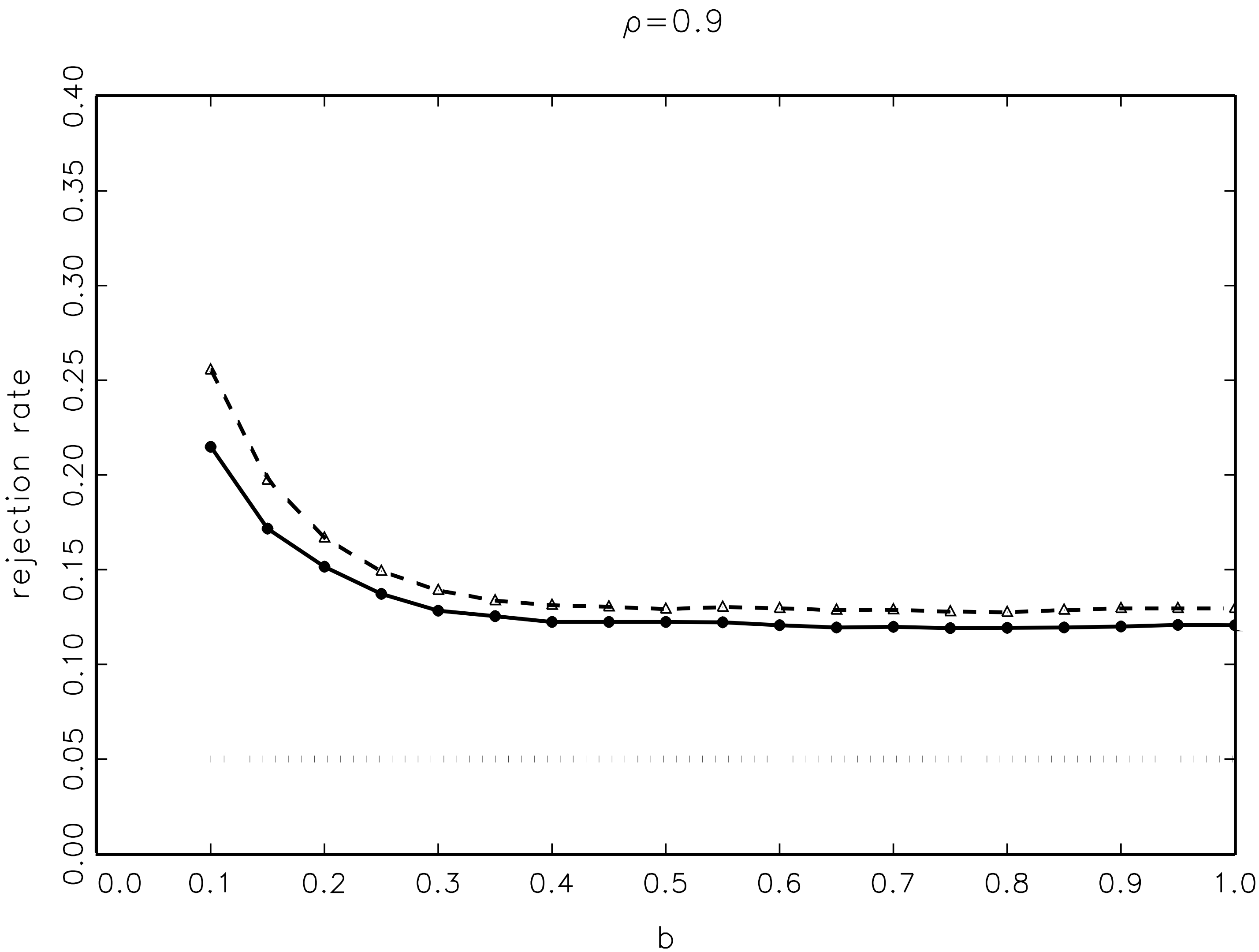


Figure 2.90: AM and ES - Random Bernoulli ($p = 0.5$), Bartlett, $T = 100$

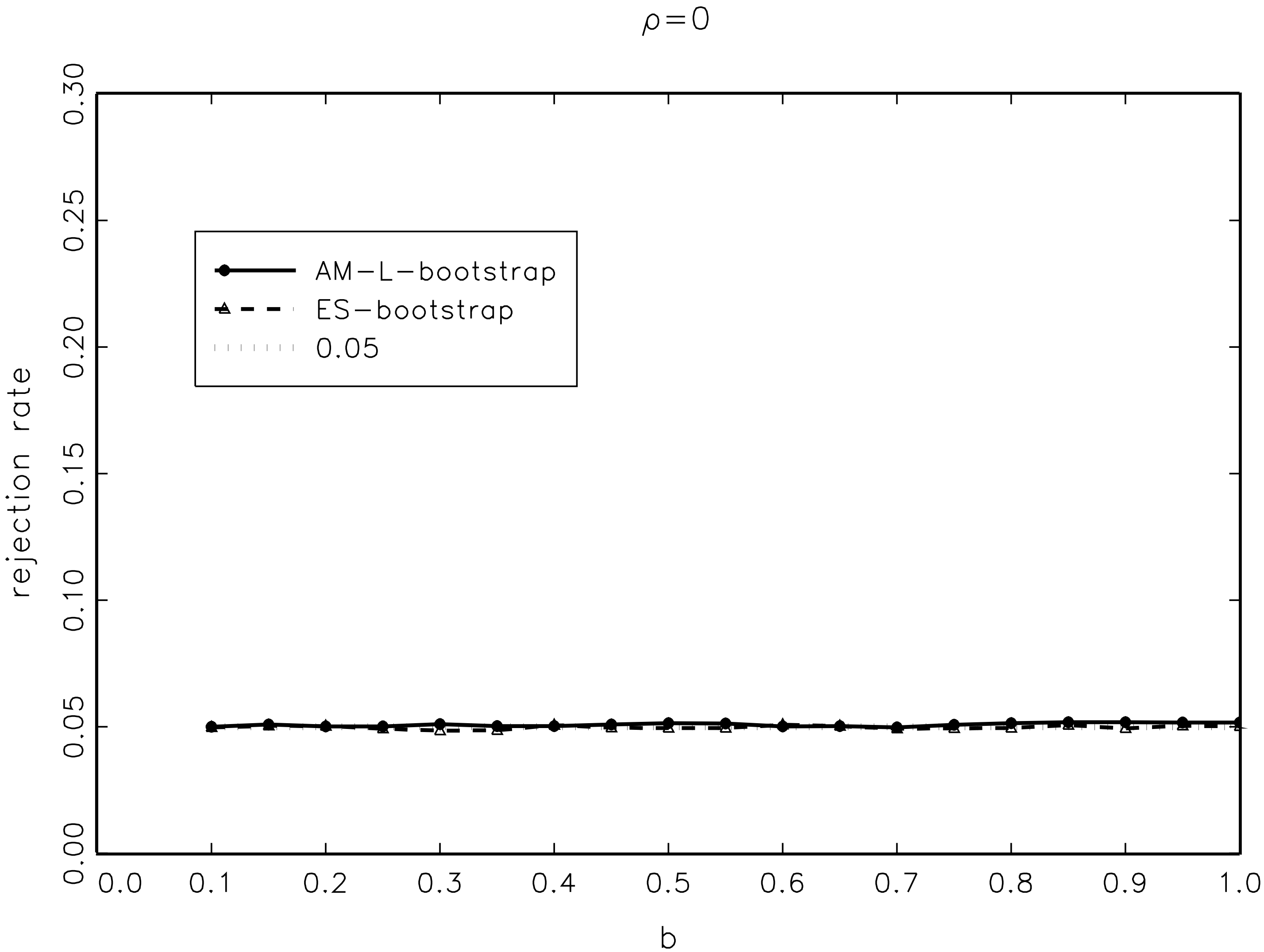


Figure 2.90: (cont'd)

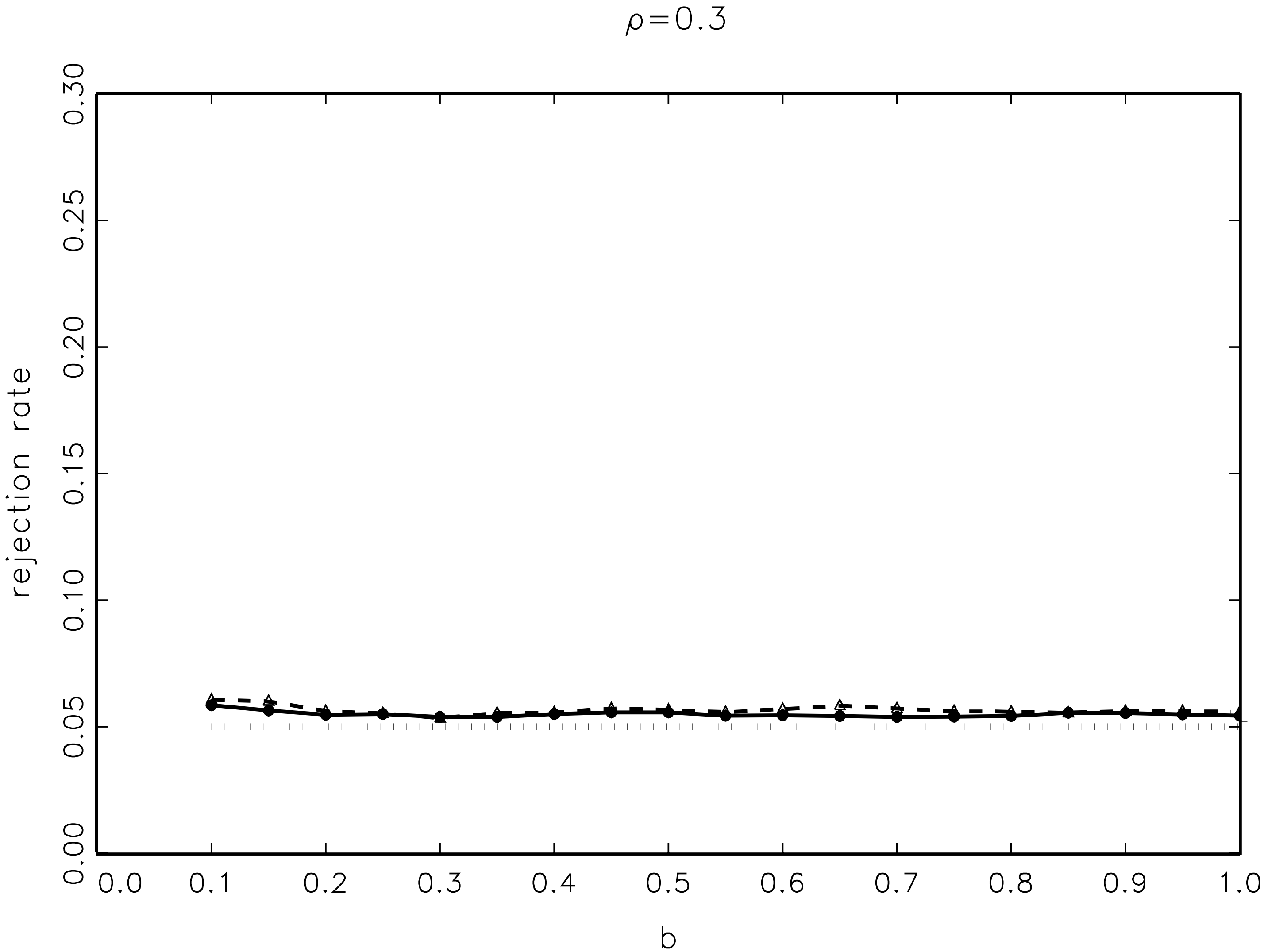


Figure 2.90: (cont'd)

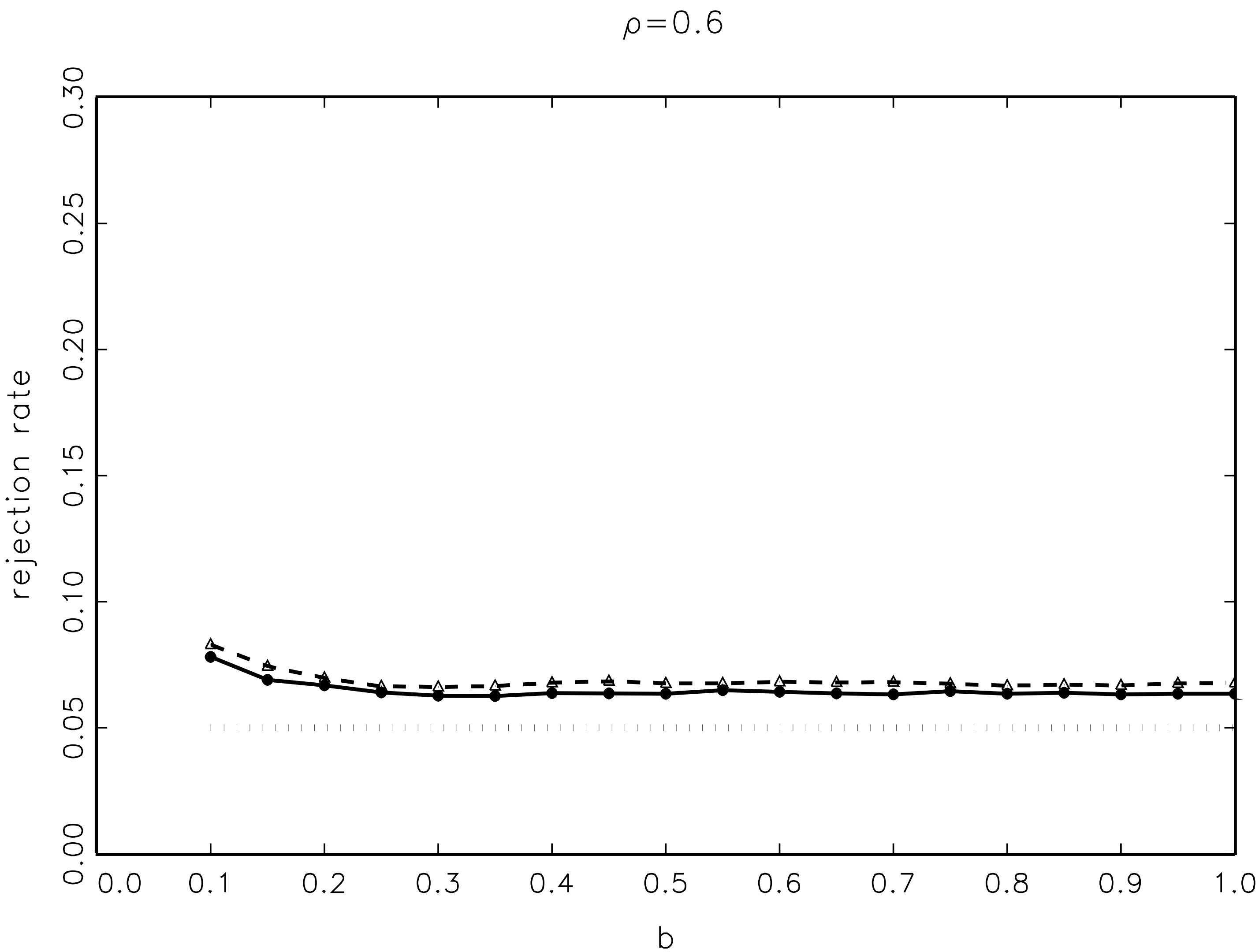


Figure 2.90: (cont'd)

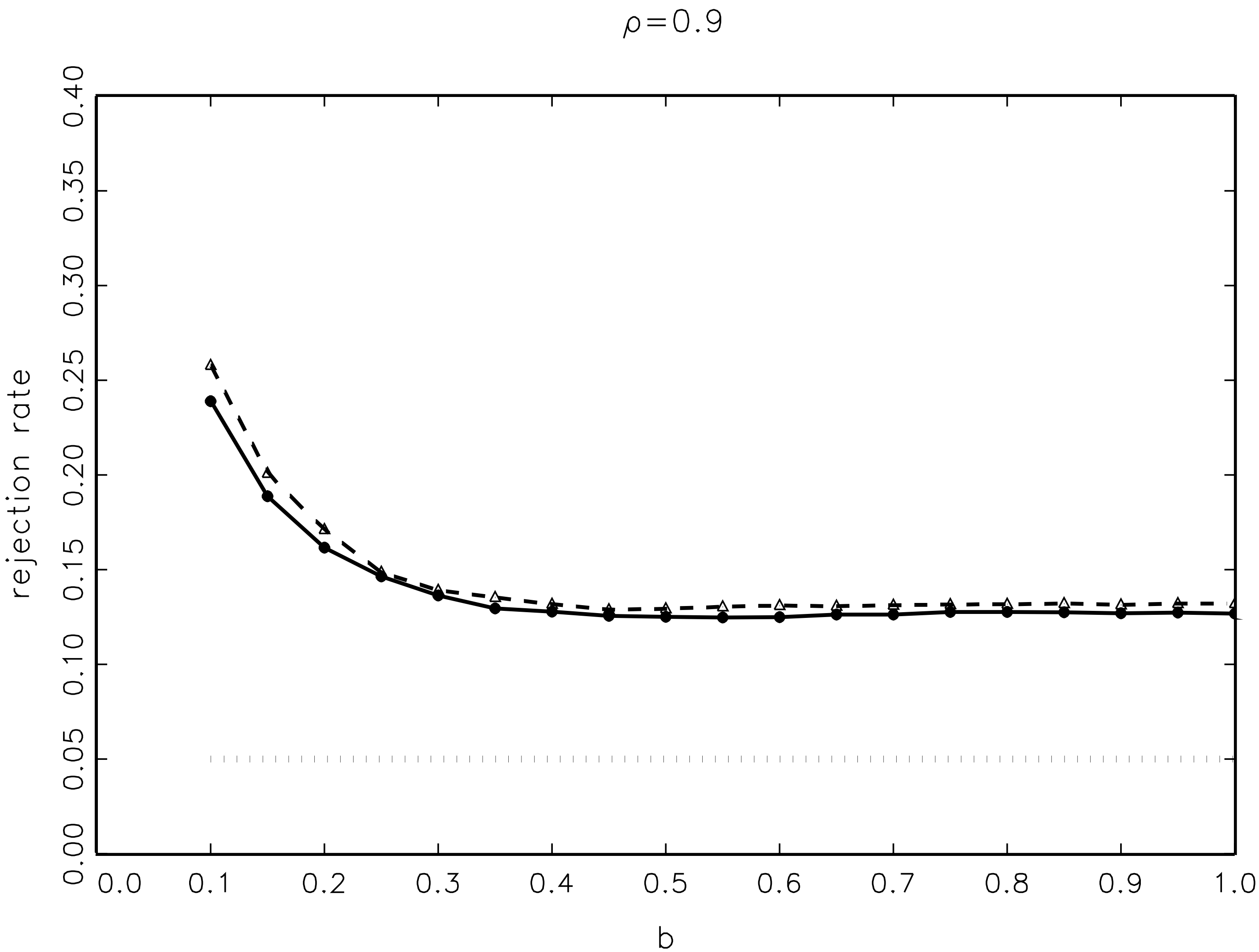


Figure 2.91: AM and ES - Random Bernoulli ($p = 0.7$), Bartlett, $T = 100$

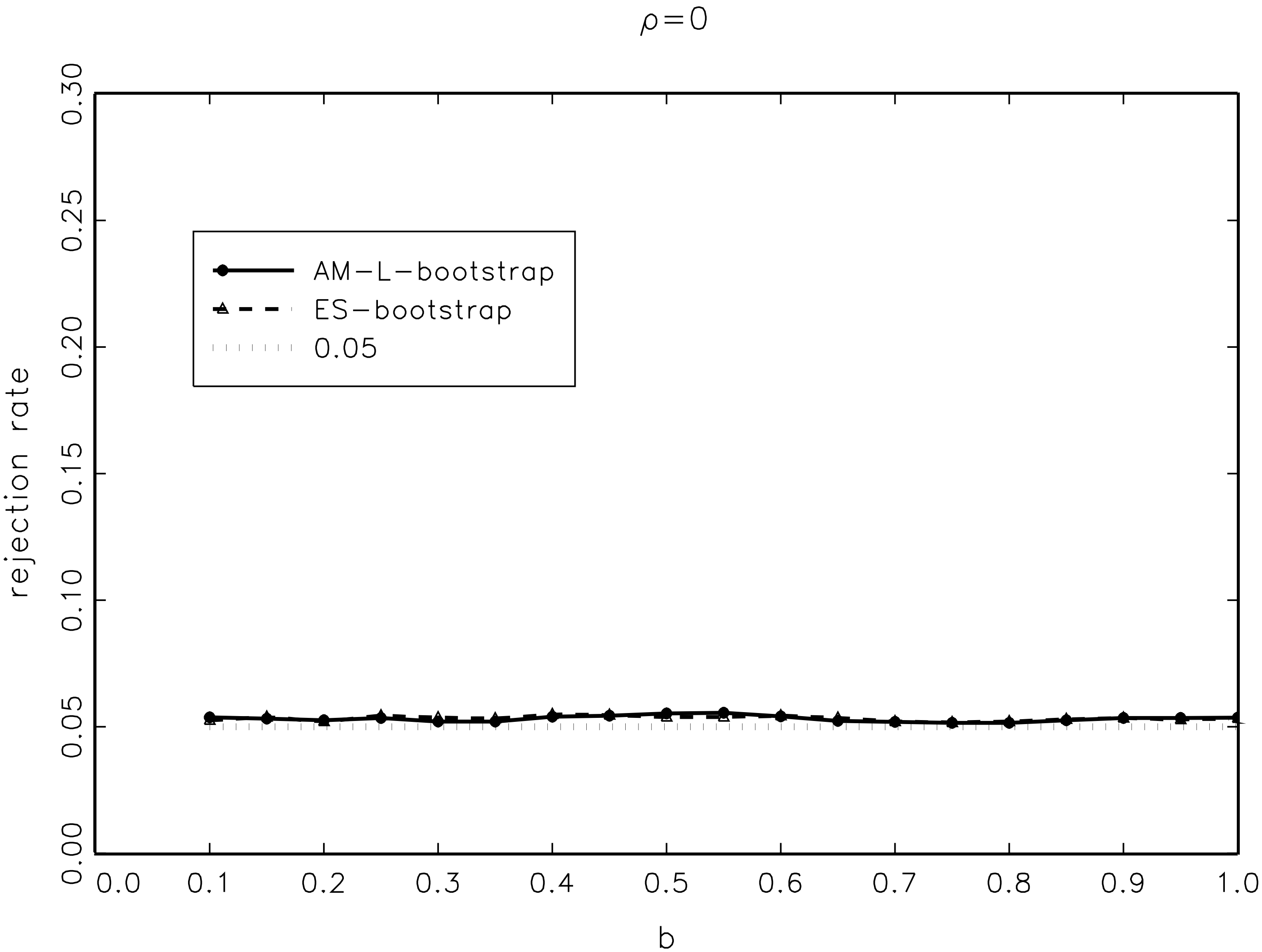


Figure 2.91: (cont'd)

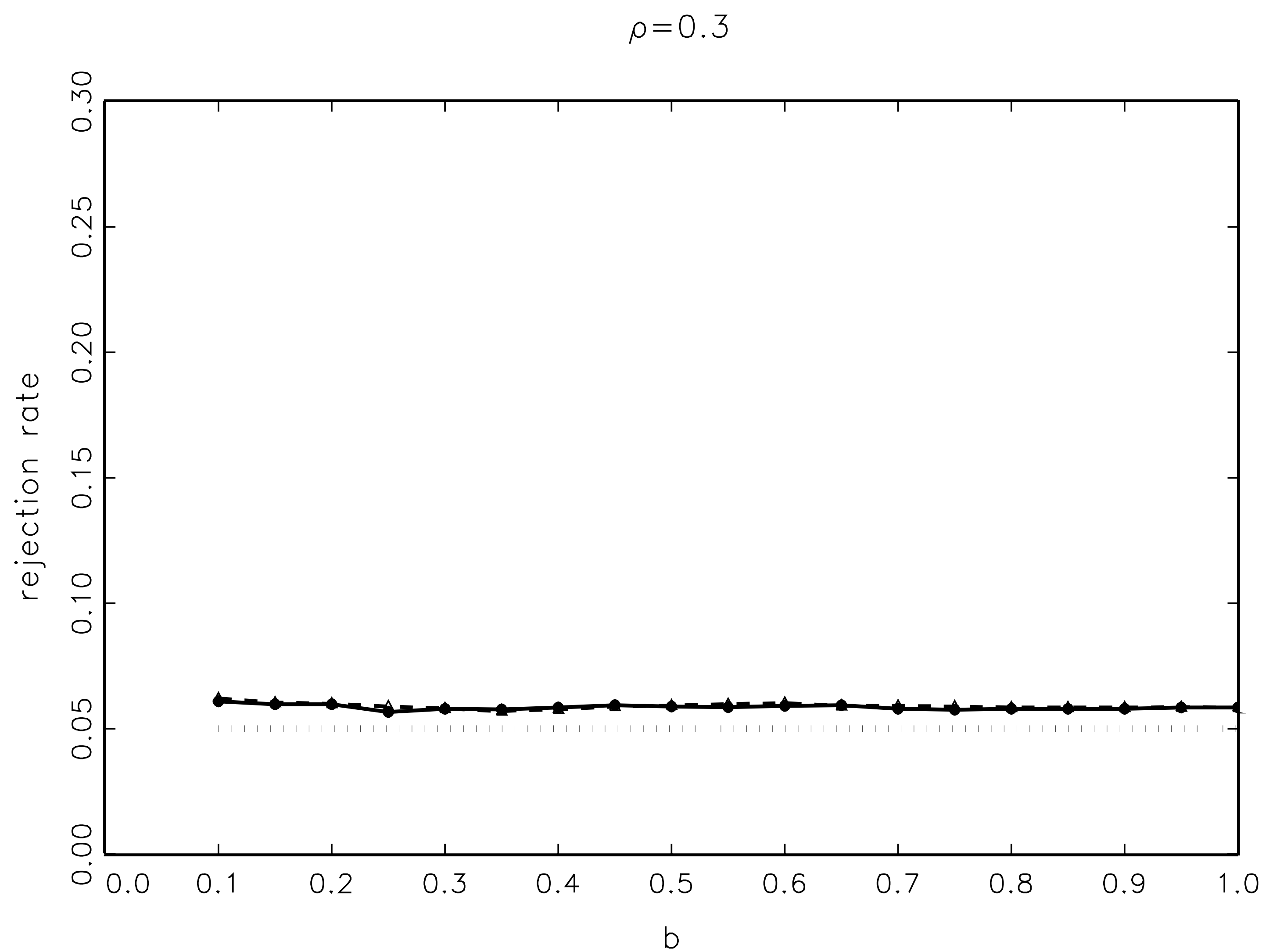


Figure 2.91: (cont'd)

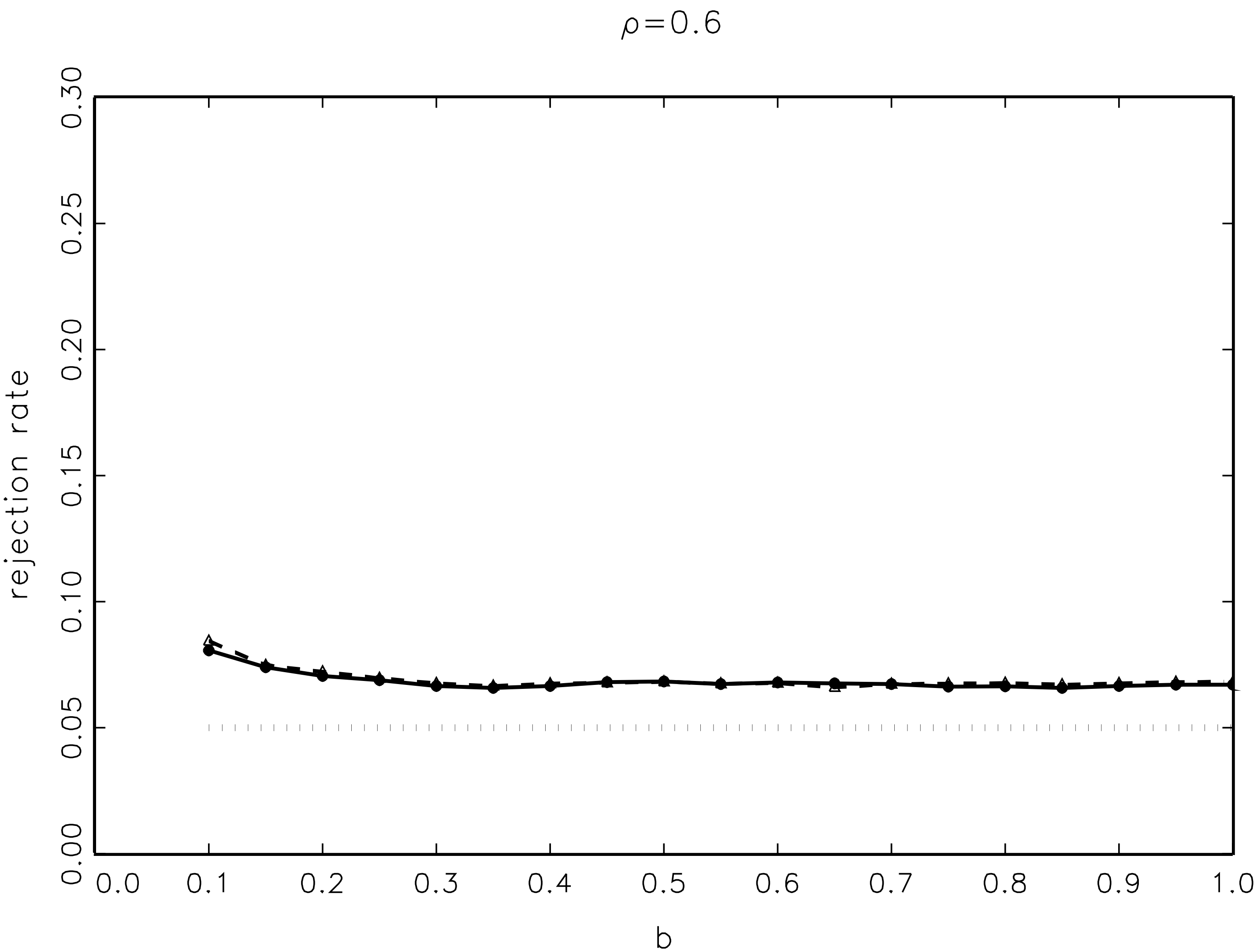


Figure 2.91: (cont'd)

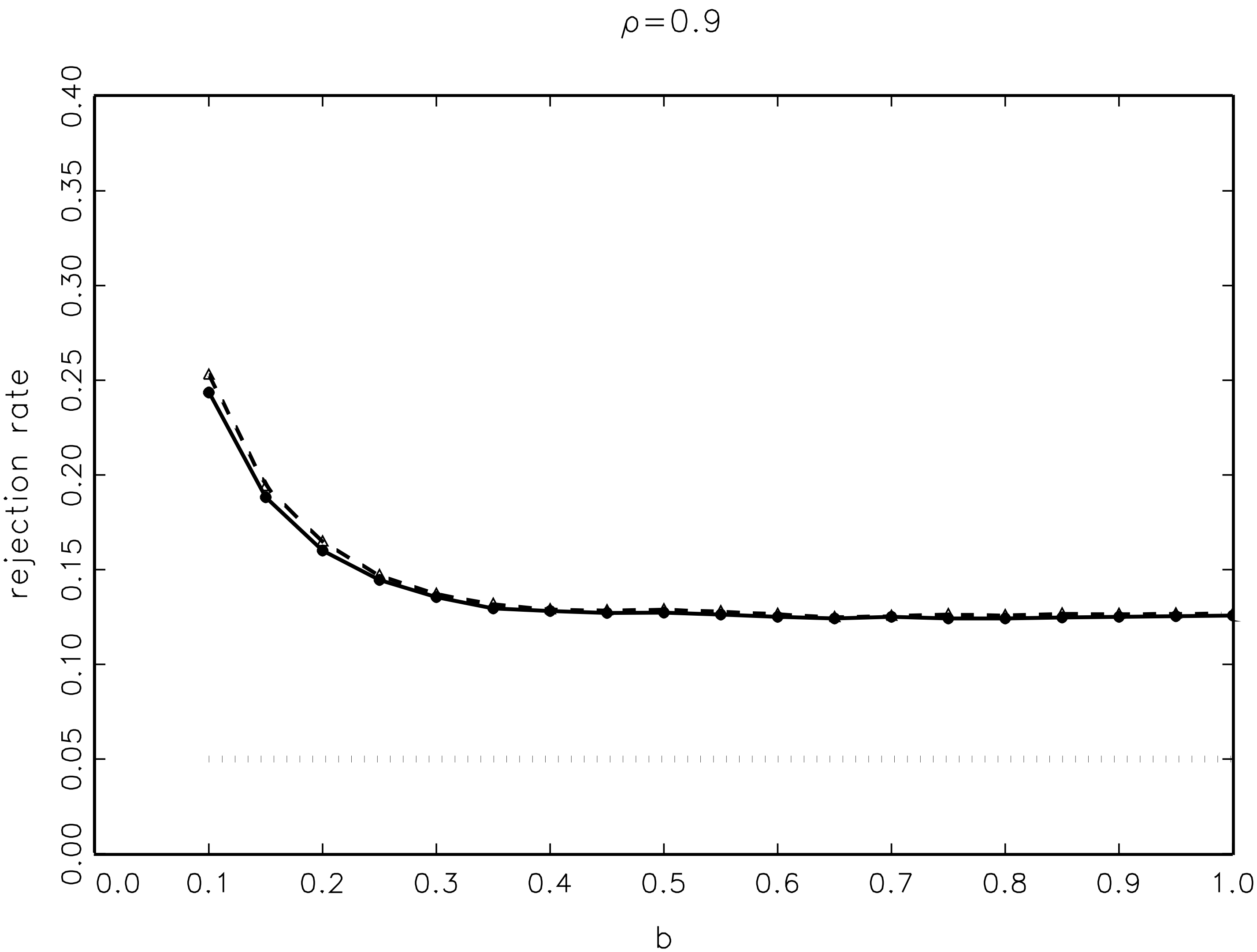


Figure 2.92: AM and ES - Random Bernoulli ($p = 0.3$), Bartlett, $T = 200$

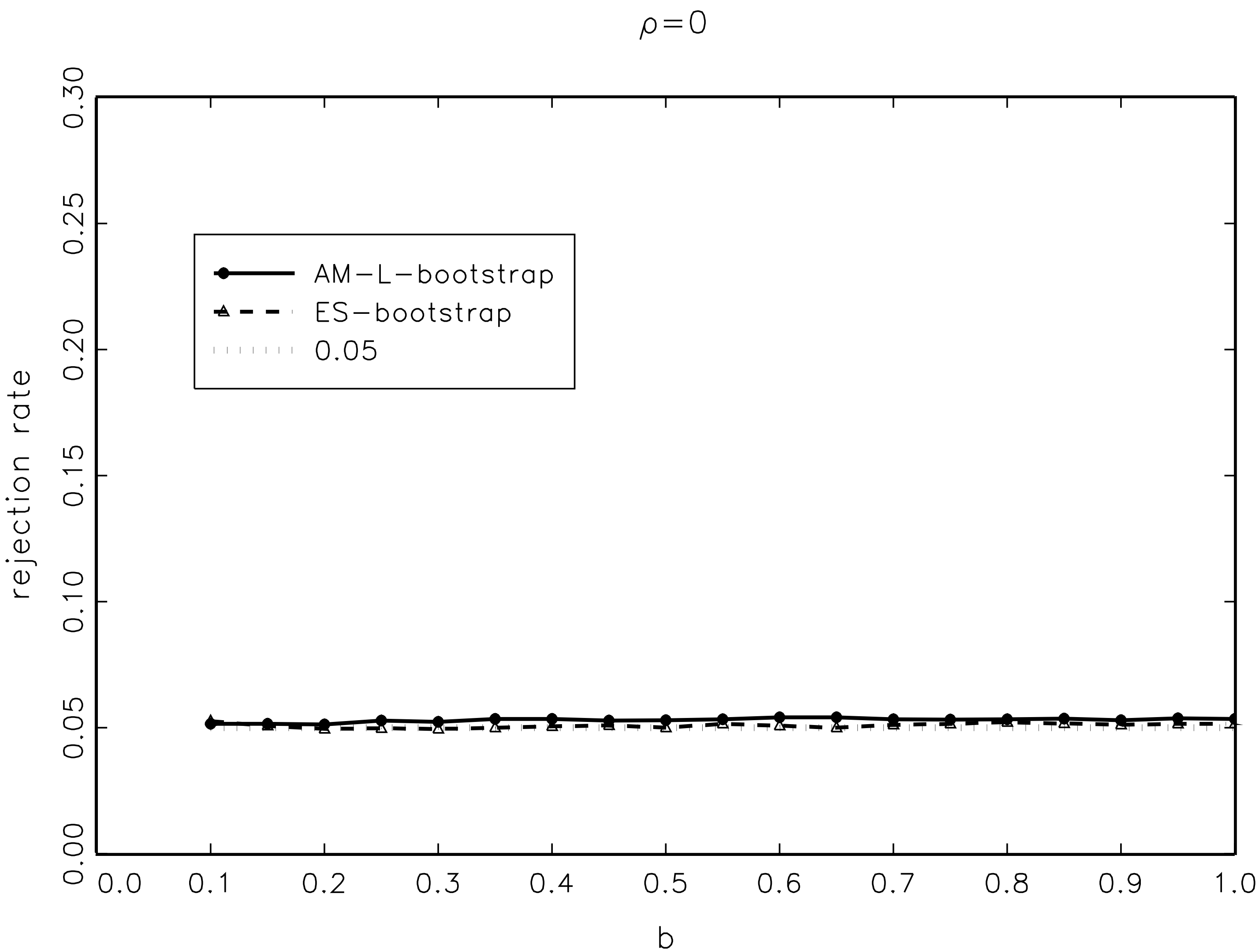


Figure 2.92: (cont'd)

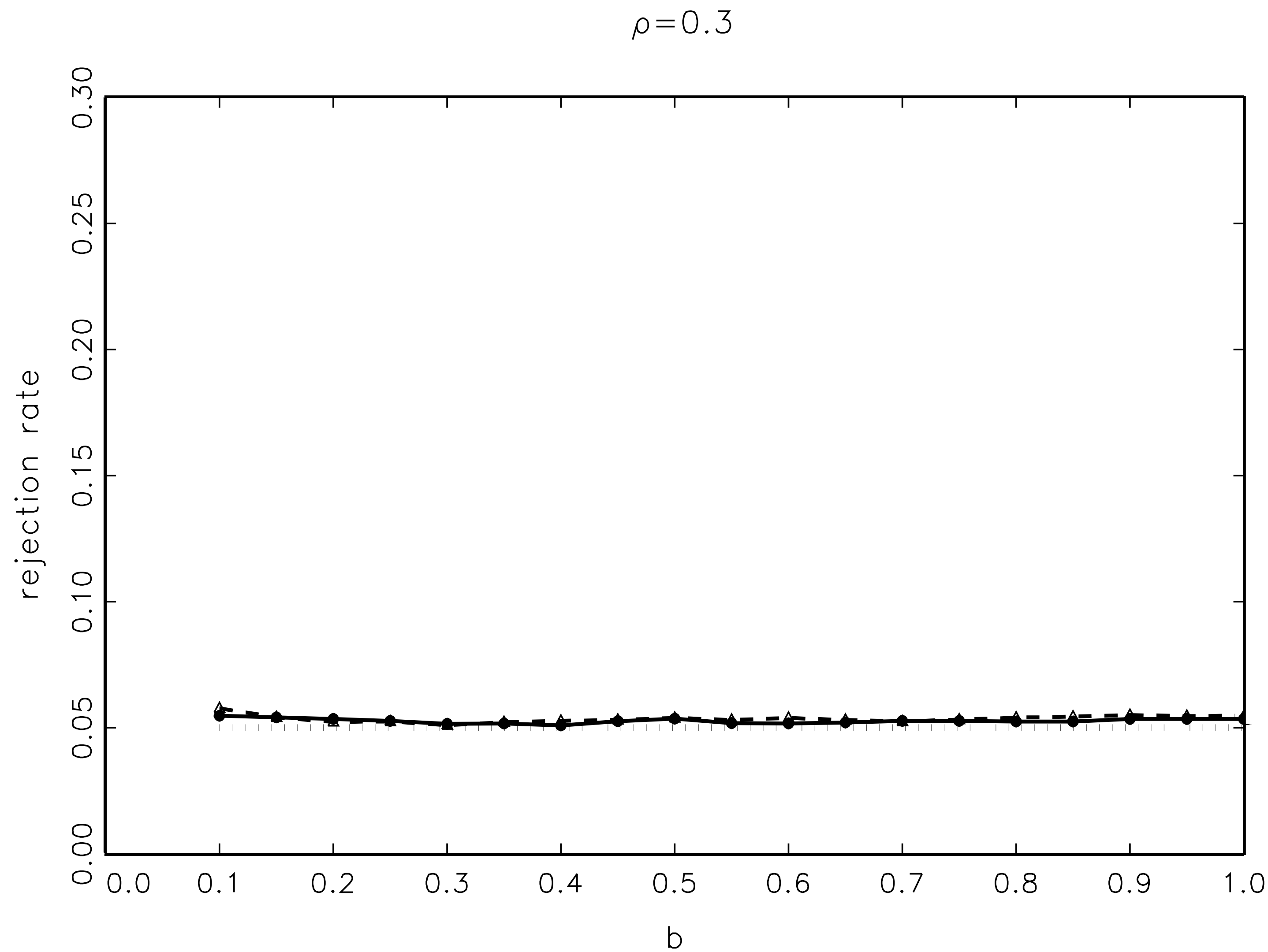


Figure 2.92: (cont'd)

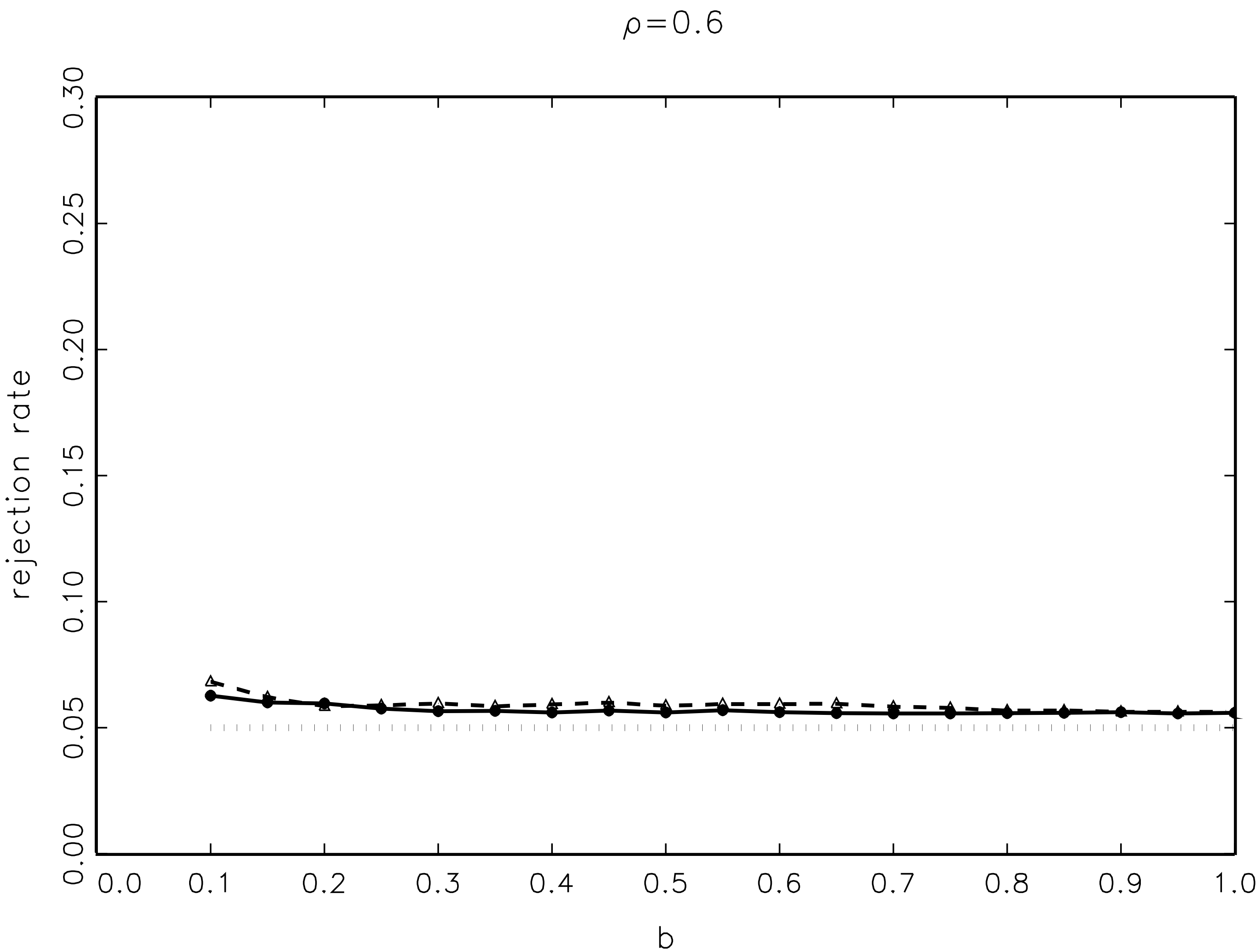


Figure 2.92: (cont'd)

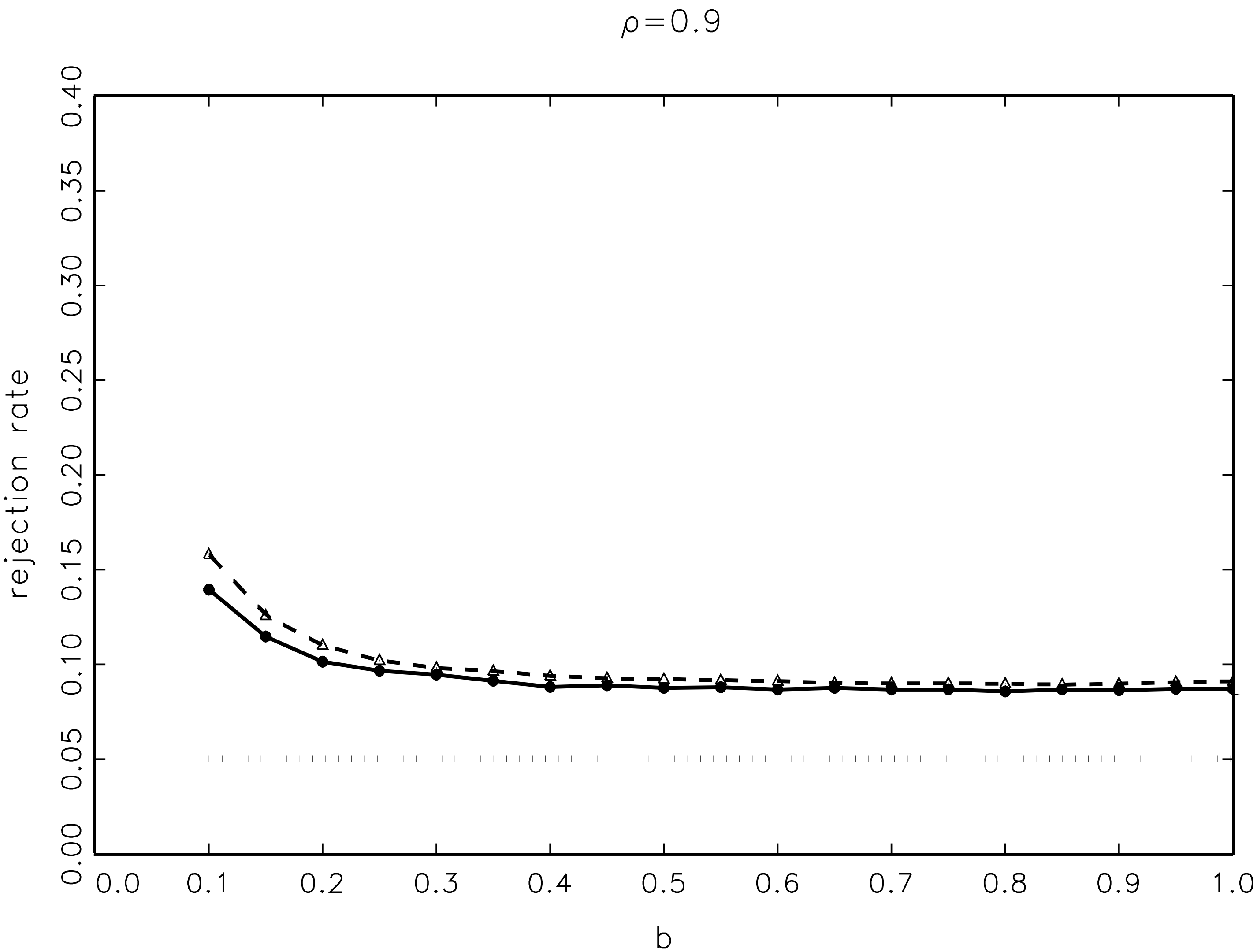


Figure 2.93: AM and ES - Random Bernoulli ($p = 0.5$), Bartlett, $T = 200$

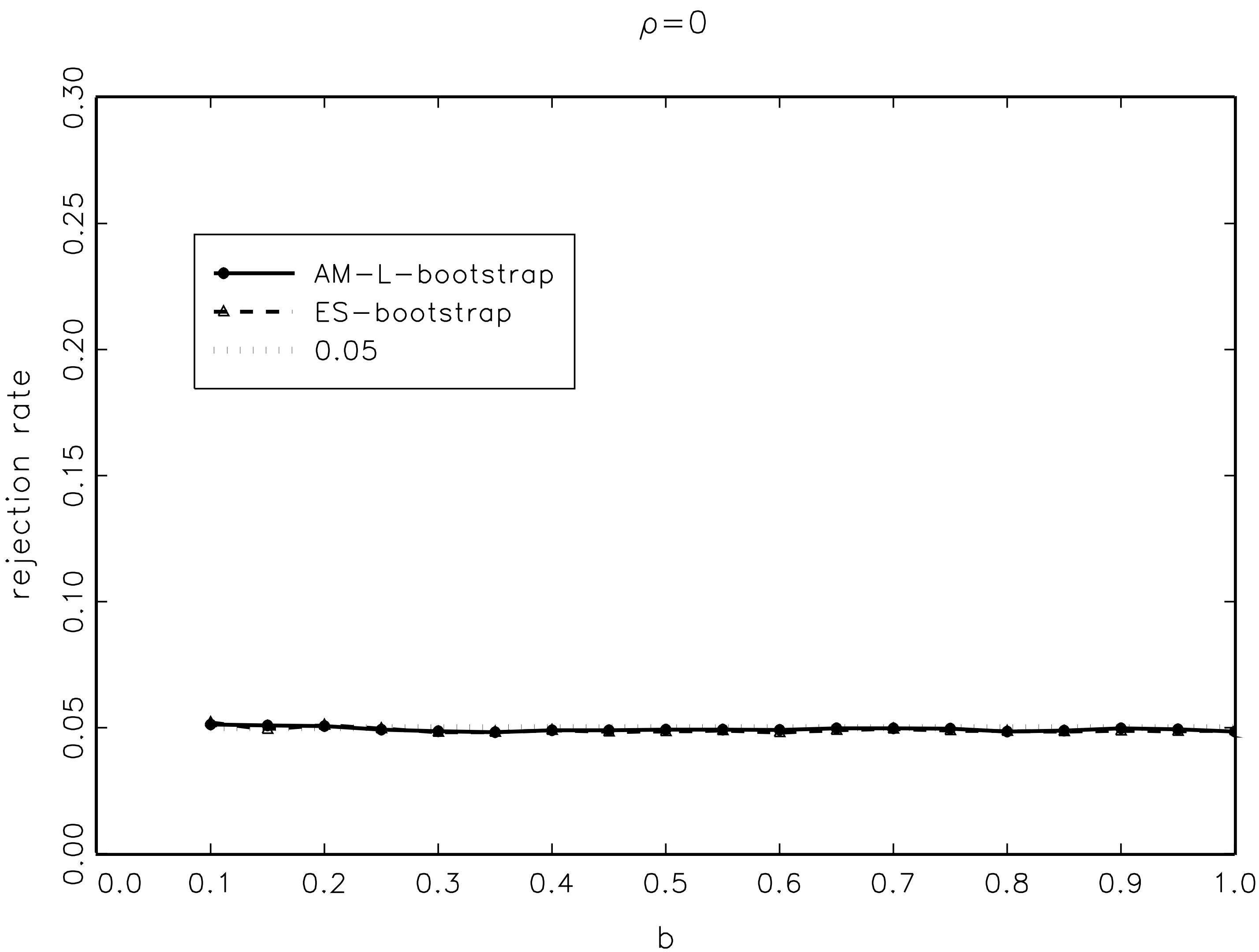


Figure 2.93: (cont'd)

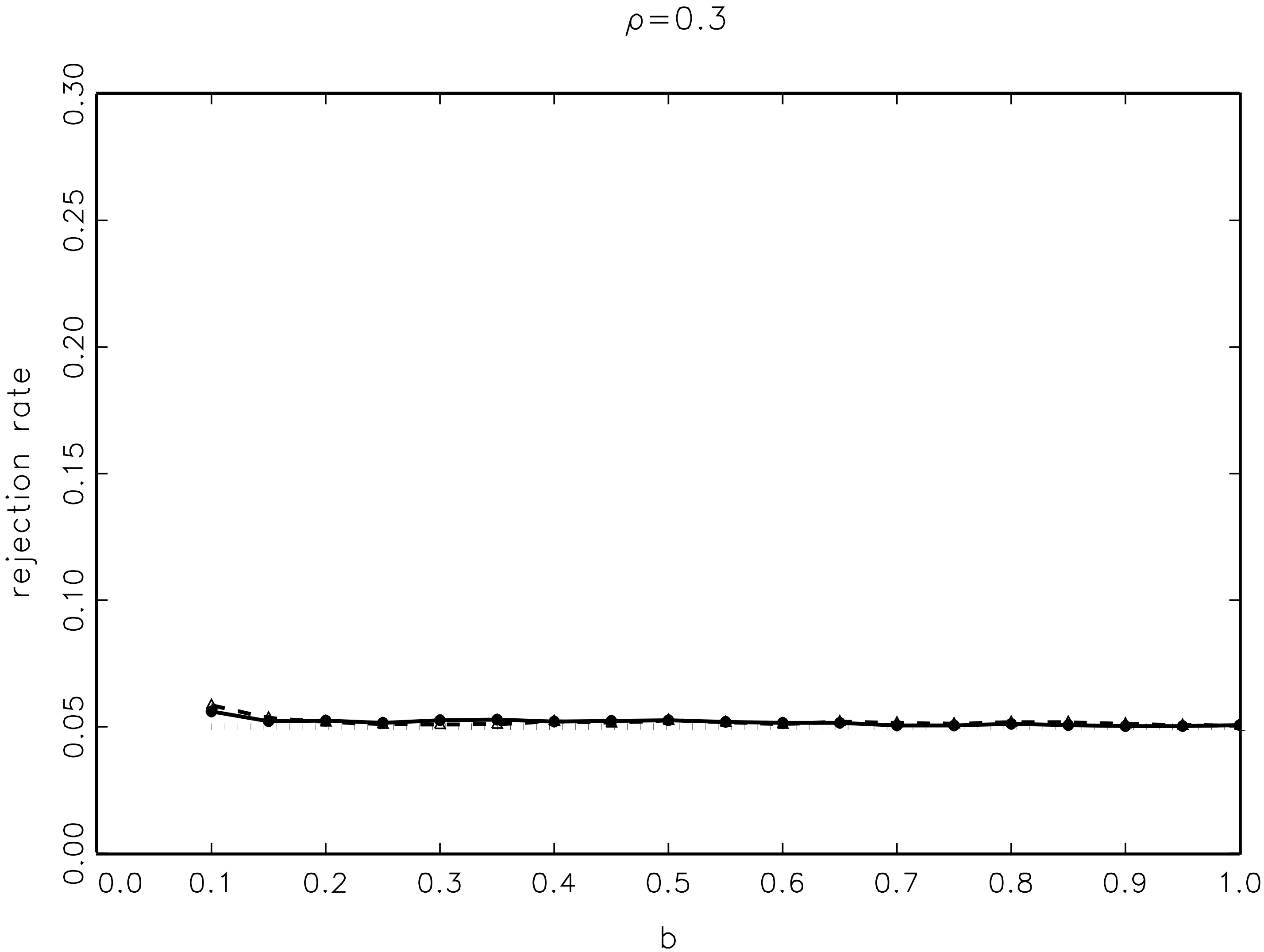


Figure 2.93: (cont'd)

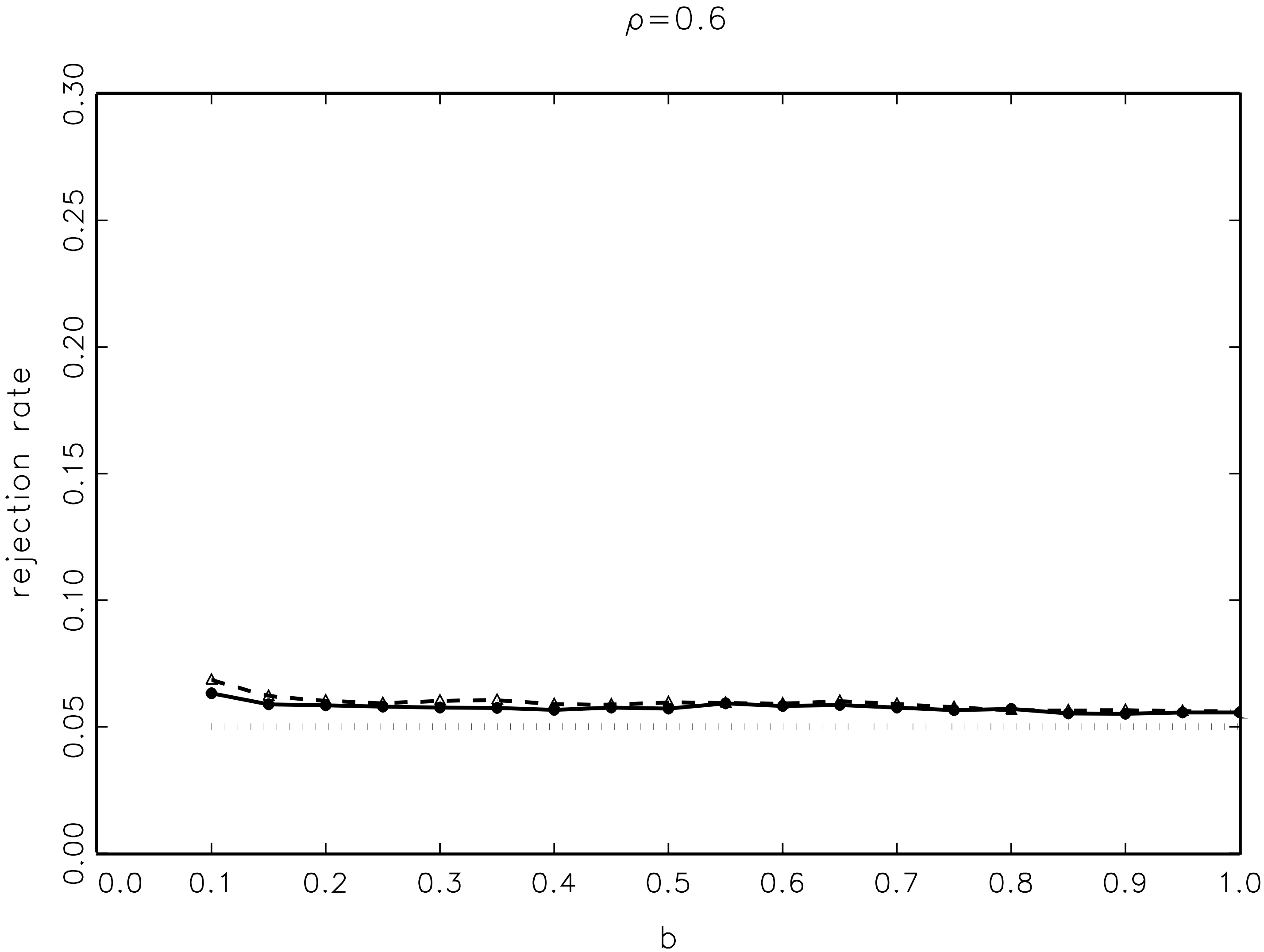


Figure 2.93: (cont'd)

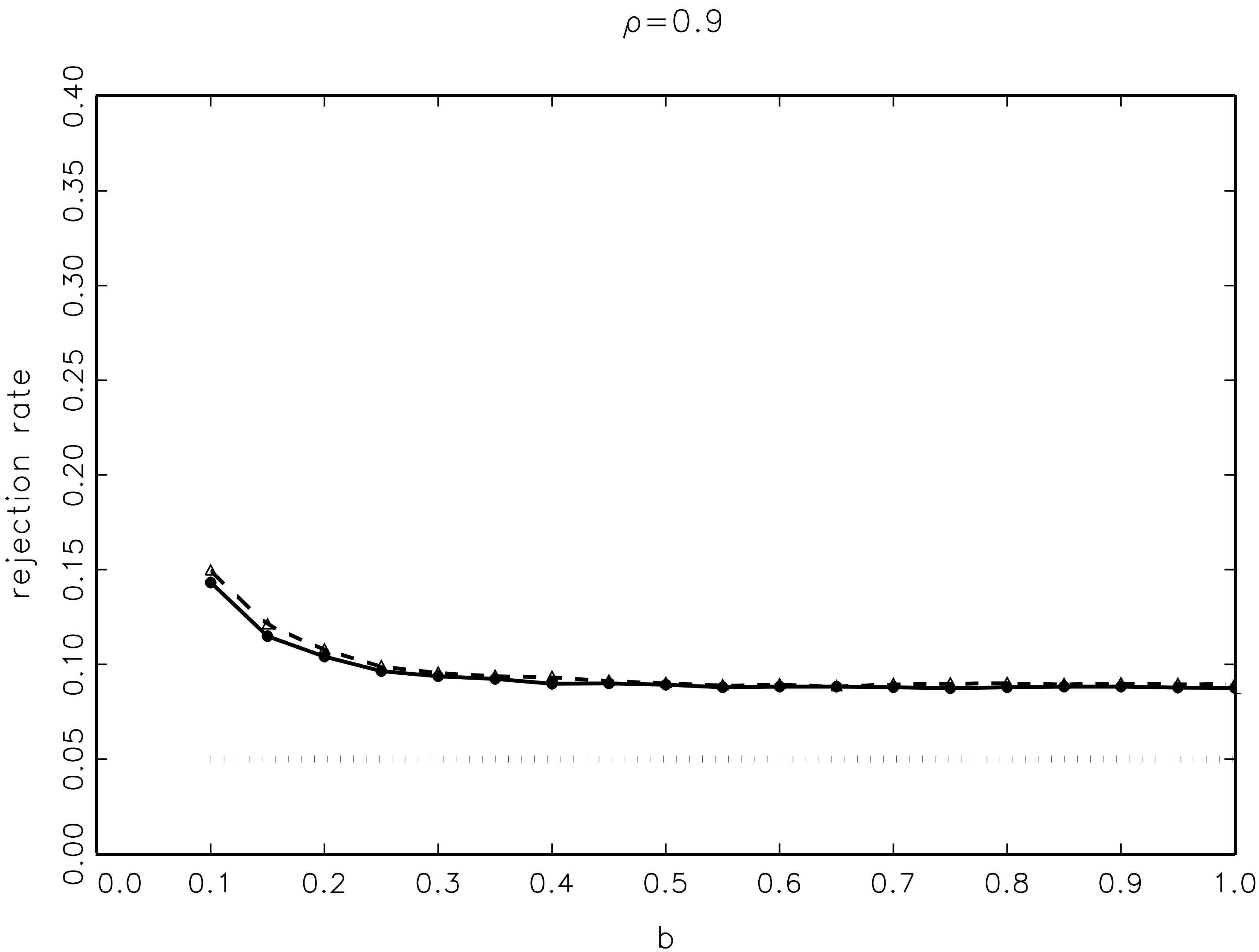


Figure 2.94: AM and ES - Random Bernoulli ($p = 0.7$), Bartlett, $T = 200$

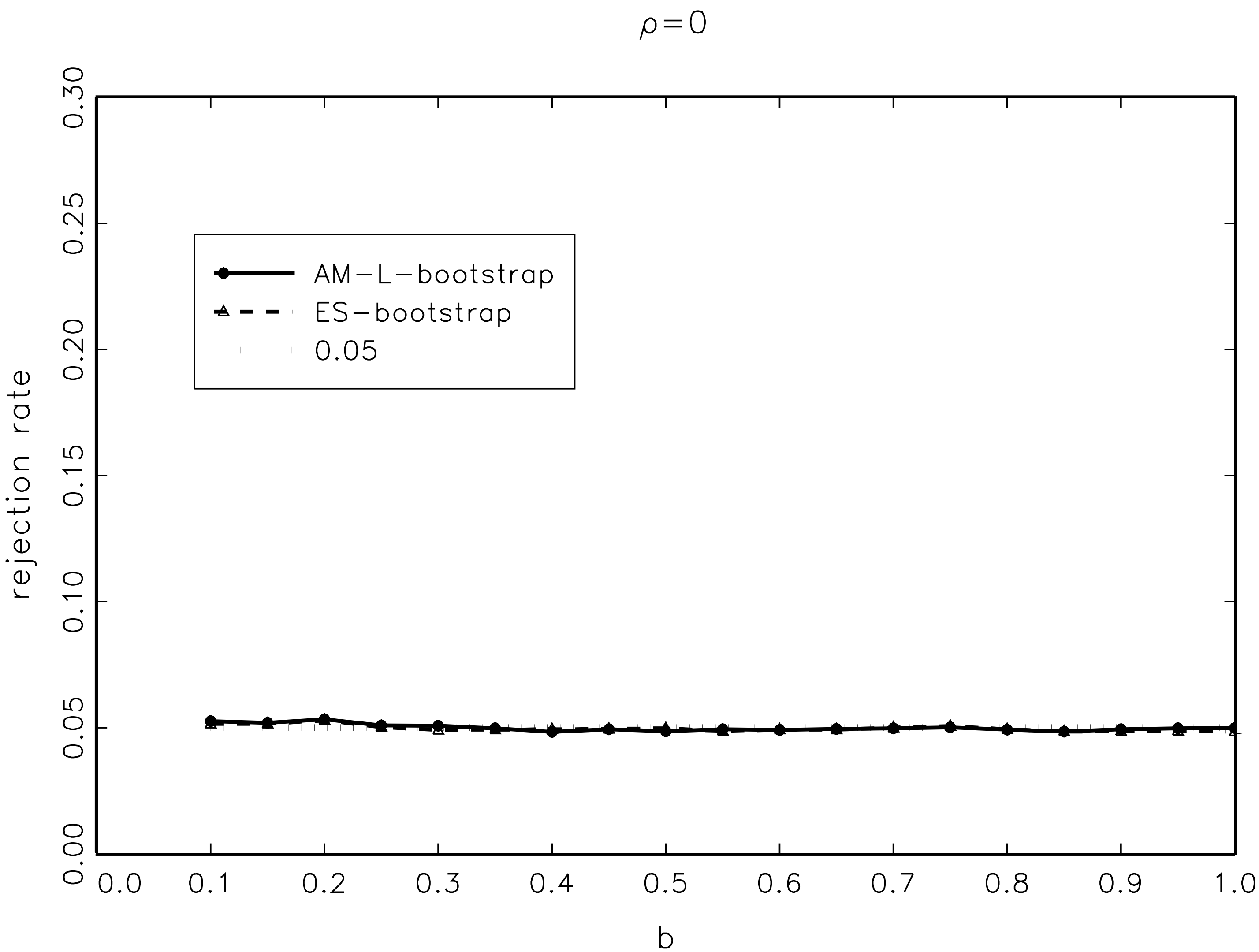


Figure 2.94: (cont'd)

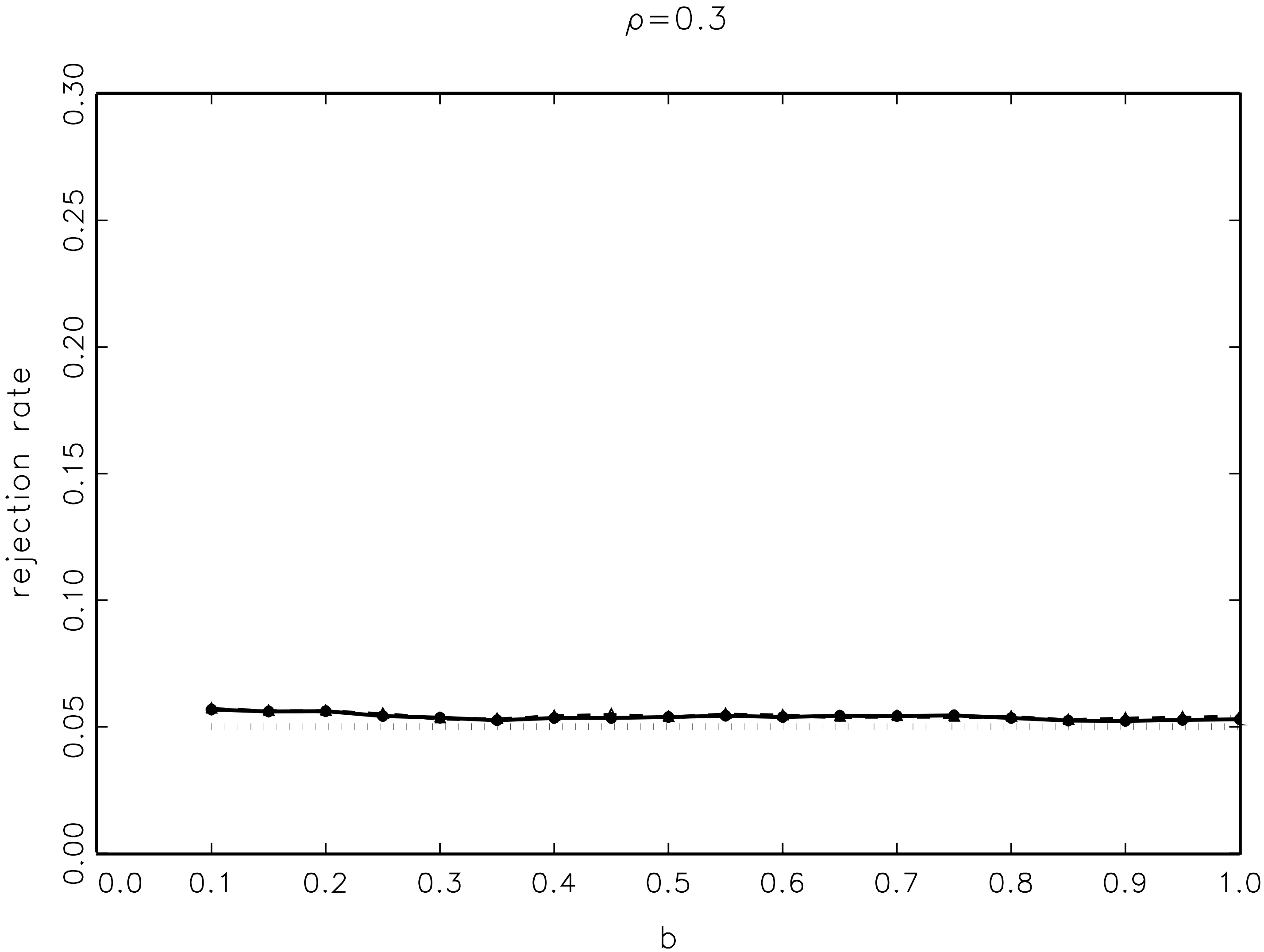


Figure 2.94: (cont'd)

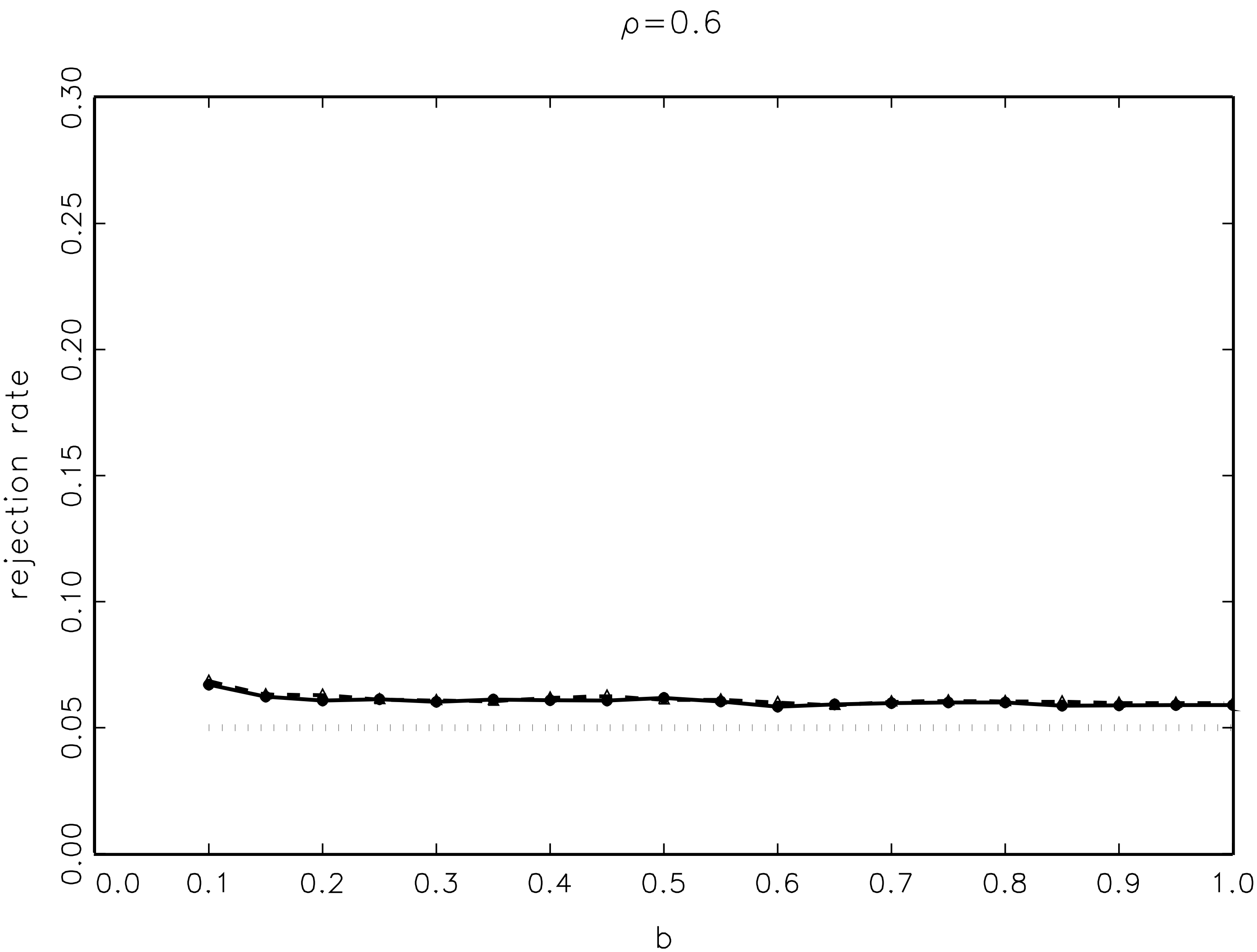
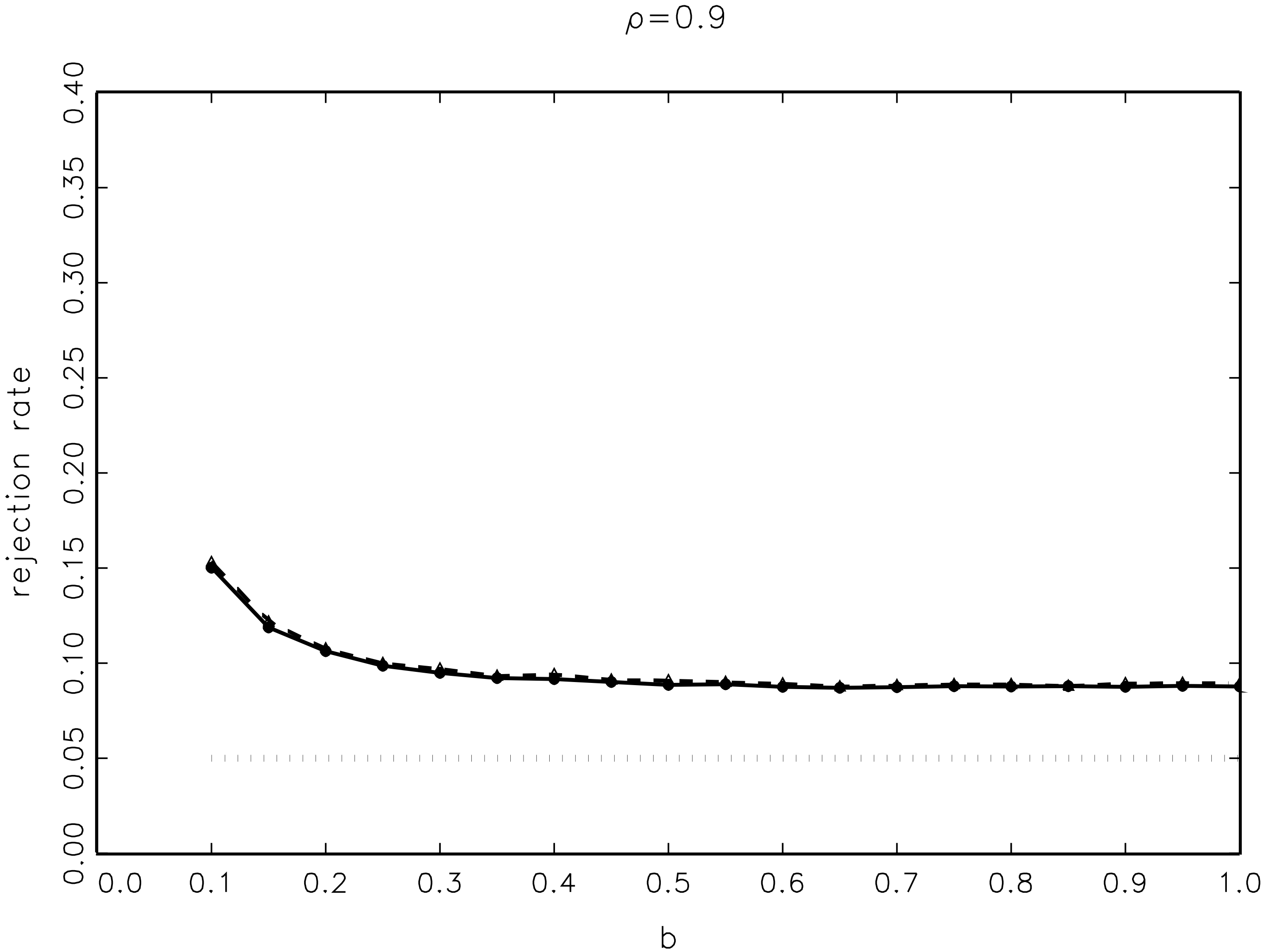


Figure 2.94: (cont'd)



CHAPTER 3

INFERENCE IN TIME SERIES MODELS USING SMOOTHED CLUSTERED STANDARD ERRORS

3.1 INTRODUCTION

This chapter proposes a long run variance estimator for conducting inference in time series regression models that combines the traditional nonparametric kernel approach, Newey and West (1987) and Andrews (1991), with a cluster approach, Bester et al. (2011). The basic idea is to divide the time periods into non-overlapping clusters with equal numbers of observations. The long run variance estimator is constructed by first aggregating within clusters and then kernel smoothing across clusters. This approach is similar in spirit to the approach proposed by Driscoll and Kraay (1998) in panel settings. Under the assumption that the time series data is weakly dependent and covariance stationary, we develop an asymptotic theory for test statistics based on this "smoothed clustered" long run variance estimator. We derive asymptotic results holding the number of clusters fixed and also treating the clusters as increasing with the sample size. Our large number of clusters results are closely linked to the fixed-b results obtained by Vogelsang (2012) for Driscoll and Kraay (1998) statistics in panel settings. We show that in the large number of clusters setting robust test statistics follow the standard fixed-b limits obtained by Kiefer and Vogelsang (2005) assuming that the kernel bandwidth is treated as a fixed proportion of the sample size. In contrast, for the fixed number of clusters case, we obtain a different asymptotic limit. While one might expect the relative accuracy of the two asymptotic approximations to depend on the number of clusters relative to the sample size, we find in a simulation study that the “fixed number of cluster” asymptotic approximation works well whether the number of clusters is small or large. The simulations also suggest that the naive *i.i.d.* bootstrap mimics the fixed number of clusters critical values.

The motivation for clustering before kernel smoothing is as follows. Averaging within clusters works well even when serial correlation is relatively strong within clusters. Given our weak dependence and covariance stationarity assumption, within cluster averages will be asymptotically independent. But, in finite samples the cluster averages will be correlated and kernel smoothing can help to reduce finite sample over-rejection problems. In fact, we find in our finite sample simulations that clustering before kernel smoothing does reduce over-rejections caused by strong serial correlation without a great cost in terms of power.

The rest of the chapter is organized as follows. In the next section the model is given and the long run variance is defined. Section 3.3 lays out the inference problem and provides asymptotic results for test statistics based on the smoothed clustered long run variance estimator. Section 3.4 explores the finite sample properties of the test statistics in a simple location model. Proofs are given in Appendix E. The case where the number of groups does not evenly divide the sample is discussed in Appendix F.

3.2 MODEL AND CLUSTERED SMOOTHED STANDARD ERRORS

Consider the time series regression model,

$$y_t = x_t'\beta + u_t, t = 1, \ldots, T,$$

where β is a $(k \times 1)$ vector of regression parameters, x_t is a $(k \times 1)$ vector of regressors, and u_t is a mean zero error process. The ordinary least squares (OLS) estimator of β is

$$\hat{\beta} = \left(\sum_{t=1}^{T} x_t x_t' \right)^{-1} \sum_{t=1}^{T} x_t y_t.$$

Divide the time periods into G contiguous, non-overlapping groups of equal size n_G such that $T = n_G G$.[1] Rewriting the OLS estimator, $\hat{\beta}$, using group notation as

$$\hat{\beta} = \left(\sum_{g=1}^{G} \sum_{t=(g-1)n_G+1}^{g\,n_G} x_t x_t' \right)^{-1} \sum_{g=1}^{G} \sum_{t=(g-1)n_G+1}^{g\,n_G} x_t y_t. \tag{3.1}$$

Conceptually, this way of rewriting $\hat{\beta}$ can be viewed as the outcome of rearranging the data into G time periods with n_G "cross-section" units per time period resulting in an artificial panel data structure. From this artificial panel perspective $\hat{\beta}$ in (3.1) is exactly the pooled OLS estimator of β. Plugging in for y_t gives

$$\begin{aligned} \hat{\beta} - \beta &= \left(\sum_{g=1}^{G} \sum_{t=(g-1)n_G+1}^{gn_G} x_t x_t' \right)^{-1} \sum_{g=1}^{G} \sum_{t=(g-1)n_G+1}^{gn_G} x_t u_t \\ &= \left(\sum_{g=1}^{G} \sum_{t=(g-1)n_G+1}^{gn_G} x_t x_t' \right)^{-1} \sum_{g=1}^{G} \sum_{t=(g-1)n_G+1}^{gn_G} v_t, \end{aligned} \tag{3.2}$$

where $v_t = x_t u_t$. Using the panel perspective, we can directly apply the variance-covariance matrix estimator proposed by Driscoll and Kraay (1998) as follows. Let $\hat{v}_t = x_t \hat{u}_t$, where $\hat{u}_t = y_t - x_t' \hat{\beta}$ are the OLS residuals. Define

$$\widehat{\overline{v}}_g = \sum_{t=(g-1)n_G+1}^{gn_G} \hat{v}_t, \quad g = 1, \ldots, G,$$

which is the sum of $\hat{v}_t$ within group g. Compute the nonparametric kernel HAC estimator using $\widehat{\overline{v}}_g$ for $g = 1, 2, ..., G$ as

$$\widehat{\overline{\Omega}} = \widehat{\overline{\Gamma}}_0 + \sum_{j=1}^{G-1} k\left(\frac{j}{M}\right) \left(\widehat{\overline{\Gamma}}_j + \widehat{\overline{\Gamma}}_j'\right),$$

where $\widehat{\overline{\Gamma}}_j = G^{-1} \sum_{g=j+1}^{G} \widehat{\overline{v}}_g \widehat{\overline{v}}_{g-j}'$ are the sample autocovariaces of $\widehat{\overline{v}}_g$. Here, $k(x)$ is a kernel function such that $k(x) = k(-x)$, $k(0) = 1$, $|k(x)| \leq 1$, $k(x)$ is continuous at $x = 0$,

[1] Cases where G does not evenly divide T is easily handled but the notation is more tedious. See Appendix F.

$\int_{-\infty}^{\infty} k^2(x) < \infty$, and M is the bandwidth parameter. Using well known algebra we can rewrite $\widehat{\widehat{\Omega}}$ as

$$\widehat{\widehat{\Omega}} = \frac{1}{G} \sum_{g=1}^{G} \sum_{h=1}^{G} k\left(\frac{|g-h|}{M}\right) \widehat{\widehat{v}}_g \widehat{\widehat{v}}_h',$$

which we call the "cluster then HAC" variance-covariance matrix estimator or CHAC for short. Notice that the CHAC estimator gives full weight for observations within clusters, a feature that the usual nonparametric kernel HAC estimator does not have. Smoothing across clusters accounts for finite sample serial correlation across clusters which is a generalization of the cluster estimator proposed by Bester, Conley, and Hansen (2011). Note that the Bester, Conley, and Hansen (2011) estimator is a special case of $\widehat{\widehat{\Omega}}$ obtained when $\widehat{\widehat{\Omega}} = \widehat{\widehat{\Gamma}}_0$, i.e. when zero weights is imposed across clusters. Also note that when $G = T$ and $n_G = 1$, the CHAC estimator becomes the usual kernel HAC estimator. Therefore, the CHAC estimator is more general and nests the traditional approach and the time series cluster approach.

Using $\widehat{\widehat{\Omega}}$ as the middle term of a sandwich variance for $\hat{\beta}$, we obtain the sample variance-covariance matrix

$$\widehat{V}_{\mathrm{CHAC}} = G\left(\sum_{g=1}^{G} \sum_{t=(g-1)n_G+1}^{g\,n_G} x_t x_t'\right)^{-1} \widehat{\widehat{\Omega}} \left(\sum_{g=1}^{G} \sum_{t=(g-1)n_G+1}^{g\,n_G} x_t x_t'\right)^{-1}.$$

3.3 INFERENCE AND ASYMPTOTIC THEORY

This section defines test statistics for testing linear restrictions on the β vector and derives the asymptotic null behavior of the tests. Results for large-G, fixed-n_G and large-n_G, fixed-G are treated separately as they require different regularity conditions. Throughout, the symbol "$\Rightarrow$"denotes weak convergence of a sequence of stochastic processes to a limiting stochastic process.

We consider testing the null hypothesis $H_0 : R\beta = r$ against $H_0 : R\beta \neq r$, where R is a $q \times k$ matrix of known constants with full rank with $q \leq k$ and r is a $q \times 1$ vector of known constants. Define the Wald statistic as

$$W_{CHAC} = \left(R\hat{\beta} - r\right)' \left[R\widehat{V}_{CHAC}R'\right]^{-1} \left(R\hat{\beta} - r\right),$$

or with the single restriction ($q = 1$) the t-statistic as

$$t_{CHAC} = \frac{\left(R\hat{\beta} - r\right)}{\sqrt{R\widehat{V}_{CHAC}R'}}.$$

3.3.1 Large-G, Fixed-n_G

Vogelsang (2012) developed fixed-b results for the panel analogues to W_{CHAC} and t_{CHAC} for the cases of a large number of time periods and a fixed number of cross-section units. Vogelsang (2012) provided conditions under which the fixed-b limits are equivalent to the standard fixed-b limits obtained by Kiefer and Vogelsang (2005) in pure time series settings. Given the natural similarities between W_{CHAC} and t_{CHAC} and the panel statistics, it is not surprising that the large-G, fixed-n_G limits of W_{CHAC} and t_{CHAC} follow the standard fixed-b limits under suitable regularity conditions. The asymptotic theory in Vogelsang (2012) mainly relies on weak dependence and covariance stationarity in time dimension. In our model because we divide the pure time series into non-overlapping clusters, as long as the original time series satisfies weak dependence and covariance stationarity, the regularity conditions used by Vogelsang (2012) hold in our model as well.

Define $\overline{v}_g = \sum_{t=(g-1)n_G+1}^{gn_G} v_t$. We make the following assumptions.

Assumption A.

1. *n_G is a fixed number and $G = n_G T$.*
2. *For $r \in (0,1]$, $G^{-1}\sum_{g=1}^{[rG]}\sum_{t=(g-1)n_G+1}^{gn_G} x_t x_t' \Rightarrow rQ_c$. Q_c is non-singular.*
3. *$E(\overline{v}_g) = 0$ and $G^{-1/2}\sum_{g=1}^{[rG]}\overline{v}_g \Rightarrow \Lambda_c \mathcal{W}_k(r)$, where $\mathcal{W}_k(r)$ is an $k \times 1$ vector of independent standard Wiener processes and $\Lambda_c \Lambda_c' = \Omega_c$ is the $k \times k$ long run variance matrix (2π times the zero frequency spectral density matrix) of $\overline{v}_g$.*

Assumption A1 is stating that we are considering the large-G, fixed-n_G case. Assumptions A2-A3 are the usual high level assumptions used to obtain fixed-b asymptotic results. Note that

$$\frac{1}{G}\sum_{g=1}^{[rG]}\sum_{t=(g-1)n_G+1}^{gn_G} x_t x_t' = \frac{1}{G}\sum_{t=1}^{[rG]n_G} x_t x_t' = \frac{n_G}{T}\sum_{t=1}^{[\frac{r}{n_G}T]n_G} x_t x_t',$$

where the second equality is obtained by plugging in $G = T/n_G$. If the second moment of x_t satisfies a law of large numbers (LLN) uniformly in r, i.e. $T^{-1}\sum_{t=1}^{[rT]} x_t x_t' \Rightarrow rQ$, then Assumption A2 is satisfied with $Q_c = n_G Q$ because $(n_G/T)\sum_{t=1}^{[r/n_G T]n_G} x_t x_t'$ is asymptotically equivalent to $(n_G/T)\sum_{t=1}^{[rT]} x_t x_t'$. Assumption A3 states that the functional central limit theorem (FCLT) holds for the scaled partial sums of $\overline{v}_g$. As with assumption A2, we can write

$$G^{-1/2}\sum_{g=1}^{[rG]} \overline{v}_g = \frac{1}{G}\sum_{g=1}^{[rG]}\sum_{t=(g-1)n_G+1}^{gn_G} v_t = n_G{}^{1/2}T^{-1/2}\sum_{t=1}^{[\frac{r}{n_G}T]n_G} v_t.$$

If v_t itself follows a FCLT so that $T^{-1/2}\sum_{t=1}^{[rT]} v_t \Rightarrow \Lambda\mathcal{W}_k(r)$, then Assumption A3 is satisfied with $\Lambda_c\Lambda_c' = n_G\Lambda\Lambda'$ because $n_G{}^{1/2}T^{-1/2}\sum_{t=1}^{[r/n_G T]n_G} v_t$ is asymptotically equivalent to $n_G^{1/2}T^{-1/2}\sum_{t=1}^{[rT]} v_t$. If we are making primitive assumptions for a FCLT such as v_t being a mean zero δ-order (for some $\delta > 2$) covariance stationary process that is α-mixing of size $-\beta/(\beta-2)$,[2] then $\overline{v}_g$ is also a mean zero δ-order (for some $\delta > 2$) covariance stationary process that is α-mixing of the same size because finite sums ($n_G < \infty$) of α-mixing processes are also α-mixing with the same size.[3] Therefore, if a FCLT holds for v_t then it will hold for $\overline{v}_g$. In general Assumptions A2-A3 are slightly weaker than assumptions usually used to obtain fixed-b results and are sufficient for the following theorem. The proof follows directly from Vogelsang (2012, Theorem 1).

[2]Phillips and Durlauf (1986) provide sufficient conditions for v_t to satisfy a FCLT.
[3]See White (2001).

Theorem 3.1. *Let $h > 0$ be an integer and let $B_h(r)$ denote a generic $h \times 1$ vector of stochastic processes. Let the random matrix, $P(b, B_h)$, be defined as follows for $b \in (0, 1]$.*

Case (i) : if $k(x)$ is twice continuously differentiable everywhere,

$P(b, B_h) \equiv \int_0^1 \int_0^1 \frac{1}{b^2} k'' \left(\frac{r-s}{b}\right) B_h(r) B_h(s)' \, drds,$

Case (ii) : if $k(x)$ is continuous, $k(x) = 0$ for $|x| \geq 1$ and $k(x)$ is twice continuously differentiable everywhere except for $|x| = 1$,

$P(b, B_h) \equiv \int \int_{|r-s|<b} \frac{1}{b^2} k'' \left(\frac{r-s}{b}\right) B_h(r) B_h(s)' \, drds$

$+\frac{k_-(1)'}{b} \int_0^{1-b} \left(B_h(r+b) B_h(r)' + B_h(r) B_h(r+b)'\right) dr,$

where $k_-(1)' = \lim_{h \to 0} [(k(1) - k(1-h))/h]$,

Case(iii) : if $k(x)$ is the Bartlett kernel,

$P(b, B_h) \equiv \frac{2}{b} \int_0^1 B_h(r) B_h(r)' \, dr - \frac{1}{b} \int_0^{1-b} \left(B_h(r+b) B_h(r)' + B_h(r) B_h(r+b)'\right) dr.$

(a) Then under Assumption A, as $G \to \infty$,

$$\sqrt{G}\left(\hat{\beta} - \beta\right) \Rightarrow Q_c^{-1} \Lambda_c \mathcal{W}_k(1).$$

(b) Let $\widetilde{\mathcal{W}}_k(r)$ denote a $k \times 1$ vector of stochastic processes defined as $\widetilde{\mathcal{W}}_k(r) \equiv \mathcal{W}_k(r) - r\mathcal{W}_k(1)$, for all $r \in (0, 1]$. Assume $M = bG$ where $b \in (0, 1]$ is fixed. Then, under Assumption A and H_0, for $G \to \infty$, n_G fixed,

$$W_{CHAC} \Rightarrow \mathcal{W}_q(1)' \left[P(b, \widetilde{\mathcal{W}}_q)\right]^{-1} \mathcal{W}_q(1)$$

or if there is one restriction ($q = 1$),

$$t_{CHAC} \Rightarrow \frac{\mathcal{W}_1(1)}{\sqrt{P(b, \widetilde{\mathcal{W}}_1)}}.$$

3.3.2 Fixed-G, Large-n_G results

When the number of clusters is fixed, the LLN and FCLT work within the clusters rather than across the clusters. To obtain an asymptotically pivotal result, it is sufficient for the LLN and FCLT to hold for the original time series. Consider the assumption:

Assumption B.

1. *G is fixed. $n_G = G^{-1}T$.*

2. *For $r \in (0,1]$, $T^{-1}\sum_{t=1}^{[rT]} x_t x_t' \Rightarrow rQ$ and Q is non-singular.*

3. *For $r \in (0,1]$, $T^{-1/2}\sum_{t=1}^{[rT]} v_t \Rightarrow \Lambda \mathcal{W}_k(r)$, where $\mathcal{W}_k(r)$ is an $k \times 1$ vector of independent standard Wiener processes and $\Omega = \Lambda\Lambda'$ is the $k \times k$ long run variance matrix (2π times the zero frequency spectral density matrix) of v_t.*

Assumption B1 restates that we are considering the case where the number of clusters is fixed and the size of each cluster is increasing with T. Assumptions B2-B3 state that a law of large numbers applies to $T^{-1}\sum_{t=1}^{[rT]} x_t x_t'$ uniformly in r and a FCLT applies to the scaled partial sum of v_t. These two assumptions are sufficient for fixed-b asymptotic theory to go through when G is fixed and $n_G \to \infty$. The following theorem states asymptotic behavior of OLS, CHAC, and W_{CHAC} (t_{CHAC} when $q = 1$) when G is fixed and $n_G \to \infty$. The proof is provided in Appendix E.

Theorem 3.2. *Let $k > 0$ be an integer and let $B_k(r)$ denote a generic $k \times 1$ vector of stochastic processes. Let the random matrix, $P(G, M, B_k)$, be defined as follows:*

$$P(G,M,B_k) = \sum_{g=1}^{G-1}\sum_{h=1}^{G-1} B_k\left(\frac{g}{G}\right)\left(2k\left(\frac{|g-h|}{M}\right) - k\left(\frac{|g-h+1|}{M}\right) - k\left(\frac{|g-h-1|}{M}\right)\right) \times B_k\left(\frac{h}{G}\right)',$$

and when $k(\cdot)$ is Bartlett kernel $P(G, M, B_k)$ can be further simplified as

$$P(G,M,B_k) = \frac{2}{M}\sum_{g=1}^{G-1} B_k\left(\frac{g}{G}\right) B_k\left(\frac{g}{G}\right)' - \frac{1}{M}\sum_{g=1}^{G-M-1}\left(B_k\left(\frac{g}{G}\right) B_k\left(\frac{g+M}{G}\right)' + B_k\left(\frac{g+M}{G}\right) B_k\left(\frac{g}{G}\right)'\right).$$

(a) Then under Assumption B, as $T \to \infty$ and $n_G \to \infty$,

$$\sqrt{T}\left(\hat{\beta} - \beta\right) \Rightarrow Q^{-1}\Lambda\mathcal{W}_k(1),$$

(b) *Let $\widetilde{\mathcal{W}}_k(r)$ denote a $k \times 1$ vector of stochastic processes defined as $\widetilde{\mathcal{W}}_k(r) \equiv \mathcal{W}_k(r) - r\mathcal{W}_k(1)$, for all $r \in (0,1]$. Under Assumption B, for G fixed, $n_G \to \infty$,*

$$\frac{G}{T}\widehat{\widehat{\Omega}} \Rightarrow \Lambda P(G, M, \widetilde{\mathcal{W}}_k)\Lambda',$$

(c) *and under H_0, as $T \to \infty$ and $n_G \to \infty$,*

$$W_{CHAC} \Rightarrow \mathcal{W}_q(1)' P(G, M, \widetilde{\mathcal{W}}_q)^{-1} \mathcal{W}_q(1)$$

and when $q = 1$,

$$t_{CHAC} \Rightarrow \frac{\mathcal{W}_1(1)}{\sqrt{P(G, M, \widetilde{\mathcal{W}}_1)}}.$$

The fixed-G asymptotic approximation of W_{CHAC} in Theorem 3.2 (c) is different from the fixed-b asymptotic approximation found in Theorem 3.1 (b) which is the usual fixed-b limit in Kiefer and Vogelsang (2005). In fact from Bester et al. (2011) we know that when $M = 1$ and a truncating kernel is used, the fixed-G limit of W_{CHAC} in Theorem 3.2 (c) simplifies to $Gq/(G-q)F_{q,G-q}$ and when $q = 1$ the limit of t_{CHAC} simplifies to $\sqrt{G/(G-1)}t_{G-1}$.

Table 3.1 tabulates the asymptotic critical values for the fixed-G limit for the case of the Bartlett kernel. The critical values were obtained via simulation methods. The Wiener processes in the limits were approximated by scaled partial sums of $1,000$ independent standard normal random variables. $50,000$ replications were used. We see from Table 3.1 that as the number of clusters, G, gets smaller and/or the bandwidth, M, gets larger, the tail of the distribution becomes fatter. As G decreases and/or M increases, less down-weighting is used when calculating CHAC and it is well known from the fixed-b literature that less down-weighting leads to fatter tails of test statistics because of systematic downward bias in the variance estimator.

Generally speaking, it is well known that using less down-weighting in conjunction with fixed-b critical values tends to alleviate over-rejection problems caused by strong serial correlation. The standard HAC estimator can only reduce down-weighting by increasing M (for a given kernel). CHAC can reduce down-weighting by not only increasing M but also by increasing the number of observations per cluster, i.e. by decreasing G. This additional flexibility in down-weighting gives the CHAC approach the ability to reduce size distortions with less loss in power than the original HAC approach. We compare the relative performance of two different weighting schemes, i.e. to choose HAC or CHAC, using a simulation study in the next section.

3.4 FINITE SAMPLE PERFORMANCE

3.4.1 Empirical Rejection Probabilities

In this section we examine the finite sample performance of the robust test statistics based on the CHAC estimator using both the fixed-G, large-n_G approximation and large-G, fixed-n_G approximation. Here we focus on the Bartlett kernel. When $G = T$, it follows that $n_G = 1$ and the CHAC estimator simplifies to the usual HAC estimator without clustering, and when we use $M = 1$, the CHAC estimator simplifies to the pure clustering approach of Bester et al. (2011). Therefore, we can make direct comparisons of those two existing approaches in our results.

We focus on the simple location model

$$\begin{aligned} y_t &= \beta + u_t, \\ u_t &= \rho u_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1}, \end{aligned} \tag{3.3}$$

where $u_0 = \varepsilon_0 = 0$, $\varepsilon_t \sim i.i.d.\, N(0,1)$ with $\rho \in \{-0.5, 0, 0.5, 0.8, 0.9\}$, $\theta \in \{-0.5, 0, 0.5\}$. We set $\beta = 0$. Results are given for sample size $T = 60$ and the number of clusters $G \in \{2, 3, 4, 5, 6, 10, 12, 15, 60\}$. Note that these values of G are factors of 60 and so the clusters evenly divide the sample. With this data generating process we test the null hypothesis that $\beta = 0$ against the alternative $\beta \neq 0$ at a nominal level of 5%. When computing the CHAC estimator, we use the Bartlett kernel with $M \in \{1, 2, \ldots, 9, 10, 12, 15, 30, 40, 50, 60\}$.

For the simple location model the CHAC based t-test is computed as

$$t_{CHAC} = \frac{\hat{\beta}}{\sqrt{G\left(\sum_{t=1}^{T} x_t^2\right)^{-1} \hat{\hat{\Omega}} \left(\sum_{t=1}^{T} x_t^2\right)^{-1}}} = \frac{\hat{\beta}}{\sqrt{\frac{G}{T^2}\hat{\hat{\Omega}}}}$$

where

$$\hat{\beta} = \left(\sum_{t=1}^{T} x_t^2\right)^{-1} \sum_{t=1}^{T} x_t y_t = T^{-1} \sum_{t=1}^{T} y_t$$

and

$$\hat{\hat{\Omega}} = \frac{1}{G} \sum_{g=1}^{G} \sum_{h=1}^{G} k\left(\frac{|g-h|}{M}\right) \hat{\hat{v}}_g \hat{\hat{v}}_h$$

where $\hat{\hat{v}}_g = \sum_{t=(g-1)n_G+1}^{gn_G} \hat{v}_t$ with $\hat{v}_t = y_t - \hat{\beta}$.

We reject the null hypothesis whenever $|t_T| > t_c$ (or reject the null whenever $t_T < t_c^l$ or $t_T > t_c^r$ if $-t_c^l \neq t_c^r$) where t_c is a critical value. Using 10,000 replications, we compute empirical rejection probabilities. From Theorem 3.1 (b), we know that under large-G, fixed-n_G asymptotic theory, $t_c = t_c^{large-G}$ is the 97.5% percentile of the standard fixed-b asymptotic distribution with $b = M/G$. Under fixed-G, large-n_G asymptotic theory $t_c = t_c^{fixed-G}$ is the 97.5% percentile of the distribution derived in Theorem 3.2 (c). We obtain $t_c^{large-G}$ and $t_c^{fixed-G}$ by simulating the corresponding distribution which is possible because both of the distributions are functions of Brownian motion.

In addition we use the bootstrap to obtain critical values. We consider the naive moving block bootstrap with block size $l = n_G$ so that the block lengths used in the resampling match the cluster sizes used to compute $\hat{\hat{\Omega}}$. We also use block size $l = 1$ (the *i.i.d.* bootstrap). Specific details about computing bootstrap critical values are as follows. Let the vector $\omega_t = (y_t, x_t')'$ collect the dependent and explanatory variables (here $x_t = 1$). Let $B_{t,n_G} = \{\omega_t, \omega_{t+1}, \ldots, \omega_{t+n_G-1}\}$ be the block of n_G consecutive observations starting at ω_t. Draw G blocks randomly with replacement from the set of overlapping blocks $\{B_{1,n_G}, \ldots, B_{T-n_G+1,n_G}\}$ and obtain a bootstrap resample of size T. Repeating this 999 times we obtain 999 bootstrap resamples which we denote $\omega_t^{\bullet B} = (y_t^{\bullet B}, x_t^{\bullet B\prime})'$, $t = 1, \ldots, T$, $B = 1, \ldots, 999$. For the *i.i.d.* bootstrap we resample T observations from the original observations with replacement and repeating this 999 times we again obtain 999 bootstrap resamples. For each bootstrap resample we compute the naive bootstrap test statistic

$$t_{CHAC}^{B} = \frac{\hat{\beta}^B - \hat{\beta}}{\sqrt{\frac{G}{T^2}\widehat{\widehat{\Omega}}_B}}$$

where

$$\hat{\beta}^B = \frac{\sum_{t=1}^{T} y_t^{\bullet B}}{T}$$

is the OLS estimator of the B^{th} bootstrap resample and

$$\widehat{\widehat{\Omega}}_B = \frac{1}{G}\sum_{g=1}^{G}\sum_{h=1}^{G} k\left(\frac{|g-h|}{M}\right)\widehat{\widehat{v}}_g^{\bullet B}\widehat{\widehat{v}}_h^{\bullet B}$$

where $\widehat{\widehat{v}}_g^{\bullet B} = \sum_{t=(g-1)n_G+1}^{gn_G} \hat{v}_t^{\bullet B}$ with $\hat{v}_t^{\bullet B} = y_t^{\bullet B} - \hat{\beta}^B$. Then the bootstrap critical values $\{t_c^l, t_c^r\}$ are the 0.025 and 0.975 quantile of the t_{CHAC}^B, $B = 1, \ldots, 999$ respectively. We denote the critical values obtained from the n_G block bootstrap as $\{t_c^{l-block}, t_c^{r-block}\}$ and from the *i.i.d.* bootstrap as $\{t_c^{l-i.i.d.}, t_c^{r-i.i.d.}\}$.

Gonçalves and Vogelsang (2011) showed that the naive moving block bootstrap with block length fixed or increasing but slower than the sample size ($l^2/T \to 0$) has the same limiting distribution as the fixed-b asymptotic distribution. This equivalence is mainly due to the fact that bootstrap resamples generated from the moving block bootstrap, which we denote $(y_t^{\bullet}, x_t^{\bullet\prime})$, satisfy (a)$^{\bullet}$ $T^{-1}\sum_{t=1}^{[rT]} x_t^{\bullet} x_t^{\bullet} \Rightarrow rQ^{\bullet}$ and (b)$^{\bullet}$ $T^{-1/2}\sum_{t=1}^{[rT]} v_t^{\bullet} \Rightarrow \Lambda^{\bullet}\mathcal{W}_k(r)$ for some $Q^{\bullet}$ and $\Lambda^{\bullet}$ where $p^{\bullet}$ denotes the probability measure induced by the bootstrap resampling, conditional on a realization of the original time series. Our asymptotic theory framework and the test statistics are not exactly the same as the ones considered in Gonçalves and Vogelsang (2011). However the results in Gonçalves and Vogelsang (2011) can still be applied. Recall that in Section 3.3.1 (large-G), we pointed out that the original time series satisfying conditions (a) $T^{-1}\sum_{t=1}^{[rT]} x_t x_t \Rightarrow rQ$ and (b) $T^{-1/2}\sum_{t=1}^{[rT]} v_t \Rightarrow \Lambda\mathcal{W}_k(r)$ is sufficient for Assumption A2 and A3. Then, if bootstrap resamples satisfy (a)$^\bullet$ and (b)$^\bullet$, then from Theorem 3.1 (b), the asymptotic distribution of t^B_{CHAC} is the standard fixed-b limit evaluated at M/G. Because the asymptotic distribution in Theorem 3.1 (b) is pivotal with respect to Λ and Q, t^B_{CHAC} and t_{CHAC} have the same limiting distributions. Similarly for Theorem 3.2, because the conditions (a)$^\bullet$ and (b)$^\bullet$ are the same as Assumptions B2 and B3, when we treat G as fixed, t^B_{CHAC} will have the fixed-G distribution as in Theorem 3.2 (c) because fixed-G asymptotic distributions are pivotal with respect to Q,Λ and $Q^\bullet$,$\Lambda^\bullet$ respectively. Therefore, if the bootstrap resamples satisfy (a)$^\bullet$ and (b)$^\bullet$, then the critical values computed from the bootstrap will be first order asymptotically equivalent to t_{CHAC} in the fixed-b sense. See Gonçalves and Vogelsang (2011) for sufficient conditions on the original time series for the bootstrap resamples to satisfy (a)$^\bullet$ and (b)$^\bullet$. The required conditions are similar to the usual weak dependence assumptions required for fixed-b asymptotic theory in Kiefer and Vogelsang (2005) to go through. Therefore, we can conjecture that when G is small, the bootstrap will mimic the fixed-G, large-n_G critical values and when G is large it will mimic the large-G fixed-n_G critical values.

Tables 3.2-3.3 reports empirical null rejection probabilities for the t_{CHAC}. Table 3.2 reports rejections using large-G, fixed-n_G critical values. Table 3.3 reports the rejection probabilities using fixed-G, large-n_G critical values. Because we are using the Bartlett kernel (which truncates), when $M = 1$, $t_{\text{CHAC}} \xrightarrow{d} \sqrt{G/G-1}\,t_{G-1}$ if G is fixed following Bester et al. (2011).

Examining the rejections in Tables 3.2-3.3, it is clear that the fixed-G asymptotic approximation has better size properties than the large-G asymptotic approximation regardless of G. When G is small , the fixed-G asymptotic approximation works well in terms of size across all ρ, θ combinations and M. When $\rho = 0.9$ and $\theta = 0.5$, using $G = 2$, the fixed-G critical value delivers empirical rejection probabilities of 0.06 for both bandwidth values of $M = 1, 2$. A null rejection of 0.06, which is very close to the nominal level 0.05, is impressive given the strong serial correlation and relatively small value of T. When G is large, the fixed-G and large-G asymptotic approximations have similar performance. Therefore, the use of fixed-G critical values is a good idea for all values of G.

Tables 3.4-3.5 reports empirical null rejection probabilities for the t_{CHAC} using the bootstrap critical values. Table 3.4 reports rejection probabilities using the overlapping n_G block bootstrap. Table 3.5 reports rejection probabilities using the *i.i.d* bootstrap. The first obvious pattern is that the *i.i.d*. bootstrap and the fixed-G rejection probabilities are nearly identical in all cases. This is true regardless of the value G. In contrast the large-G critical values and the *i.i.d* bootstrap only have similar finite sample performance once G becomes large, $G = 30$ to 60. The performance of the block bootstrap depends on the strength of the serial correlation. Using middle sized blocks can result in less size distortion than either the *i.i.d*. bootstrap or the fixed-G critical values when the serial correlation is strong. When serial correlation is weak, then use of the block bootstrap can result in over-rejections that do not occur with the *i.i.d*. bootstrap (see the $\rho = 0$, $\theta = 0$ case). Similar comparisons between the block bootstrap and the *i.i.d*. bootstrap were found by Gonçalves and Vogelsang (2011) for the non-clustered HAC case.

It may seem surprising at first that i) even when G is large, the fixed-G approximation works well and ii) the *i.i.d*. bootstrap mimics the fixed-G limit even when when G is large including the $G = T$ case. However, a closer look at the tabulated fixed-G critical values in Table 3.1 indicates these results are not surprising. If we took $G = 60$ critical values from Table 3.1 and compared them to the critical values tabulated by Kiefer and Vogelsang (2005), we would see that the critical values are very close to each other. This suggests that the critical values of the fixed-G random variable approaches that of the large-G (i.e. standard fixed-b) random variable as G increases. It is this apparent continuity in G that explains the patterns in the finite sample simulations. The patterns are not so surprising upon closer examination of the theory.

The simulation study shows that using a relatively small value of G can substantially reduce size distortions. In the next subsection we investigate the impact of the choice of G and M on power to assess the power loss that is expected to be incurred when controlling over-rejection problems.

3.4.2 Size Adjusted Power

We now report some power calculations to investigate the impact of G and M on power. We compute size-adjusted power so that we can make power comparisons independent of over-rejection problems. The size-corrections we employ obviously cannot be used in empirical applications. We use the same data generating process as in Section 3.4.1. We set $\theta = 0$ and focus on AR(1) errors with $\rho \in \{0, 0.5, 0.8, 0.9\}$. As with Section 3.4.1, the null is $\beta = 0$, and the sample size is $T = 60$. Again we use $G \in \{2, 3, 4, 5, 6, 10, 12, 15, 60\}$. For a given value of ρ and combination of G, M we first simulate the finite sample null critical values of t_{CHAC} using $10,000$ replications and then compute size-adjusted power by obtaining the rejection probabilities for a grid of values of $\beta \in (0, 7]$ again using $10,000$ replications.

Table 3.6 reports the area above the size adjusted power curve which is conceptually the average of Type II error across alternatives. The area is divided by $\max \beta$ so that the total square area is normalized to 1. From Table 3.6 we can see the general power-size trade-off. Decreasing G, with M fixed, or increasing M, with G fixed, always increases type II error, and we know from Tables 3.2-3.5 that over-rejection problems are decreasing in these scenarios. When M and G change together, it is more difficult to see clear patterns in the size-power trade-off. For example, if we want to compare the $G = 60, M = 30$ case (i.e. the usual HAC estimator with bandwidth equal to half the sample size) with the $G = 30, M = 9$ case, when $\rho = 0.5$, the change in the power-size trade-off it is not so obvious because decreasing G will increase size while decreasing power whereas the decrease in M has the opposite effect. From Table 3.6 we see that between these two cases, the $G = 30, M = 9$ case has higher average power. More specifically, the Type II error for $G = 30, M = 9$ is 0.60 while for $G = 60, M = 30$ the type II error is 0.63. Referring back to Table 3.5 for these two cases, notice that $G = 30, M = 9$ has size of 0.069 while $G = 60, M = 30$ has size of 0.070. Although the difference in size is small, there are improvements in both size and power by dividing the sample into 30 clusters. By further decreasing G to 15 and M to 3, the type II error decreases to 0.595, but the size remains at 0.069. So, we have cases where dividing time series into clusters and smoothing can reduce the over-rejection problem without a great cost in terms of power and sometimes it is possible to increase power without inducing more over-rejections. Compared to the usual HAC approach, clustering with smoothing is usually a better option than simply moving around the bandwidth.

Figure 3.1 depicts power for some interesting cases where clustering and smoothing provides greater power while not increasing over-rejections for $\rho \in \{0.5, 0.8, 0.9\}$. The benchmark case is the size-adjusted power curve for the $G = 60, M = 30$ case. The other combinations of G and M give tests with similar size to the $G = 60, M = 30$. We see that there indeed is room for improvement in power while holding size constant through clustering and smoothing. For all ρ, the $G = 60$ case (the usual HAC estimator), has the lowest power compared to other combinations of G and M which have similar size.

Figures 3.2-3.27 also compare the power-size trade off between the usual HAC estimator ($G = 60$) for a range of bandwidths with the CHAC estimator. The label for each power curve indicates the size of that test. Figure 3.2 considers the case of $\rho = 0.5$ and CHAC implemented with a range of values of G but always with $M = 1$. The case where no clustering or smoothing is being used is the $G = 60, M = 1$ which serves as the benchmark (see the light blue line). Size is quite inflated in this case: 0.267. If a researcher wants to reduce this over-rejection, then using the usual HAC estimator would need to increase M, which are depicted by the grey lines. On the other hand, with CHAC the researcher can choose to divide the time series into clusters (while still using $M = 1$) to reduce the over-rejection problem (see the pink lines). In the second graph of Figure 3.2, we can see that by increasing M to 20 or larger the researcher reduces size to about 0.069 for the HAC test. All of these HAC tests are dominated by using CHAC with $G = 6$ and $M = 1$ in terms of both size and power. Cases where the CHAC approach dominates the usual HAC approach are found regularly when CHAC is implemented with other values of M and for $\rho = 0.8, 0.9$. For example, in Figure 3.3, we see that when we match the size between HAC and CHAC, then there exists a value of g such that CHAC with that $G = g, M = 2$ has better power than HAC. If we match the power of the HAC and CHAC tests, there is always some value of g such that CHAC with $G = g, M = 2$ has better size. As the remaining power figures show, there rarely are situations where there is not a CHAC estimator that is either better in size when holding power fixed or has more power when holding size distortions fixed.

3.5 CONCLUSION AND REMAINING WORK

In this chapter we analyzed smoothed clustered standard errors in time series regression models. We find that asymptotic approximation generated under the assumption of a fixed number of clusters, G, works well even for the large values of G. Even under strong serial correlation, the over-rejection problem is relatively small when a small number of clusters is used. Also because fixed-G asymptotic approximation can be simply obtained by the *i.i.d.* naive bootstrap in practice, in empirical work with strong serial correlation, smoothed clustered standard errors can be useful. A simulation study shows that in general, there rarely are situations where there is not a CHAC estimator that is either better in size when holding power fixed or has more power when holding size distortions fixed compared to the usual HAC estimators. What this chapter does not address is a systematic method for choosing the number of clusters and the bandwidth used to implement the CHAC estimator. Developing a data-dependent method to choose G and M remains as an important topic of ongoing research.

For certain applications clustering and smoothing can be natural given the structure of the data. For example, for a financial market that is not open on weekends it is natural to cluster within the week and smooth across weeks. Tables 3.7-3.8 shows a small simulation study with three months of daily data generated from the DGP in Section 3.4.1 where every weekend is missing resulting in $T = 60$ observations. The regression model is estimated by deleting the missing observations. This can be considered as constructing AM series or ES regression model in Chapter 2 and then calculating the test statistic based on CHAC standard error. Many of the patterns exhibited in the simulations without missing data are seen in Tables 3.7-3.8. There is one interesting pattern that is surprising. When $G = 12$, i.e. each cluster has length 5, the rejection rates are smaller than those of $G = 10$ when $\rho \geq 0$. In the other tables of results we always see higher rejection rates using $G = 12$ compared to $G = 10$ when $\rho \geq 0$. It seems reasonable to conjecture that the exact match of 5 observations per cluster with the number of observations per week has something to do with this pattern. The application of clustered standard errors to times series with missing observations remains as an interesting topic for future work.

Table 3.1: Critical Values : Fixed G

G	M	1%	2.5%	5%	10%	50%	90%	95%	97.5%	99%
2	1	−45.991	−17.920	−8.992	−4.390	−0.010	4.375	8.874	17.942	46.230
2	2	−65.041	−25.342	−12.716	−6.208	−0.014	6.187	12.550	25.374	65.379
3	1	−8.710	−5.323	−3.605	−2.325	−0.008	2.305	3.563	5.227	8.680
3	2	−11.315	−7.057	−4.702	−2.997	−0.009	2.980	4.618	6.805	11.286
3	3	−13.858	−8.642	−5.759	−3.671	−0.012	3.650	5.656	8.334	13.823
4	1	−5.303	−3.670	−2.724	−1.917	−0.008	1.896	2.723	3.676	5.214
4	2	−6.945	−4.716	−3.428	−2.349	−0.008	2.346	3.409	4.679	6.769
4	3	−8.005	−5.603	−4.038	−2.782	−0.010	2.764	4.045	5.518	7.931
4	4	−9.243	−6.470	−4.663	−3.212	−0.012	3.191	4.671	6.371	9.158
5	1	−4.143	−3.120	−2.407	−1.732	−0.007	1.720	2.381	3.124	4.240
5	2	−5.272	−3.857	−2.907	−2.056	−0.008	2.050	2.886	3.829	5.322
5	3	−6.288	−4.540	−3.407	−2.403	−0.009	2.390	3.397	4.492	6.246
5	4	−7.010	−5.136	−3.874	−2.720	−0.010	2.710	3.847	5.092	7.069
5	5	−7.837	−5.742	−4.331	−3.041	−0.011	3.029	4.301	5.693	7.903
6	1	−3.693	−2.837	−2.230	−1.628	−0.007	1.623	2.209	2.805	3.641
6	2	−4.526	−3.396	−2.615	−1.887	−0.007	1.872	2.598	3.349	4.514
6	3	−5.356	−3.980	−3.022	−2.159	−0.008	2.152	3.026	3.915	5.301
6	4	−6.006	−4.507	−3.430	−2.434	−0.009	2.410	3.400	4.427	5.945
6	5	−6.619	−4.942	−3.775	−2.684	−0.010	2.671	3.754	4.883	6.564
6	6	−7.251	−5.414	−4.136	−2.940	−0.011	2.926	4.112	5.349	7.190
7	1	−3.405	−2.658	−2.114	−1.569	−0.006	1.554	2.108	2.651	3.401
7	2	−4.057	−3.114	−2.431	−1.778	−0.007	1.768	2.413	3.089	4.110
7	3	−4.752	−3.576	−2.779	−2.004	−0.008	1.992	2.764	3.570	4.770
7	4	−5.401	−4.032	−3.128	−2.239	−0.009	2.225	3.099	4.003	5.358
7	5	−5.949	−4.436	−3.461	−2.470	−0.009	2.448	3.409	4.436	5.913
7	6	−6.394	−4.823	−3.749	−2.681	−0.010	2.666	3.709	4.816	6.358
7	7	−6.906	−5.209	−4.050	−2.896	−0.011	2.879	4.006	5.202	6.868

Table 3.1: (cont'd)

G	M	1%	2.5%	5%	10%	50%	90%	95%	97.5%	99%
8	1	−3.210	−2.526	−2.048	−1.522	−0.006	1.516	2.035	2.530	3.208
8	2	−3.749	−2.915	−2.311	−1.704	−0.007	1.693	2.298	2.908	3.746
8	3	−4.346	−3.324	−2.611	−1.899	−0.007	1.895	2.593	3.324	4.351
8	4	−4.945	−3.739	−2.907	−2.100	−0.008	2.095	2.893	3.720	4.878
8	5	−5.457	−4.109	−3.205	−2.300	−0.009	2.289	3.185	4.082	5.364
8	6	−5.876	−4.454	−3.474	−2.485	−0.009	2.484	3.452	4.445	5.799
8	7	−6.279	−4.788	−3.731	−2.673	−0.010	2.675	3.698	4.767	6.228
8	8	−6.712	−5.118	−3.989	−2.857	−0.011	2.860	3.954	5.096	6.658
9	1	−3.099	−2.467	−1.997	−1.498	−0.006	1.482	1.980	2.465	3.084
9	2	−3.559	−2.779	−2.222	−1.648	−0.007	1.630	2.212	2.773	3.553
9	3	−4.079	−3.138	−2.491	−1.820	−0.007	1.808	2.462	3.137	4.057
9	4	−4.630	−3.513	−2.743	−1.994	−0.008	1.988	2.723	3.512	4.554
9	5	−5.091	−3.880	−3.015	−2.167	−0.008	2.156	2.988	3.859	5.017
9	6	−5.493	−4.196	−3.268	−2.342	−0.009	2.332	3.223	4.165	5.439
9	7	−5.878	−4.481	−3.496	−2.510	−0.009	2.498	3.470	4.456	5.801
9	8	−6.227	−4.769	−3.716	−2.674	−0.010	2.662	3.692	4.743	6.137
9	9	−6.605	−5.058	−3.941	−2.836	−0.010	2.823	3.916	5.031	6.509
10	1	−2.989	−2.401	−1.954	−1.470	−0.006	1.463	1.951	2.394	2.986
10	2	−3.383	−2.692	−2.149	−1.606	−0.007	1.594	2.144	2.680	3.434
10	3	−3.876	−3.000	−2.382	−1.749	−0.007	1.749	2.371	3.021	3.883
10	4	−4.310	−3.325	−2.613	−1.911	−0.007	1.908	2.606	3.358	4.348
10	5	−4.761	−3.655	−2.849	−2.069	−0.008	2.063	2.839	3.663	4.733
10	6	−5.156	−3.943	−3.072	−2.237	−0.008	2.223	3.065	3.948	5.144
10	7	−5.497	−4.222	−3.296	−2.391	−0.009	2.369	3.289	4.231	5.520
10	8	−5.827	−4.472	−3.498	−2.538	−0.010	2.520	3.491	4.494	5.868
10	9	−6.134	−4.730	−3.698	−2.685	−0.010	2.673	3.690	4.747	6.178
10	10	−6.465	−4.986	−3.898	−2.830	−0.011	2.818	3.889	5.004	6.512

Table 3.1: (cont'd)

G	M	1%	2.5%	5%	10%	50%	90%	95%	97.5%	99%
11	1	−2.910	−2.350	−1.913	−1.447	−0.006	1.442	1.904	2.333	2.916
11	2	−3.314	−2.612	−2.101	−1.565	−0.007	1.563	2.081	2.581	3.267
11	3	−3.724	−2.899	−2.303	−1.708	−0.007	1.690	2.284	2.868	3.669
11	4	−4.112	−3.198	−2.519	−1.846	−0.007	1.839	2.491	3.193	4.068
11	5	−4.536	−3.470	−2.746	−1.995	−0.008	1.983	2.711	3.469	4.485
11	6	−4.899	−3.750	−2.959	−2.137	−0.008	2.121	2.919	3.741	4.868
11	7	−5.240	−4.017	−3.157	−2.285	−0.009	2.264	3.112	4.000	5.171
11	8	−5.548	−4.260	−3.346	−2.422	−0.009	2.402	3.307	4.259	5.488
11	9	−5.842	−4.495	−3.524	−2.553	−0.010	2.534	3.495	4.469	5.789
11	10	−6.143	−4.713	−3.710	−2.688	−0.010	2.666	3.675	4.708	6.055
11	11	−6.443	−4.943	−3.892	−2.819	−0.011	2.796	3.855	4.937	6.351
12	1	−2.867	−2.311	−1.889	−1.428	−0.006	1.427	1.884	2.304	2.840
12	2	−3.173	−2.541	−2.057	−1.537	−0.007	1.538	2.044	2.526	3.167
12	3	−3.553	−2.806	−2.236	−1.662	−0.007	1.661	2.226	2.800	3.533
12	4	−3.932	−3.066	−2.426	−1.793	−0.007	1.785	2.420	3.057	3.899
12	5	−4.325	−3.328	−2.621	−1.923	−0.007	1.921	2.613	3.321	4.271
12	6	−4.667	−3.609	−2.814	−2.055	−0.008	2.053	2.804	3.586	4.617
12	7	−5.003	−3.854	−3.008	−2.188	−0.008	2.175	3.000	3.826	4.935
12	8	−5.300	−4.096	−3.188	−2.321	−0.009	2.309	3.174	4.037	5.230
12	9	−5.587	−4.289	−3.365	−2.446	−0.010	2.426	3.349	4.266	5.512
12	10	−5.862	−4.509	−3.541	−2.563	−0.010	2.548	3.508	4.467	5.819
12	11	−6.129	−4.720	−3.703	−2.684	−0.010	2.664	3.682	4.676	6.097
12	12	−6.402	−4.930	−3.868	−2.803	−0.011	2.783	3.846	4.884	6.368

Table 3.1: (cont'd)

G	M	1%	2.5%	5%	10%	50%	90%	95%	97.5%	99%
13	1	−2.795	−2.281	−1.869	−1.424	−0.006	1.415	1.864	2.273	2.811
13	2	−3.100	−2.499	−2.018	−1.521	−0.007	1.516	2.016	2.496	3.095
13	3	−3.444	−2.726	−2.191	−1.630	−0.007	1.624	2.184	2.724	3.416
13	4	−3.797	−2.975	−2.370	−1.751	−0.007	1.743	2.354	2.971	3.801
13	5	−4.168	−3.216	−2.556	−1.874	−0.007	1.860	2.532	3.217	4.132
13	6	−4.481	−3.460	−2.724	−1.992	−0.008	1.984	2.721	3.448	4.446
13	7	−4.790	−3.686	−2.913	−2.117	−0.008	2.100	2.891	3.684	4.775
13	8	−5.094	−3.915	−3.090	−2.236	−0.009	2.214	3.064	3.894	5.053
13	9	−5.350	−4.128	−3.248	−2.355	−0.009	2.332	3.230	4.123	5.300
13	10	−5.612	−4.328	−3.408	−2.473	−0.009	2.448	3.393	4.315	5.586
13	11	−5.859	−4.528	−3.559	−2.584	−0.010	2.563	3.540	4.511	5.828
13	12	−6.107	−4.706	−3.717	−2.694	−0.010	2.675	3.690	4.705	6.084
13	13	−6.356	−4.898	−3.869	−2.804	−0.011	2.784	3.841	4.897	6.333
14	1	−2.765	−2.265	−1.846	−1.413	−0.006	1.405	1.843	2.250	2.748
14	2	−3.030	−2.447	−1.995	−1.500	−0.006	1.495	1.979	2.437	3.046
14	3	−3.345	−2.657	−2.151	−1.605	−0.007	1.591	2.140	2.653	3.358
14	4	−3.683	−2.893	−2.315	−1.715	−0.007	1.701	2.289	2.886	3.665
14	5	−3.994	−3.121	−2.477	−1.827	−0.007	1.814	2.451	3.114	4.003
14	6	−4.314	−3.341	−2.640	−1.934	−0.007	1.922	2.624	3.344	4.298
14	7	−4.609	−3.574	−2.817	−2.048	−0.008	2.031	2.787	3.551	4.587
14	8	−4.906	−3.783	−2.982	−2.158	−0.008	2.146	2.959	3.769	4.844
14	9	−5.176	−3.972	−3.128	−2.275	−0.009	2.262	3.103	3.970	5.096
14	10	−5.383	−4.154	−3.278	−2.387	−0.009	2.365	3.249	4.173	5.358
14	11	−5.611	−4.340	−3.423	−2.488	−0.009	2.473	3.402	4.350	5.603
14	12	−5.865	−4.522	−3.566	−2.587	−0.010	2.573	3.546	4.528	5.857
14	13	−6.095	−4.707	−3.715	−2.694	−0.010	2.677	3.683	4.703	6.091
14	14	−6.325	−4.885	−3.856	−2.795	−0.010	2.778	3.822	4.881	6.321

Table 3.1: (cont'd)

G	M	1%	2.5%	5%	10%	50%	90%	95%	97.5%	99%
15	1	−2.717	−2.230	−1.834	−1.401	−0.006	1.393	1.834	2.235	2.723
15	2	−2.965	−2.420	−1.965	−1.488	−0.006	1.475	1.952	2.410	2.971
15	3	−3.255	−2.613	−2.106	−1.577	−0.007	1.574	2.089	2.601	3.249
15	4	−3.558	−2.822	−2.257	−1.679	−0.007	1.675	2.241	2.811	3.564
15	5	−3.875	−3.021	−2.417	−1.786	−0.007	1.777	2.394	3.029	3.847
15	6	−4.177	−3.245	−2.570	−1.889	−0.007	1.881	2.551	3.237	4.155
15	7	−4.430	−3.445	−2.733	−1.991	−0.008	1.990	2.711	3.433	4.424
15	8	−4.727	−3.656	−2.887	−2.097	−0.008	2.086	2.863	3.638	4.681
15	9	−4.980	−3.851	−3.023	−2.208	−0.009	2.192	3.010	3.834	4.949
15	10	−5.205	−4.010	−3.171	−2.315	−0.009	2.299	3.159	4.004	5.176
15	11	−5.416	−4.196	−3.312	−2.412	−0.009	2.397	3.293	4.167	5.422
15	12	−5.618	−4.372	−3.443	−2.511	−0.010	2.498	3.428	4.339	5.619
15	13	−5.826	−4.531	−3.578	−2.606	−0.010	2.596	3.555	4.510	5.844
15	14	−6.036	−4.702	−3.704	−2.704	−0.010	2.691	3.688	4.665	6.059
15	15	−6.248	−4.867	−3.834	−2.799	−0.011	2.785	3.817	4.829	6.271
20	1	−2.606	−2.160	−1.780	−1.369	−0.006	1.360	1.781	2.156	2.604
20	2	−2.786	−2.290	−1.875	−1.433	−0.006	1.426	1.871	2.289	2.780
20	3	−2.998	−2.443	−1.986	−1.499	−0.006	1.492	1.969	2.431	2.990
20	4	−3.227	−2.590	−2.099	−1.572	−0.006	1.565	2.080	2.577	3.221
20	5	−3.446	−2.747	−2.209	−1.642	−0.007	1.640	2.197	2.732	3.440
20	6	−3.686	−2.895	−2.320	−1.722	−0.007	1.718	2.300	2.903	3.642
20	7	−3.900	−3.050	−2.438	−1.802	−0.007	1.795	2.420	3.068	3.880
20	8	−4.138	−3.211	−2.551	−1.881	−0.007	1.875	2.538	3.221	4.101
20	9	−4.343	−3.357	−2.666	−1.959	−0.008	1.950	2.651	3.370	4.327
20	10	−4.530	−3.514	−2.788	−2.036	−0.008	2.025	2.769	3.520	4.520
20	11	−4.724	−3.661	−2.905	−2.118	−0.008	2.101	2.880	3.668	4.723
20	12	−4.943	−3.793	−3.013	−2.197	−0.009	2.177	2.994	3.800	4.890

Table 3.1: (cont'd)

G	M	1%	2.5%	5%	10%	50%	90%	95%	97.5%	99%
20	13	−5.093	−3.927	−3.126	−2.273	−0.009	2.254	3.096	3.944	5.061
20	14	−5.305	−4.063	−3.230	−2.347	−0.009	2.329	3.195	4.071	5.256
20	15	−5.441	−4.196	−3.328	−2.422	−0.009	2.405	3.296	4.202	5.403
20	20	−6.235	−4.801	−3.823	−2.779	−0.011	2.762	3.785	4.815	6.221
30	1	−2.490	−2.088	−1.731	−1.343	−0.006	1.332	1.732	2.089	2.501
30	2	−2.615	−2.166	−1.794	−1.381	−0.006	1.373	1.796	2.172	2.618
30	3	−2.745	−2.266	−1.867	−1.422	−0.006	1.416	1.857	2.267	2.749
30	4	−2.892	−2.368	−1.936	−1.469	−0.006	1.463	1.925	2.365	2.895
30	5	−3.035	−2.464	−2.013	−1.516	−0.006	1.511	1.992	2.457	3.041
30	6	−3.192	−2.571	−2.087	−1.563	−0.007	1.560	2.069	2.555	3.193
30	7	−3.355	−2.673	−2.163	−1.614	−0.007	1.610	2.150	2.661	3.349
30	8	−3.491	−2.766	−2.238	−1.669	−0.007	1.657	2.223	2.767	3.474
30	9	−3.651	−2.880	−2.312	−1.718	−0.007	1.711	2.293	2.881	3.624
30	10	−3.795	−2.985	−2.392	−1.768	−0.007	1.763	2.366	2.994	3.770
30	11	−3.945	−3.087	−2.464	−1.822	−0.007	1.815	2.441	3.101	3.918
30	12	−4.083	−3.194	−2.545	−1.875	−0.007	1.866	2.520	3.189	4.075
30	13	−4.228	−3.293	−2.622	−1.923	−0.008	1.917	2.598	3.289	4.224
30	14	−4.357	−3.404	−2.701	−1.974	−0.008	1.973	2.683	3.383	4.358
30	15	−4.482	−3.508	−2.779	−2.026	−0.008	2.022	2.762	3.480	4.488
30	20	−5.104	−3.968	−3.140	−2.292	−0.009	2.273	3.115	3.954	5.070
30	25	−5.617	−4.397	−3.480	−2.534	−0.010	2.519	3.453	4.372	5.634
30	30	−6.138	−4.799	−3.804	−2.771	−0.010	2.752	3.780	4.782	6.170

Table 3.1: (cont'd)

G	M	1%	2.5%	5%	10%	50%	90%	95%	97.5%	99%
40	1	−2.462	−2.055	−1.709	−1.330	−0.006	1.320	1.711	2.067	2.450
40	2	−2.546	−2.121	−1.756	−1.359	−0.006	1.347	1.754	2.127	2.525
40	3	−2.630	−2.189	−1.802	−1.390	−0.006	1.382	1.798	2.194	2.635
40	4	−2.741	−2.267	−1.859	−1.421	−0.006	1.418	1.854	2.266	2.728
40	5	−2.854	−2.342	−1.914	−1.453	−0.006	1.448	1.907	2.328	2.836
40	6	−2.951	−2.415	−1.972	−1.490	−0.006	1.485	1.958	2.399	2.946
40	7	−3.059	−2.489	−2.028	−1.525	−0.006	1.521	2.013	2.475	3.058
40	8	−3.184	−2.564	−2.085	−1.564	−0.006	1.557	2.063	2.551	3.173
40	9	−3.291	−2.644	−2.140	−1.601	−0.006	1.596	2.120	2.630	3.283
40	10	−3.400	−2.715	−2.195	−1.639	−0.007	1.631	2.177	2.707	3.402
40	11	−3.520	−2.806	−2.249	−1.678	−0.007	1.667	2.232	2.787	3.505
40	12	−3.635	−2.879	−2.310	−1.716	−0.007	1.708	2.289	2.873	3.619
40	13	−3.728	−2.960	−2.368	−1.755	−0.007	1.747	2.339	2.955	3.716
40	14	−3.848	−3.037	−2.423	−1.794	−0.007	1.785	2.399	3.035	3.828
40	15	−3.970	−3.111	−2.481	−1.832	−0.007	1.824	2.456	3.114	3.929
40	20	−4.486	−3.494	−2.773	−2.029	−0.008	2.018	2.747	3.481	4.447
40	25	−4.953	−3.840	−3.054	−2.222	−0.009	2.210	3.014	3.838	4.913
40	30	−5.373	−4.167	−3.310	−2.413	−0.009	2.394	3.277	4.160	5.340
40	35	−5.753	−4.478	−3.557	−2.590	−0.010	2.575	3.523	4.462	5.754
40	40	−6.167	−4.784	−3.805	−2.768	−0.011	2.749	3.762	4.772	6.150

Table 3.1: (cont'd)

G	M	1%	2.5%	5%	10%	50%	90%	95%	97.5%	99%
60	1	−2.410	−2.026	−1.697	−1.316	−0.006	1.309	1.686	2.025	2.416
60	2	−2.473	−2.067	−1.719	−1.334	−0.006	1.323	1.714	2.069	2.450
60	3	−2.538	−2.113	−1.751	−1.355	−0.006	1.344	1.746	2.113	2.513
60	4	−2.596	−2.161	−1.784	−1.377	−0.006	1.368	1.780	2.160	2.580
60	5	−2.663	−2.209	−1.817	−1.399	−0.006	1.390	1.814	2.203	2.650
60	6	−2.734	−2.256	−1.855	−1.421	−0.006	1.413	1.850	2.255	2.722
60	7	−2.806	−2.304	−1.890	−1.441	−0.006	1.435	1.882	2.309	2.792
60	8	−2.876	−2.353	−1.927	−1.465	−0.006	1.458	1.920	2.348	2.857
60	9	−2.943	−2.407	−1.965	−1.488	−0.006	1.483	1.953	2.391	2.928
60	10	−3.017	−2.456	−2.005	−1.513	−0.006	1.507	1.992	2.441	3.011
60	11	−3.094	−2.506	−2.039	−1.536	−0.007	1.532	2.029	2.495	3.083
60	12	−3.171	−2.555	−2.078	−1.560	−0.007	1.555	2.062	2.548	3.166
60	13	−3.248	−2.612	−2.114	−1.585	−0.007	1.581	2.101	2.595	3.242
60	14	−3.333	−2.660	−2.154	−1.611	−0.007	1.605	2.139	2.653	3.314
60	15	−3.403	−2.707	−2.189	−1.638	−0.007	1.630	2.178	2.700	3.386
60	20	−3.770	−2.963	−2.379	−1.765	−0.007	1.759	2.363	2.975	3.740
60	25	−4.155	−3.229	−2.570	−1.897	−0.007	1.884	2.555	3.234	4.105
60	30	−4.481	−3.491	−2.776	−2.026	−0.008	2.015	2.748	3.467	4.447
60	35	−4.791	−3.714	−2.957	−2.155	−0.008	2.144	2.931	3.703	4.767
60	40	−5.080	−3.943	−3.137	−2.288	−0.009	2.270	3.098	3.927	5.037
60	45	−5.357	−4.164	−3.308	−2.412	−0.009	2.391	3.271	4.147	5.324
60	50	−5.613	−4.363	−3.468	−2.531	−0.010	2.517	3.441	4.351	5.590
60	55	−5.883	−4.565	−3.636	−2.653	−0.010	2.634	3.601	4.558	5.857
60	60	−6.136	−4.771	−3.798	−2.772	−0.011	2.749	3.760	4.765	6.118

Table 3.1: (cont'd)

G	M	1%	2.5%	5%	10%	50%	90%	95%	97.5%	99%
80	1	−2.386	−2.009	−1.684	−1.313	−0.006	1.299	1.678	2.014	2.397
80	2	−2.437	−2.042	−1.705	−1.325	−0.006	1.314	1.698	2.049	2.428
80	3	−2.472	−2.077	−1.725	−1.338	−0.006	1.327	1.723	2.079	2.467
80	4	−2.520	−2.106	−1.750	−1.354	−0.006	1.342	1.748	2.117	2.515
80	5	−2.567	−2.146	−1.776	−1.372	−0.006	1.359	1.773	2.145	2.560
80	6	−2.621	−2.179	−1.800	−1.390	−0.006	1.378	1.796	2.184	2.613
80	7	−2.670	−2.215	−1.824	−1.406	−0.006	1.396	1.821	2.218	2.670
80	8	−2.731	−2.255	−1.854	−1.421	−0.006	1.412	1.848	2.256	2.729
80	9	−2.785	−2.294	−1.881	−1.437	−0.006	1.430	1.873	2.290	2.777
80	10	−2.836	−2.330	−1.907	−1.454	−0.006	1.448	1.898	2.327	2.831
80	11	−2.893	−2.366	−1.934	−1.471	−0.006	1.465	1.922	2.360	2.878
80	12	−2.939	−2.403	−1.960	−1.487	−0.006	1.482	1.953	2.390	2.935
80	13	−2.991	−2.443	−1.989	−1.507	−0.006	1.502	1.980	2.428	2.998
80	14	−3.040	−2.483	−2.020	−1.524	−0.006	1.518	2.006	2.467	3.055
80	15	−3.105	−2.517	−2.047	−1.542	−0.007	1.536	2.035	2.504	3.110
80	20	−3.394	−2.709	−2.189	−1.638	−0.007	1.628	2.174	2.700	3.391
80	25	−3.674	−2.904	−2.332	−1.733	−0.007	1.729	2.306	2.904	3.658
80	30	−3.943	−3.095	−2.474	−1.830	−0.007	1.822	2.456	3.103	3.939
80	35	−4.204	−3.297	−2.621	−1.927	−0.008	1.922	2.600	3.293	4.178
80	40	−4.485	−3.490	−2.776	−2.025	−0.008	2.016	2.742	3.476	4.443
80	45	−4.714	−3.661	−2.912	−2.125	−0.008	2.110	2.882	3.652	4.681
80	50	−4.929	−3.833	−3.043	−2.222	−0.009	2.206	3.011	3.823	4.901
80	55	−5.140	−3.997	−3.186	−2.317	−0.009	2.299	3.143	3.989	5.130
80	60	−5.342	−4.168	−3.307	−2.409	−0.009	2.390	3.269	4.148	5.323
80	65	−5.544	−4.320	−3.425	−2.500	−0.010	2.485	3.390	4.296	5.523
80	70	−5.753	−4.464	−3.554	−2.591	−0.010	2.571	3.515	4.452	5.732
80	75	−5.940	−4.618	−3.673	−2.679	−0.010	2.658	3.638	4.614	5.954
80	80	−6.131	−4.766	−3.795	−2.768	−0.010	2.748	3.757	4.763	6.148

Table 3.1: (cont'd)

G	M	1%	2.5%	5%	10%	50%	90%	95%	97.5%	99%
120	1	−2.362	−1.996	−1.675	−1.304	−0.006	1.293	1.667	2.004	2.367
120	2	−2.392	−2.012	−1.685	−1.312	−0.006	1.301	1.678	2.020	2.398
120	3	−2.426	−2.038	−1.702	−1.321	−0.006	1.311	1.693	2.044	2.419
120	4	−2.457	−2.058	−1.715	−1.331	−0.006	1.319	1.710	2.063	2.449
120	5	−2.491	−2.086	−1.733	−1.343	−0.006	1.332	1.729	2.087	2.475
120	6	−2.518	−2.108	−1.749	−1.354	−0.006	1.343	1.746	2.113	2.502
120	7	−2.554	−2.135	−1.764	−1.365	−0.006	1.354	1.765	2.133	2.536
120	8	−2.587	−2.156	−1.783	−1.378	−0.006	1.367	1.780	2.155	2.572
120	9	−2.622	−2.179	−1.798	−1.389	−0.006	1.379	1.793	2.177	2.606
120	10	−2.653	−2.204	−1.818	−1.399	−0.006	1.390	1.810	2.199	2.643
120	11	−2.693	−2.227	−1.834	−1.411	−0.006	1.401	1.829	2.226	2.680
120	12	−2.730	−2.254	−1.853	−1.420	−0.006	1.411	1.847	2.251	2.713
120	13	−2.767	−2.277	−1.873	−1.429	−0.006	1.424	1.865	2.276	2.752
120	14	−2.797	−2.304	−1.889	−1.441	−0.006	1.435	1.880	2.296	2.790
120	15	−2.836	−2.327	−1.908	−1.452	−0.006	1.447	1.901	2.323	2.827
120	20	−3.011	−2.456	−2.002	−1.511	−0.006	1.507	1.991	2.440	3.012
120	25	−3.199	−2.579	−2.097	−1.571	−0.007	1.567	2.082	2.565	3.202
120	30	−3.395	−2.706	−2.190	−1.638	−0.007	1.628	2.175	2.701	3.378
120	35	−3.585	−2.836	−2.283	−1.701	−0.007	1.694	2.266	2.834	3.559
120	40	−3.764	−2.961	−2.379	−1.765	−0.007	1.758	2.357	2.970	3.735
120	45	−3.951	−3.098	−2.475	−1.831	−0.007	1.823	2.452	3.101	3.911
120	50	−4.142	−3.224	−2.569	−1.895	−0.007	1.887	2.550	3.224	4.096
120	55	−4.307	−3.358	−2.671	−1.961	−0.008	1.953	2.648	3.344	4.266
120	60	−4.471	−3.493	−2.772	−2.026	−0.008	2.016	2.740	3.471	4.427
120	65	−4.617	−3.599	−2.862	−2.092	−0.008	2.079	2.835	3.595	4.597
120	70	−4.790	−3.720	−2.952	−2.158	−0.008	2.142	2.923	3.709	4.743
120	75	−4.934	−3.835	−3.044	−2.222	−0.009	2.208	3.008	3.830	4.899
120	80	−5.090	−3.944	−3.132	−2.286	−0.009	2.268	3.097	3.929	5.043

Table 3.1: (cont'd)

G	M	1%	2.5%	5%	10%	50%	90%	95%	97.5%	99%
120	85	−5.224	−4.052	−3.227	−2.349	−0.009	2.327	3.181	4.035	5.193
120	90	−5.353	−4.164	−3.308	−2.409	−0.009	2.390	3.266	4.141	5.328
120	95	−5.480	−4.268	−3.387	−2.470	−0.010	2.452	3.358	4.248	5.441
120	100	−5.599	−4.355	−3.465	−2.531	−0.010	2.512	3.438	4.343	5.575
120	105	−5.741	−4.461	−3.550	−2.593	−0.010	2.571	3.518	4.457	5.722
120	110	−5.871	−4.568	−3.631	−2.651	−0.010	2.629	3.600	4.562	5.866
120	115	−6.006	−4.665	−3.709	−2.709	−0.010	2.687	3.679	4.664	5.996
120	120	−6.131	−4.765	−3.789	−2.769	−0.010	2.743	3.760	4.768	6.127

Table 3.2: Large G, Empirical null rejection probabilities, 5% level, $T = 60$

ρ	θ	M	$G \to \infty$ critical value values of G									
			2	3	4	5	6	10	12	15	30	60
-0.5	-0.5	1	0.212	0.094	0.055	0.034	0.021	0.006	0.003	0.001	0.000	0.000
		2	0.217	0.093	0.056	0.037	0.026	0.010	0.006	0.004	0.001	0.000
		3		0.094	0.054	0.037	0.025	0.013	0.008	0.005	0.001	0.000
		4			0.055	0.035	0.025	0.014	0.010	0.008	0.002	0.000
		5				0.036	0.023	0.013	0.009	0.008	0.003	0.000
		6					0.023	0.013	0.009	0.008	0.004	0.000
		7						0.012	0.009	0.008	0.004	0.000
		8						0.012	0.010	0.008	0.005	0.001
		9						0.012	0.009	0.008	0.006	0.001
		10						0.012	0.009	0.008	0.005	0.001
		12							0.009	0.008	0.006	0.001
		15								0.008	0.006	0.002
		20									0.006	0.003
		25									0.005	0.003
		30									0.006	0.003
		40										0.004
		50										0.003
		60										0.003

Table 3.2: (cont'd)

ρ	θ	M	$G \to \infty$ critical value									
			values of G									
			2	3	4	5	6	10	12	15	30	60
−0.5	0	1	0.239	0.125	0.094	0.076	0.061	0.043	0.035	0.031	0.018	0.002
		2	0.245	0.123	0.090	0.075	0.059	0.046	0.040	0.037	0.025	0.016
		3		0.124	0.086	0.072	0.060	0.049	0.041	0.040	0.029	0.015
		4			0.087	0.068	0.059	0.050	0.042	0.041	0.033	0.021
		5				0.069	0.057	0.048	0.043	0.040	0.034	0.022
		6					0.057	0.048	0.043	0.041	0.035	0.025
		7						0.047	0.044	0.043	0.035	0.026
		8						0.046	0.043	0.041	0.036	0.029
		9						0.047	0.042	0.042	0.037	0.030
		10						0.047	0.041	0.040	0.036	0.031
		12							0.042	0.041	0.038	0.032
		15								0.041	0.039	0.032
		20									0.038	0.034
		25									0.037	0.036
		30									0.037	0.037
		40										0.035
		50										0.034
		60										0.034

Table 3.2: (cont'd)

ρ	θ	M	$G \to \infty$ critical value, values of G: 2	3	4	5	6	10	12	15	30	60
-0.5	0.5	1	0.242	0.135	0.101	0.088	0.071	0.063	0.058	0.057	0.054	0.053
		2	0.246	0.130	0.097	0.082	0.071	0.057	0.056	0.053	0.051	0.051
		3		0.132	0.092	0.079	0.072	0.057	0.052	0.052	0.049	0.051
		4			0.093	0.076	0.069	0.059	0.055	0.051	0.049	0.049
		5				0.077	0.067	0.058	0.054	0.052	0.049	0.049
		6					0.067	0.057	0.056	0.053	0.048	0.049
		7						0.056	0.054	0.051	0.047	0.049
		8						0.055	0.053	0.051	0.048	0.049
		9						0.055	0.052	0.050	0.047	0.049
		10						0.055	0.052	0.051	0.048	0.048
		12							0.052	0.049	0.049	0.047
		15								0.049	0.051	0.048
		20									0.049	0.048
		25									0.046	0.049
		30									0.048	0.050
		40										0.049
		50										0.048
		60										0.047

Table 3.2: (cont'd)

ρ	θ	M	$G \to \infty$ critical value									
			values of G									
			2	3	4	5	6	10	12	15	30	60
0	−0.5	1	0.226	0.115	0.081	0.059	0.044	0.026	0.018	0.011	0.002	0.000
		2	0.232	0.112	0.081	0.064	0.046	0.032	0.023	0.020	0.006	0.001
		3		0.115	0.076	0.062	0.048	0.033	0.027	0.025	0.011	0.003
		4			0.077	0.059	0.049	0.034	0.028	0.026	0.013	0.004
		5				0.060	0.046	0.034	0.028	0.026	0.016	0.006
		6					0.046	0.034	0.028	0.028	0.019	0.009
		7						0.035	0.029	0.027	0.019	0.010
		8						0.034	0.027	0.028	0.020	0.010
		9						0.033	0.028	0.029	0.022	0.012
		10						0.033	0.027	0.028	0.021	0.013
		12							0.027	0.027	0.023	0.016
		15								0.027	0.024	0.017
		20									0.023	0.019
		25									0.022	0.021
		30									0.023	0.021
		40										0.020
		50										0.020
		60										0.019

Table 3.2: (cont'd)

ρ	θ	M	$G \to \infty$ critical value									
			values of G									
			2	3	4	5	6	10	12	15	30	60
0	0	1	0.242	0.135	0.101	0.088	0.071	0.063	0.058	0.057	0.054	0.053
		2	0.246	0.130	0.097	0.082	0.071	0.057	0.056	0.053	0.051	0.051
		3		0.132	0.092	0.079	0.072	0.057	0.052	0.052	0.049	0.051
		4			0.093	0.076	0.069	0.059	0.055	0.051	0.049	0.049
		5				0.077	0.067	0.058	0.054	0.052	0.049	0.049
		6					0.067	0.057	0.056	0.053	0.048	0.049
		7						0.056	0.054	0.051	0.047	0.049
		8						0.055	0.053	0.051	0.048	0.049
		9						0.055	0.052	0.050	0.047	0.049
		10						0.055	0.052	0.051	0.048	0.048
		12							0.052	0.049	0.049	0.047
		15								0.049	0.051	0.048
		20									0.049	0.048
		25									0.046	0.049
		30									0.048	0.050
		40										0.049
		50										0.048
		60										0.047

Table 3.2: (cont'd)

ρ	θ	M	$G \to \infty$ critical value values of G 2	3	4	5	6	10	12	15	30	60
0	0.5	1	0.242	0.136	0.104	0.090	0.079	0.071	0.070	0.071	0.090	0.152
		2	0.248	0.132	0.100	0.083	0.076	0.064	0.062	0.060	0.068	0.090
		3		0.136	0.097	0.079	0.076	0.061	0.058	0.058	0.061	0.072
		4			0.099	0.078	0.071	0.062	0.060	0.057	0.059	0.067
		5				0.079	0.070	0.062	0.059	0.057	0.056	0.064
		6					0.070	0.060	0.059	0.058	0.056	0.061
		7						0.058	0.059	0.058	0.055	0.060
		8						0.059	0.057	0.057	0.055	0.059
		9						0.059	0.057	0.056	0.055	0.057
		10						0.060	0.057	0.056	0.054	0.057
		12							0.056	0.056	0.056	0.056
		15								0.055	0.056	0.056
		20									0.054	0.055
		25									0.054	0.056
		30									0.054	0.056
		40										0.055
		50										0.054
		60										0.055

Table 3.2: (cont'd)

ρ	θ	M	$G \to \infty$ critical value									
			values of G									
			2	3	4	5	6	10	12	15	30	60
0.5	−0.5	1	0.242	0.135	0.101	0.088	0.071	0.063	0.058	0.057	0.054	0.053
		2	0.246	0.130	0.097	0.082	0.071	0.057	0.056	0.053	0.051	0.051
		3		0.132	0.092	0.079	0.072	0.057	0.052	0.052	0.049	0.051
		4			0.093	0.076	0.069	0.059	0.055	0.051	0.049	0.049
		5				0.077	0.067	0.058	0.054	0.052	0.049	0.049
		6					0.067	0.057	0.056	0.053	0.048	0.049
		7						0.056	0.054	0.051	0.047	0.049
		8						0.055	0.053	0.051	0.048	0.049
		9						0.055	0.052	0.050	0.047	0.049
		10						0.055	0.052	0.051	0.048	0.048
		12							0.052	0.049	0.049	0.047
		15								0.049	0.051	0.048
		20									0.049	0.048
		25									0.046	0.049
		30									0.048	0.050
		40										0.049
		50										0.048
		60										0.047

Table 3.2: (cont'd)

ρ	θ	M	$G \to \infty$ critical value: values of G									
			2	3	4	5	6	10	12	15	30	60
0.5	0	1	0.248	0.145	0.111	0.101	0.092	0.096	0.099	0.110	0.177	0.269
		2	0.254	0.141	0.105	0.090	0.086	0.077	0.074	0.079	0.110	0.178
		3		0.142	0.101	0.089	0.083	0.074	0.073	0.072	0.088	0.134
		4			0.102	0.086	0.083	0.073	0.069	0.070	0.080	0.110
		5				0.087	0.079	0.073	0.071	0.070	0.075	0.097
		6					0.080	0.070	0.070	0.070	0.072	0.088
		7						0.070	0.069	0.070	0.071	0.084
		8						0.070	0.068	0.069	0.070	0.079
		9						0.070	0.068	0.067	0.069	0.077
		10						0.071	0.067	0.068	0.068	0.075
		12							0.068	0.067	0.069	0.072
		15								0.068	0.070	0.071
		20									0.066	0.068
		25									0.067	0.070
		30									0.068	0.069
		40										0.066
		50										0.067
		60										0.068

Table 3.2: (cont'd)

ρ	θ	M	$G \to \infty$ critical value values of G									
			2	3	4	5	6	10	12	15	30	60
0.5	0.5	1	0.250	0.146	0.113	0.101	0.097	0.100	0.105	0.122	0.203	0.338
		2	0.256	0.143	0.105	0.093	0.087	0.079	0.080	0.083	0.121	0.204
		3		0.144	0.102	0.091	0.086	0.075	0.074	0.076	0.095	0.149
		4			0.104	0.089	0.083	0.076	0.073	0.073	0.084	0.121
		5				0.090	0.081	0.074	0.074	0.073	0.079	0.105
		6					0.083	0.072	0.072	0.073	0.077	0.095
		7						0.072	0.070	0.074	0.073	0.089
		8						0.072	0.070	0.072	0.072	0.084
		9						0.072	0.072	0.070	0.072	0.082
		10						0.073	0.070	0.070	0.072	0.080
		12							0.072	0.070	0.073	0.077
		15								0.072	0.072	0.073
		20									0.069	0.072
		25									0.070	0.073
		30									0.071	0.072
		40										0.069
		50										0.070
		60										0.071

Table 3.2: (cont'd)

ρ	θ	M	$G \to \infty$ critical value									
			values of G									
			2	3	4	5	6	10	12	15	30	60
0.8	−0.5	1	0.265	0.168	0.143	0.137	0.142	0.180	0.199	0.226	0.310	0.385
		2	0.271	0.158	0.130	0.119	0.115	0.125	0.134	0.150	0.227	0.311
		3		0.160	0.127	0.117	0.113	0.111	0.113	0.124	0.178	0.262
		4			0.129	0.115	0.109	0.107	0.109	0.112	0.150	0.227
		5				0.116	0.109	0.105	0.106	0.106	0.134	0.201
		6					0.110	0.105	0.104	0.106	0.123	0.178
		7						0.103	0.103	0.104	0.118	0.162
		8						0.102	0.102	0.102	0.112	0.148
		9						0.105	0.101	0.101	0.109	0.140
		10						0.106	0.101	0.100	0.107	0.134
		12							0.103	0.101	0.105	0.123
		15								0.103	0.102	0.113
		20									0.102	0.106
		25									0.101	0.104
		30									0.102	0.102
		40										0.101
		50										0.101
		60										0.101

Table 3.2: (cont'd)

ρ	θ	M	$G \to \infty$ critical value									
			values of G									
			2	3	4	5	6	10	12	15	30	60
0.8	0	1	0.266	0.171	0.152	0.146	0.158	0.210	0.237	0.277	0.407	0.540
		2	0.273	0.163	0.137	0.127	0.125	0.138	0.151	0.175	0.277	0.408
		3		0.166	0.131	0.122	0.122	0.121	0.128	0.139	0.215	0.329
		4			0.133	0.121	0.119	0.117	0.120	0.124	0.175	0.279
		5				0.122	0.118	0.115	0.118	0.120	0.153	0.243
		6					0.119	0.115	0.115	0.117	0.139	0.215
		7						0.113	0.115	0.116	0.130	0.192
		8						0.113	0.113	0.114	0.124	0.175
		9						0.115	0.114	0.114	0.122	0.163
		10						0.116	0.113	0.113	0.119	0.153
		12							0.114	0.112	0.117	0.140
		15								0.114	0.115	0.127
		20									0.114	0.119
		25									0.112	0.117
		30									0.114	0.115
		40										0.113
		50										0.112
		60										0.114

Table 3.2: (cont'd)

ρ	θ	M	$G \to \infty$ critical value									
			values of G									
			2	3	4	5	6	10	12	15	30	60
0.8	0.5	1	0.266	0.172	0.153	0.148	0.159	0.215	0.242	0.281	0.422	0.563
		2	0.273	0.163	0.137	0.128	0.126	0.139	0.154	0.178	0.284	0.424
		3		0.166	0.132	0.123	0.123	0.123	0.130	0.141	0.220	0.339
		4			0.134	0.121	0.119	0.120	0.122	0.126	0.178	0.285
		5				0.123	0.118	0.117	0.119	0.121	0.155	0.247
		6					0.119	0.116	0.118	0.118	0.141	0.219
		7						0.114	0.117	0.117	0.133	0.196
		8						0.114	0.115	0.116	0.127	0.179
		9						0.116	0.115	0.116	0.123	0.165
		10						0.117	0.114	0.115	0.121	0.155
		12							0.116	0.114	0.119	0.141
		15								0.116	0.116	0.130
		20									0.115	0.121
		25									0.114	0.118
		30									0.117	0.117
		40										0.115
		50										0.115
		60										0.117

Table 3.2: (cont'd)

ρ	θ	M	$G \to \infty$ critical value values of G 2	3	4	5	6	10	12	15	30	60
0.9	−0.5	1	0.295	0.218	0.211	0.226	0.249	0.330	0.367	0.402	0.519	0.605
		2	0.301	0.202	0.182	0.181	0.186	0.226	0.252	0.286	0.407	0.520
		3		0.204	0.181	0.172	0.174	0.190	0.203	0.227	0.334	0.458
		4			0.182	0.173	0.169	0.176	0.184	0.199	0.287	0.408
		5				0.175	0.171	0.171	0.175	0.184	0.254	0.369
		6					0.172	0.167	0.172	0.175	0.229	0.335
		7						0.166	0.167	0.172	0.210	0.309
		8						0.167	0.166	0.170	0.199	0.288
		9						0.167	0.166	0.166	0.191	0.270
		10						0.168	0.166	0.165	0.183	0.255
		12							0.168	0.166	0.175	0.228
		15								0.168	0.172	0.204
		20									0.165	0.183
		25									0.166	0.174
		30									0.168	0.172
		40										0.165
		50										0.166
		60										0.168

Table 3.2: (cont'd)

ρ	θ	M	$G \to \infty$ critical value									
			values of G									
			2	3	4	5	6	10	12	15	30	60
0.9	0	1	0.298	0.222	0.216	0.238	0.263	0.355	0.397	0.439	0.574	0.688
		2	0.305	0.208	0.184	0.183	0.192	0.237	0.266	0.306	0.442	0.575
		3		0.211	0.184	0.177	0.178	0.199	0.213	0.239	0.360	0.501
		4			0.186	0.176	0.176	0.180	0.190	0.207	0.306	0.442
		5				0.178	0.176	0.176	0.179	0.190	0.268	0.399
		6					0.178	0.172	0.177	0.181	0.239	0.361
		7						0.173	0.174	0.177	0.221	0.333
		8						0.173	0.173	0.174	0.208	0.307
		9						0.174	0.174	0.173	0.199	0.288
		10						0.175	0.174	0.173	0.191	0.269
		12							0.176	0.174	0.181	0.241
		15								0.175	0.176	0.214
		20									0.173	0.192
		25									0.174	0.180
		30									0.175	0.176
		40										0.173
		50										0.174
		60										0.175

Table 3.2: (cont'd)

ρ	θ	M	$G \to \infty$ critical value									
			values of G									
			2	3	4	5	6	10	12	15	30	60
0.9	0.5	1	0.300	0.225	0.218	0.238	0.263	0.357	0.399	0.443	0.582	0.700
		2	0.305	0.209	0.183	0.183	0.193	0.239	0.269	0.308	0.446	0.584
		3		0.212	0.185	0.177	0.178	0.200	0.214	0.241	0.364	0.506
		4			0.186	0.177	0.177	0.182	0.192	0.208	0.309	0.447
		5				0.179	0.177	0.177	0.180	0.194	0.270	0.402
		6					0.179	0.174	0.176	0.182	0.242	0.365
		7						0.174	0.174	0.178	0.222	0.335
		8						0.173	0.174	0.176	0.209	0.310
		9						0.176	0.175	0.173	0.200	0.290
		10						0.177	0.175	0.174	0.194	0.270
		12							0.177	0.175	0.182	0.242
		15								0.176	0.177	0.215
		20									0.174	0.194
		25									0.175	0.180
		30									0.177	0.177
		40										0.174
		50										0.174
		60										0.177

Table 3.3: Fixed G, Empirical null rejection probabilities, 5% level, $T = 60$

ρ	θ	M	fixed G critical value									
			values of G									
			2	3	4	5	6	10	12	15	30	60
−0.5	−0.5	1	0.041	0.033	0.027	0.018	0.012	0.004	0.002	0.001	0.000	0.000
		2	0.041	0.032	0.029	0.020	0.016	0.008	0.005	0.003	0.001	0.000
		3		0.032	0.027	0.021	0.015	0.011	0.007	0.005	0.001	0.000
		4			0.027	0.020	0.016	0.011	0.009	0.006	0.002	0.000
		5				0.020	0.016	0.011	0.009	0.007	0.003	0.000
		6					0.016	0.011	0.008	0.007	0.004	0.000
		7						0.010	0.008	0.007	0.004	0.000
		8						0.010	0.008	0.007	0.005	0.001
		9						0.010	0.009	0.007	0.006	0.001
		10						0.010	0.009	0.008	0.005	0.002
		12							0.009	0.008	0.006	0.001
		15								0.007	0.006	0.003
		20									0.006	0.003
		25									0.005	0.003
		30									0.006	0.003
		40										0.004
		50										0.003
		60										0.004

Table 3.3: (cont'd)

			fixed G critical value									
ρ	θ	M	values of G									
			2	3	4	5	6	10	12	15	30	60
−0.5	0	1	0.050	0.045	0.047	0.043	0.038	0.034	0.030	0.028	0.017	0.001
		2	0.050	0.044	0.046	0.044	0.041	0.039	0.034	0.033	0.025	0.015
		3		0.044	0.046	0.046	0.043	0.042	0.037	0.038	0.029	0.015
		4			0.046	0.045	0.042	0.042	0.038	0.037	0.033	0.021
		5				0.045	0.041	0.040	0.039	0.038	0.034	0.022
		6					0.041	0.041	0.038	0.039	0.036	0.026
		7						0.041	0.040	0.040	0.036	0.027
		8						0.041	0.038	0.039	0.037	0.030
		9						0.040	0.039	0.040	0.036	0.031
		10						0.040	0.039	0.038	0.036	0.031
		12							0.039	0.040	0.037	0.032
		15								0.040	0.038	0.033
		20									0.038	0.035
		25									0.037	0.036
		30									0.037	0.037
		40										0.037
		50										0.037
		60										0.035

Table 3.3: (cont'd)

ρ	θ	M	fixed G critical value									
			values of G									
			2	3	4	5	6	10	12	15	30	60
−0.5	0.5	1	0.050	0.050	0.051	0.050	0.049	0.050	0.049	0.050	0.051	0.052
		2	0.050	0.048	0.051	0.051	0.049	0.048	0.050	0.050	0.050	0.050
		3		0.048	0.049	0.050	0.048	0.049	0.047	0.050	0.049	0.051
		4			0.049	0.050	0.049	0.049	0.050	0.049	0.048	0.049
		5				0.050	0.050	0.049	0.049	0.049	0.049	0.050
		6					0.050	0.051	0.050	0.050	0.048	0.049
		7						0.049	0.050	0.049	0.048	0.049
		8						0.050	0.050	0.047	0.048	0.050
		9						0.049	0.050	0.048	0.047	0.050
		10						0.049	0.050	0.048	0.047	0.049
		12							0.049	0.047	0.048	0.048
		15								0.048	0.051	0.049
		20									0.049	0.048
		25									0.048	0.049
		30									0.048	0.051
		40										0.051
		50										0.050
		60										0.048

Table 3.3: (cont'd)

ρ	θ	M	fixed G critical value									
			values of G									
			2	3	4	5	6	10	12	15	30	60
0	−0.5	1	0.044	0.041	0.038	0.034	0.030	0.020	0.015	0.009	0.002	0.000
		2	0.044	0.040	0.041	0.037	0.033	0.026	0.021	0.017	0.005	0.001
		3		0.040	0.040	0.038	0.033	0.028	0.023	0.023	0.010	0.003
		4			0.040	0.038	0.033	0.029	0.024	0.024	0.013	0.004
		5				0.038	0.032	0.029	0.025	0.025	0.016	0.007
		6					0.032	0.029	0.024	0.025	0.019	0.009
		7						0.030	0.025	0.026	0.019	0.010
		8						0.029	0.025	0.026	0.021	0.011
		9						0.029	0.026	0.027	0.021	0.012
		10						0.029	0.025	0.027	0.021	0.014
		12							0.025	0.027	0.023	0.016
		15								0.026	0.023	0.018
		20									0.023	0.019
		25									0.023	0.021
		30									0.023	0.021
		40										0.021
		50										0.021
		60										0.020

Table 3.3: (cont'd)

ρ	θ	M	fixed G critical value									
			values of G									
			2	3	4	5	6	10	12	15	30	60
0	0	1	0.050	0.050	0.051	0.050	0.049	0.050	0.049	0.050	0.051	0.052
		2	0.050	0.048	0.051	0.051	0.049	0.048	0.050	0.050	0.050	0.050
		3		0.048	0.049	0.050	0.048	0.049	0.047	0.050	0.049	0.051
		4			0.049	0.050	0.049	0.049	0.050	0.049	0.048	0.049
		5				0.050	0.050	0.049	0.049	0.049	0.049	0.050
		6					0.050	0.051	0.050	0.050	0.048	0.049
		7						0.049	0.050	0.049	0.048	0.049
		8						0.050	0.050	0.047	0.048	0.050
		9						0.049	0.050	0.048	0.047	0.050
		10						0.049	0.050	0.048	0.047	0.049
		12							0.049	0.047	0.048	0.048
		15								0.048	0.051	0.049
		20									0.049	0.048
		25									0.048	0.049
		30									0.048	0.051
		40										0.051
		50										0.050
		60										0.048

Table 3.3: (cont'd)

ρ	θ	M	fixed G critical value									
			values of G									
			2	3	4	5	6	10	12	15	30	60
0	0.5	1	0.049	0.050	0.052	0.054	0.053	0.057	0.058	0.064	0.085	0.148
		2	0.049	0.046	0.053	0.052	0.052	0.053	0.056	0.056	0.065	0.089
		3		0.046	0.051	0.051	0.053	0.053	0.053	0.055	0.060	0.072
		4			0.051	0.050	0.053	0.052	0.055	0.054	0.059	0.067
		5				0.050	0.053	0.054	0.055	0.054	0.056	0.065
		6					0.053	0.054	0.054	0.055	0.056	0.062
		7						0.052	0.053	0.054	0.055	0.060
		8						0.052	0.053	0.054	0.055	0.061
		9						0.053	0.054	0.053	0.054	0.059
		10						0.053	0.054	0.054	0.054	0.058
		12							0.053	0.055	0.056	0.057
		15								0.054	0.056	0.057
		20									0.055	0.056
		25									0.056	0.056
		30									0.055	0.056
		40										0.056
		50										0.056
		60										0.056

Table 3.3: (cont'd)

ρ	θ	M	fixed G critical value									
			values of G									
			2	3	4	5	6	10	12	15	30	60
0.5	−0.5	1	0.050	0.050	0.051	0.050	0.049	0.050	0.049	0.050	0.051	0.052
		2	0.050	0.048	0.051	0.051	0.049	0.048	0.050	0.050	0.050	0.050
		3		0.048	0.049	0.050	0.048	0.049	0.047	0.050	0.049	0.051
		4			0.049	0.050	0.049	0.049	0.050	0.049	0.048	0.049
		5				0.050	0.050	0.049	0.049	0.049	0.049	0.050
		6					0.050	0.051	0.050	0.050	0.048	0.049
		7						0.049	0.050	0.049	0.048	0.049
		8						0.050	0.050	0.047	0.048	0.050
		9						0.049	0.050	0.048	0.047	0.050
		10						0.049	0.050	0.048	0.047	0.049
		12							0.049	0.047	0.048	0.048
		15								0.048	0.051	0.049
		20									0.049	0.048
		25									0.048	0.049
		30									0.048	0.051
		40										0.051
		50										0.050
		60										0.048

Table 3.3: (cont'd)

ρ	θ	M	fixed G critical value									
			values of G									
			2	3	4	5	6	10	12	15	30	60
0.5	0	1	0.049	0.054	0.055	0.060	0.062	0.078	0.086	0.100	0.171	0.265
		2	0.049	0.052	0.056	0.058	0.060	0.066	0.069	0.073	0.108	0.176
		3		0.052	0.054	0.058	0.058	0.064	0.066	0.069	0.087	0.134
		4			0.054	0.057	0.058	0.063	0.064	0.067	0.079	0.110
		5				0.057	0.057	0.062	0.066	0.066	0.075	0.098
		6					0.057	0.061	0.065	0.065	0.073	0.089
		7						0.061	0.063	0.066	0.071	0.084
		8						0.063	0.064	0.066	0.070	0.080
		9						0.064	0.065	0.064	0.068	0.078
		10						0.064	0.065	0.066	0.068	0.077
		12							0.065	0.066	0.069	0.074
		15								0.067	0.068	0.072
		20									0.067	0.068
		25									0.069	0.070
		30									0.069	0.069
		40										0.068
		50										0.070
		60										0.070

Table 3.3: (cont'd)

ρ	θ	M	fixed G critical value									
			values of G									
			2	3	4	5	6	10	12	15	30	60
0.5	0.5	1	0.049	0.056	0.057	0.060	0.065	0.083	0.093	0.110	0.198	0.335
		2	0.049	0.054	0.056	0.059	0.061	0.069	0.073	0.078	0.119	0.202
		3		0.054	0.054	0.058	0.059	0.065	0.068	0.073	0.094	0.148
		4			0.054	0.058	0.059	0.065	0.067	0.069	0.083	0.121
		5				0.058	0.060	0.064	0.068	0.069	0.080	0.105
		6					0.060	0.063	0.066	0.068	0.077	0.096
		7						0.065	0.065	0.070	0.074	0.090
		8						0.065	0.066	0.069	0.073	0.086
		9						0.065	0.068	0.067	0.072	0.083
		10						0.065	0.068	0.068	0.071	0.082
		12							0.069	0.069	0.072	0.078
		15								0.070	0.071	0.075
		20									0.069	0.072
		25									0.071	0.073
		30									0.072	0.072
		40										0.071
		50										0.073
		60										0.073

Table 3.3: (cont'd)

ρ	θ	M	fixed G critical value									
			values of G									
			2	3	4	5	6	10	12	15	30	60
0.8	-0.5	1	0.053	0.062	0.075	0.084	0.101	0.157	0.180	0.212	0.305	0.382
		2	0.053	0.060	0.070	0.076	0.085	0.112	0.126	0.142	0.225	0.309
		3		0.060	0.070	0.076	0.083	0.099	0.105	0.119	0.176	0.261
		4			0.070	0.077	0.084	0.095	0.100	0.107	0.149	0.226
		5				0.077	0.086	0.093	0.099	0.102	0.135	0.202
		6					0.086	0.093	0.097	0.101	0.124	0.178
		7						0.092	0.096	0.100	0.118	0.163
		8						0.094	0.096	0.098	0.112	0.151
		9						0.096	0.097	0.098	0.108	0.142
		10						0.096	0.098	0.098	0.106	0.136
		12							0.099	0.100	0.104	0.125
		15								0.101	0.101	0.115
		20									0.102	0.107
		25									0.103	0.104
		30									0.103	0.102
		40										0.103
		50										0.103
		60										0.104

Table 3.3: (cont'd)

ρ	θ	M	fixed G critical value									
			values of G									
			2	3	4	5	6	10	12	15	30	60
0.8	0	1	0.054	0.064	0.080	0.091	0.113	0.184	0.220	0.261	0.400	0.537
		2	0.054	0.063	0.074	0.081	0.093	0.124	0.141	0.167	0.275	0.406
		3		0.063	0.077	0.082	0.089	0.111	0.119	0.134	0.214	0.328
		4			0.077	0.083	0.091	0.105	0.113	0.120	0.175	0.278
		5				0.083	0.094	0.102	0.110	0.114	0.153	0.243
		6					0.094	0.105	0.107	0.112	0.140	0.216
		7						0.103	0.107	0.111	0.131	0.193
		8						0.102	0.108	0.109	0.125	0.178
		9						0.105	0.108	0.110	0.121	0.165
		10						0.105	0.109	0.110	0.118	0.155
		12							0.110	0.111	0.116	0.141
		15								0.112	0.114	0.130
		20									0.114	0.120
		25									0.114	0.117
		30									0.115	0.115
		40										0.116
		50										0.116
		60										0.116

Table 3.3: (cont'd)

ρ	θ	M	fixed G critical value									
			values of G									
			2	3	4	5	6	10	12	15	30	60
0.8	0.5	1	0.054	0.064	0.081	0.093	0.117	0.188	0.225	0.268	0.416	0.560
		2	0.054	0.064	0.075	0.080	0.093	0.124	0.144	0.170	0.282	0.422
		3		0.064	0.076	0.083	0.090	0.112	0.121	0.136	0.218	0.338
		4			0.076	0.084	0.092	0.107	0.114	0.121	0.178	0.285
		5				0.084	0.095	0.103	0.111	0.116	0.155	0.248
		6					0.095	0.106	0.108	0.114	0.142	0.220
		7						0.104	0.109	0.113	0.133	0.196
		8						0.104	0.110	0.110	0.127	0.180
		9						0.107	0.109	0.111	0.122	0.168
		10						0.107	0.111	0.112	0.120	0.157
		12							0.111	0.113	0.118	0.144
		15								0.115	0.116	0.132
		20									0.116	0.122
		25									0.116	0.118
		30									0.118	0.117
		40										0.117
		50										0.118
		60										0.119

Table 3.3: (cont'd)

ρ	θ	M	fixed G critical value									
			values of G									
			2	3	4	5	6	10	12	15	30	60
0.9	−0.5	1	0.060	0.086	0.125	0.155	0.196	0.304	0.346	0.390	0.515	0.601
		2	0.060	0.083	0.106	0.127	0.148	0.209	0.240	0.276	0.404	0.518
		3		0.083	0.110	0.124	0.136	0.173	0.193	0.220	0.333	0.457
		4			0.110	0.128	0.136	0.161	0.175	0.193	0.286	0.408
		5				0.128	0.142	0.156	0.167	0.179	0.254	0.370
		6					0.142	0.153	0.161	0.170	0.230	0.336
		7						0.154	0.159	0.168	0.211	0.310
		8						0.156	0.159	0.165	0.199	0.289
		9						0.158	0.160	0.161	0.190	0.272
		10						0.158	0.162	0.162	0.182	0.257
		12							0.162	0.165	0.174	0.231
		15								0.166	0.171	0.206
		20									0.166	0.184
		25									0.168	0.175
		30									0.169	0.172
		40										0.168
		50										0.170
		60										0.171

Table 3.3: (cont'd)

ρ	θ	M	fixed G critical value									
			values of G									
			2	3	4	5	6	10	12	15	30	60
0.9	0	1	0.060	0.088	0.127	0.164	0.205	0.327	0.376	0.425	0.568	0.686
		2	0.060	0.083	0.109	0.129	0.152	0.220	0.255	0.296	0.440	0.573
		3		0.083	0.113	0.128	0.142	0.182	0.201	0.231	0.360	0.500
		4			0.113	0.133	0.140	0.166	0.181	0.202	0.306	0.442
		5				0.133	0.146	0.161	0.172	0.186	0.269	0.400
		6					0.146	0.160	0.167	0.175	0.241	0.362
		7						0.161	0.165	0.173	0.222	0.334
		8						0.164	0.165	0.170	0.208	0.309
		9						0.165	0.168	0.167	0.198	0.291
		10						0.165	0.170	0.170	0.190	0.272
		12							0.171	0.172	0.180	0.243
		15								0.173	0.176	0.216
		20									0.173	0.192
		25									0.175	0.180
		30									0.176	0.176
		40										0.175
		50										0.177
		60										0.177

Table 3.3: (cont'd)

ρ	θ	M	fixed G critical value									
			values of G									
			2	3	4	5	6	10	12	15	30	60
0.9	0.5	1	0.060	0.088	0.126	0.165	0.206	0.331	0.379	0.429	0.577	0.697
		2	0.060	0.084	0.108	0.131	0.153	0.221	0.256	0.299	0.444	0.582
		3		0.084	0.112	0.128	0.142	0.182	0.202	0.234	0.363	0.505
		4			0.112	0.133	0.141	0.166	0.181	0.202	0.308	0.447
		5				0.133	0.147	0.162	0.173	0.187	0.270	0.403
		6					0.147	0.161	0.168	0.176	0.243	0.365
		7						0.161	0.165	0.174	0.222	0.336
		8						0.164	0.167	0.170	0.209	0.311
		9						0.165	0.170	0.168	0.199	0.293
		10						0.165	0.171	0.171	0.192	0.273
		12							0.171	0.173	0.180	0.245
		15								0.175	0.177	0.217
		20									0.174	0.194
		25									0.177	0.180
		30									0.178	0.177
		40										0.176
		50										0.179
		60										0.179

Table 3.4: Empirical null rejection probabilities with block bootstrap critical values, 5% level, $T = 60$

ρ	θ	M	block bootstrap critical value									
			values of G									
			2	3	4	5	6	10	12	15	30	60
-0.5	-0.5	1	0.087	0.065	0.046	0.034	0.023	0.009	0.005	0.003	0.000	0.000
		2	0.087	0.067	0.049	0.038	0.029	0.014	0.010	0.007	0.001	0.000
		3		0.067	0.050	0.037	0.030	0.018	0.013	0.010	0.002	0.000
		4			0.050	0.038	0.029	0.018	0.015	0.011	0.004	0.000
		5				0.038	0.028	0.018	0.015	0.013	0.005	0.000
		6					0.028	0.018	0.015	0.014	0.006	0.000
		7						0.018	0.015	0.014	0.007	0.001
		8						0.018	0.014	0.013	0.009	0.001
		9						0.018	0.015	0.014	0.009	0.002
		10						0.018	0.015	0.014	0.010	0.002
		12							0.015	0.013	0.010	0.002
		15								0.013	0.010	0.003
		20									0.010	0.003
		25									0.009	0.003
		30									0.009	0.003
		40										0.004
		50										0.004
		60										0.004

Table 3.4: (cont'd)

ρ	θ	M	block bootstrap critical value									
			values of G									
			2	3	4	5	6	10	12	15	30	60
−0.5	0	1	0.115	0.082	0.072	0.064	0.054	0.045	0.040	0.035	0.021	0.002
		2	0.115	0.081	0.069	0.064	0.055	0.051	0.044	0.042	0.029	0.016
		3		0.081	0.070	0.063	0.057	0.051	0.044	0.045	0.034	0.016
		4			0.070	0.066	0.055	0.052	0.047	0.044	0.036	0.023
		5				0.066	0.057	0.050	0.047	0.045	0.039	0.025
		6					0.057	0.052	0.047	0.045	0.040	0.027
		7						0.053	0.047	0.046	0.041	0.029
		8						0.052	0.047	0.046	0.040	0.030
		9						0.052	0.046	0.046	0.040	0.031
		10						0.052	0.047	0.045	0.041	0.031
		12							0.046	0.045	0.042	0.032
		15								0.046	0.042	0.034
		20									0.041	0.036
		25									0.043	0.037
		30									0.042	0.037
		40										0.038
		50										0.037
		60										0.036

Table 3.4: (cont'd)

ρ	θ	M	block bootstrap critical value									
			values of G									
			2	3	4	5	6	10	12	15	30	60
−0.5	0.5	1	0.120	0.083	0.078	0.071	0.063	0.059	0.056	0.055	0.053	0.054
		2	0.120	0.082	0.075	0.072	0.063	0.059	0.056	0.055	0.051	0.053
		3		0.082	0.075	0.068	0.065	0.057	0.053	0.055	0.052	0.052
		4			0.075	0.069	0.064	0.060	0.054	0.052	0.052	0.051
		5				0.069	0.064	0.057	0.055	0.052	0.051	0.051
		6					0.064	0.059	0.055	0.052	0.052	0.051
		7						0.059	0.056	0.052	0.051	0.051
		8						0.058	0.055	0.054	0.051	0.050
		9						0.057	0.053	0.052	0.051	0.051
		10						0.057	0.054	0.053	0.051	0.050
		12							0.055	0.054	0.049	0.049
		15								0.053	0.051	0.049
		20									0.050	0.050
		25									0.051	0.050
		30									0.049	0.050
		40										0.049
		50										0.050
		60										0.050

Table 3.4: (cont'd)

ρ	θ	M	block bootstrap critical value									
			values of G									
			2	3	4	5	6	10	12	15	30	60
0	-0.5	1	0.105	0.076	0.064	0.053	0.044	0.028	0.023	0.015	0.003	0.000
		2	0.105	0.078	0.065	0.057	0.048	0.038	0.030	0.025	0.009	0.001
		3		0.078	0.066	0.057	0.048	0.037	0.032	0.031	0.014	0.003
		4			0.066	0.057	0.047	0.041	0.033	0.030	0.019	0.005
		5				0.057	0.048	0.040	0.035	0.032	0.020	0.007
		6					0.048	0.039	0.035	0.032	0.022	0.009
		7						0.039	0.035	0.031	0.022	0.010
		8						0.039	0.035	0.032	0.024	0.012
		9						0.039	0.034	0.032	0.024	0.013
		10						0.039	0.034	0.032	0.025	0.014
		12							0.034	0.033	0.026	0.017
		15								0.032	0.025	0.019
		20									0.025	0.019
		25									0.025	0.020
		30									0.025	0.021
		40										0.021
		50										0.022
		60										0.022

Table 3.4: (cont'd)

ρ	θ	M	block bootstrap critical value									
			values of G									
			2	3	4	5	6	10	12	15	30	60
0	0	1	0.120	0.083	0.078	0.071	0.063	0.059	0.056	0.055	0.053	0.054
		2	0.120	0.082	0.075	0.072	0.063	0.059	0.056	0.055	0.051	0.053
		3		0.082	0.075	0.068	0.065	0.057	0.053	0.055	0.052	0.052
		4			0.075	0.069	0.064	0.060	0.054	0.052	0.052	0.051
		5				0.069	0.064	0.057	0.055	0.052	0.051	0.051
		6					0.064	0.059	0.055	0.052	0.052	0.051
		7						0.059	0.056	0.052	0.051	0.051
		8						0.058	0.055	0.054	0.051	0.050
		9						0.057	0.053	0.052	0.051	0.051
		10						0.057	0.054	0.053	0.051	0.050
		12							0.055	0.054	0.049	0.049
		15								0.053	0.051	0.049
		20									0.050	0.050
		25									0.051	0.050
		30									0.049	0.050
		40										0.049
		50										0.050
		60										0.050

Table 3.4: (cont'd)

ρ	θ	M	block bootstrap critical value									
			values of G									
			2	3	4	5	6	10	12	15	30	60
0	0.5	1	0.121	0.084	0.076	0.074	0.066	0.064	0.062	0.066	0.086	0.150
		2	0.121	0.083	0.073	0.073	0.064	0.060	0.060	0.059	0.068	0.091
		3		0.083	0.073	0.071	0.065	0.060	0.057	0.058	0.062	0.076
		4			0.073	0.071	0.065	0.061	0.057	0.056	0.059	0.070
		5				0.071	0.066	0.059	0.057	0.055	0.058	0.067
		6					0.066	0.061	0.056	0.055	0.057	0.064
		7						0.061	0.056	0.057	0.056	0.062
		8						0.061	0.056	0.055	0.055	0.061
		9						0.060	0.056	0.056	0.055	0.060
		10						0.060	0.055	0.057	0.054	0.059
		12							0.056	0.055	0.055	0.059
		15								0.054	0.057	0.057
		20									0.056	0.057
		25									0.055	0.058
		30									0.055	0.058
		40										0.055
		50										0.056
		60										0.057

Table 3.4: (cont'd)

ρ	θ	M	block bootstrap critical value									
			values of G									
			2	3	4	5	6	10	12	15	30	60
0.5	−0.5	1	0.120	0.083	0.078	0.071	0.063	0.059	0.056	0.055	0.053	0.054
		2	0.120	0.082	0.075	0.072	0.063	0.059	0.056	0.055	0.051	0.053
		3		0.082	0.075	0.068	0.065	0.057	0.053	0.055	0.052	0.052
		4			0.075	0.069	0.064	0.060	0.054	0.052	0.052	0.051
		5				0.069	0.064	0.057	0.055	0.052	0.051	0.051
		6					0.064	0.059	0.055	0.052	0.052	0.051
		7						0.059	0.056	0.052	0.051	0.051
		8						0.058	0.055	0.054	0.051	0.050
		9						0.057	0.053	0.052	0.051	0.051
		10						0.057	0.054	0.053	0.051	0.050
		12							0.055	0.054	0.049	0.049
		15								0.053	0.051	0.049
		20									0.050	0.050
		25									0.051	0.050
		30									0.049	0.050
		40										0.049
		50										0.050
		60										0.050

Table 3.4: (cont'd)

ρ	θ	M	block bootstrap critical value									
			values of G									
			2	3	4	5	6	10	12	15	30	60
0.5	0	1	0.120	0.083	0.076	0.075	0.073	0.083	0.087	0.101	0.168	0.265
		2	0.120	0.080	0.071	0.072	0.066	0.070	0.070	0.073	0.109	0.177
		3		0.080	0.071	0.072	0.069	0.066	0.066	0.068	0.089	0.135
		4			0.071	0.072	0.068	0.067	0.067	0.065	0.078	0.114
		5				0.072	0.068	0.068	0.067	0.065	0.075	0.100
		6					0.068	0.067	0.064	0.065	0.073	0.092
		7						0.066	0.067	0.065	0.071	0.087
		8						0.066	0.065	0.066	0.070	0.083
		9						0.066	0.065	0.064	0.069	0.080
		10						0.066	0.064	0.064	0.069	0.078
		12							0.063	0.065	0.069	0.076
		15								0.064	0.067	0.073
		20									0.066	0.071
		25									0.067	0.071
		30									0.067	0.070
		40										0.068
		50										0.069
		60										0.070

Table 3.4: (cont'd)

ρ	θ	M	block bootstrap critical value									
			values of G									
			2	3	4	5	6	10	12	15	30	60
0.5	0.5	1	0.120	0.082	0.074	0.073	0.072	0.085	0.091	0.108	0.196	0.332
		2	0.120	0.080	0.068	0.070	0.067	0.070	0.072	0.076	0.117	0.205
		3		0.080	0.070	0.070	0.068	0.068	0.068	0.069	0.094	0.153
		4			0.070	0.070	0.068	0.068	0.066	0.067	0.082	0.126
		5				0.070	0.068	0.067	0.067	0.065	0.077	0.110
		6					0.068	0.067	0.066	0.065	0.076	0.101
		7						0.067	0.068	0.066	0.073	0.094
		8						0.066	0.067	0.066	0.072	0.089
		9						0.067	0.067	0.066	0.072	0.083
		10						0.067	0.066	0.065	0.072	0.082
		12							0.066	0.065	0.071	0.080
		15								0.065	0.070	0.077
		20									0.068	0.072
		25									0.070	0.074
		30									0.069	0.072
		40										0.073
		50										0.072
		60										0.074

Table 3.4: (cont'd)

ρ	θ	M	block bootstrap critical value									
			values of G									
			2	3	4	5	6	10	12	15	30	60
0.8	−0.5	1	0.115	0.081	0.082	0.087	0.100	0.145	0.168	0.203	0.301	0.379
		2	0.115	0.078	0.073	0.081	0.083	0.106	0.117	0.134	0.222	0.309
		3		0.078	0.075	0.079	0.083	0.094	0.099	0.111	0.175	0.262
		4			0.075	0.081	0.084	0.091	0.094	0.102	0.148	0.230
		5				0.081	0.085	0.092	0.093	0.098	0.131	0.202
		6					0.085	0.091	0.094	0.098	0.120	0.182
		7						0.092	0.093	0.096	0.114	0.165
		8						0.092	0.093	0.095	0.109	0.153
		9						0.091	0.094	0.095	0.107	0.144
		10						0.091	0.093	0.095	0.106	0.137
		12							0.094	0.095	0.102	0.126
		15								0.095	0.100	0.115
		20									0.098	0.107
		25									0.099	0.105
		30									0.100	0.103
		40										0.100
		50										0.102
		60										0.103

Table 3.4: (cont'd)

ρ	θ	M	block bootstrap critical value									
			values of G									
			2	3	4	5	6	10	12	15	30	60
0.8	0	1	0.108	0.076	0.080	0.087	0.102	0.164	0.201	0.244	0.394	0.537
		2	0.108	0.072	0.071	0.077	0.082	0.110	0.127	0.153	0.269	0.407
		3		0.072	0.071	0.074	0.082	0.098	0.107	0.121	0.206	0.330
		4			0.071	0.076	0.082	0.092	0.100	0.110	0.170	0.278
		5				0.076	0.083	0.092	0.097	0.104	0.149	0.243
		6					0.083	0.093	0.098	0.102	0.134	0.217
		7						0.094	0.097	0.103	0.127	0.197
		8						0.096	0.097	0.102	0.122	0.180
		9						0.095	0.098	0.101	0.116	0.165
		10						0.095	0.098	0.102	0.114	0.156
		12							0.098	0.102	0.111	0.143
		15								0.102	0.107	0.132
		20									0.107	0.121
		25									0.110	0.118
		30									0.110	0.115
		40										0.115
		50										0.116
		60										0.117

Table 3.4: (cont'd)

ρ	θ	M	block bootstrap critical value									
			values of G									
			2	3	4	5	6	10	12	15	30	60
0.8	0.5	1	0.106	0.073	0.080	0.085	0.101	0.167	0.201	0.251	0.409	0.559
		2	0.106	0.071	0.068	0.075	0.081	0.109	0.126	0.155	0.276	0.421
		3		0.071	0.071	0.073	0.080	0.098	0.108	0.122	0.209	0.341
		4			0.071	0.073	0.081	0.091	0.099	0.110	0.173	0.286
		5				0.073	0.083	0.092	0.098	0.105	0.150	0.249
		6					0.083	0.093	0.099	0.103	0.136	0.221
		7						0.094	0.098	0.103	0.129	0.200
		8						0.095	0.097	0.101	0.123	0.182
		9						0.096	0.098	0.102	0.118	0.170
		10						0.096	0.098	0.101	0.115	0.161
		12							0.098	0.103	0.112	0.147
		15								0.103	0.110	0.132
		20									0.110	0.122
		25									0.110	0.119
		30									0.112	0.117
		40										0.118
		50										0.118
		60										0.118

Table 3.4: (cont'd)

ρ	θ	M	block bootstrap critical value									
			values of G									
			2	3	4	5	6	10	12	15	30	60
0.9	−0.5	1	0.099	0.083	0.101	0.126	0.159	0.264	0.313	0.367	0.505	0.601
		2	0.099	0.077	0.091	0.106	0.119	0.182	0.210	0.256	0.397	0.518
		3		0.077	0.092	0.104	0.114	0.150	0.168	0.200	0.327	0.456
		4			0.092	0.106	0.114	0.139	0.150	0.175	0.281	0.408
		5				0.106	0.117	0.136	0.143	0.160	0.245	0.369
		6					0.117	0.134	0.140	0.153	0.219	0.338
		7						0.136	0.139	0.149	0.203	0.313
		8						0.138	0.140	0.146	0.189	0.291
		9						0.138	0.141	0.148	0.180	0.272
		10						0.138	0.142	0.147	0.174	0.257
		12							0.143	0.148	0.164	0.232
		15								0.150	0.158	0.206
		20									0.157	0.183
		25									0.160	0.173
		30									0.161	0.169
		40										0.167
		50										0.168
		60										0.170

Table 3.4: (cont'd)

ρ	θ	M	block bootstrap critical value									
			values of G									
			2	3	4	5	6	10	12	15	30	60
0.9	0	1	0.095	0.074	0.099	0.126	0.160	0.282	0.339	0.399	0.564	0.687
		2	0.095	0.071	0.085	0.101	0.117	0.184	0.219	0.270	0.432	0.573
		3		0.071	0.085	0.102	0.112	0.150	0.172	0.206	0.350	0.499
		4			0.085	0.103	0.113	0.137	0.152	0.177	0.298	0.445
		5				0.103	0.115	0.136	0.145	0.162	0.260	0.399
		6					0.115	0.134	0.141	0.154	0.232	0.364
		7						0.135	0.140	0.151	0.210	0.333
		8						0.137	0.140	0.148	0.195	0.308
		9						0.139	0.142	0.147	0.186	0.291
		10						0.139	0.143	0.148	0.178	0.272
		12							0.144	0.149	0.169	0.245
		15								0.151	0.161	0.217
		20									0.162	0.192
		25									0.166	0.181
		30									0.167	0.176
		40										0.174
		50										0.177
		60										0.177

Table 3.4: (cont'd)

ρ	θ	M	block bootstrap critical value									
			values of G									
			2	3	4	5	6	10	12	15	30	60
0.9	0.5	1	0.093	0.071	0.099	0.124	0.158	0.285	0.341	0.406	0.570	0.697
		2	0.093	0.068	0.084	0.099	0.116	0.183	0.220	0.271	0.437	0.582
		3		0.068	0.085	0.098	0.110	0.149	0.171	0.207	0.353	0.504
		4			0.085	0.101	0.110	0.137	0.153	0.178	0.301	0.447
		5				0.101	0.114	0.134	0.145	0.161	0.261	0.405
		6					0.114	0.134	0.140	0.154	0.232	0.366
		7						0.134	0.139	0.149	0.211	0.337
		8						0.136	0.139	0.147	0.196	0.311
		9						0.137	0.141	0.146	0.187	0.291
		10						0.137	0.142	0.146	0.178	0.275
		12							0.144	0.149	0.170	0.247
		15								0.151	0.164	0.218
		20									0.163	0.193
		25									0.167	0.182
		30									0.168	0.176
		40										0.176
		50										0.178
		60										0.180

Table 3.5: Empirical null rejection probabilities with *i.i.d.* bootstrap critical values, 5% level, $T = 60$

ρ	θ	M	iid bootstrap critical value									
			values of G									
			2	3	4	5	6	10	12	15	30	60
-0.5	-0.5	1	0.041	0.035	0.026	0.019	0.013	0.004	0.002	0.001	0.000	0.000
		2	0.041	0.033	0.029	0.021	0.017	0.009	0.005	0.004	0.001	0.000
		3		0.033	0.029	0.021	0.016	0.012	0.007	0.005	0.001	0.000
		4			0.029	0.021	0.017	0.011	0.009	0.007	0.003	0.000
		5				0.021	0.017	0.012	0.009	0.008	0.003	0.000
		6					0.017	0.011	0.008	0.008	0.004	0.000
		7						0.012	0.009	0.007	0.005	0.001
		8						0.011	0.009	0.007	0.006	0.001
		9						0.011	0.008	0.008	0.006	0.002
		10						0.011	0.009	0.008	0.005	0.002
		12							0.008	0.009	0.006	0.002
		15								0.008	0.006	0.003
		20									0.006	0.003
		25									0.006	0.003
		30									0.006	0.003
		40										0.004
		50										0.004
		60										0.004

Table 3.5: (cont'd)

ρ	θ	M	iid bootstrap critical value									
			values of G									
			2	3	4	5	6	10	12	15	30	60
−0.5	0	1	0.048	0.045	0.048	0.044	0.039	0.034	0.030	0.028	0.018	0.002
		2	0.048	0.046	0.047	0.045	0.040	0.040	0.036	0.035	0.025	0.016
		3		0.046	0.046	0.045	0.044	0.042	0.039	0.039	0.030	0.016
		4			0.046	0.045	0.042	0.043	0.039	0.039	0.033	0.023
		5				0.045	0.041	0.041	0.040	0.038	0.035	0.025
		6					0.041	0.042	0.040	0.040	0.036	0.027
		7						0.042	0.040	0.041	0.036	0.029
		8						0.042	0.039	0.040	0.036	0.030
		9						0.041	0.040	0.040	0.037	0.031
		10						0.041	0.040	0.039	0.037	0.031
		12							0.039	0.040	0.038	0.032
		15								0.040	0.039	0.034
		20									0.039	0.036
		25									0.039	0.037
		30									0.039	0.037
		40										0.038
		50										0.037
		60										0.036

Table 3.5: (cont'd)

ρ	θ	M	iid bootstrap critical value									
			values of G									
			2	3	4	5	6	10	12	15	30	60
-0.5	0.5	1	0.050	0.051	0.051	0.051	0.049	0.051	0.050	0.052	0.052	0.054
		2	0.050	0.050	0.050	0.051	0.049	0.051	0.051	0.052	0.052	0.053
		3		0.050	0.051	0.048	0.049	0.052	0.049	0.051	0.051	0.052
		4			0.051	0.049	0.048	0.051	0.050	0.049	0.051	0.051
		5				0.049	0.048	0.051	0.049	0.050	0.050	0.051
		6					0.048	0.051	0.050	0.051	0.048	0.051
		7						0.051	0.050	0.049	0.049	0.051
		8						0.051	0.050	0.047	0.049	0.050
		9						0.050	0.049	0.047	0.050	0.051
		10						0.050	0.050	0.048	0.050	0.050
		12							0.049	0.048	0.050	0.049
		15								0.047	0.049	0.049
		20									0.049	0.050
		25									0.049	0.050
		30									0.049	0.050
		40										0.049
		50										0.050
		60										0.050

Table 3.5: (cont'd)

ρ	θ	M	iid bootstrap critical value									
			values of G									
			2	3	4	5	6	10	12	15	30	60
0	−0.5	1	0.043	0.041	0.040	0.035	0.029	0.020	0.014	0.010	0.002	0.000
		2	0.043	0.041	0.039	0.039	0.033	0.027	0.021	0.018	0.006	0.001
		3		0.041	0.039	0.038	0.032	0.029	0.024	0.023	0.012	0.003
		4			0.039	0.038	0.035	0.030	0.026	0.025	0.013	0.005
		5				0.038	0.032	0.029	0.026	0.025	0.015	0.007
		6					0.032	0.031	0.025	0.026	0.018	0.009
		7						0.030	0.026	0.026	0.019	0.010
		8						0.030	0.026	0.027	0.020	0.012
		9						0.030	0.026	0.026	0.021	0.013
		10						0.030	0.026	0.026	0.022	0.014
		12							0.026	0.026	0.022	0.017
		15								0.026	0.024	0.019
		20									0.025	0.019
		25									0.024	0.020
		30									0.024	0.021
		40										0.021
		50										0.022
		60										0.022

Table 3.5: (cont'd)

ρ	θ	M	iid bootstrap critical value									
			values of G									
			2	3	4	5	6	10	12	15	30	60
0	0	1	0.050	0.051	0.051	0.051	0.049	0.051	0.050	0.052	0.052	0.054
		2	0.050	0.050	0.050	0.051	0.049	0.051	0.051	0.052	0.052	0.053
		3		0.050	0.051	0.048	0.049	0.052	0.049	0.051	0.051	0.052
		4			0.051	0.049	0.048	0.051	0.050	0.049	0.051	0.051
		5				0.049	0.048	0.051	0.049	0.050	0.050	0.051
		6					0.048	0.051	0.050	0.051	0.048	0.051
		7						0.051	0.050	0.049	0.049	0.051
		8						0.051	0.050	0.047	0.049	0.050
		9						0.050	0.049	0.047	0.050	0.051
		10						0.050	0.050	0.048	0.050	0.050
		12							0.049	0.048	0.050	0.049
		15								0.047	0.049	0.049
		20									0.049	0.050
		25									0.049	0.050
		30									0.049	0.050
		40										0.049
		50										0.050
		60										0.050

Table 3.5: (cont'd)

ρ	θ	M	iid bootstrap critical value									
			values of G									
			2	3	4	5	6	10	12	15	30	60
0	0.5	1	0.048	0.050	0.053	0.055	0.056	0.059	0.061	0.064	0.087	0.150
		2	0.048	0.049	0.052	0.052	0.052	0.056	0.057	0.058	0.069	0.091
		3		0.049	0.052	0.051	0.054	0.055	0.055	0.057	0.062	0.076
		4			0.052	0.053	0.053	0.053	0.055	0.054	0.059	0.070
		5				0.053	0.053	0.055	0.054	0.055	0.058	0.067
		6					0.053	0.055	0.055	0.055	0.058	0.064
		7						0.053	0.054	0.055	0.057	0.062
		8						0.053	0.054	0.054	0.056	0.061
		9						0.055	0.053	0.054	0.056	0.060
		10						0.055	0.054	0.056	0.055	0.059
		12							0.054	0.055	0.057	0.059
		15								0.055	0.057	0.057
		20									0.055	0.057
		25									0.055	0.058
		30									0.056	0.058
		40										0.055
		50										0.056
		60										0.057

Table 3.5: (cont'd)

ρ	θ	M	iid bootstrap critical value									
			values of G									
			2	3	4	5	6	10	12	15	30	60
0.5	-0.5	1	0.050	0.051	0.051	0.051	0.049	0.051	0.050	0.052	0.052	0.054
		2	0.050	0.050	0.050	0.051	0.049	0.051	0.051	0.052	0.052	0.053
		3		0.050	0.051	0.048	0.049	0.052	0.049	0.051	0.051	0.052
		4			0.051	0.049	0.048	0.051	0.050	0.049	0.051	0.051
		5				0.049	0.048	0.051	0.049	0.050	0.050	0.051
		6					0.048	0.051	0.050	0.051	0.048	0.051
		7						0.051	0.050	0.049	0.049	0.051
		8						0.051	0.050	0.047	0.049	0.050
		9						0.050	0.049	0.047	0.050	0.051
		10						0.050	0.050	0.048	0.050	0.050
		12							0.049	0.048	0.050	0.049
		15								0.047	0.049	0.049
		20									0.049	0.050
		25									0.049	0.050
		30									0.049	0.050
		40										0.049
		50										0.050
		60										0.050

Table 3.5: (cont'd)

ρ	θ	M	iid bootstrap critical value									
			values of G									
			2	3	4	5	6	10	12	15	30	60
0.5	0	1	0.050	0.055	0.055	0.060	0.065	0.080	0.085	0.102	0.173	0.265
		2	0.050	0.053	0.055	0.058	0.060	0.068	0.071	0.075	0.110	0.177
		3		0.053	0.055	0.057	0.059	0.065	0.066	0.070	0.090	0.135
		4			0.055	0.058	0.059	0.065	0.065	0.069	0.081	0.114
		5				0.058	0.058	0.065	0.067	0.067	0.077	0.100
		6					0.058	0.063	0.065	0.067	0.075	0.092
		7						0.063	0.064	0.067	0.074	0.087
		8						0.065	0.063	0.065	0.072	0.083
		9						0.066	0.064	0.064	0.070	0.080
		10						0.066	0.065	0.066	0.071	0.078
		12							0.066	0.065	0.071	0.076
		15								0.065	0.069	0.073
		20									0.068	0.071
		25									0.068	0.071
		30									0.068	0.070
		40										0.068
		50										0.069
		60										0.070

Table 3.5: (cont'd)

ρ	θ	M	iid bootstrap critical value									
			values of G									
			2	3	4	5	6	10	12	15	30	60
0.5	0.5	1	0.050	0.056	0.057	0.061	0.067	0.086	0.095	0.115	0.201	0.332
		2	0.050	0.055	0.056	0.059	0.060	0.072	0.074	0.081	0.122	0.205
		3		0.055	0.056	0.060	0.060	0.066	0.071	0.074	0.098	0.153
		4			0.056	0.059	0.062	0.067	0.068	0.070	0.087	0.126
		5				0.059	0.060	0.066	0.067	0.069	0.080	0.110
		6					0.060	0.064	0.067	0.070	0.079	0.101
		7						0.066	0.066	0.069	0.076	0.094
		8						0.068	0.067	0.068	0.075	0.089
		9						0.068	0.067	0.068	0.074	0.083
		10						0.068	0.068	0.070	0.072	0.082
		12							0.068	0.068	0.072	0.080
		15								0.070	0.071	0.077
		20									0.071	0.072
		25									0.072	0.074
		30									0.072	0.072
		40										0.073
		50										0.072
		60										0.074

Table 3.5: (cont'd)

ρ	θ	M	iid bootstrap critical value									
			values of G									
			2	3	4	5	6	10	12	15	30	60
0.8	−0.5	1	0.052	0.063	0.073	0.088	0.101	0.157	0.181	0.212	0.303	0.379
		2	0.052	0.061	0.068	0.076	0.084	0.112	0.126	0.145	0.226	0.309
		3		0.061	0.071	0.076	0.084	0.100	0.106	0.120	0.178	0.262
		4			0.071	0.077	0.084	0.097	0.099	0.106	0.151	0.230
		5				0.077	0.084	0.093	0.096	0.102	0.135	0.202
		6					0.084	0.094	0.096	0.101	0.123	0.182
		7						0.094	0.095	0.099	0.116	0.165
		8						0.094	0.095	0.098	0.112	0.153
		9						0.096	0.094	0.098	0.107	0.144
		10						0.096	0.095	0.097	0.106	0.137
		12							0.097	0.097	0.106	0.126
		15								0.099	0.102	0.115
		20									0.099	0.107
		25									0.101	0.105
		30									0.102	0.103
		40										0.100
		50										0.102
		60										0.103

Table 3.5: (cont'd)

ρ	θ	M	iid bootstrap critical value									
			values of G									
			2	3	4	5	6	10	12	15	30	60
0.8	0	1	0.054	0.064	0.079	0.095	0.117	0.186	0.220	0.261	0.400	0.537
		2	0.054	0.064	0.073	0.083	0.092	0.126	0.143	0.169	0.276	0.407
		3		0.064	0.077	0.083	0.092	0.111	0.119	0.135	0.215	0.330
		4			0.077	0.084	0.093	0.106	0.111	0.121	0.179	0.278
		5				0.084	0.094	0.105	0.110	0.114	0.154	0.243
		6					0.094	0.105	0.108	0.113	0.142	0.217
		7						0.105	0.107	0.112	0.134	0.197
		8						0.105	0.106	0.109	0.127	0.180
		9						0.108	0.108	0.109	0.123	0.165
		10						0.108	0.110	0.110	0.120	0.156
		12							0.111	0.111	0.118	0.143
		15								0.112	0.113	0.132
		20									0.113	0.121
		25									0.116	0.118
		30									0.116	0.115
		40										0.115
		50										0.116
		60										0.117

Table 3.5: (cont'd)

ρ	θ	M	iid bootstrap critical value									
			values of G									
			2	3	4	5	6	10	12	15	30	60
0.8	0.5	1	0.053	0.066	0.080	0.097	0.118	0.190	0.225	0.269	0.414	0.559
		2	0.053	0.063	0.074	0.083	0.094	0.128	0.147	0.172	0.283	0.421
		3		0.063	0.077	0.085	0.092	0.112	0.124	0.138	0.218	0.341
		4			0.077	0.085	0.092	0.107	0.114	0.122	0.180	0.286
		5				0.085	0.094	0.108	0.110	0.115	0.158	0.249
		6					0.094	0.106	0.111	0.115	0.146	0.221
		7						0.108	0.110	0.111	0.136	0.200
		8						0.108	0.109	0.113	0.130	0.182
		9						0.109	0.110	0.113	0.124	0.170
		10						0.109	0.113	0.113	0.121	0.161
		12							0.112	0.113	0.118	0.147
		15								0.115	0.116	0.132
		20									0.116	0.122
		25									0.116	0.119
		30									0.117	0.117
		40										0.118
		50										0.118
		60										0.118

Table 3.5: (cont'd)

ρ	θ	M	iid bootstrap critical value									
			values of G									
			2	3	4	5	6	10	12	15	30	60
0.9	−0.5	1	0.060	0.085	0.123	0.157	0.195	0.306	0.344	0.388	0.512	0.601
		2	0.060	0.083	0.106	0.126	0.146	0.209	0.241	0.278	0.404	0.518
		3		0.083	0.109	0.122	0.138	0.173	0.193	0.222	0.335	0.456
		4			0.109	0.126	0.135	0.160	0.174	0.191	0.288	0.408
		5				0.126	0.141	0.156	0.165	0.177	0.254	0.369
		6					0.141	0.154	0.160	0.169	0.230	0.338
		7						0.155	0.158	0.165	0.213	0.313
		8						0.157	0.157	0.162	0.198	0.291
		9						0.157	0.160	0.161	0.188	0.272
		10						0.157	0.161	0.162	0.182	0.257
		12							0.162	0.164	0.173	0.232
		15								0.166	0.168	0.206
		20									0.165	0.183
		25									0.167	0.173
		30									0.169	0.169
		40										0.167
		50										0.168
		60										0.170

Table 3.5: (cont'd)

ρ	θ	M	iid bootstrap critical value									
			values of G									
			2	3	4	5	6	10	12	15	30	60
0.9	0	1	0.060	0.089	0.127	0.164	0.205	0.329	0.374	0.426	0.568	0.687
		2	0.060	0.085	0.108	0.131	0.152	0.221	0.253	0.297	0.440	0.573
		3		0.085	0.113	0.127	0.142	0.181	0.202	0.235	0.361	0.499
		4			0.113	0.131	0.141	0.168	0.181	0.200	0.307	0.445
		5				0.131	0.146	0.163	0.171	0.185	0.270	0.399
		6					0.146	0.159	0.167	0.177	0.242	0.364
		7						0.161	0.166	0.172	0.223	0.333
		8						0.164	0.165	0.170	0.207	0.308
		9						0.166	0.167	0.169	0.196	0.291
		10						0.166	0.168	0.170	0.190	0.272
		12							0.169	0.171	0.180	0.245
		15								0.174	0.175	0.217
		20									0.173	0.192
		25									0.176	0.181
		30									0.176	0.176
		40										0.174
		50										0.177
		60										0.177

Table 3.5: (cont'd)

ρ	θ	M	iid bootstrap critical value									
			values of G									
			2	3	4	5	6	10	12	15	30	60
0.9	0.5	1	0.060	0.087	0.127	0.166	0.205	0.332	0.378	0.432	0.575	0.697
		2	0.060	0.086	0.110	0.130	0.153	0.224	0.255	0.299	0.445	0.582
		3		0.086	0.115	0.127	0.141	0.183	0.203	0.235	0.364	0.504
		4			0.115	0.133	0.141	0.167	0.181	0.203	0.309	0.447
		5				0.133	0.146	0.164	0.173	0.187	0.272	0.405
		6					0.146	0.161	0.168	0.177	0.245	0.366
		7						0.162	0.165	0.173	0.224	0.337
		8						0.166	0.165	0.170	0.209	0.311
		9						0.167	0.167	0.168	0.198	0.291
		10						0.167	0.169	0.169	0.191	0.275
		12							0.170	0.172	0.181	0.247
		15								0.174	0.175	0.218
		20									0.175	0.193
		25									0.177	0.182
		30									0.178	0.176
		40										0.176
		50										0.178
		60										0.180

Table 3.6: Average Type II Error, 5% level, $T = 60$

ρ	M	Values of G									
		2	3	4	5	6	10	12	15	30	60
0	1	0.878	0.783	0.704	0.658	0.630	0.576	0.567	0.557	0.546	0.539
	2	0.878	0.783	0.719	0.683	0.656	0.595	0.586	0.578	0.550	0.538
	3		0.783	0.713	0.684	0.668	0.615	0.597	0.588	0.556	0.544
	4			0.713	0.678	0.672	0.633	0.616	0.597	0.565	0.549
	5				0.678	0.669	0.638	0.625	0.608	0.571	0.550
	6					0.669	0.643	0.638	0.619	0.579	0.557
	7						0.643	0.633	0.623	0.582	0.560
	8						0.646	0.635	0.625	0.589	0.564
	9						0.641	0.638	0.626	0.596	0.571
	10						0.641	0.634	0.626	0.601	0.570
	12							0.630	0.628	0.614	0.577
	15								0.631	0.626	0.585
	20									0.626	0.600
	25									0.627	0.614
	30									0.627	0.623
	40										0.627
	50										0.627
	60										0.625

Table 3.6: (cont'd)

ρ	M	Values of G									
		2	3	4	5	6	10	12	15	30	60
0.5	1	0.875	0.770	0.692	0.650	0.631	0.572	0.563	0.558	0.540	0.537
	2	0.875	0.784	0.712	0.671	0.656	0.598	0.586	0.576	0.552	0.542
	3		0.784	0.713	0.677	0.665	0.611	0.604	0.595	0.563	0.545
	4			0.713	0.674	0.671	0.634	0.621	0.603	0.573	0.550
	5				0.674	0.667	0.638	0.629	0.611	0.580	0.556
	6					0.667	0.647	0.639	0.627	0.585	0.564
	7						0.648	0.641	0.633	0.592	0.567
	8						0.648	0.649	0.633	0.601	0.572
	9						0.644	0.646	0.633	0.600	0.576
	10						0.644	0.644	0.642	0.611	0.579
	12							0.643	0.644	0.621	0.585
	15								0.642	0.634	0.597
	20									0.640	0.611
	25									0.638	0.621
	30									0.637	0.630
	40										0.639
	50										0.638
	60										0.636

Table 3.6: (cont'd)

ρ	M	Values of G									
		2	3	4	5	6	10	12	15	30	60
0.8	1	0.869	0.748	0.681	0.640	0.611	0.564	0.556	0.555	0.538	0.527
	2	0.869	0.758	0.705	0.658	0.646	0.603	0.583	0.573	0.548	0.537
	3		0.758	0.697	0.670	0.666	0.625	0.612	0.597	0.560	0.543
	4			0.697	0.675	0.676	0.636	0.631	0.613	0.572	0.549
	5				0.675	0.673	0.645	0.639	0.629	0.588	0.554
	6					0.673	0.655	0.652	0.637	0.597	0.561
	7						0.659	0.652	0.645	0.605	0.566
	8						0.656	0.654	0.646	0.613	0.572
	9						0.659	0.659	0.651	0.623	0.577
	10						0.659	0.661	0.655	0.631	0.587
	12							0.661	0.657	0.635	0.595
	15								0.657	0.646	0.610
	20									0.651	0.629
	25									0.656	0.639
	30									0.656	0.644
	40										0.651
	50										0.655
	60										0.655

Table 3.6: (cont'd)

ρ	M	Values of G									
		2	3	4	5	6	10	12	15	30	60
0.9	1	0.867	0.739	0.659	0.613	0.580	0.550	0.542	0.536	0.526	0.515
	2	0.867	0.752	0.689	0.656	0.637	0.586	0.571	0.558	0.539	0.525
	3		0.752	0.688	0.670	0.653	0.622	0.597	0.584	0.548	0.535
	4			0.688	0.669	0.654	0.636	0.626	0.607	0.559	0.540
	5				0.669	0.661	0.648	0.638	0.624	0.571	0.546
	6					0.661	0.652	0.644	0.637	0.584	0.549
	7						0.654	0.650	0.644	0.594	0.554
	8						0.655	0.648	0.652	0.606	0.560
	9						0.657	0.652	0.652	0.617	0.565
	10						0.657	0.654	0.649	0.625	0.570
	12							0.654	0.652	0.636	0.583
	15								0.653	0.647	0.602
	20									0.647	0.626
	25									0.653	0.638
	30									0.654	0.646
	40										0.647
	50										0.653
	60										0.654

Table 3.7: Daily Data, Weekends Missing, G Block Bootstrap, 5% level

ρ	θ	M	overlapping G block bootstrap critical value									
			values of G									
			2	3	4	5	6	10	12	15	30	60
−0.5	0	1	0.115	0.084	0.069	0.061	0.061	0.042	0.051	0.035	0.025	0.004
		2	0.115	0.082	0.069	0.061	0.061	0.044	0.052	0.043	0.034	0.019
		3		0.082	0.070	0.062	0.061	0.046	0.053	0.044	0.038	0.021
		4			0.070	0.061	0.061	0.046	0.051	0.045	0.041	0.029
		5				0.061	0.059	0.046	0.053	0.047	0.042	0.030
		6					0.059	0.046	0.053	0.046	0.043	0.032
		7						0.046	0.054	0.046	0.044	0.033
		8						0.046	0.054	0.048	0.043	0.034
		9						0.045	0.053	0.047	0.043	0.036
		10						0.045	0.052	0.047	0.046	0.035
		12							0.053	0.047	0.043	0.038
		15								0.046	0.043	0.037
		20									0.044	0.040
		25									0.043	0.041
		30									0.044	0.040
		40										0.040
		50										0.039
		60										0.040

Table 3.7: (cont'd)

ρ	θ	M	overlapping G block bootstrap critical value									
			values of G									
			2	3	4	5	6	10	12	15	30	60
0	0	1	0.123	0.085	0.076	0.069	0.065	0.059	0.057	0.057	0.055	0.052
		2	0.123	0.084	0.075	0.068	0.064	0.057	0.056	0.055	0.055	0.050
		3		0.084	0.076	0.068	0.064	0.060	0.055	0.053	0.055	0.053
		4			0.076	0.067	0.065	0.058	0.055	0.053	0.053	0.053
		5				0.067	0.063	0.058	0.057	0.053	0.052	0.052
		6					0.063	0.058	0.058	0.054	0.051	0.053
		7						0.058	0.058	0.055	0.052	0.052
		8						0.058	0.057	0.055	0.052	0.053
		9						0.058	0.057	0.054	0.053	0.053
		10						0.058	0.057	0.054	0.054	0.051
		12							0.058	0.053	0.053	0.051
		15								0.054	0.055	0.053
		20									0.053	0.051
		25									0.052	0.053
		30									0.052	0.052
		40										0.053
		50										0.052
		60										0.051

Table 3.7: (cont'd)

			overlapping G block bootstrap critical value									
ρ	θ	M	values of G									
			2	3	4	5	6	10	12	15	30	60
0.1	0	1	0.126	0.083	0.077	0.070	0.066	0.061	0.056	0.062	0.063	0.071
		2	0.126	0.082	0.074	0.069	0.065	0.060	0.056	0.056	0.059	0.061
		3		0.082	0.075	0.070	0.064	0.060	0.056	0.055	0.057	0.059
		4			0.075	0.070	0.065	0.060	0.056	0.055	0.054	0.058
		5				0.070	0.064	0.059	0.058	0.055	0.055	0.056
		6					0.064	0.059	0.059	0.056	0.054	0.057
		7						0.059	0.059	0.056	0.055	0.055
		8						0.059	0.057	0.056	0.056	0.055
		9						0.060	0.057	0.054	0.055	0.055
		10						0.060	0.057	0.055	0.055	0.054
		12							0.058	0.055	0.056	0.054
		15								0.055	0.057	0.055
		20									0.055	0.054
		25									0.054	0.055
		30									0.053	0.055
		40										0.054
		50										0.054
		60										0.053

Table 3.7: (cont'd)

ρ	θ	M	overlapping G block bootstrap critical value									
			values of G									
			2	3	4	5	6	10	12	15	30	60
0.3	0	1	0.126	0.083	0.076	0.074	0.067	0.068	0.056	0.073	0.086	0.124
		2	0.126	0.083	0.074	0.071	0.065	0.063	0.055	0.061	0.069	0.089
		3		0.083	0.074	0.071	0.064	0.061	0.055	0.060	0.063	0.075
		4			0.074	0.071	0.064	0.063	0.056	0.059	0.060	0.068
		5				0.071	0.064	0.062	0.056	0.059	0.059	0.064
		6					0.064	0.063	0.057	0.058	0.059	0.063
		7						0.063	0.056	0.059	0.059	0.061
		8						0.061	0.056	0.058	0.059	0.060
		9						0.063	0.057	0.058	0.058	0.059
		10						0.063	0.058	0.058	0.057	0.057
		12							0.057	0.058	0.057	0.057
		15								0.059	0.058	0.058
		20									0.057	0.061
		25									0.057	0.060
		30									0.058	0.059
		40										0.058
		50										0.059
		60										0.057

Table 3.7: (cont'd)

ρ	θ	M	overlapping G block bootstrap critical value									
			values of G									
			2	3	4	5	6	10	12	15	30	60
0.5	0	1	0.123	0.085	0.074	0.075	0.066	0.075	0.059	0.084	0.129	0.206
		2	0.123	0.082	0.070	0.073	0.065	0.066	0.058	0.067	0.088	0.135
		3		0.082	0.071	0.071	0.064	0.063	0.056	0.064	0.075	0.105
		4			0.071	0.072	0.063	0.063	0.057	0.063	0.066	0.091
		5				0.072	0.063	0.064	0.057	0.061	0.065	0.082
		6					0.063	0.066	0.056	0.062	0.063	0.077
		7						0.066	0.057	0.062	0.063	0.072
		8						0.067	0.057	0.061	0.063	0.070
		9						0.066	0.057	0.061	0.063	0.067
		10						0.066	0.058	0.062	0.063	0.066
		12							0.057	0.063	0.064	0.064
		15								0.063	0.063	0.064
		20									0.063	0.065
		25									0.064	0.066
		30									0.064	0.064
		40										0.063
		50										0.063
		60										0.063

Table 3.7: (cont'd)

ρ	θ	M	overlapping G block bootstrap critical value									
			values of G									
			2	3	4	5	6	10	12	15	30	60
0.8	0	1	0.114	0.073	0.073	0.080	0.079	0.122	0.123	0.179	0.312	0.452
		2	0.114	0.074	0.069	0.071	0.068	0.087	0.089	0.114	0.198	0.327
		3		0.074	0.069	0.073	0.069	0.080	0.079	0.093	0.150	0.254
		4			0.069	0.074	0.069	0.080	0.075	0.087	0.125	0.209
		5				0.074	0.069	0.078	0.075	0.084	0.111	0.180
		6					0.069	0.079	0.075	0.082	0.101	0.158
		7						0.078	0.077	0.083	0.096	0.145
		8						0.079	0.077	0.083	0.092	0.135
		9						0.079	0.074	0.084	0.089	0.125
		10						0.079	0.075	0.083	0.089	0.119
		12							0.076	0.082	0.089	0.110
		15								0.083	0.086	0.100
		20									0.087	0.095
		25									0.089	0.094
		30									0.088	0.095
		40										0.094
		50										0.095
		60										0.095

Table 3.7: (cont'd)

ρ	θ	M	overlapping G block bootstrap critical value									
			values of G									
			2	3	4	5	6	10	12	15	30	60
0.9	0	1	0.099	0.069	0.081	0.100	0.117	0.219	0.242	0.319	0.483	0.616
		2	0.099	0.067	0.075	0.083	0.090	0.138	0.156	0.204	0.346	0.496
		3		0.067	0.075	0.084	0.087	0.117	0.125	0.156	0.275	0.415
		4			0.075	0.087	0.091	0.112	0.114	0.137	0.226	0.360
		5				0.087	0.094	0.109	0.112	0.124	0.194	0.318
		6					0.094	0.110	0.110	0.121	0.175	0.286
		7						0.111	0.109	0.120	0.160	0.258
		8						0.112	0.111	0.121	0.150	0.240
		9						0.114	0.113	0.119	0.146	0.224
		10						0.114	0.113	0.118	0.141	0.207
		12							0.113	0.120	0.135	0.185
		15								0.122	0.130	0.166
		20									0.130	0.151
		25									0.133	0.143
		30									0.136	0.140
		40										0.139
		50										0.141
		60										0.143

Table 3.8: Daily Data, Weekends Missing, *i.i.d*. Bootstrap, 5% level

			iid bootstrap critical value									
ρ	θ	M	values of G									
			2	3	4	5	6	10	12	15	30	60
-0.5	0	1	0.051	0.050	0.050	0.043	0.047	0.037	0.048	0.032	0.021	0.004
		2	0.051	0.051	0.049	0.044	0.050	0.040	0.049	0.039	0.031	0.019
		3		0.051	0.048	0.046	0.049	0.042	0.048	0.040	0.035	0.021
		4			0.048	0.044	0.050	0.043	0.049	0.041	0.036	0.029
		5				0.044	0.050	0.041	0.049	0.041	0.038	0.030
		6					0.050	0.042	0.049	0.041	0.041	0.032
		7						0.041	0.050	0.042	0.039	0.033
		8						0.041	0.048	0.041	0.041	0.034
		9						0.040	0.050	0.042	0.041	0.036
		10						0.040	0.048	0.042	0.042	0.035
		12							0.048	0.042	0.041	0.038
		15								0.043	0.040	0.037
		20									0.041	0.040
		25									0.041	0.041
		30									0.042	0.040
		40										0.040
		50										0.039
		60										0.040

Table 3.8: (cont'd)

ρ	θ	M	iid bootstrap critical value									
			values of G									
			2	3	4	5	6	10	12	15	30	60
0	0	1	0.048	0.051	0.052	0.051	0.051	0.054	0.053	0.053	0.053	0.052
		2	0.048	0.050	0.051	0.052	0.050	0.051	0.052	0.053	0.054	0.050
		3		0.050	0.050	0.051	0.050	0.052	0.051	0.053	0.053	0.053
		4			0.050	0.050	0.051	0.052	0.051	0.052	0.051	0.053
		5				0.050	0.051	0.052	0.051	0.053	0.051	0.052
		6					0.051	0.053	0.053	0.054	0.051	0.053
		7						0.052	0.052	0.052	0.051	0.052
		8						0.051	0.052	0.051	0.051	0.053
		9						0.051	0.051	0.051	0.051	0.053
		10						0.051	0.052	0.051	0.051	0.051
		12							0.051	0.050	0.053	0.051
		15								0.050	0.053	0.053
		20									0.053	0.051
		25									0.051	0.053
		30									0.052	0.052
		40										0.053
		50										0.052
		60										0.051

Table 3.8: (cont'd)

ρ	θ	M	iid bootstrap critical value									
			values of G									
			2	3	4	5	6	10	12	15	30	60
0.1	0	1	0.046	0.051	0.051	0.052	0.050	0.055	0.051	0.058	0.063	0.071
		2	0.046	0.050	0.050	0.051	0.049	0.053	0.051	0.056	0.058	0.061
		3		0.050	0.050	0.051	0.051	0.054	0.051	0.055	0.054	0.059
		4			0.050	0.051	0.050	0.054	0.051	0.054	0.054	0.058
		5				0.051	0.050	0.053	0.053	0.055	0.053	0.056
		6					0.050	0.053	0.053	0.055	0.053	0.057
		7						0.054	0.051	0.054	0.053	0.055
		8						0.054	0.051	0.055	0.054	0.055
		9						0.054	0.051	0.052	0.055	0.055
		10						0.054	0.051	0.053	0.054	0.054
		12							0.051	0.052	0.055	0.054
		15								0.052	0.055	0.055
		20									0.054	0.054
		25									0.053	0.055
		30									0.054	0.055
		40										0.054
		50										0.054
		60										0.053

Table 3.8: (cont'd)

ρ	θ	M	iid bootstrap critical value									
			values of G									
			2	3	4	5	6	10	12	15	30	60
0.3	0	1	0.047	0.053	0.050	0.054	0.050	0.059	0.050	0.067	0.087	0.124
		2	0.047	0.052	0.048	0.053	0.048	0.057	0.049	0.058	0.067	0.089
		3		0.052	0.048	0.055	0.051	0.057	0.051	0.056	0.062	0.075
		4			0.048	0.054	0.053	0.057	0.052	0.055	0.060	0.068
		5				0.054	0.051	0.057	0.051	0.058	0.057	0.064
		6					0.051	0.058	0.053	0.058	0.056	0.063
		7						0.057	0.051	0.058	0.057	0.061
		8						0.058	0.052	0.058	0.058	0.060
		9						0.058	0.051	0.057	0.058	0.059
		10						0.058	0.052	0.057	0.059	0.057
		12							0.051	0.057	0.059	0.057
		15								0.057	0.059	0.058
		20									0.057	0.061
		25									0.057	0.060
		30									0.056	0.059
		40										0.058
		50										0.059
		60										0.057

Table 3.8: (cont'd)

ρ	θ	M	iid bootstrap critical value									
			values of G									
			2	3	4	5	6	10	12	15	30	60
0.5	0	1	0.044	0.053	0.049	0.057	0.051	0.069	0.058	0.085	0.131	0.206
		2	0.044	0.053	0.049	0.056	0.052	0.061	0.053	0.065	0.088	0.135
		3		0.053	0.049	0.057	0.053	0.061	0.052	0.061	0.076	0.105
		4			0.049	0.057	0.054	0.059	0.053	0.061	0.069	0.091
		5				0.057	0.052	0.059	0.055	0.062	0.065	0.082
		6					0.052	0.060	0.056	0.063	0.063	0.077
		7						0.060	0.055	0.061	0.063	0.072
		8						0.060	0.054	0.062	0.064	0.070
		9						0.059	0.055	0.062	0.065	0.067
		10						0.059	0.056	0.062	0.065	0.066
		12							0.056	0.061	0.065	0.064
		15								0.061	0.063	0.064
		20									0.062	0.065
		25									0.062	0.066
		30									0.062	0.064
		40										0.063
		50										0.063
		60										0.063

Table 3.8: (cont'd)

ρ	θ	M	iid bootstrap critical value									
			values of G									
			2	3	4	5	6	10	12	15	30	60
0.8	0	1	0.050	0.061	0.062	0.080	0.080	0.139	0.138	0.194	0.320	0.452
		2	0.050	0.061	0.062	0.071	0.071	0.097	0.098	0.127	0.206	0.327
		3		0.061	0.063	0.072	0.070	0.087	0.087	0.104	0.157	0.254
		4			0.063	0.071	0.071	0.086	0.082	0.093	0.134	0.209
		5				0.071	0.070	0.085	0.082	0.091	0.118	0.180
		6					0.070	0.087	0.081	0.090	0.108	0.158
		7						0.087	0.082	0.091	0.101	0.145
		8						0.087	0.084	0.091	0.097	0.135
		9						0.088	0.083	0.092	0.095	0.125
		10						0.088	0.086	0.092	0.094	0.119
		12							0.086	0.091	0.094	0.110
		15								0.092	0.094	0.100
		20									0.093	0.095
		25									0.095	0.094
		30									0.095	0.095
		40										0.094
		50										0.095
		60										0.095

Table 3.8: (cont'd)

ρ	θ	M	iid bootstrap critical value									
			values of G									
			2	3	4	5	6	10	12	15	30	60
0.9	0	1	0.057	0.072	0.095	0.126	0.145	0.255	0.273	0.340	0.490	0.616
		2	0.057	0.072	0.088	0.101	0.110	0.165	0.182	0.229	0.356	0.496
		3		0.072	0.089	0.103	0.105	0.139	0.145	0.177	0.284	0.415
		4			0.089	0.103	0.109	0.131	0.135	0.154	0.236	0.360
		5				0.103	0.110	0.129	0.128	0.145	0.206	0.318
		6					0.110	0.128	0.127	0.139	0.184	0.286
		7						0.127	0.127	0.136	0.171	0.258
		8						0.130	0.126	0.135	0.162	0.240
		9						0.131	0.128	0.134	0.154	0.224
		10						0.131	0.129	0.134	0.150	0.207
		12							0.130	0.137	0.143	0.185
		15								0.140	0.139	0.166
		20									0.138	0.151
		25									0.141	0.143
		30									0.142	0.140
		40										0.139
		50										0.141
		60										0.143

Figure 3.1: Size Adjusted Power Comparision based on $G = 60$, $M = 30$ case

For interpretation of the references to color in this and all other figures, the reader is referred to the electronic version of this dissertation.

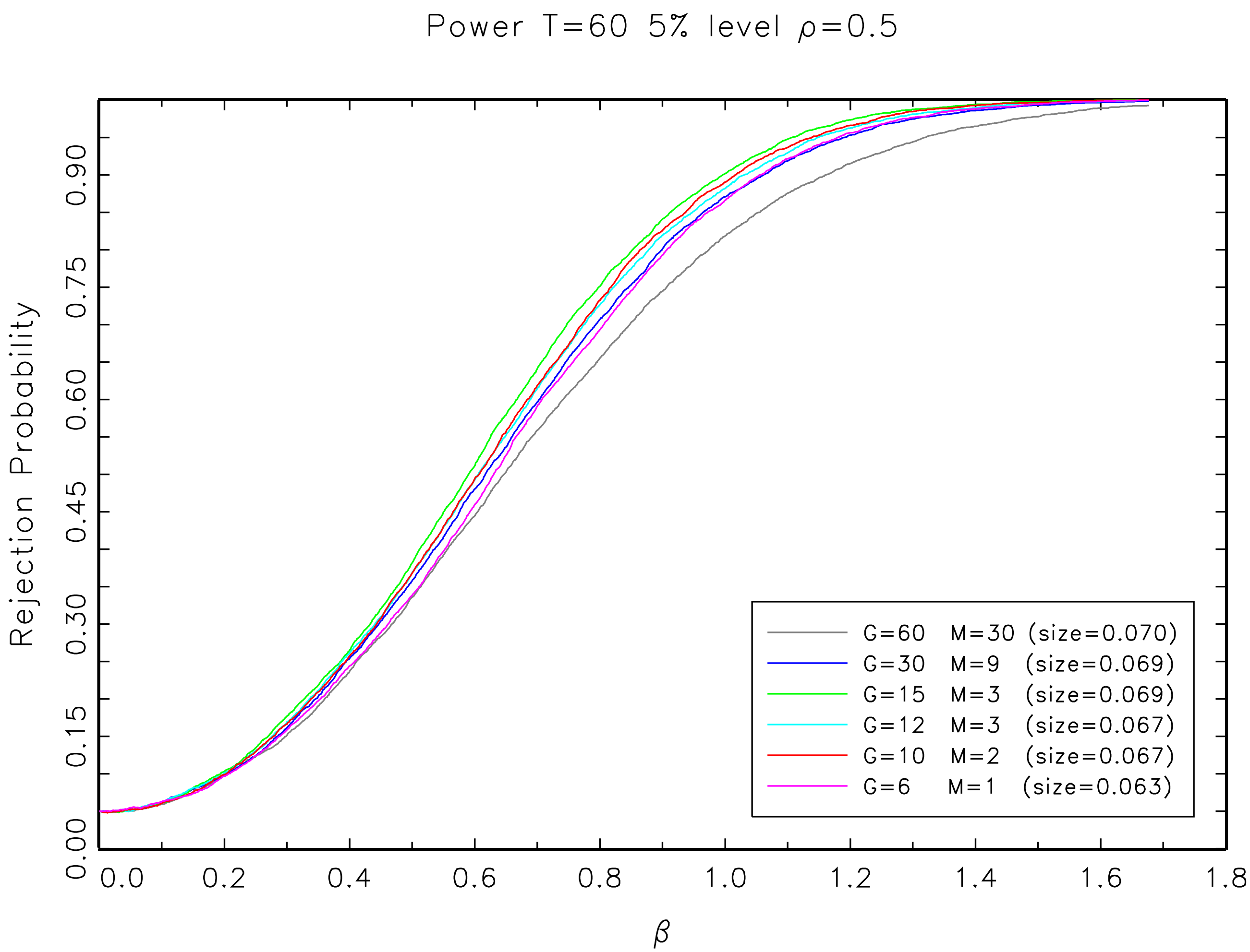


Figure 3.1: (cont'd)

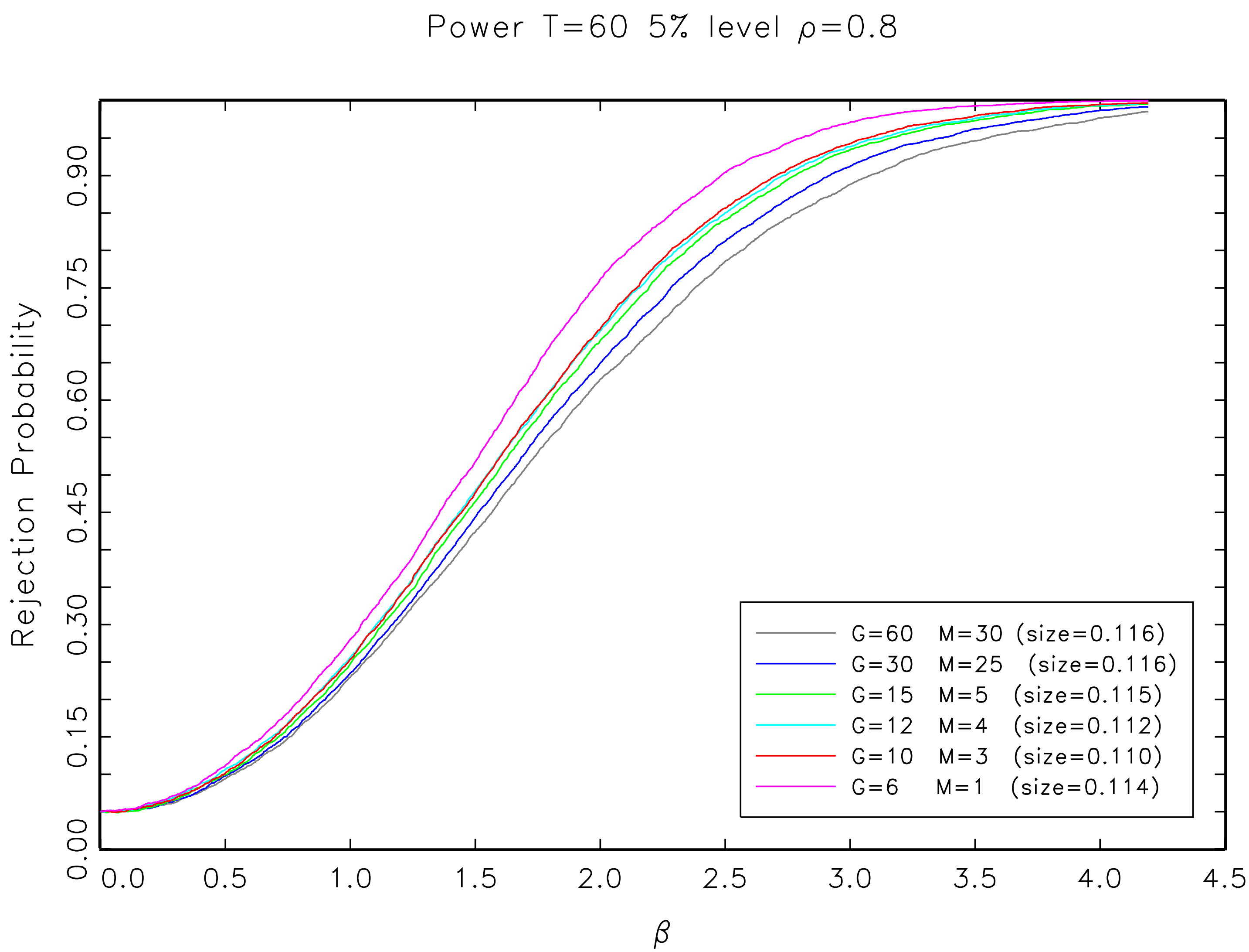


Figure 3.1: (cont'd)

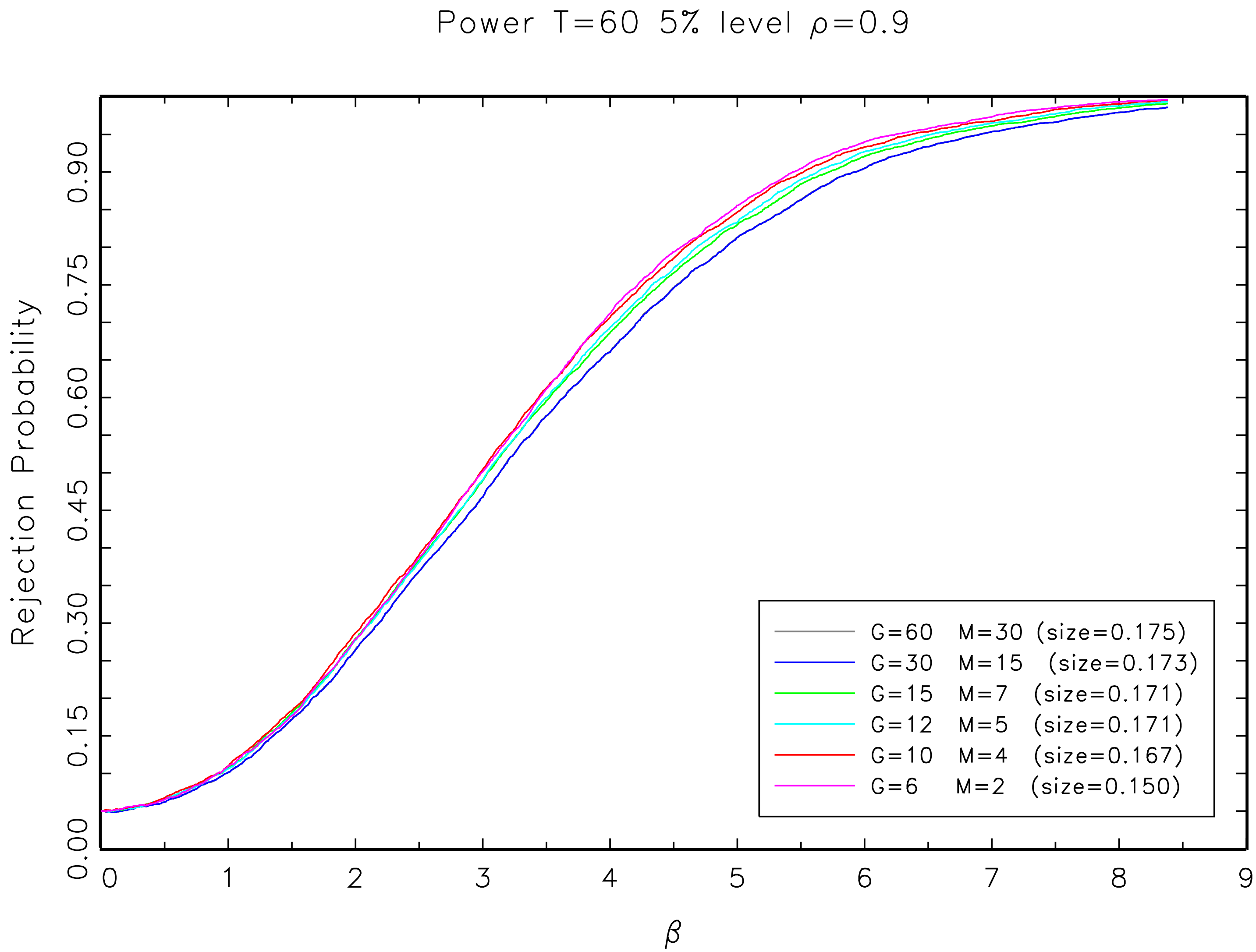


Figure 3.2: Size Adjusted Power Comparision- Clustering V.S. Smoothing, $\rho = 0.5$, $M = 1$

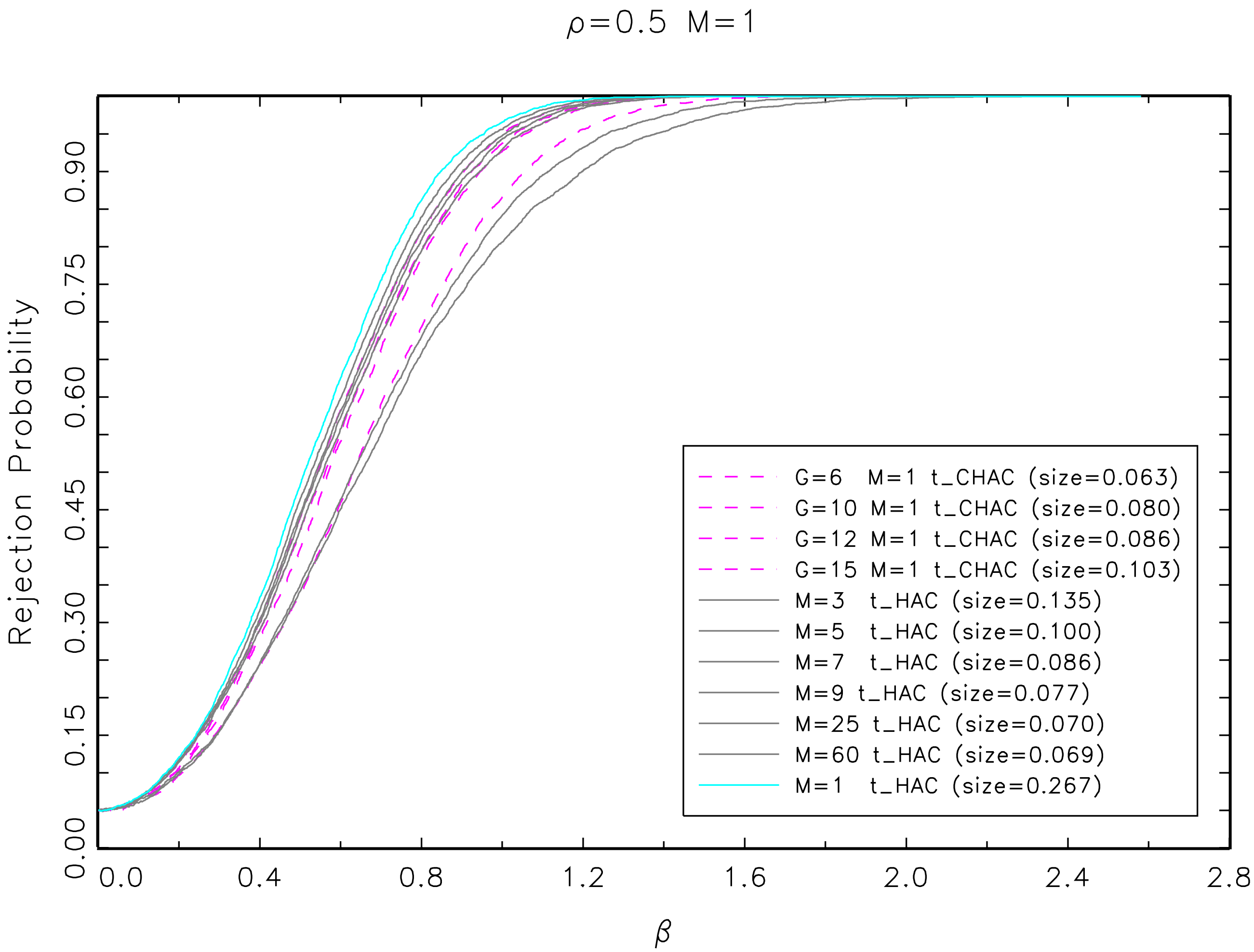


Figure 3.2: (cont'd)

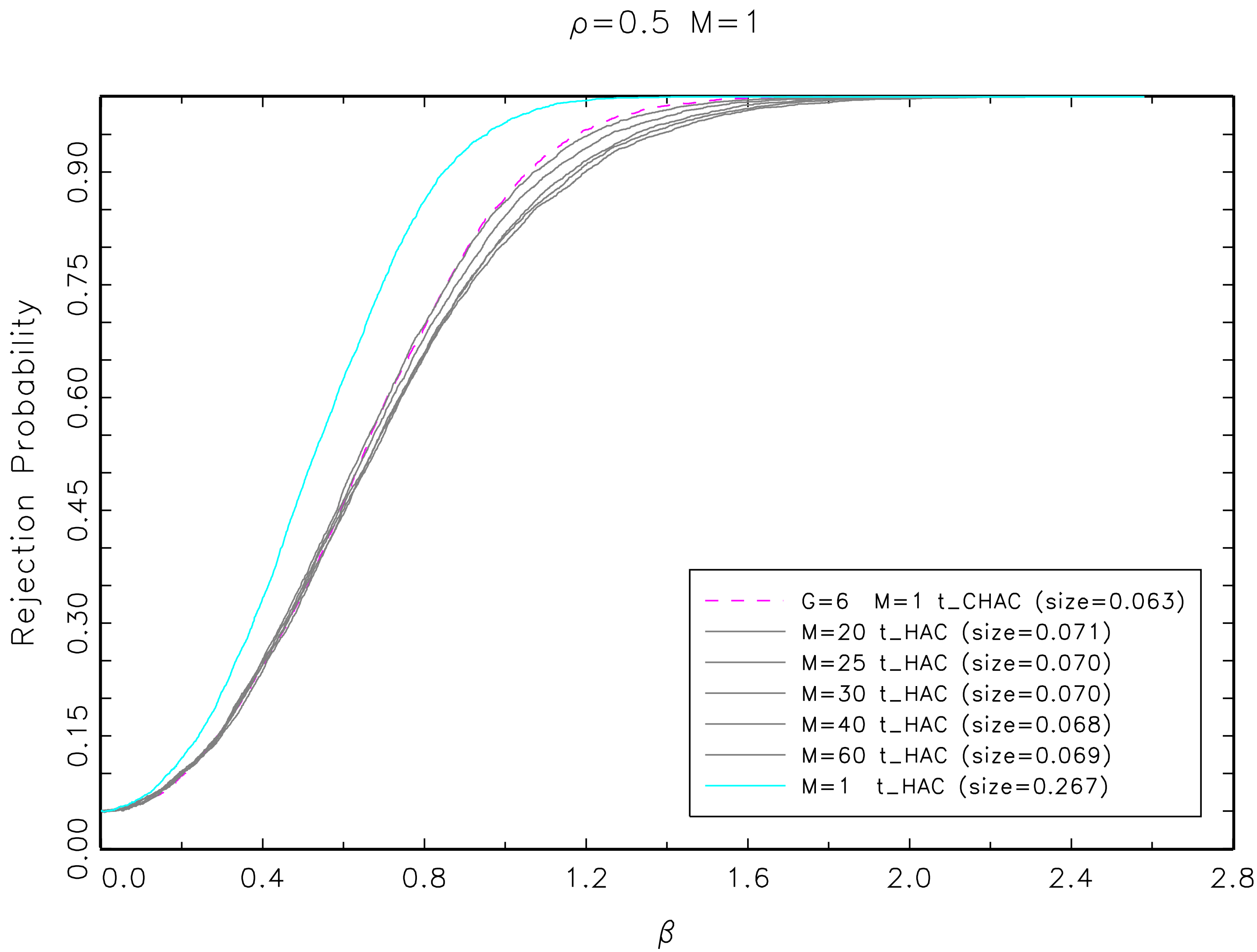


Figure 3.3: Size Adjusted Power Comparision- Clustering V.S. Smoothing, $\rho = 0.5$, $M = 2$

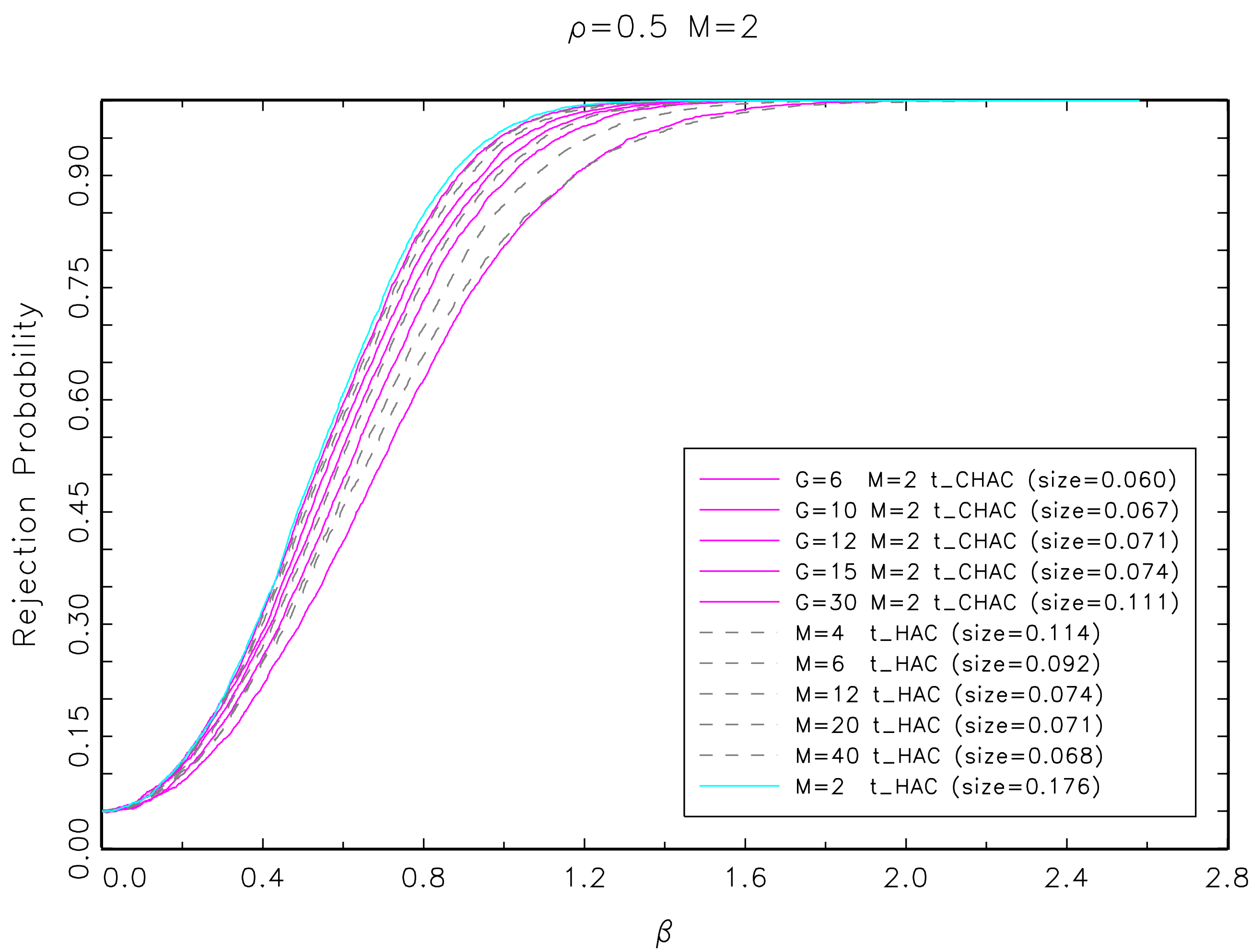


Figure 3.3: (cont'd)

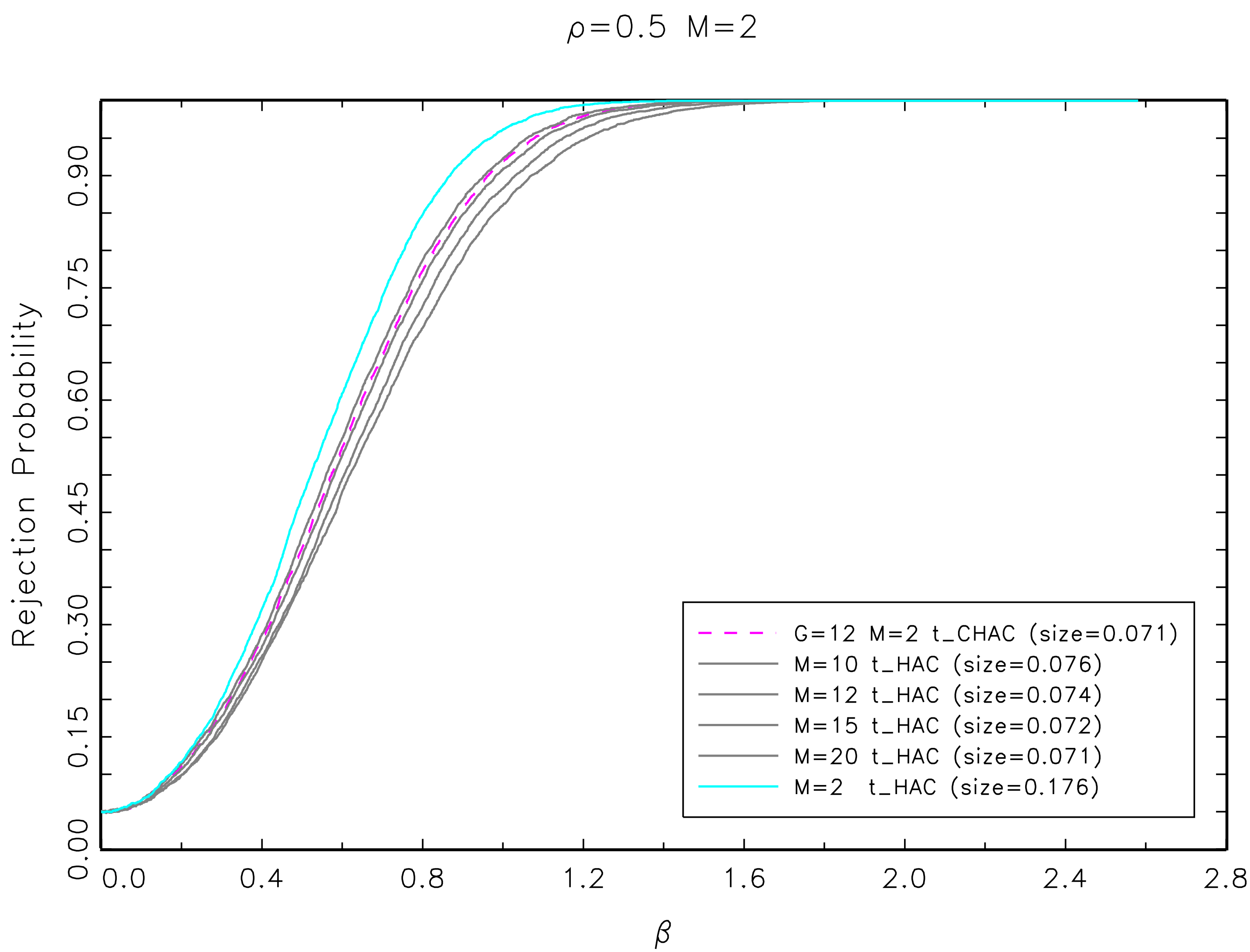


Figure 3.4: Size Adjusted Power Comparision- Clustering V.S. Smoothing, $\rho = 0.5$, $M = 3$

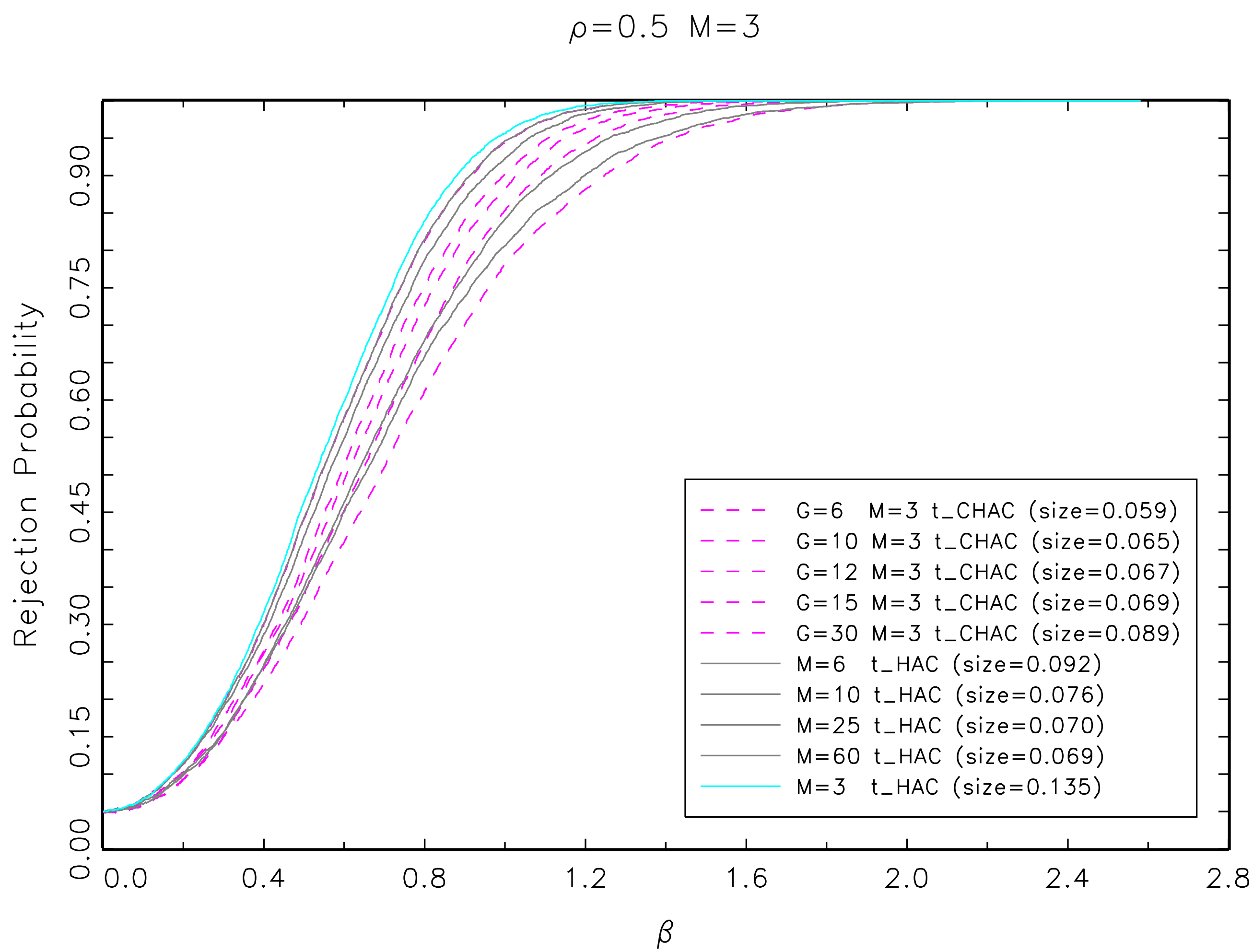


Figure 3.4: (cont'd)

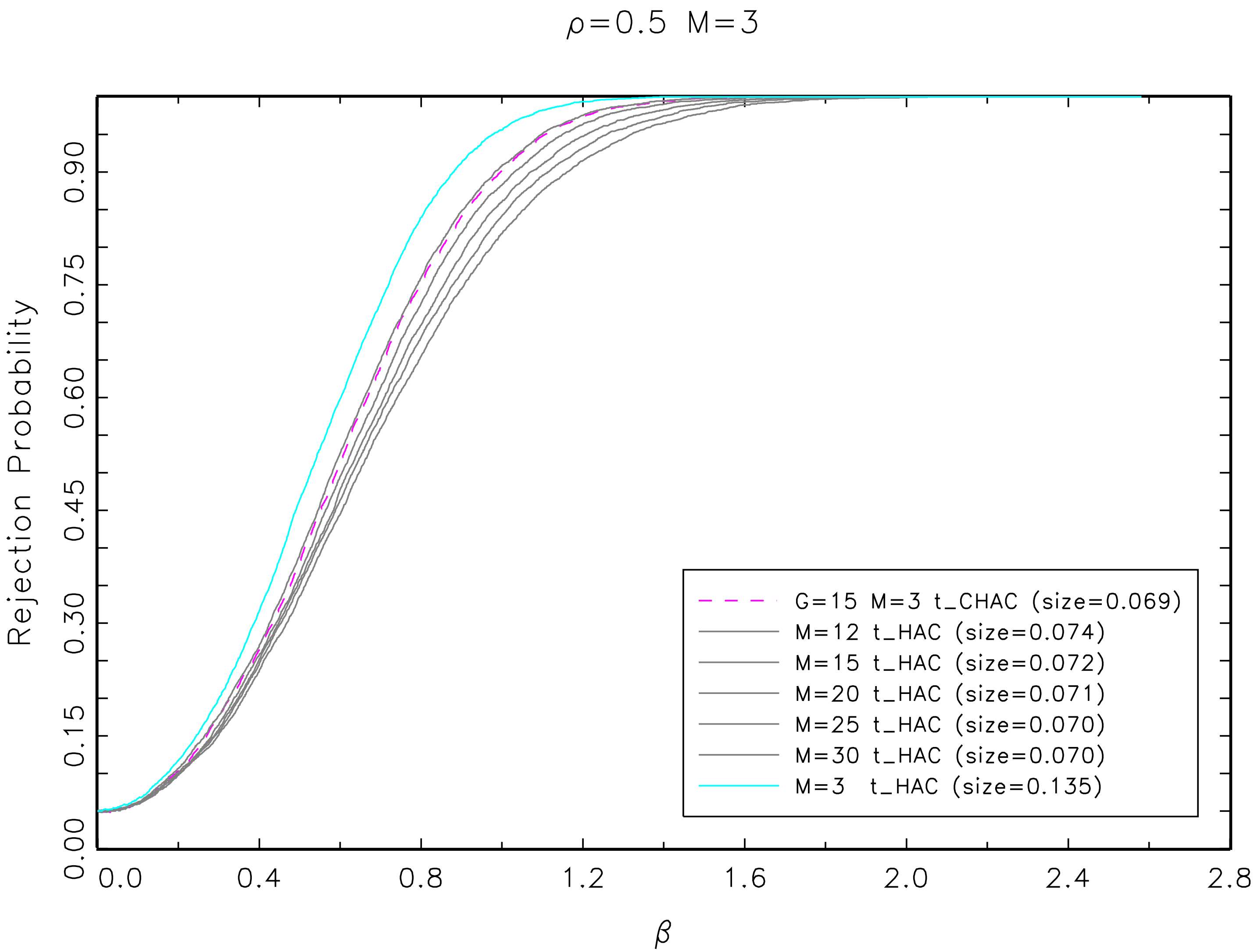


Figure 3.4: (cont'd)

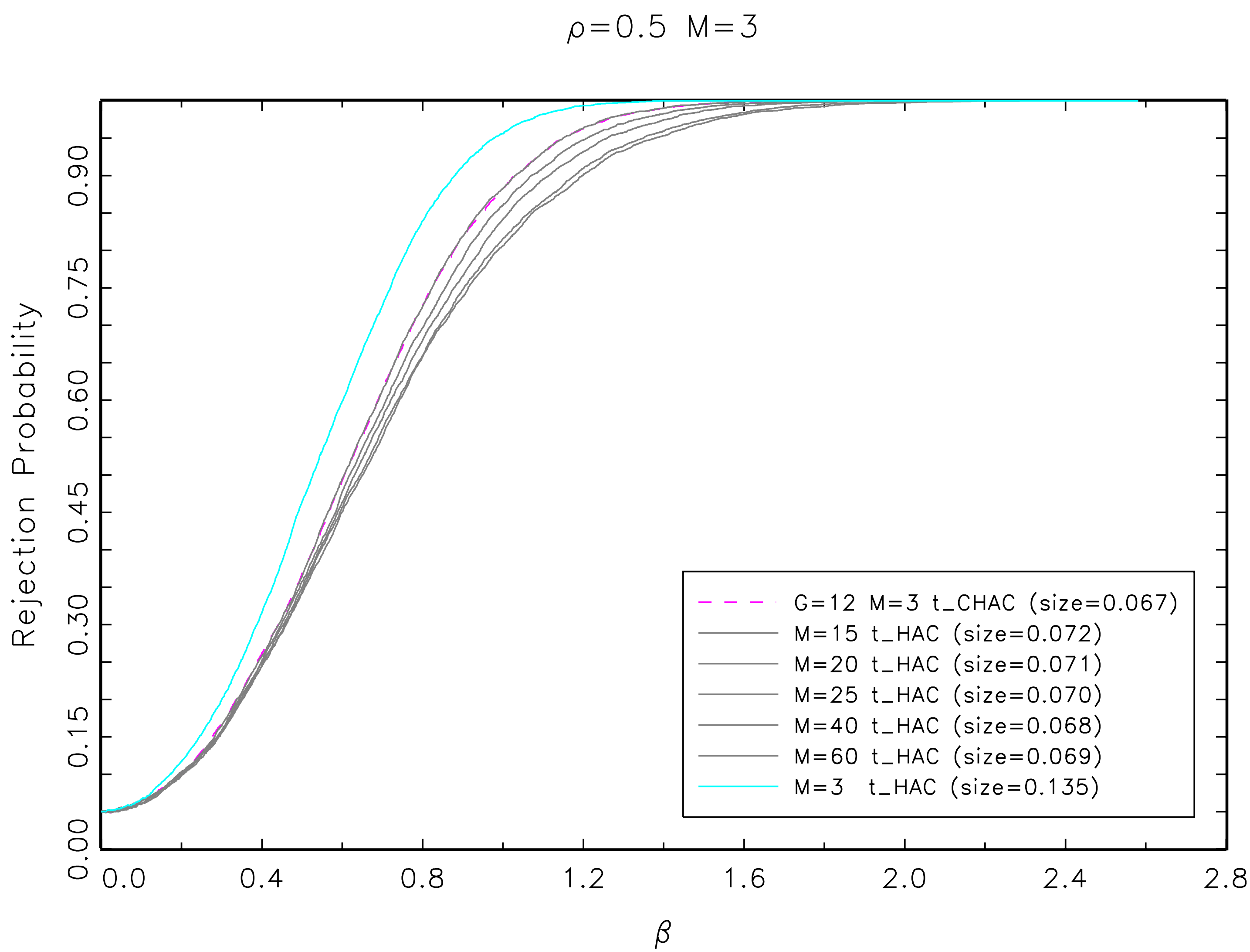


Figure 3.4: (cont'd)

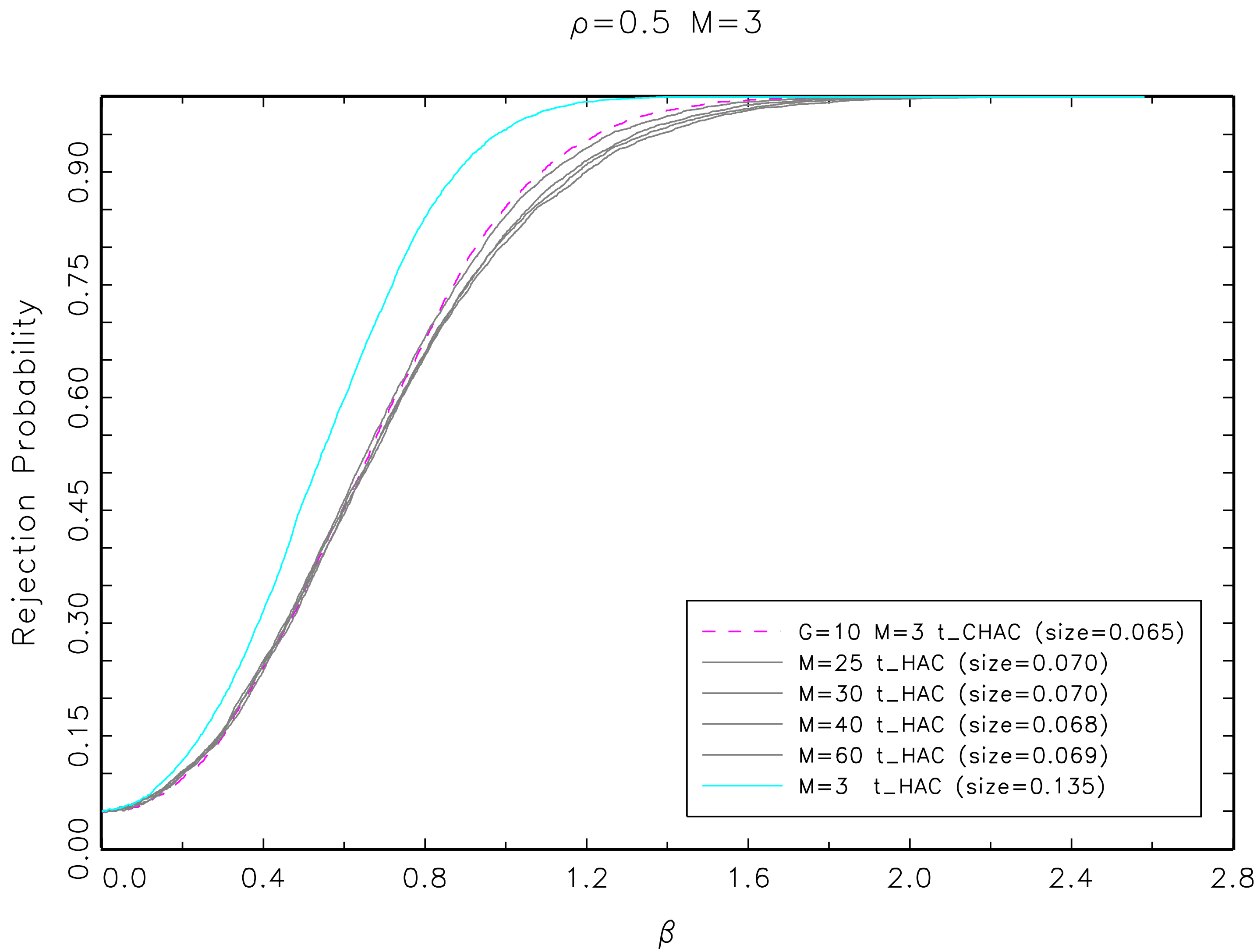


Figure 3.5: Size Adjusted Power Comparision- Clustering V.S. Smoothing, $\rho = 0.5$, $M = 4$

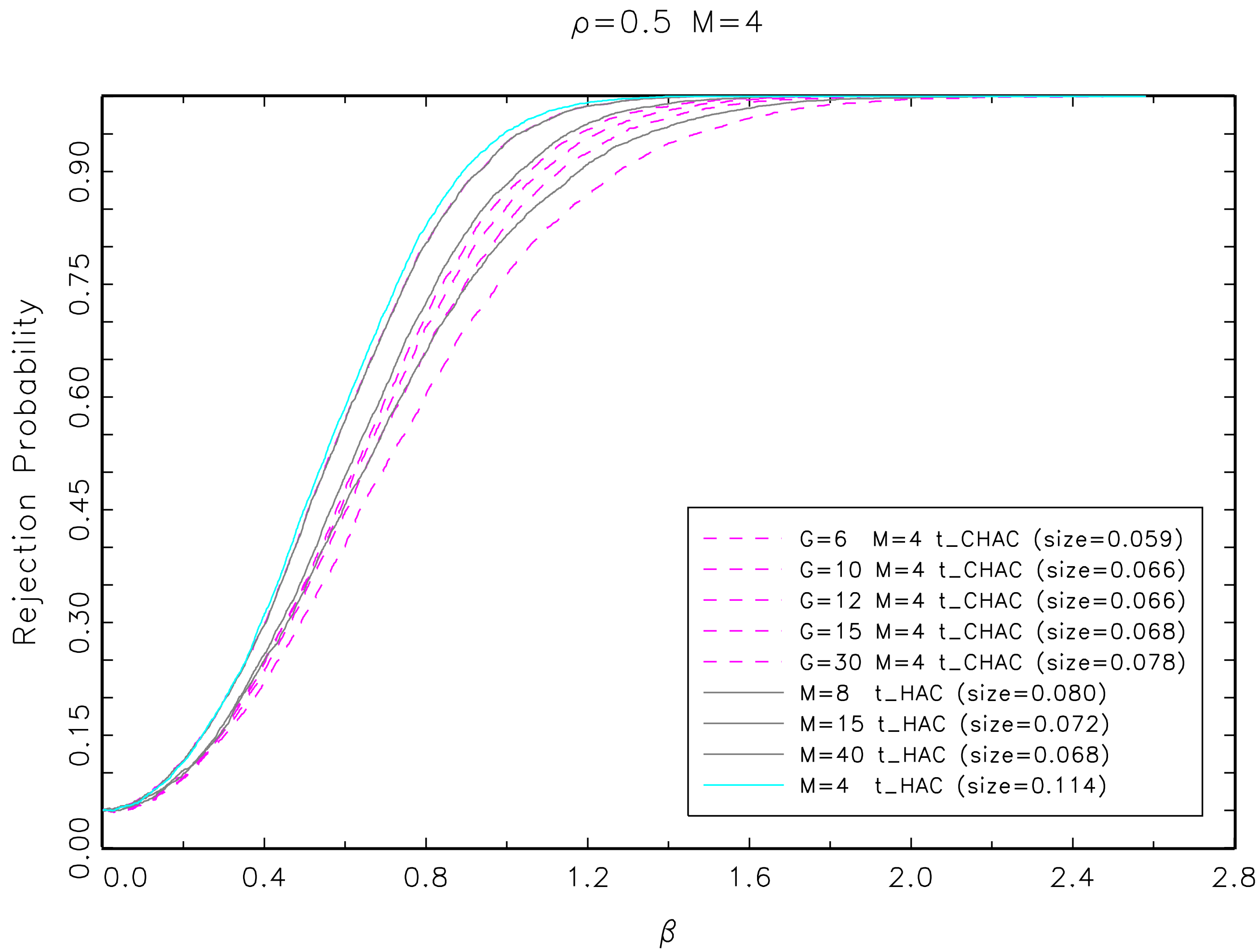


Figure 3.5: (cont'd)

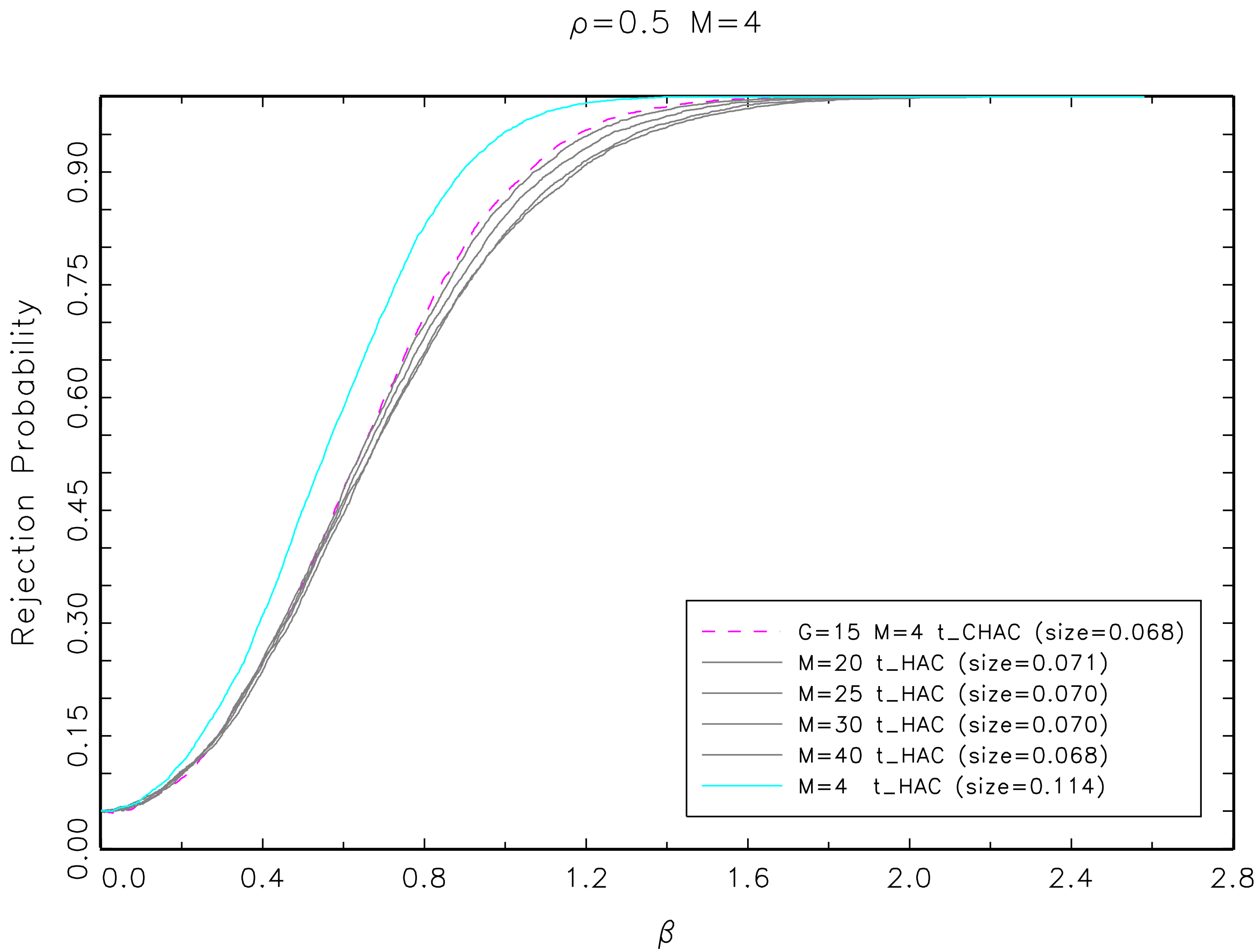


Figure 3.5: (cont'd)

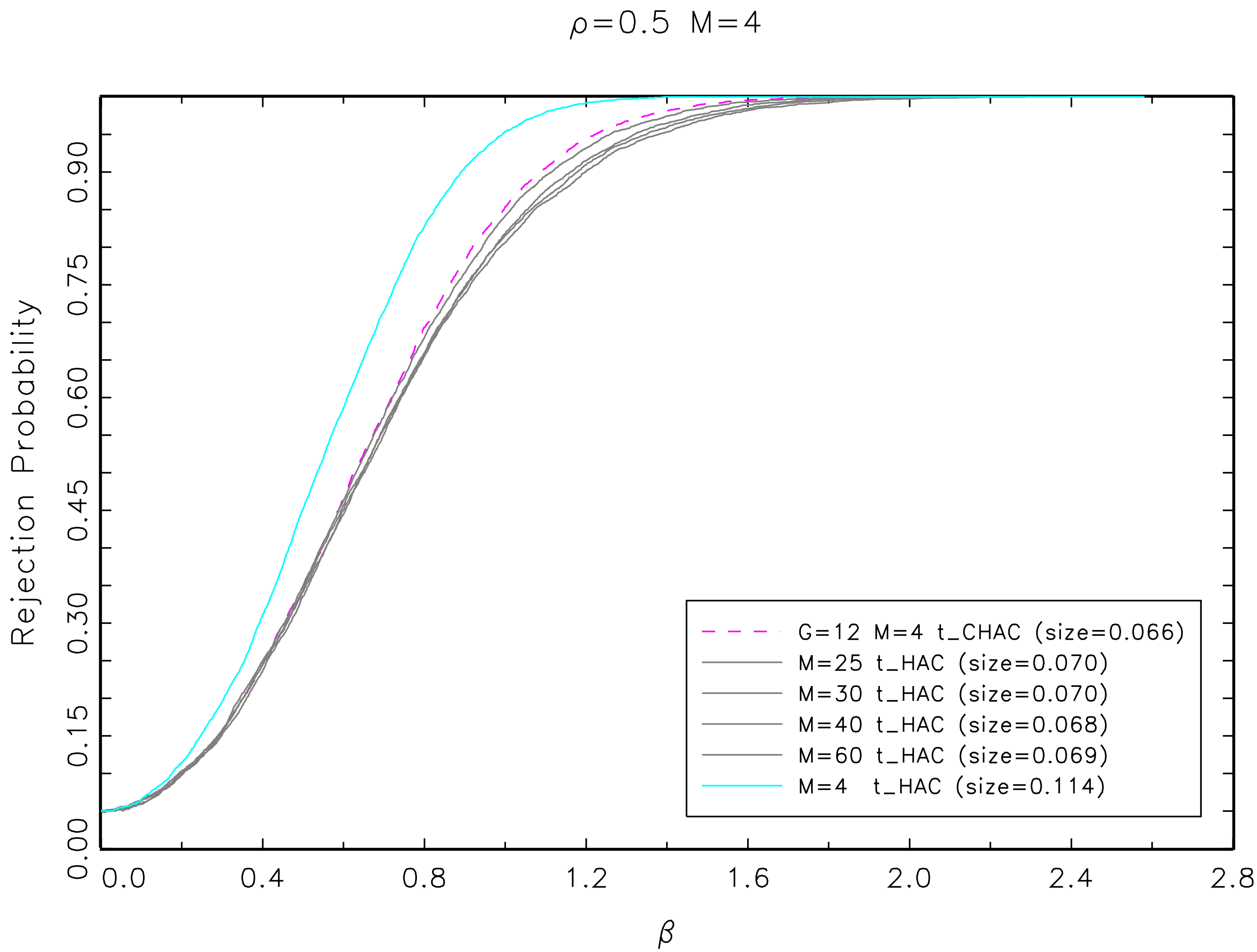


Figure 3.6: Size Adjusted Power Comparision- Clustering V.S. Smoothing, $\rho = 0.5$, $M = 5$

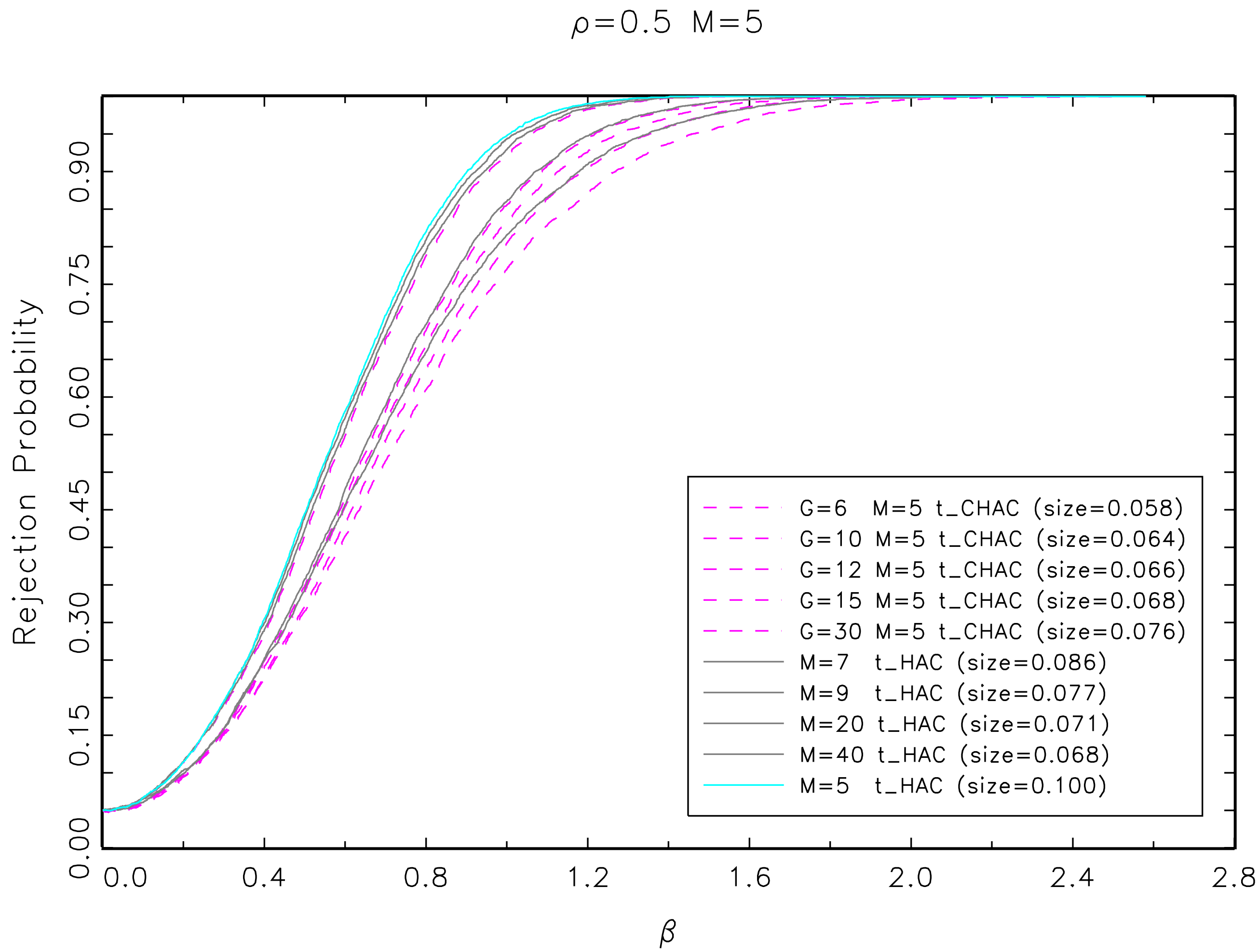


Figure 3.6: (cont'd)

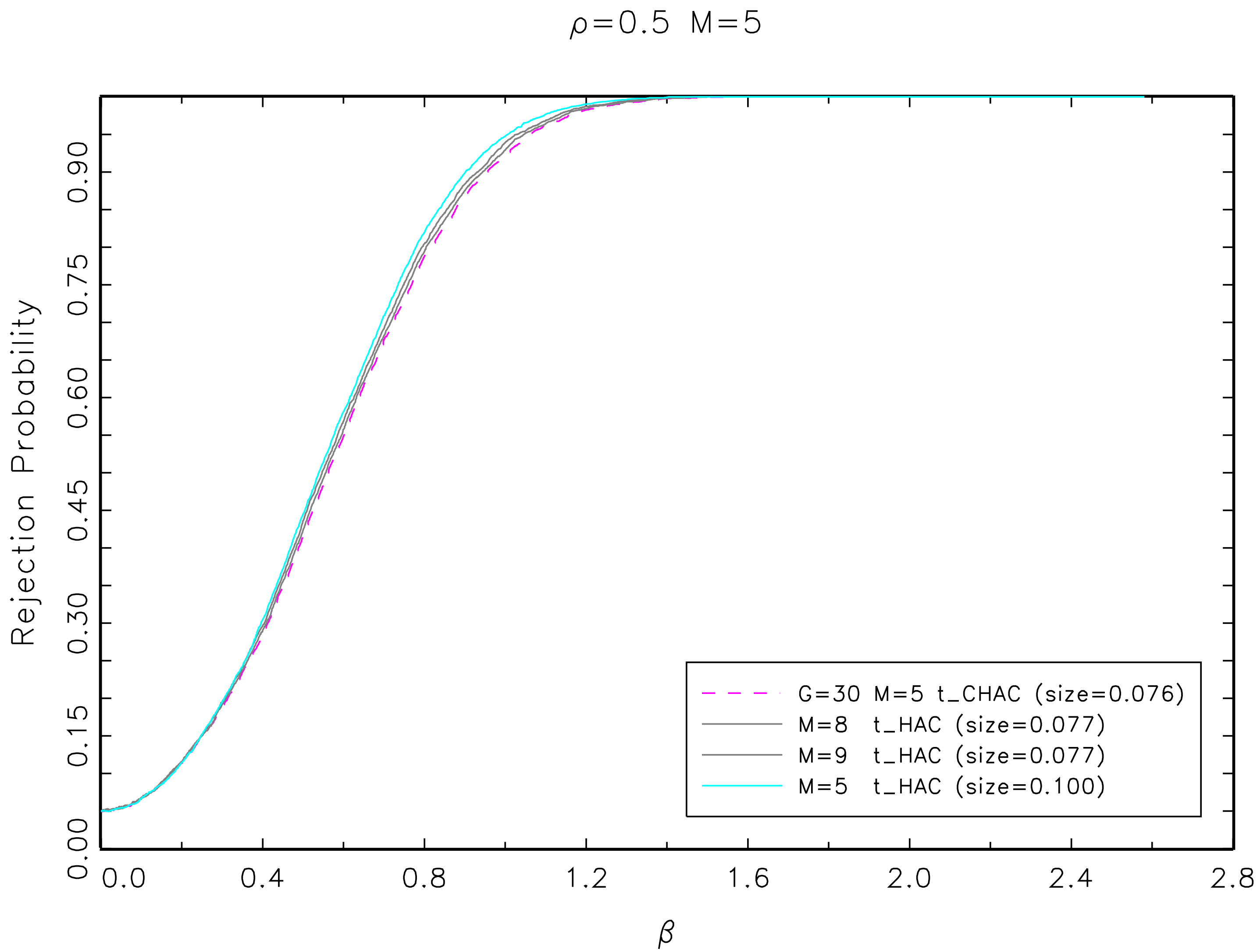


Figure 3.6: (cont'd)

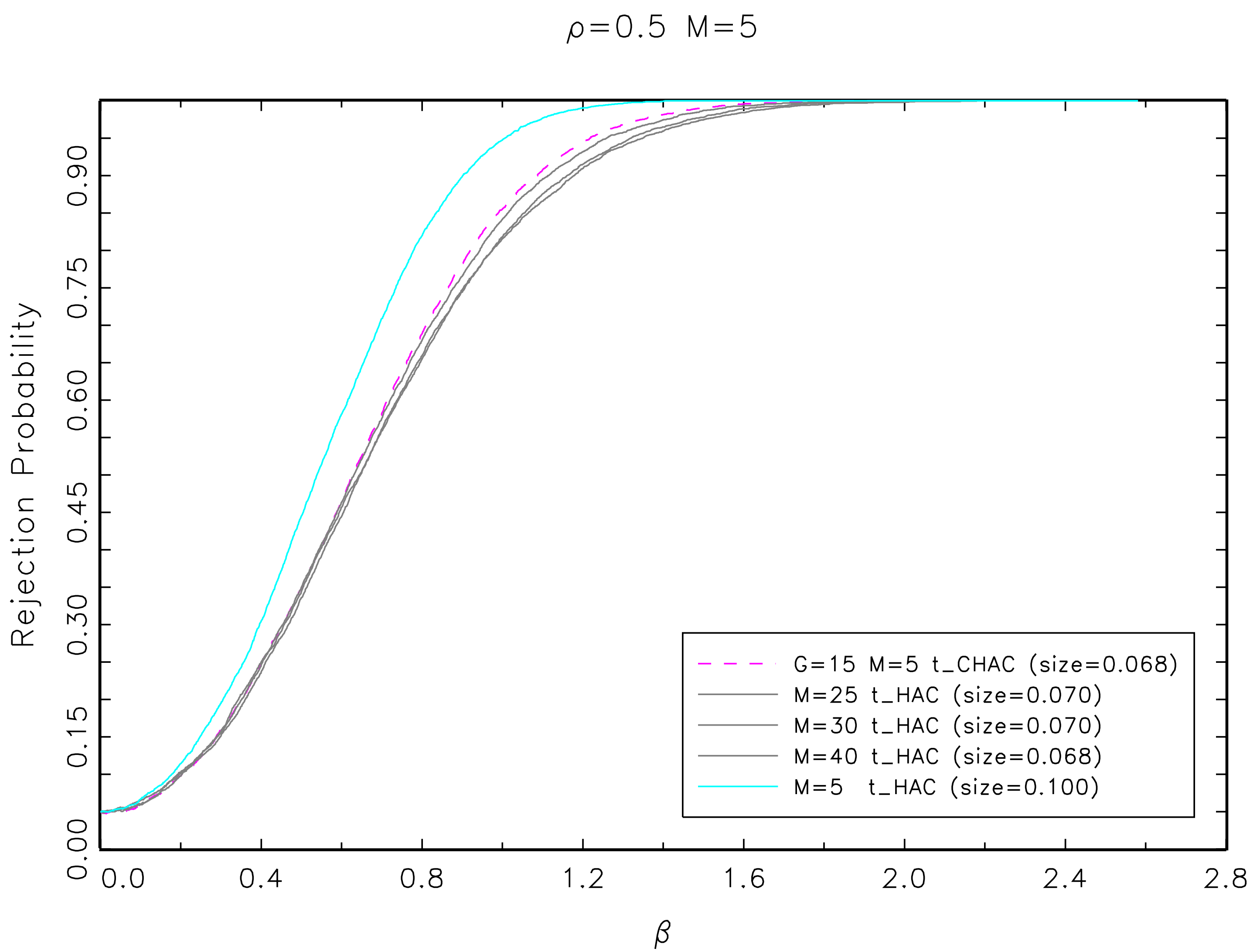


Figure 3.6: (cont'd)

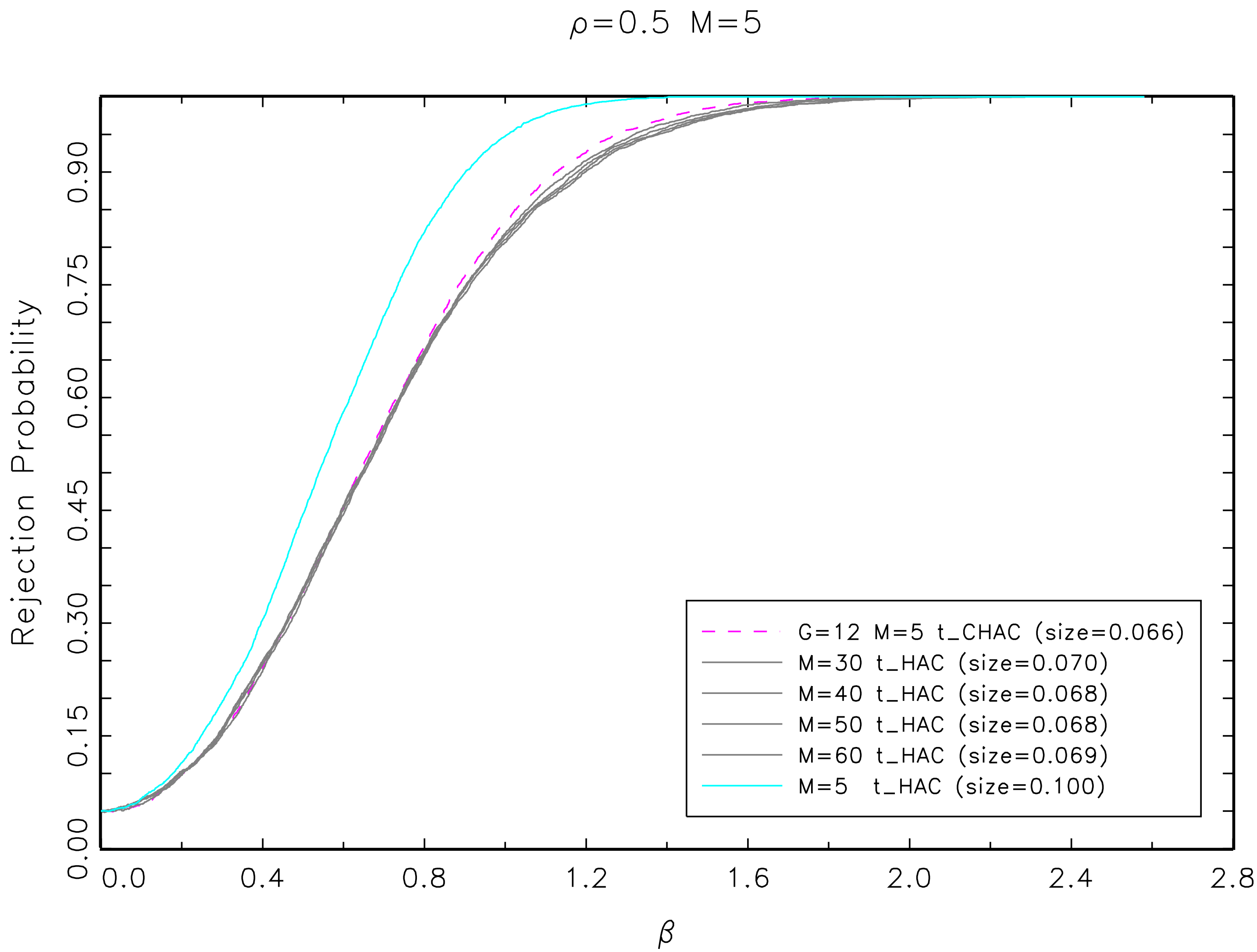


Figure 3.7: Size Adjusted Power Comparision- Clustering V.S. Smoothing, $\rho = 0.5$, $M = 6$

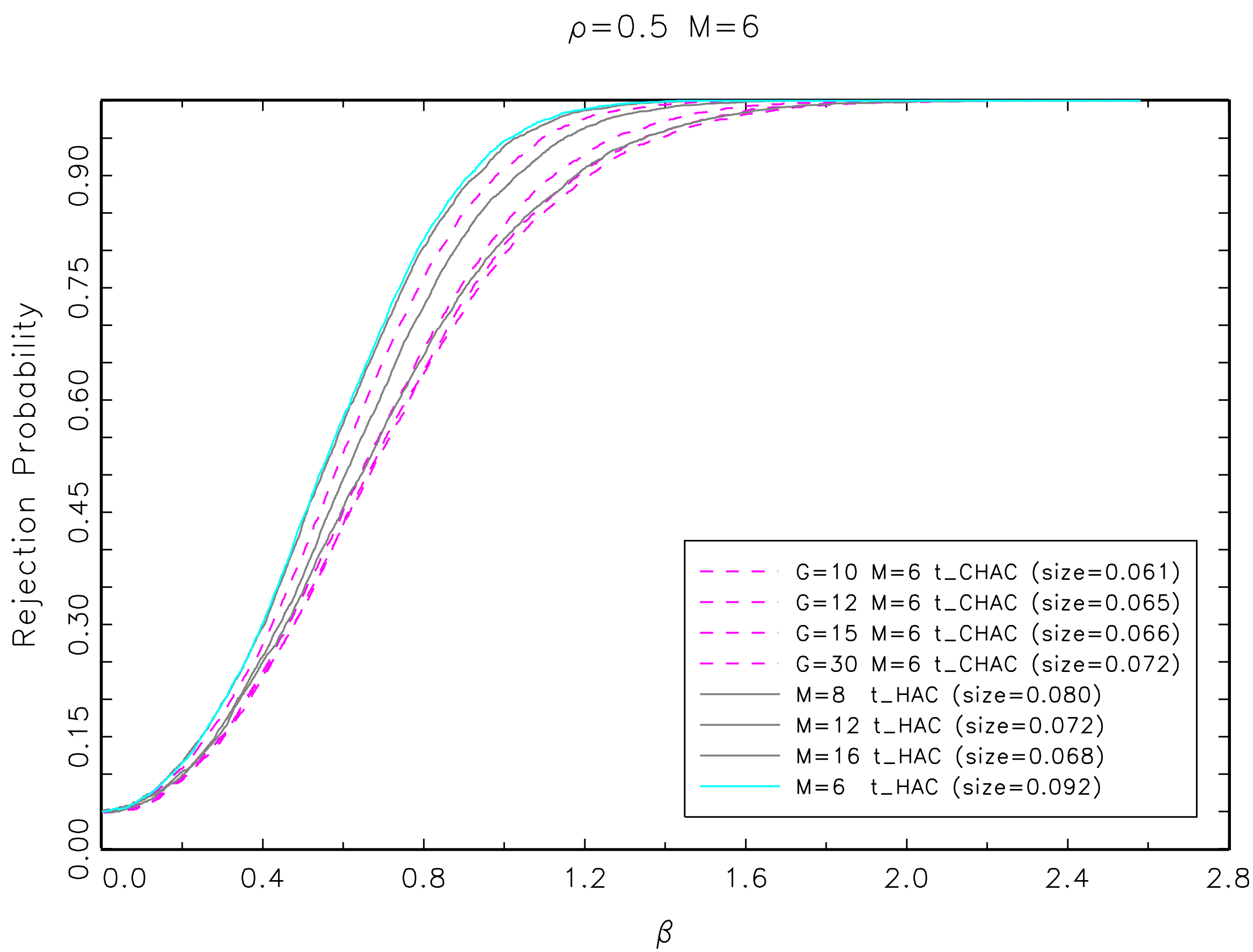


Figure 3.8: Size Adjusted Power Comparision- Clustering V.S. Smoothing, $\rho = 0.5$, $M = 7$

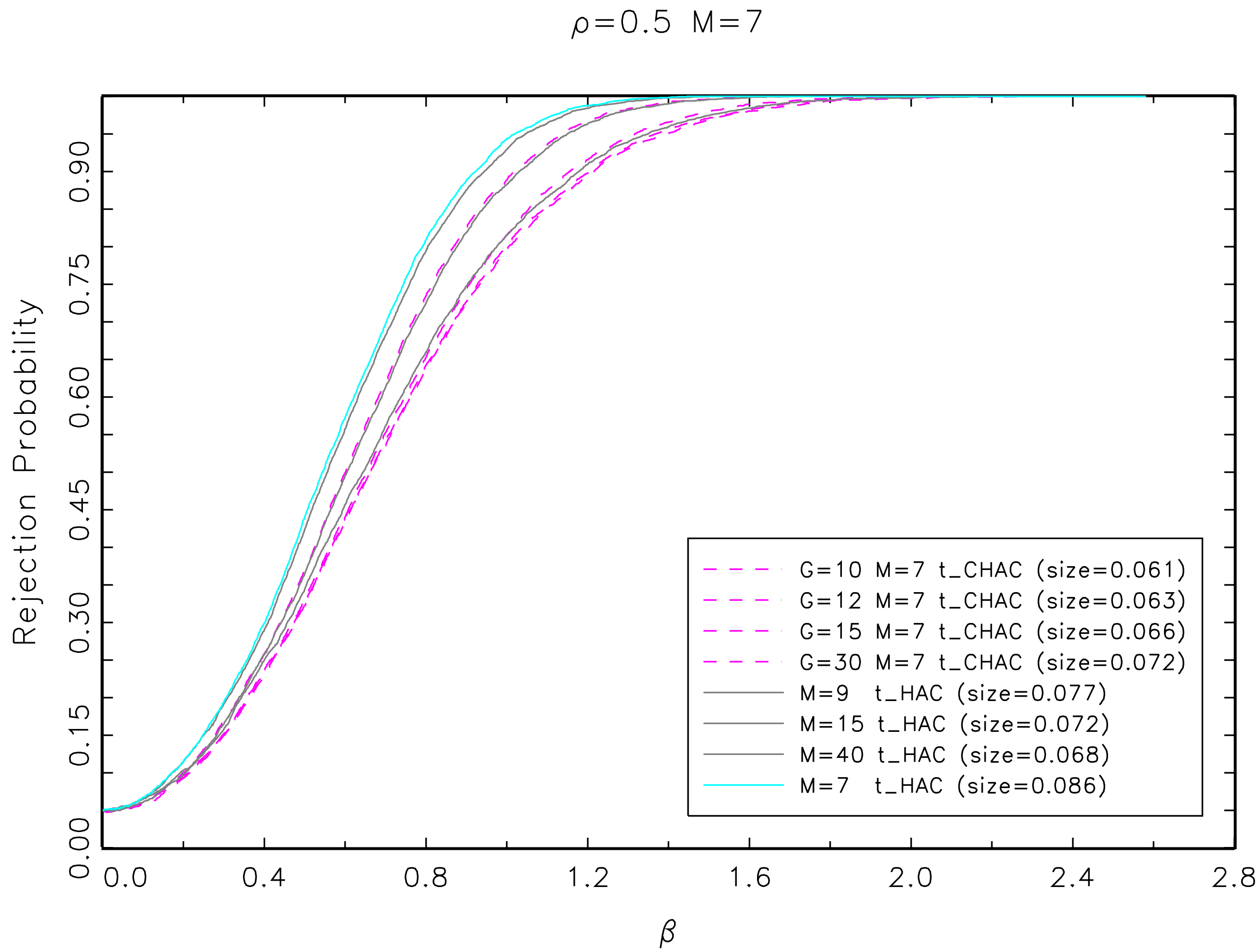


Figure 3.9: Size Adjusted Power Comparision- Clustering V.S. Smoothing, $\rho = 0.5$, $M = 8$

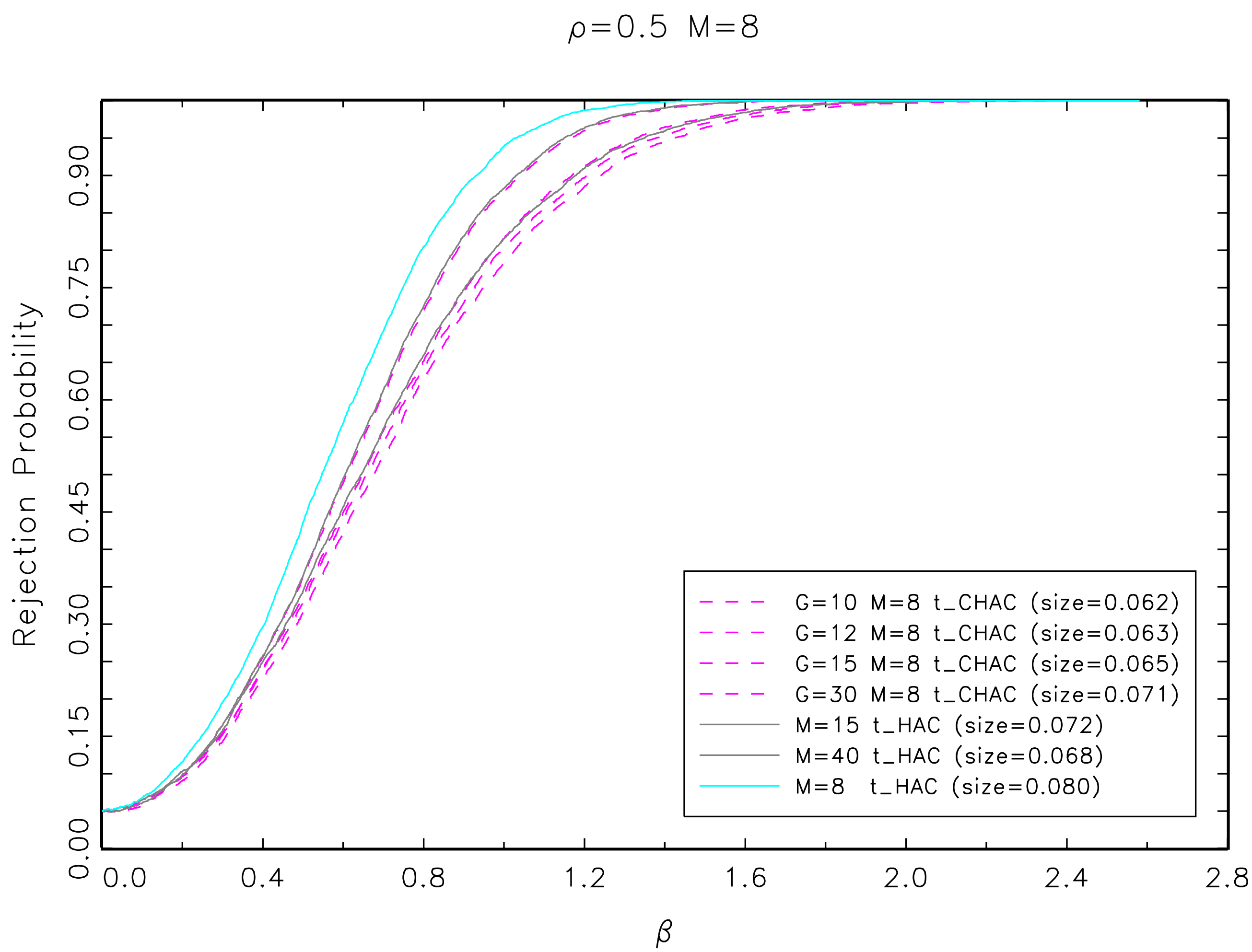


Figure 3.10: Size Adjusted Power Comparision- Clustering V.S. Smoothing, $\rho = 0.8$, $M = 1$

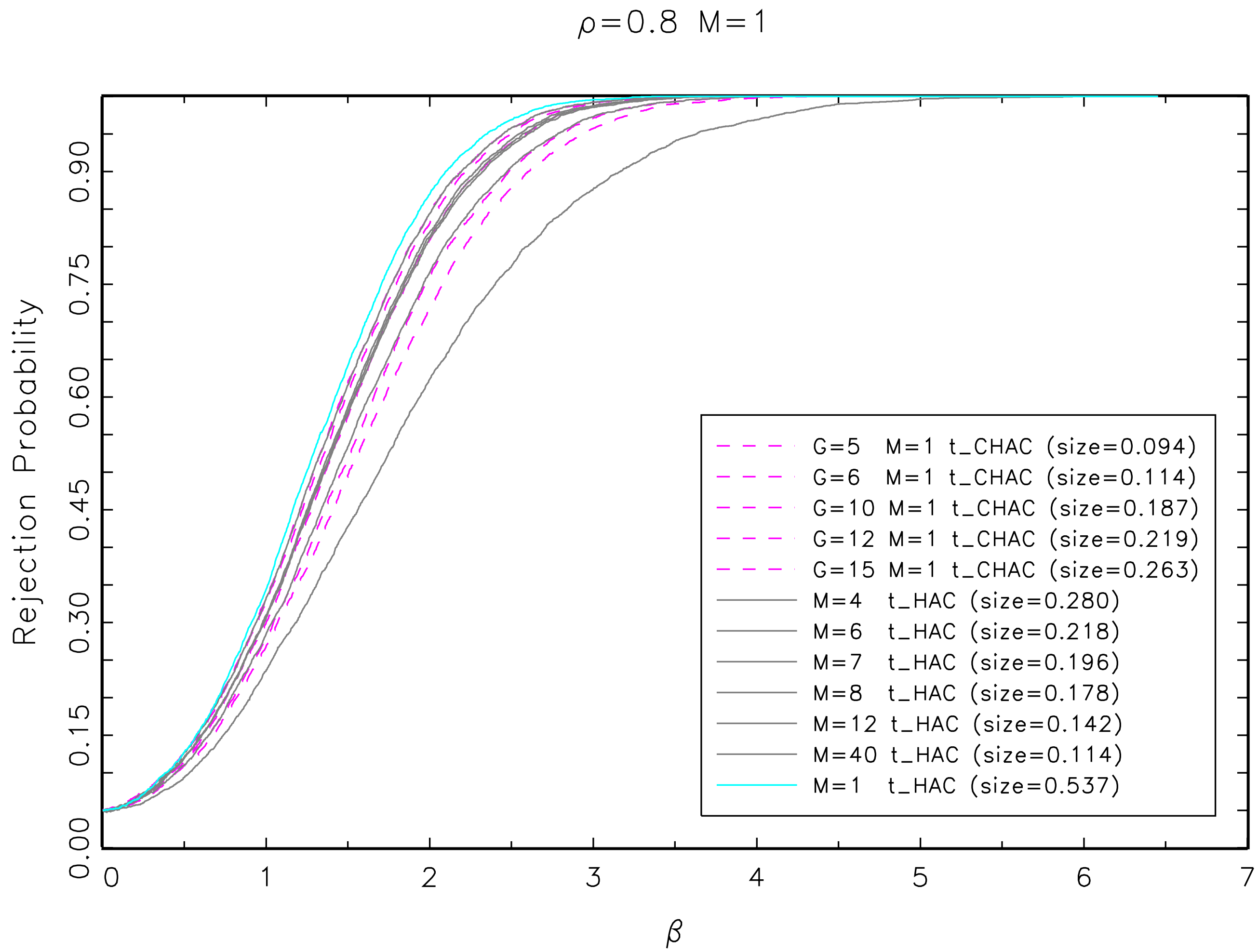


Figure 3.10: (cont'd)

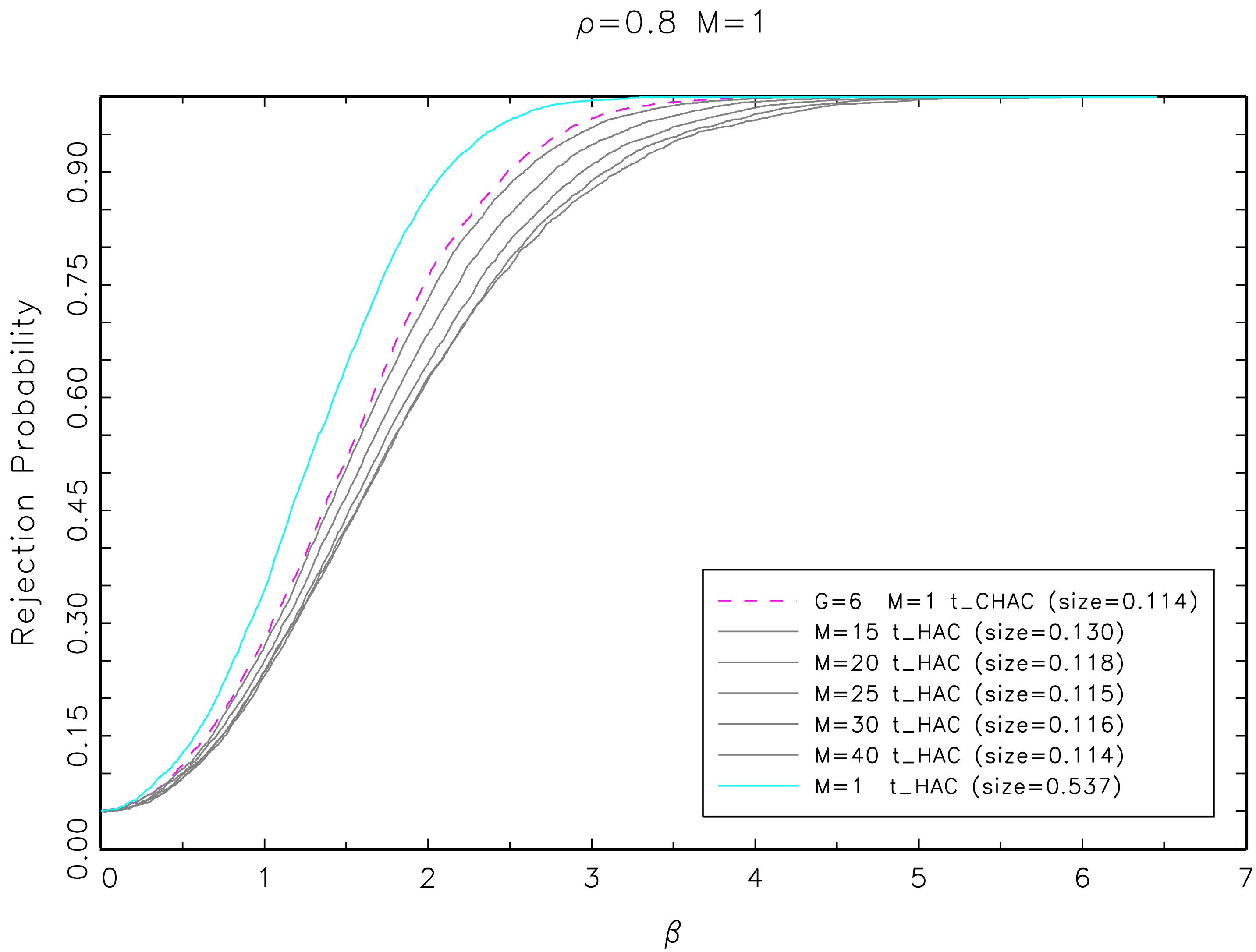


Figure 3.10: (cont'd)

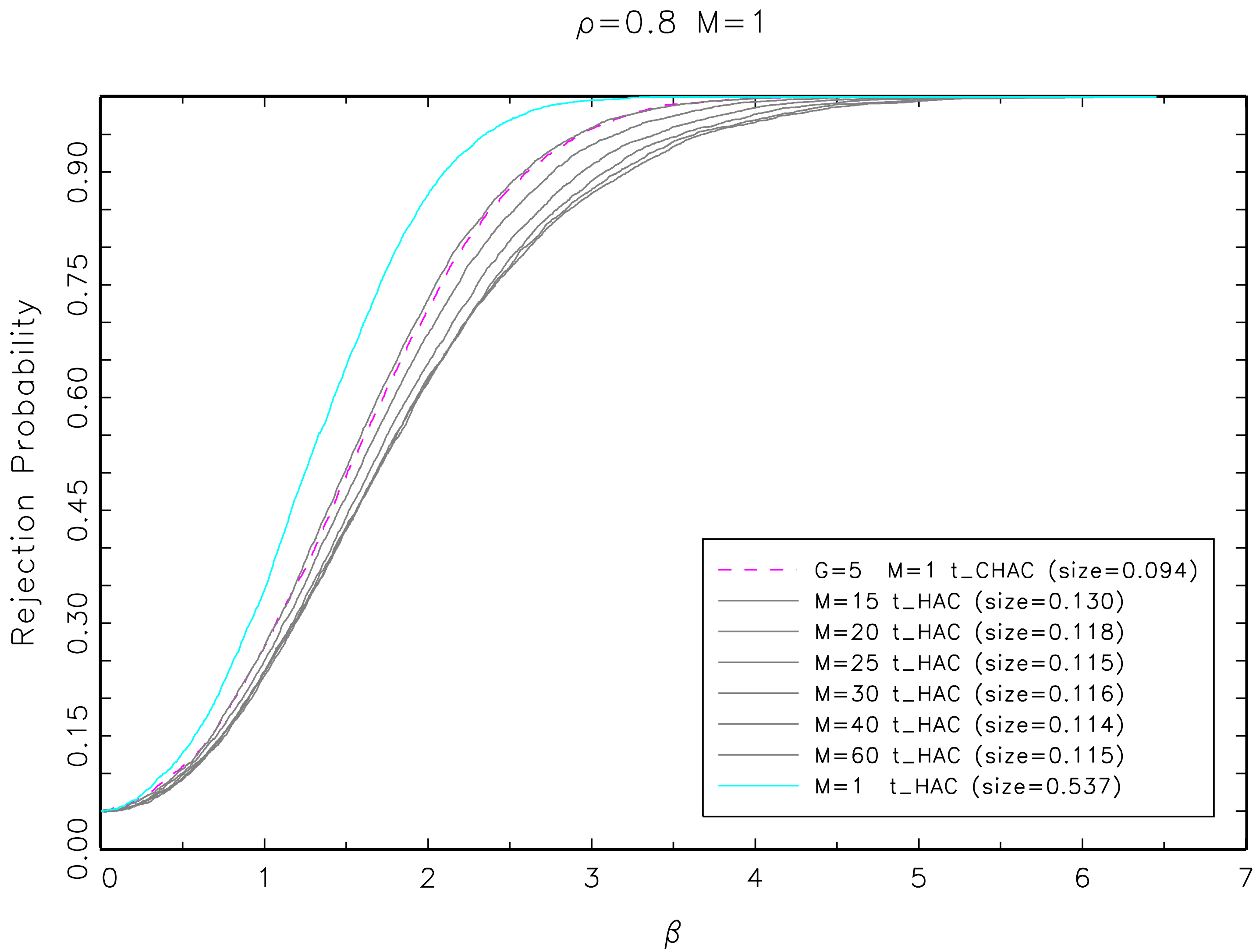


Figure 3.11: Size Adjusted Power Comparision- Clustering V.S. Smoothing, $\rho = 0.8$, $M = 2$

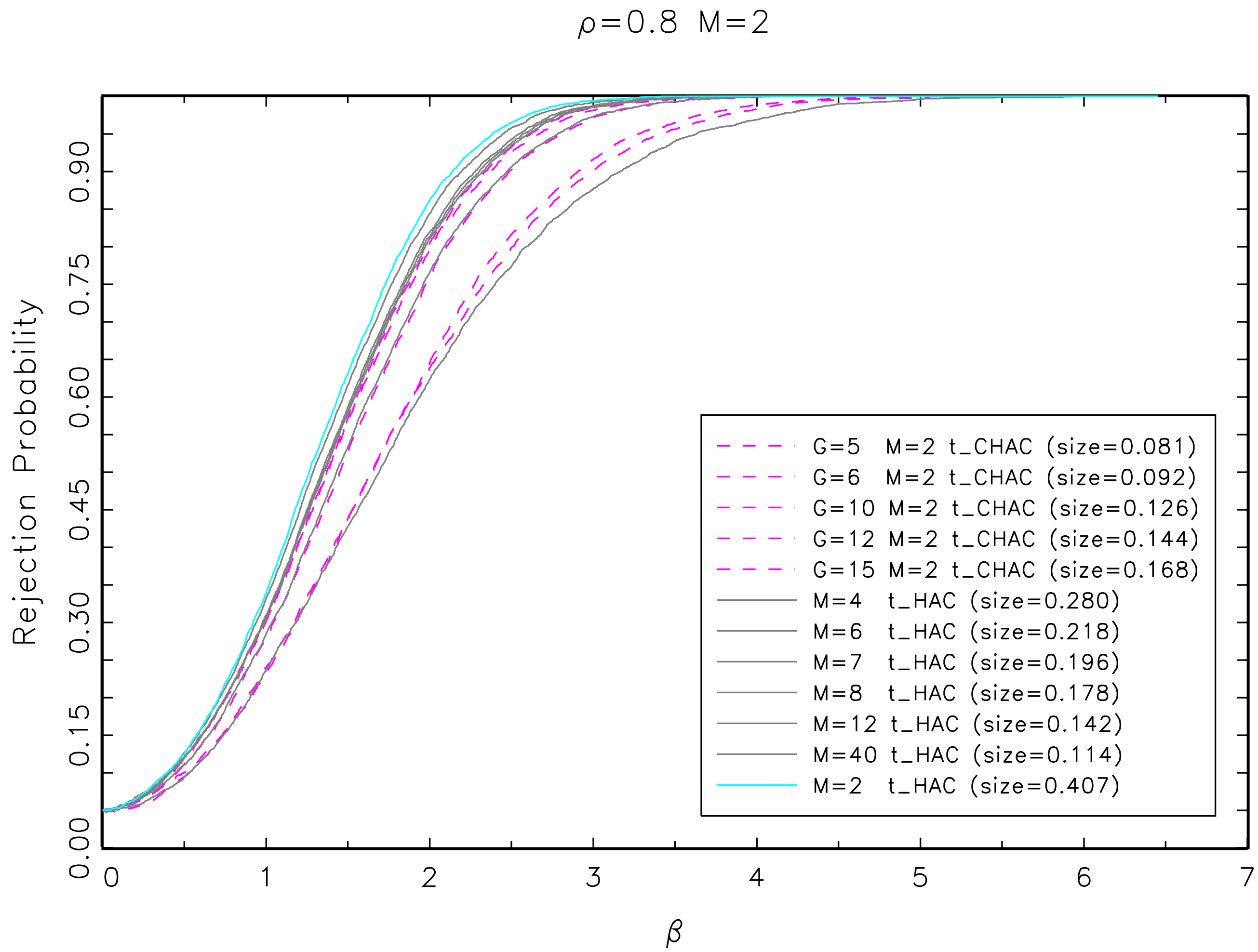


Figure 3.11: (cont'd)

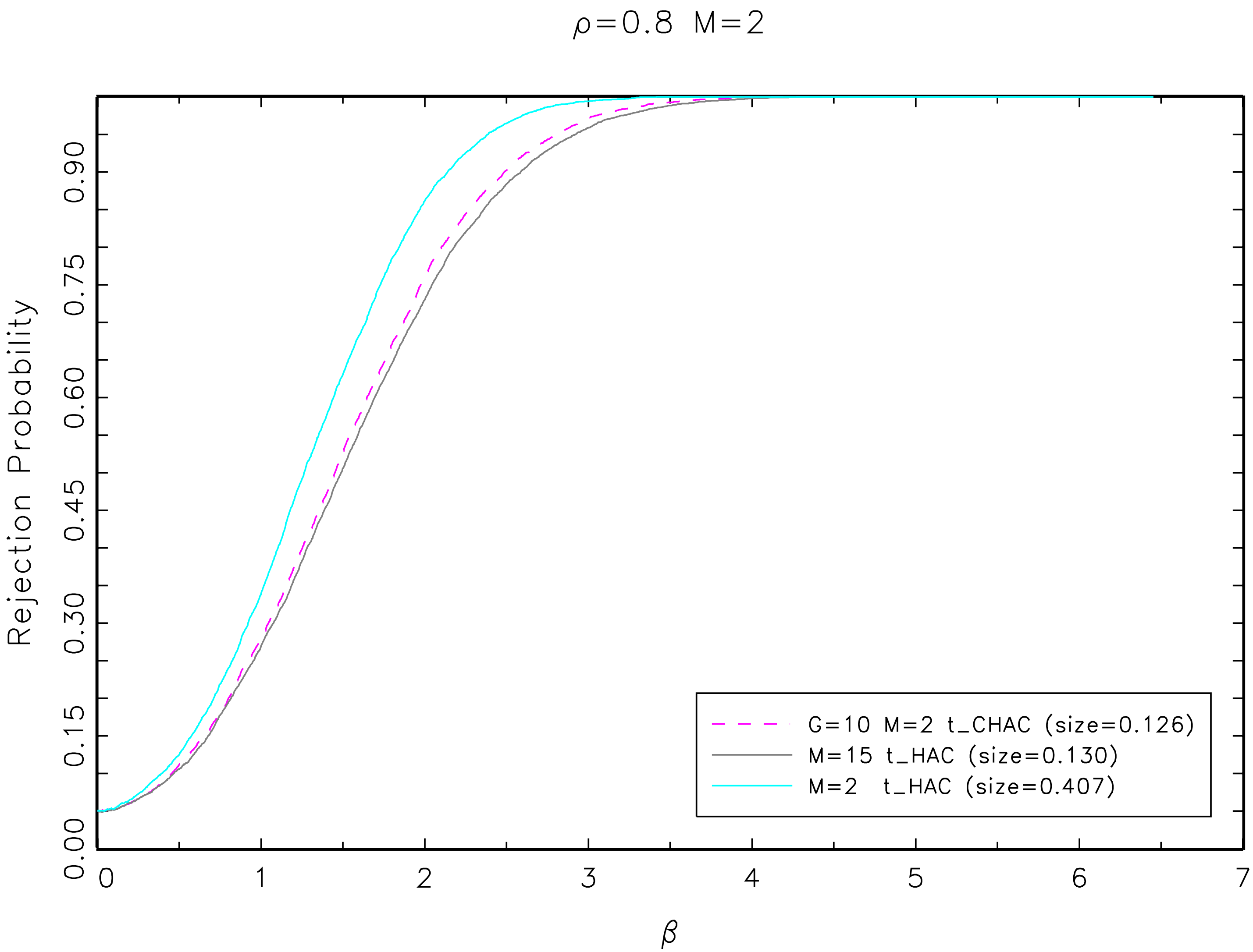


Figure 3.11: (cont'd)

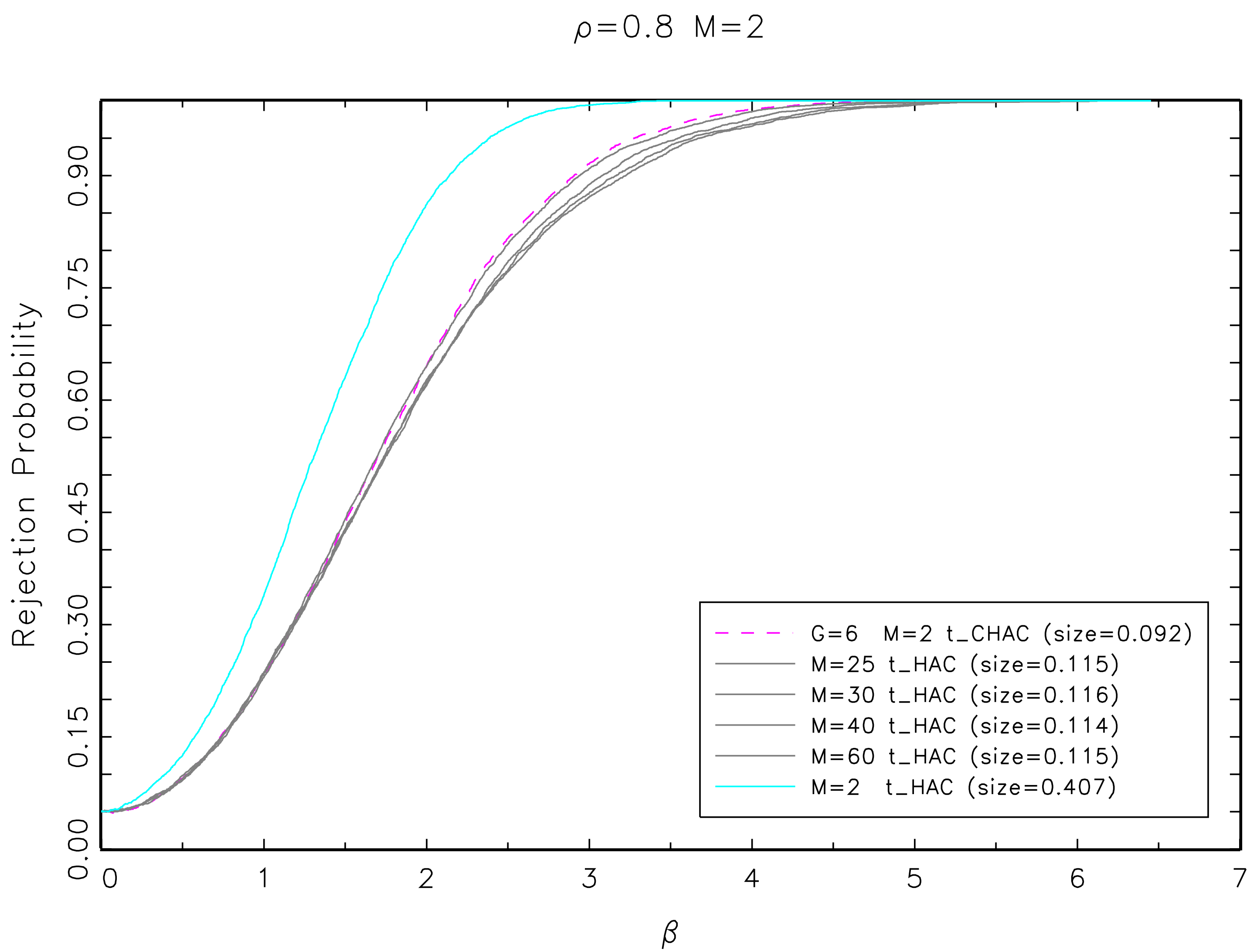


Figure 3.12: Size Adjusted Power Comparision- Clustering V.S. Smoothing, $\rho = 0.8$, $M = 3$

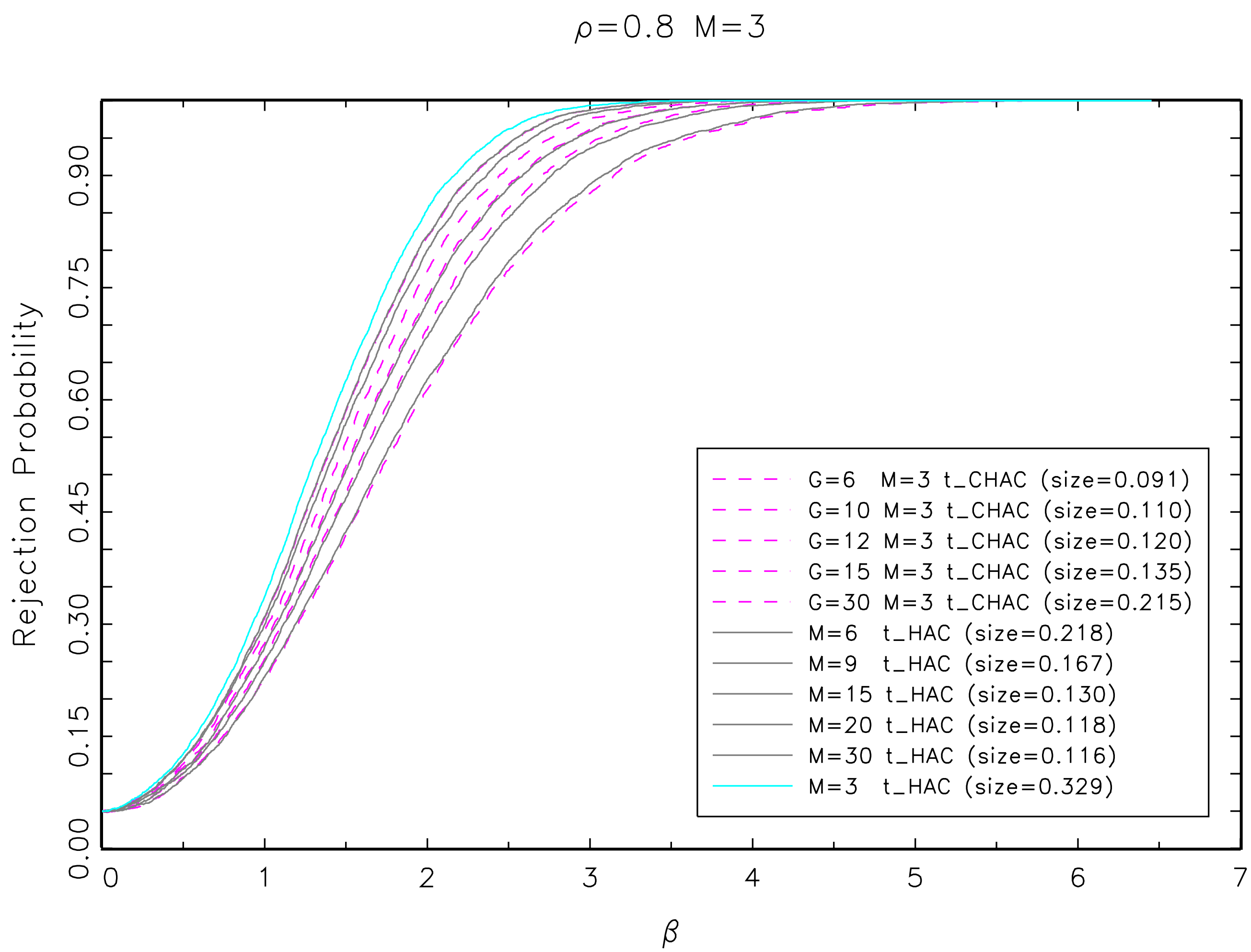


Figure 3.12: (cont'd)

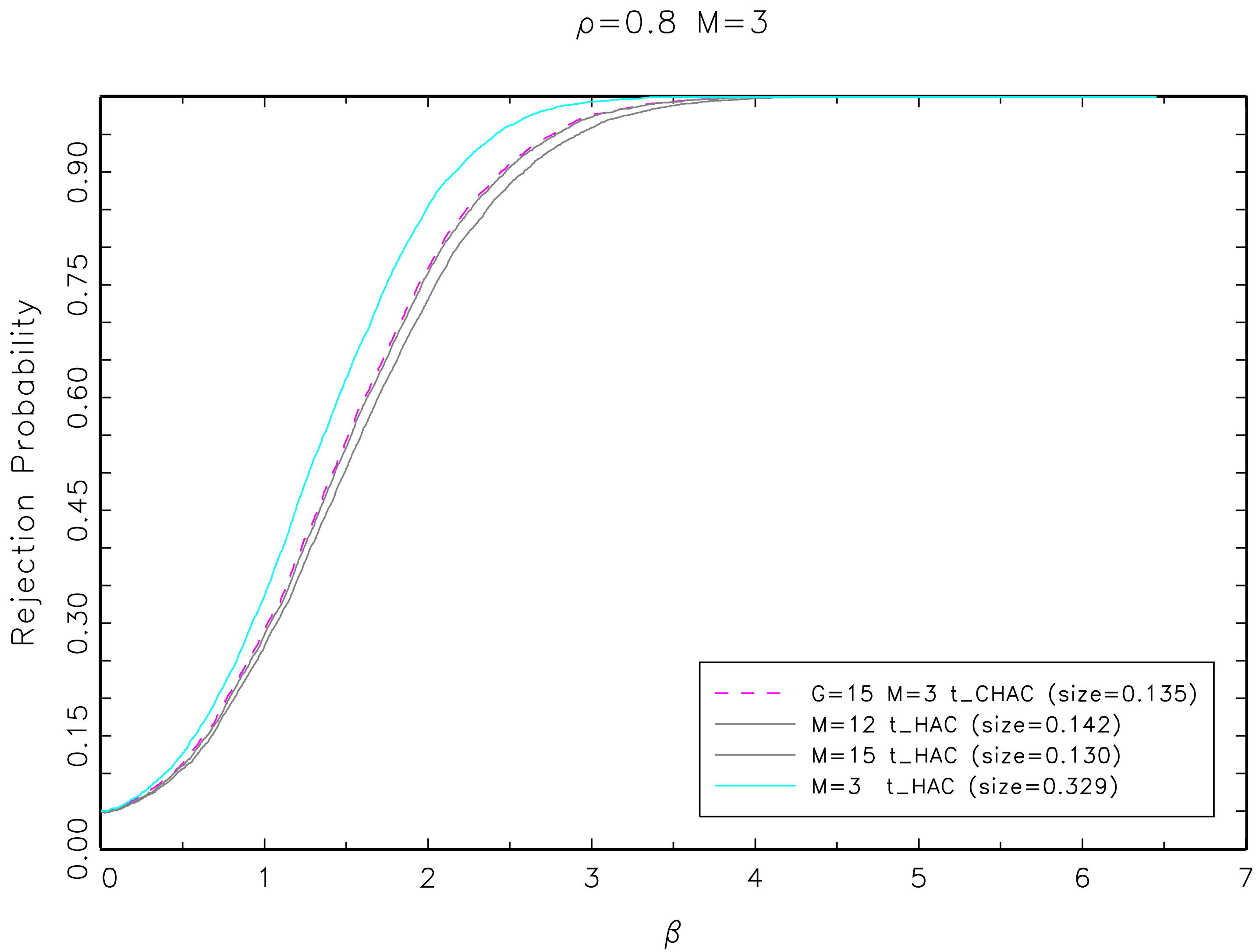


Figure 3.12: (cont'd)

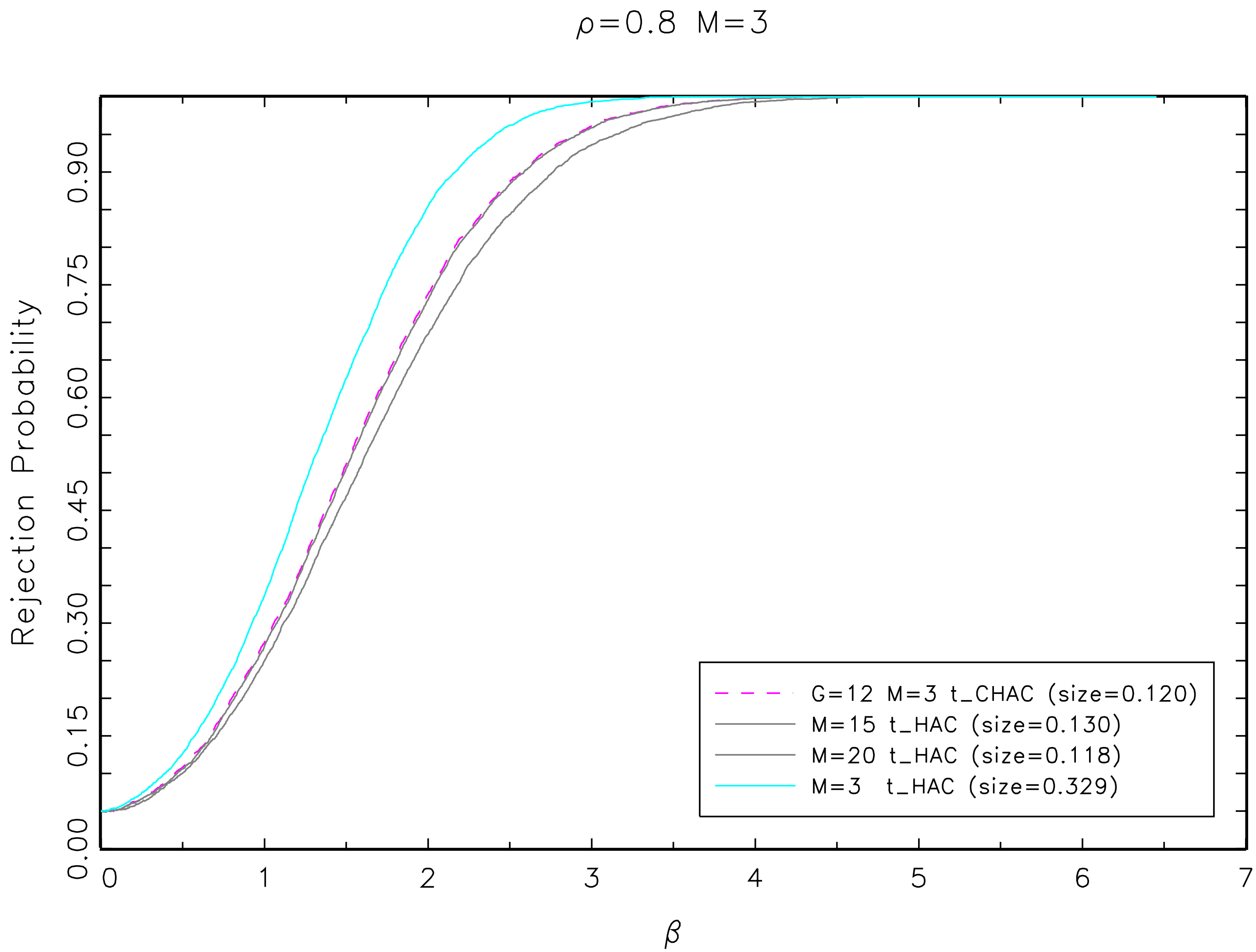


Figure 3.12: (cont'd)

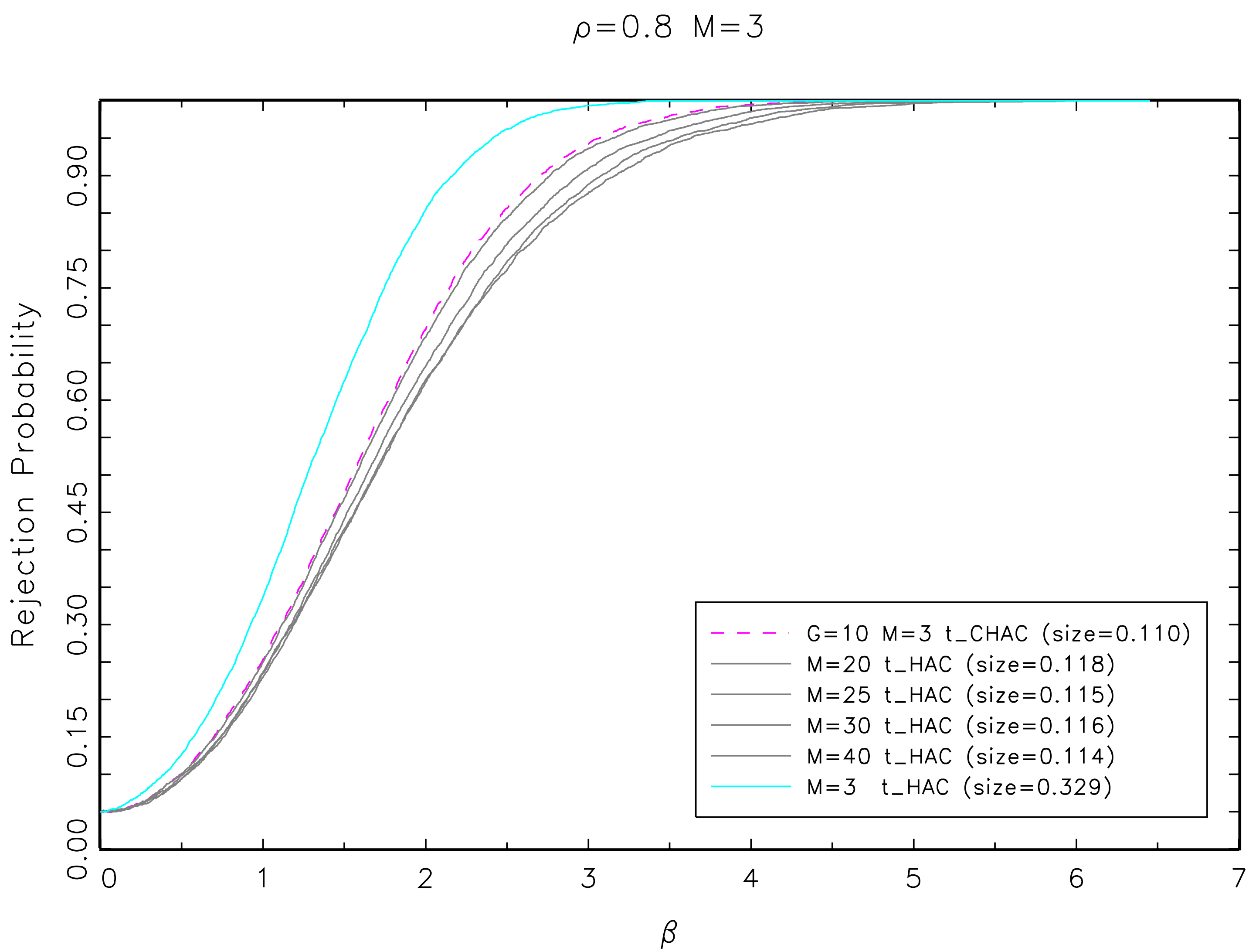


Figure 3.13: Size Adjusted Power Comparision- Clustering V.S. Smoothing, $\rho = 0.8$, $M = 4$

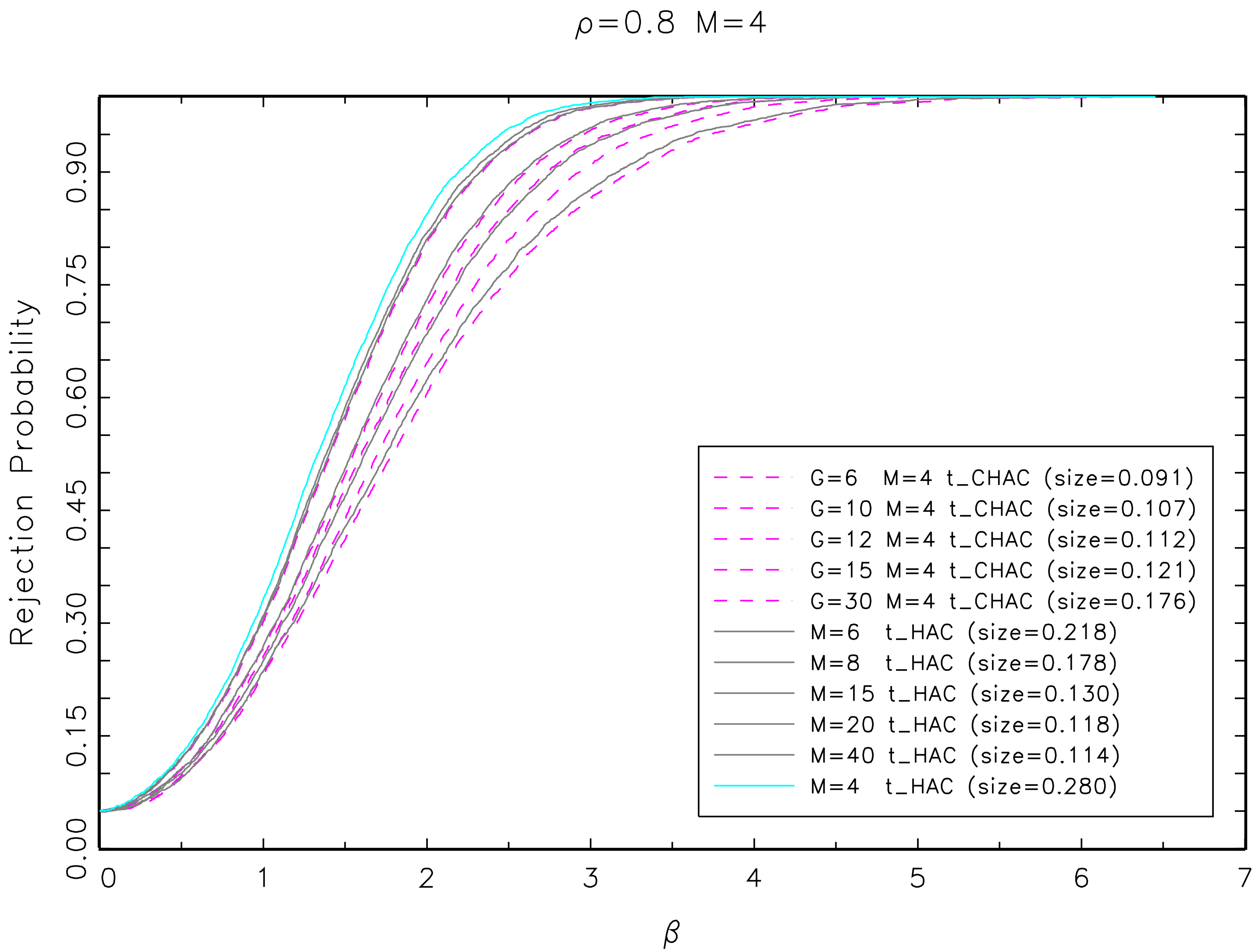


Figure 3.13: (cont'd)

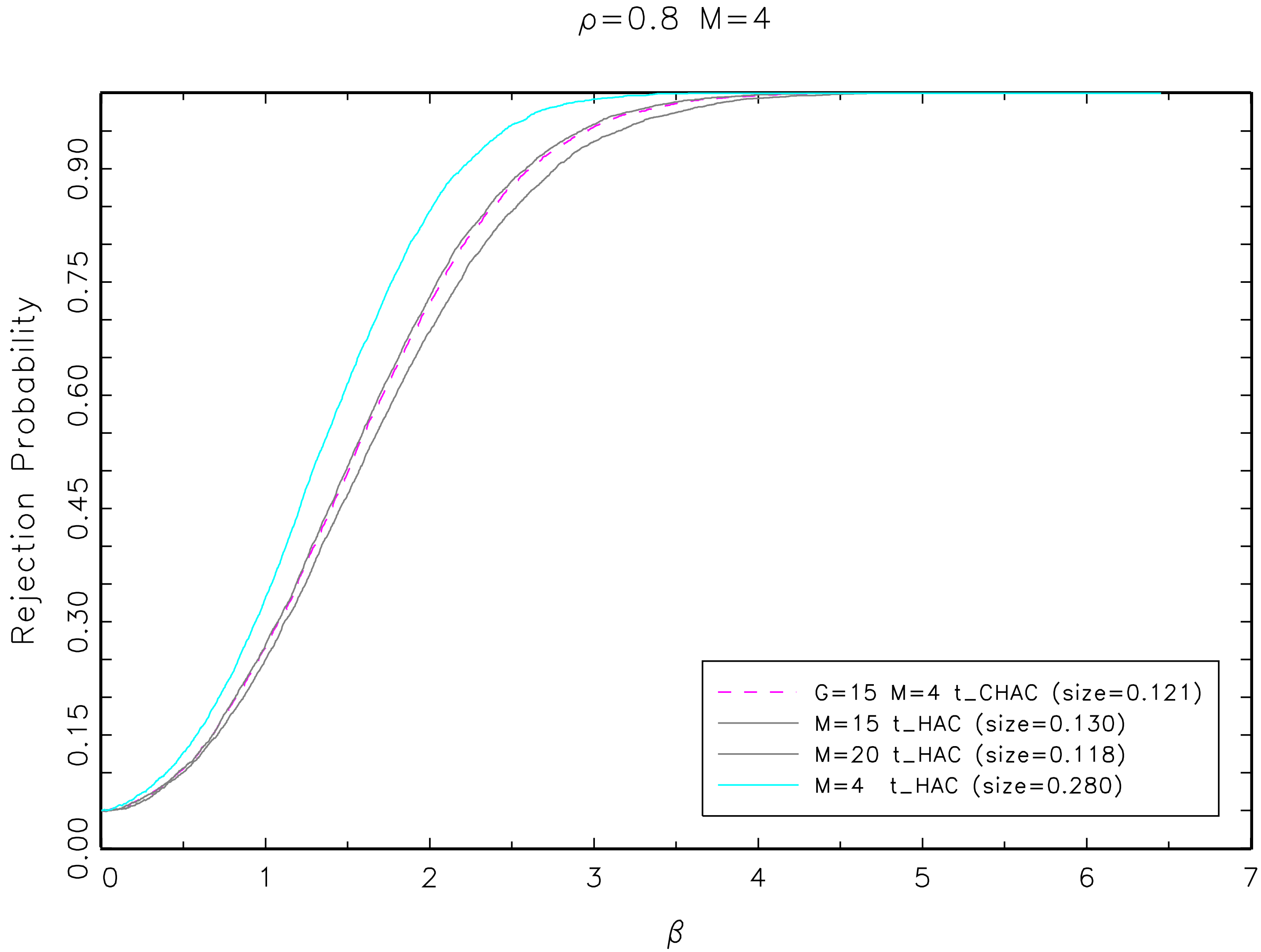


Figure 3.13: (cont'd)

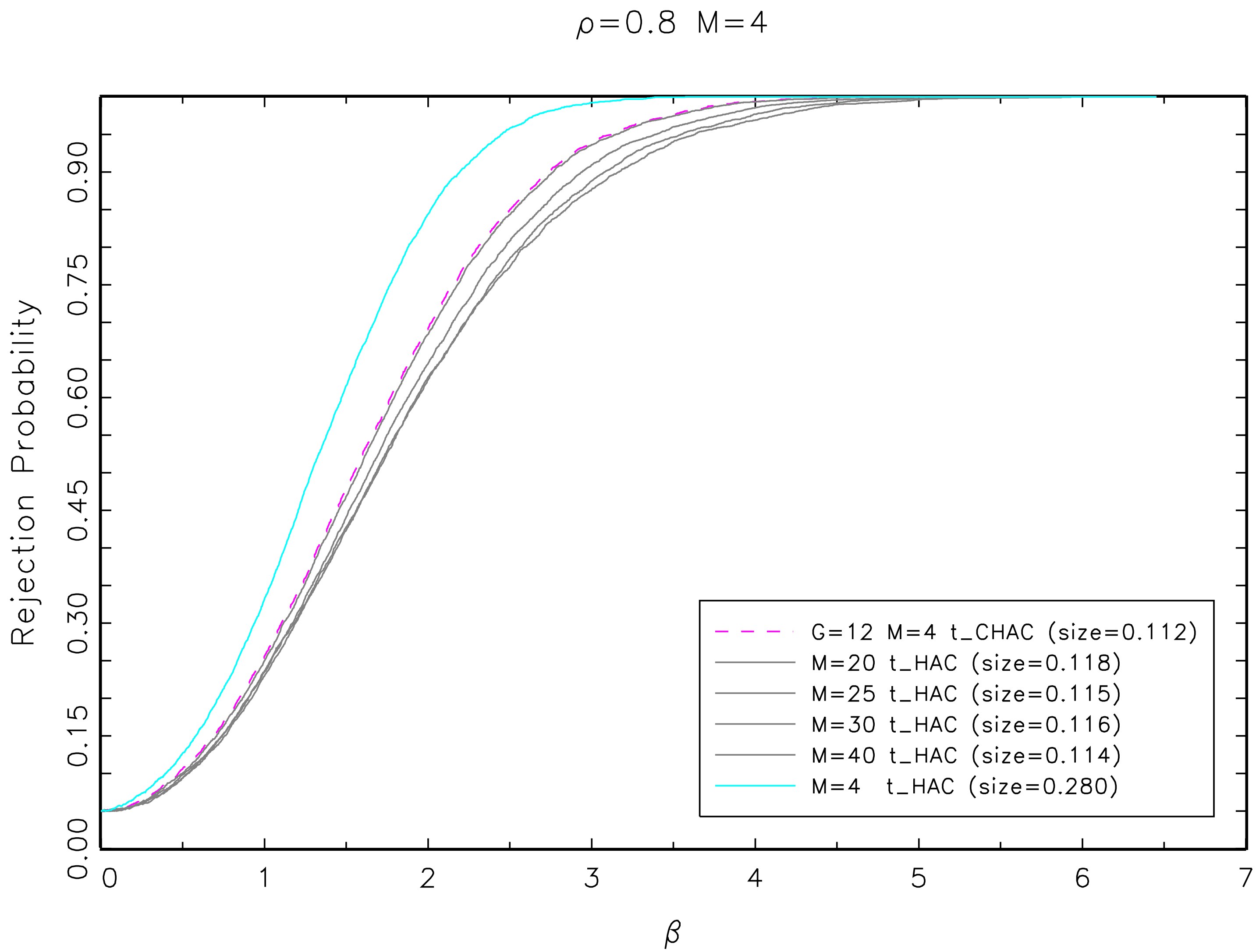


Figure 3.13: (cont'd)

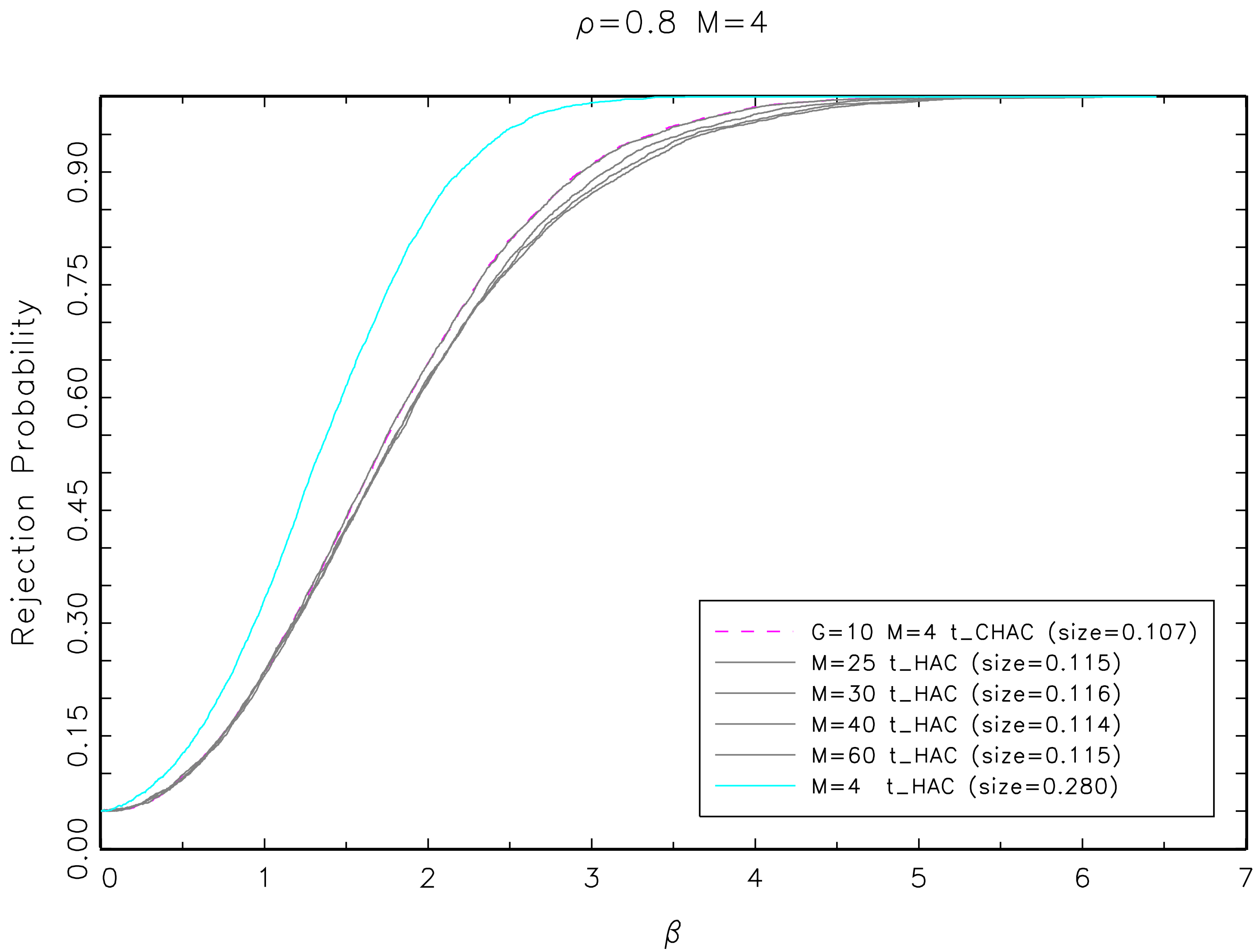


Figure 3.14: Size Adjusted Power Comparision- Clustering V.S. Smoothing, $\rho = 0.8$, $M = 5$

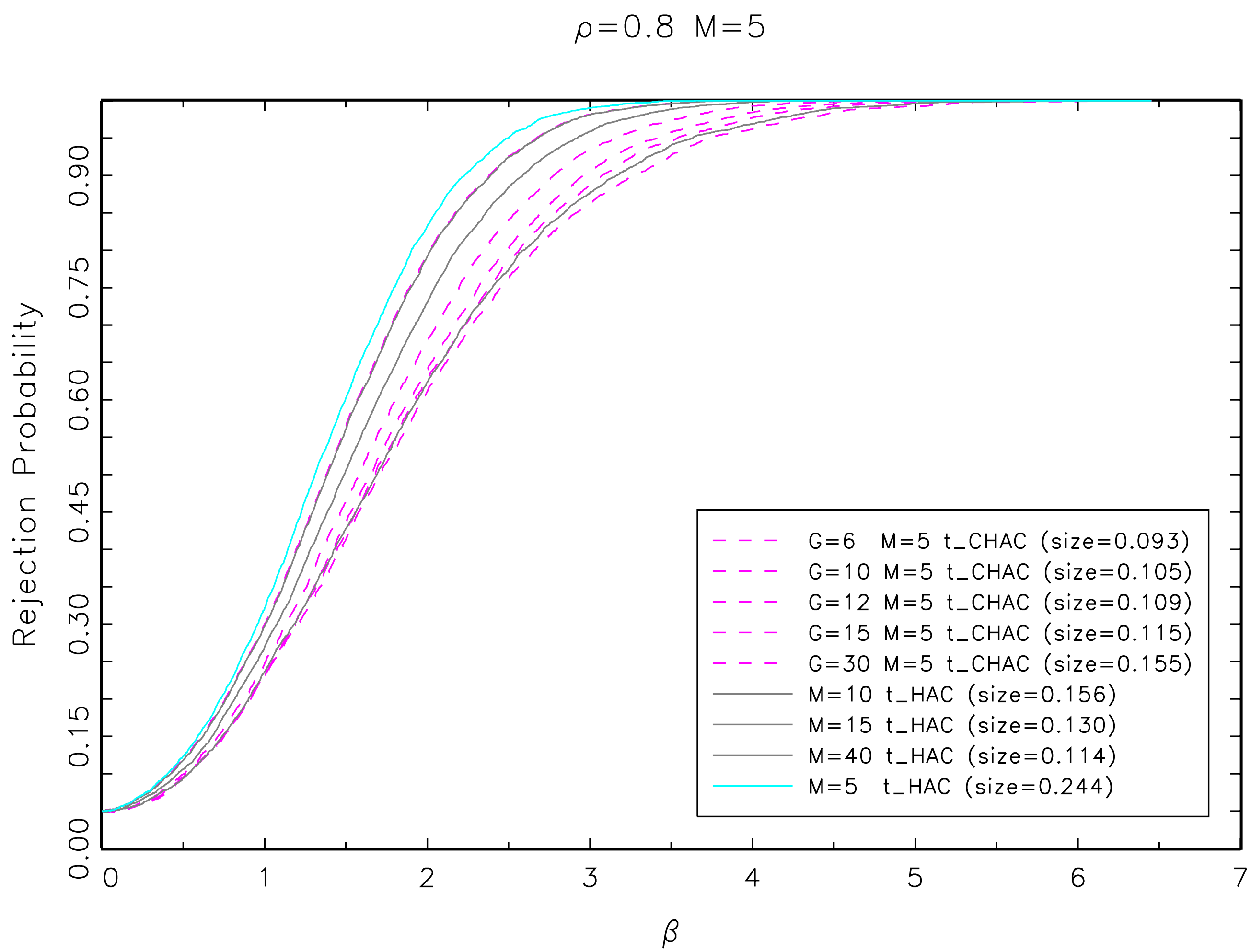


Figure 3.14: (cont'd)

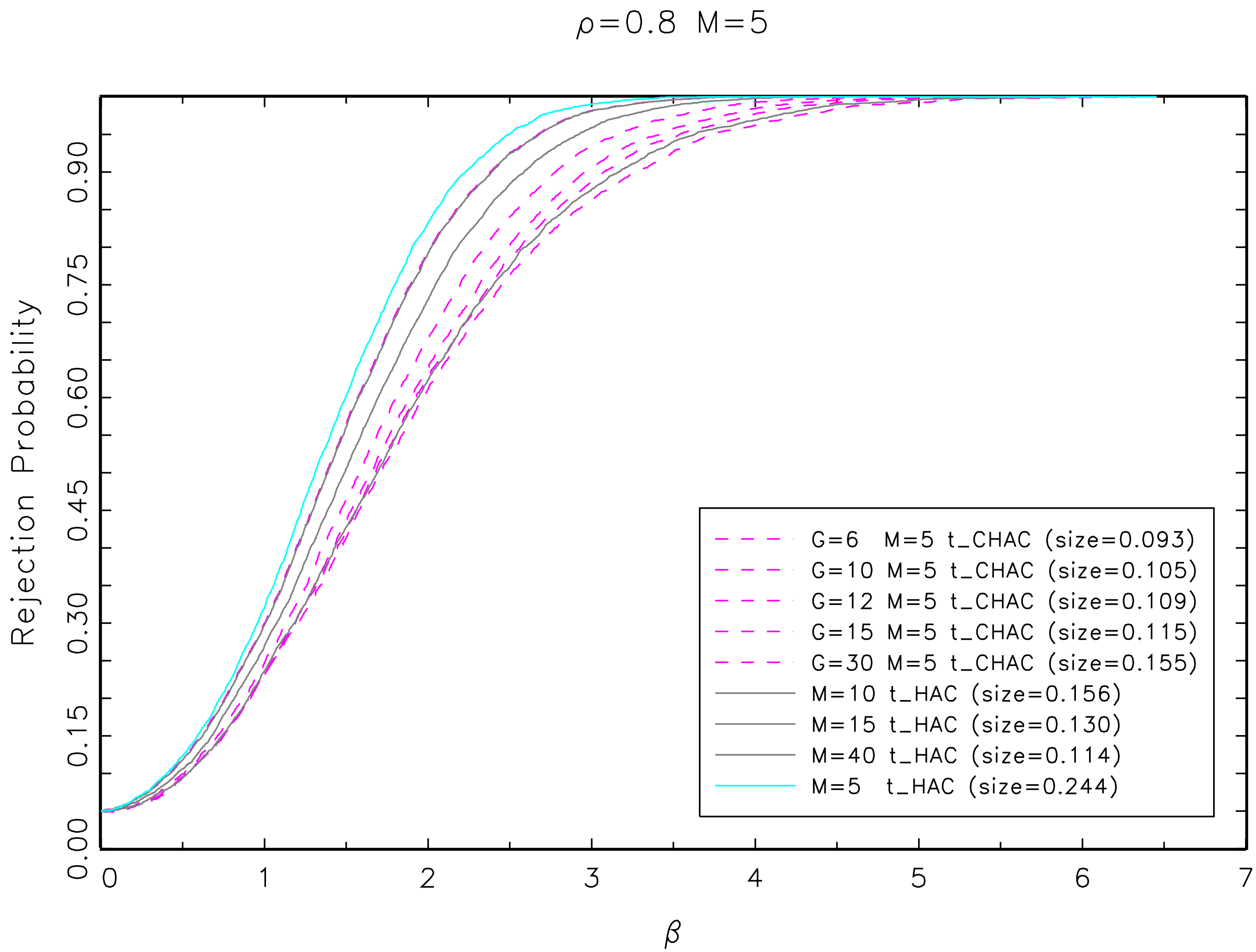


Figure 3.15: Size Adjusted Power Comparision- Clustering V.S. Smoothing, $\rho = 0.8$, $M = 6$

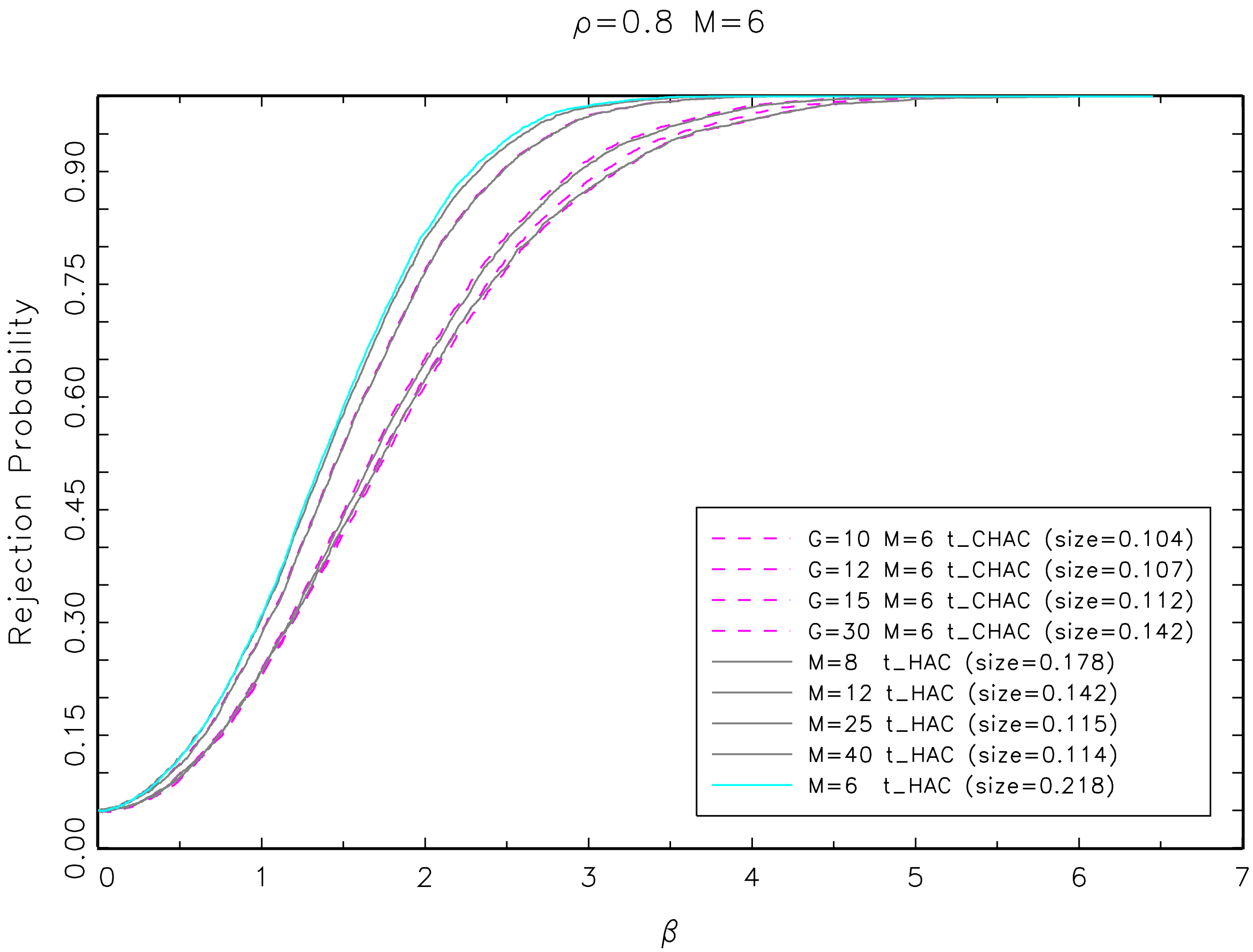


Figure 3.16: Size Adjusted Power Comparision- Clustering V.S. Smoothing, $\rho = 0.8$, $M = 7$

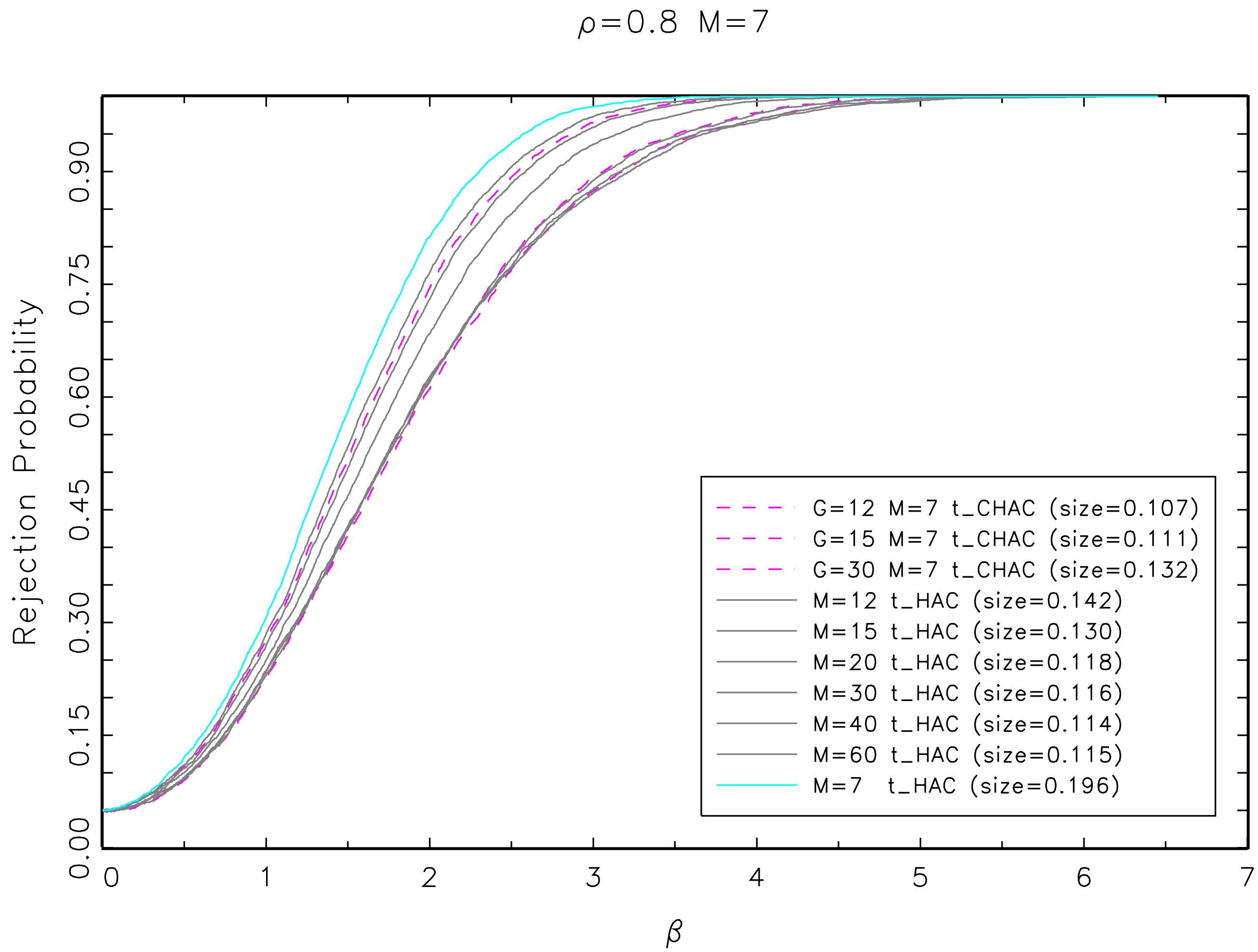


Figure 3.17: Size Adjusted Power Comparision- Clustering V.S. Smoothing, $\rho = 0.8$, $M = 8$

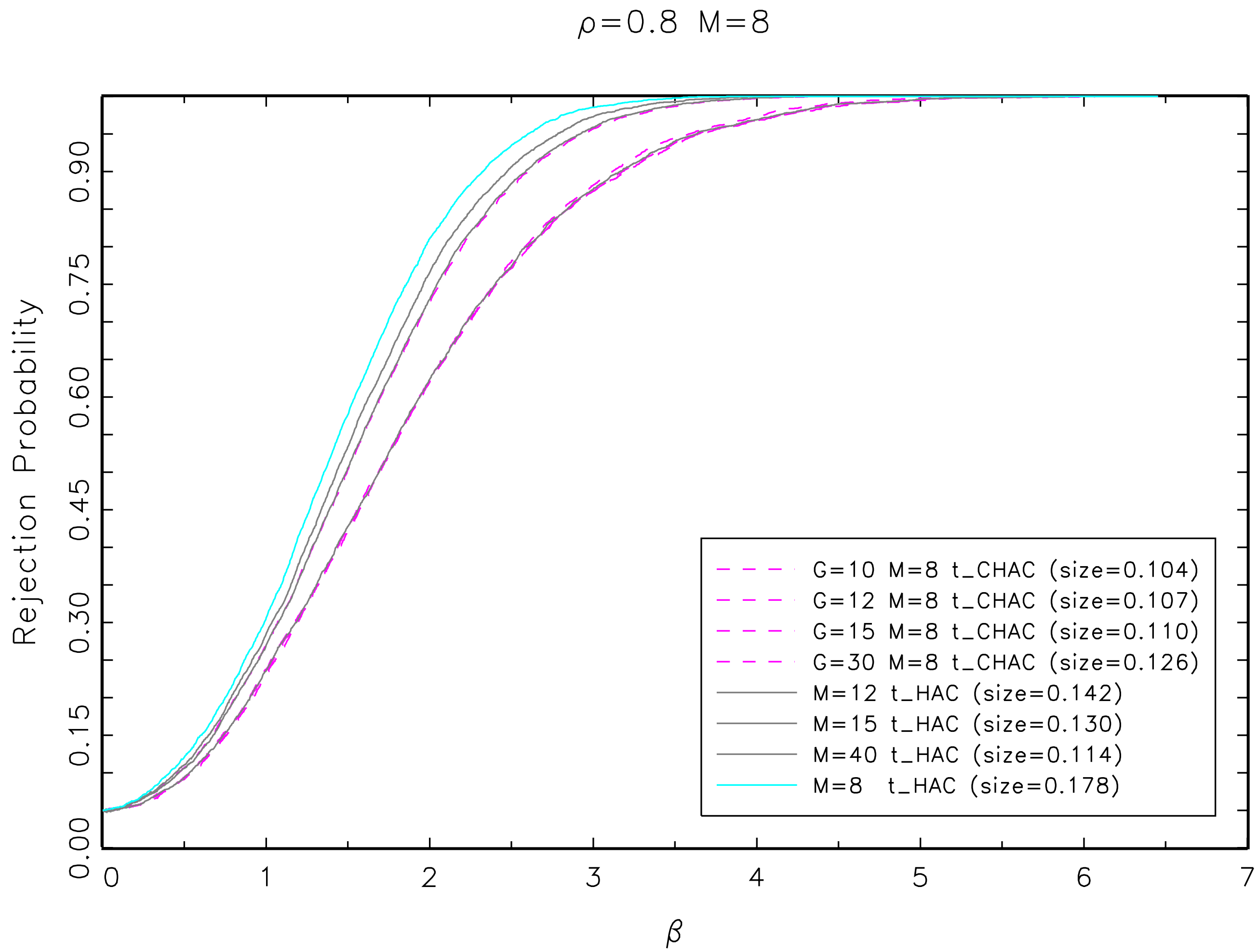


Figure 3.18: Size Adjusted Power Comparision- Clustering V.S. Smoothing, $\rho = 0.8$, $M = 9$

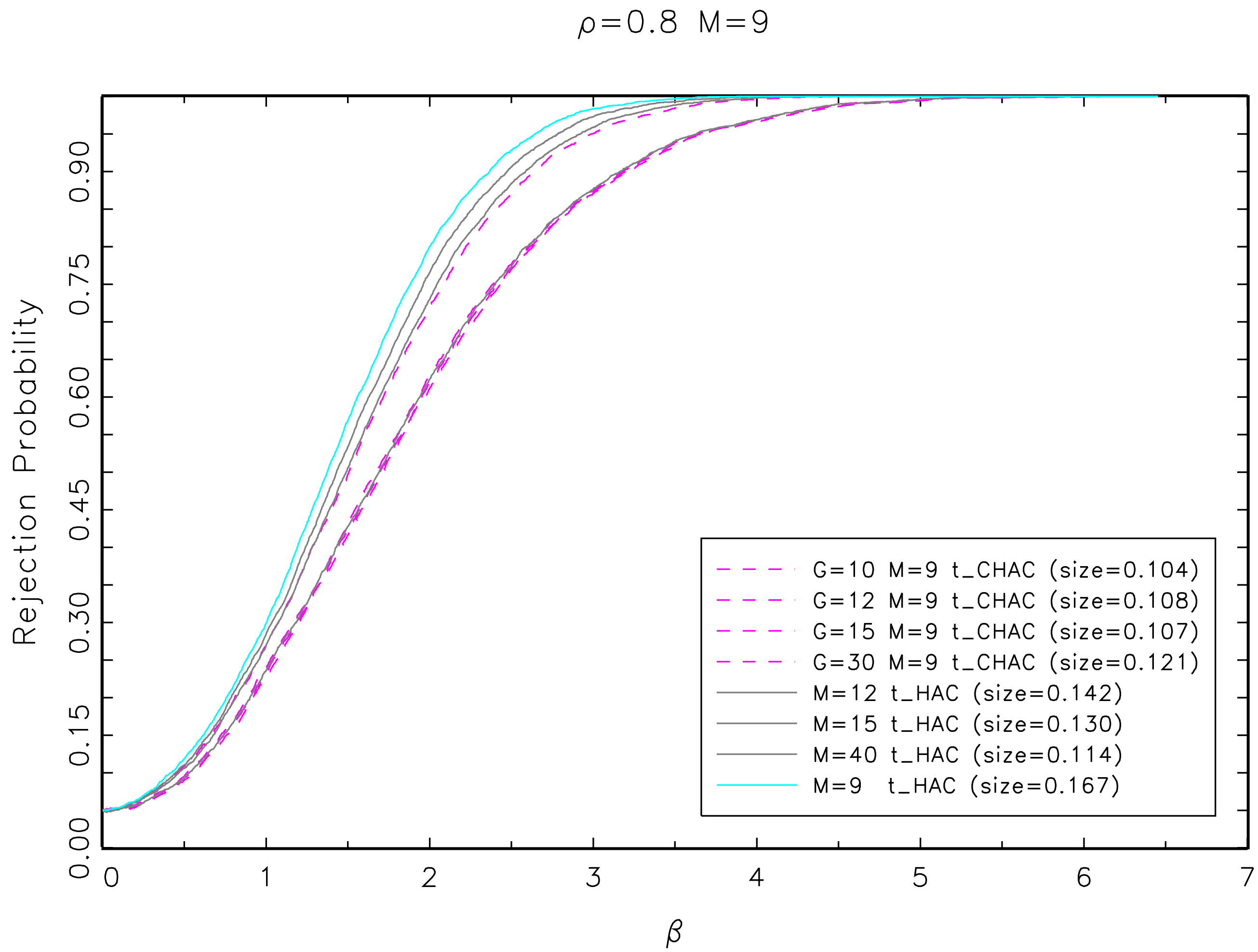


Figure 3.19: Size Adjusted Power Comparision- Clustering V.S. Smoothing, $\rho = 0.9$, $M = 1$

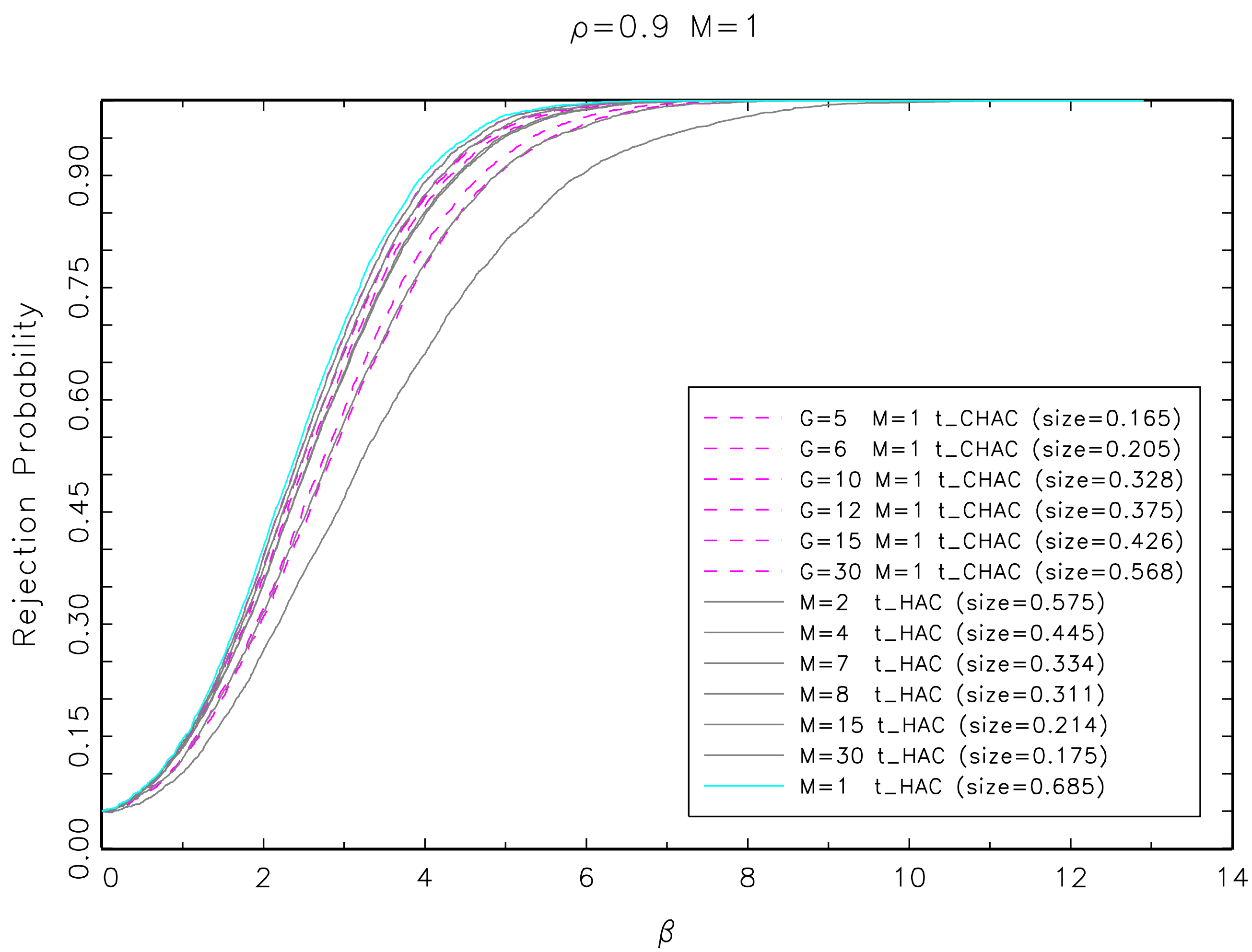


Figure 3.19: (cont'd)

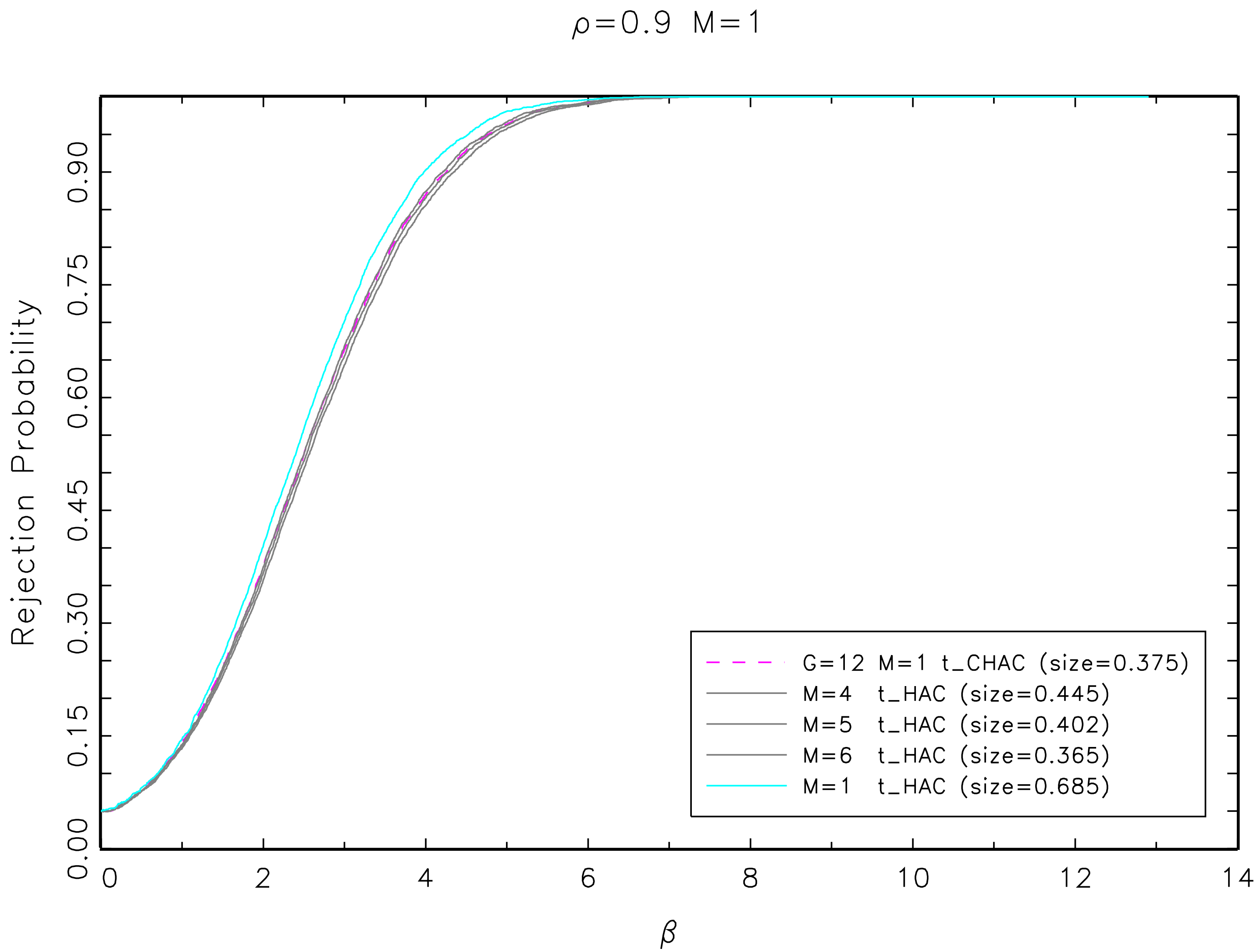


Figure 3.19: (cont'd)

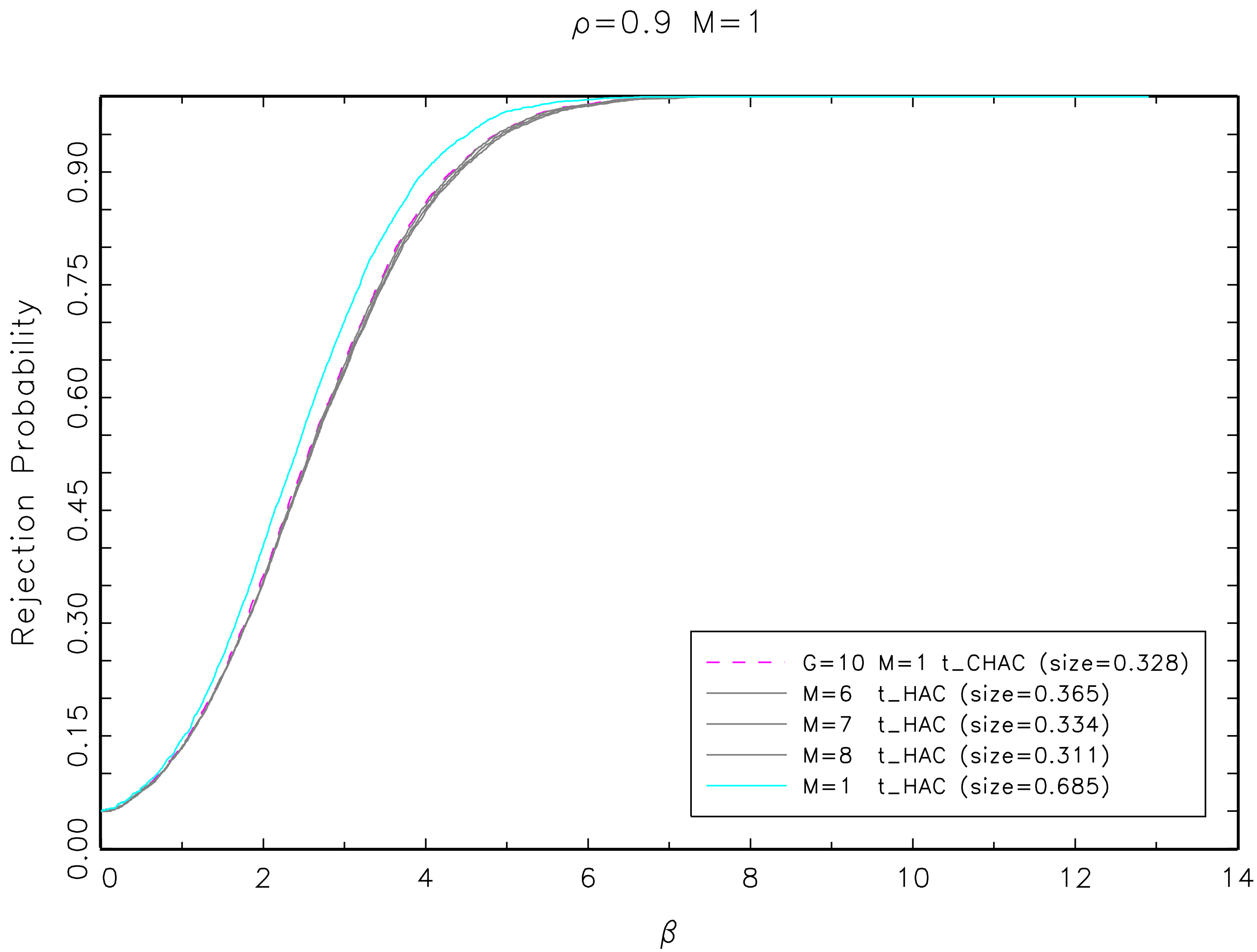


Figure 3.19: (cont'd)

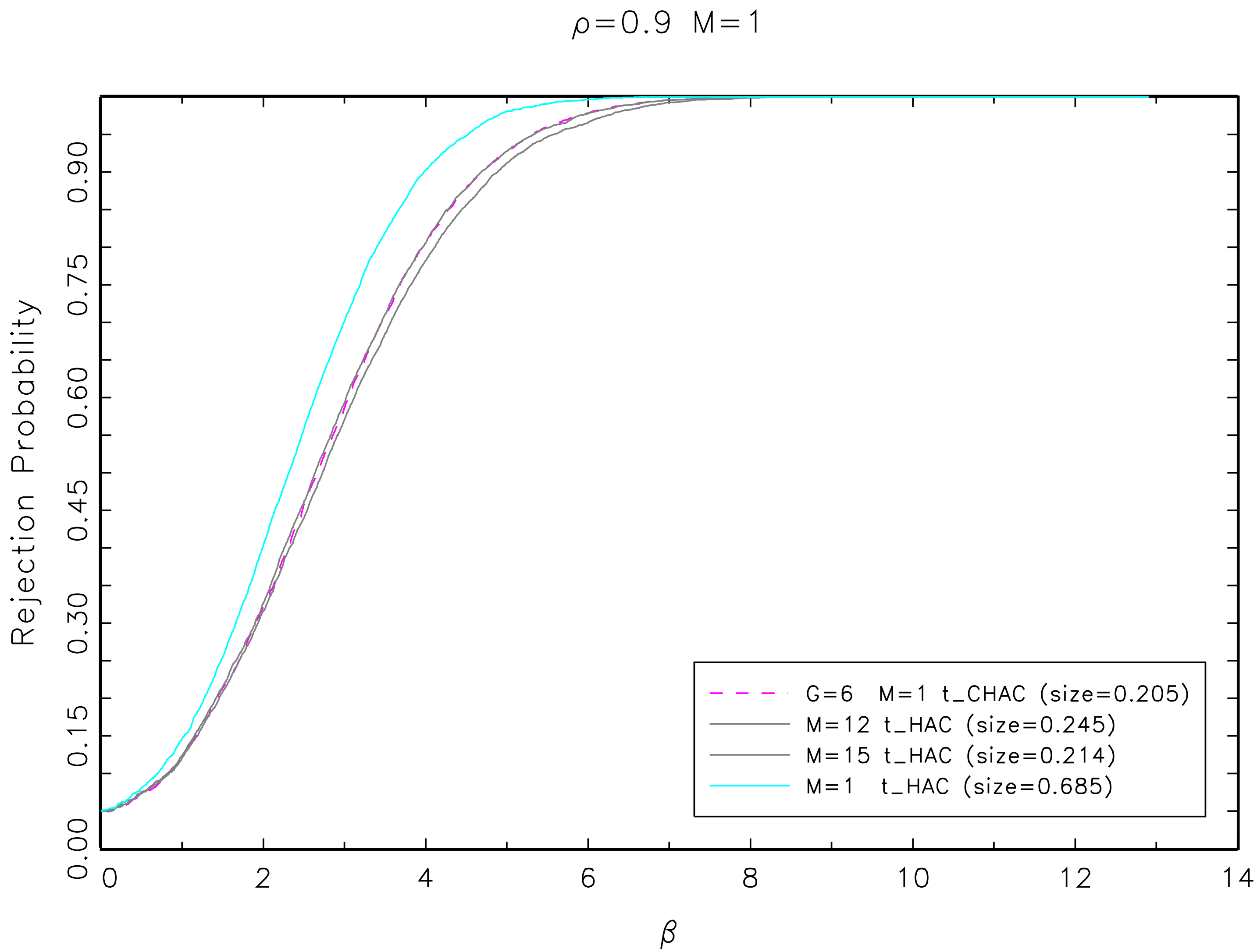


Figure 3.20: Size Adjusted Power Comparision- Clustering V.S. Smoothing, $\rho = 0.9$, $M = 2$

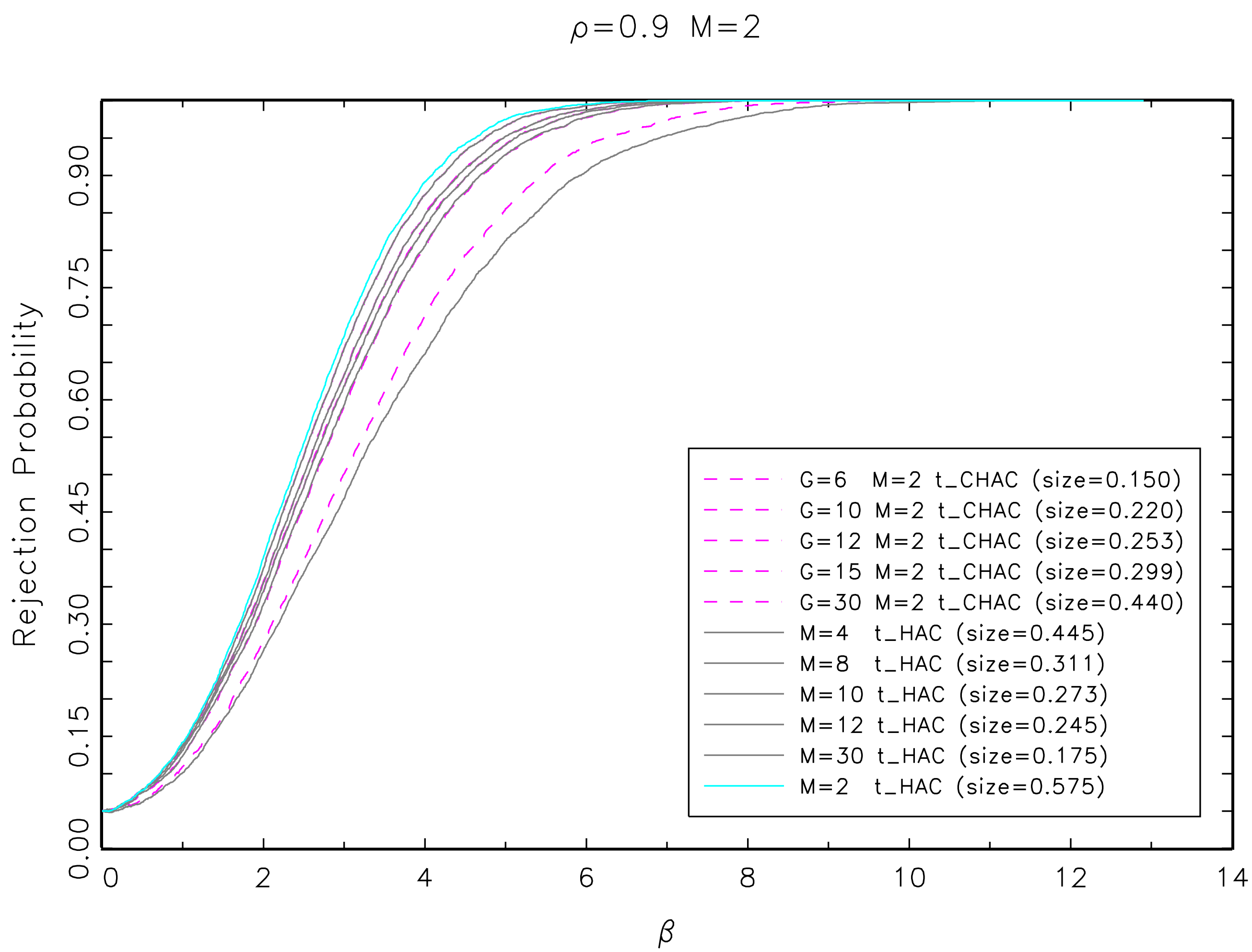


Figure 3.21: Size Adjusted Power Comparision- Clustering V.S. Smoothing, $\rho = 0.9$, $M = 3$

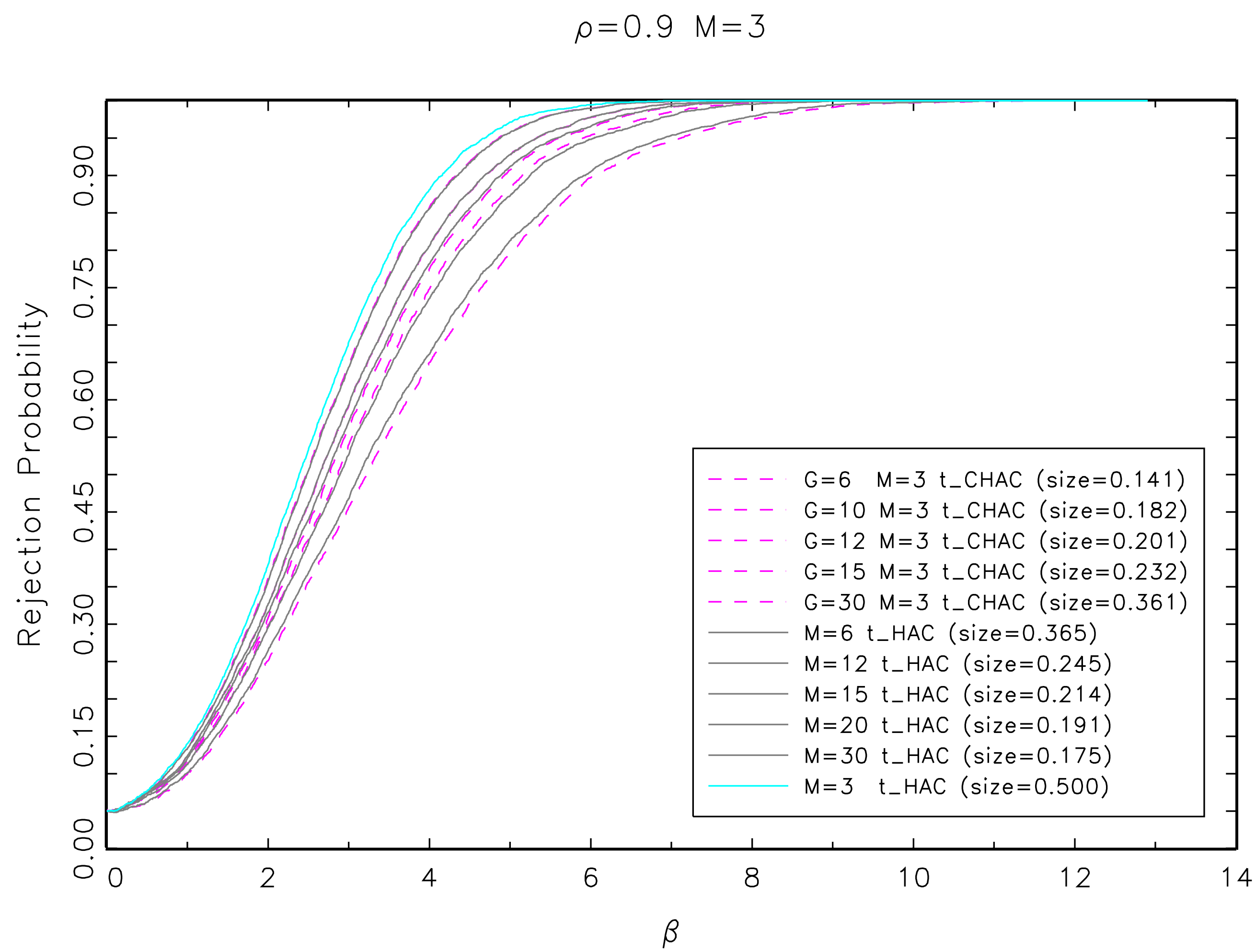


Figure 3.21: (cont'd)

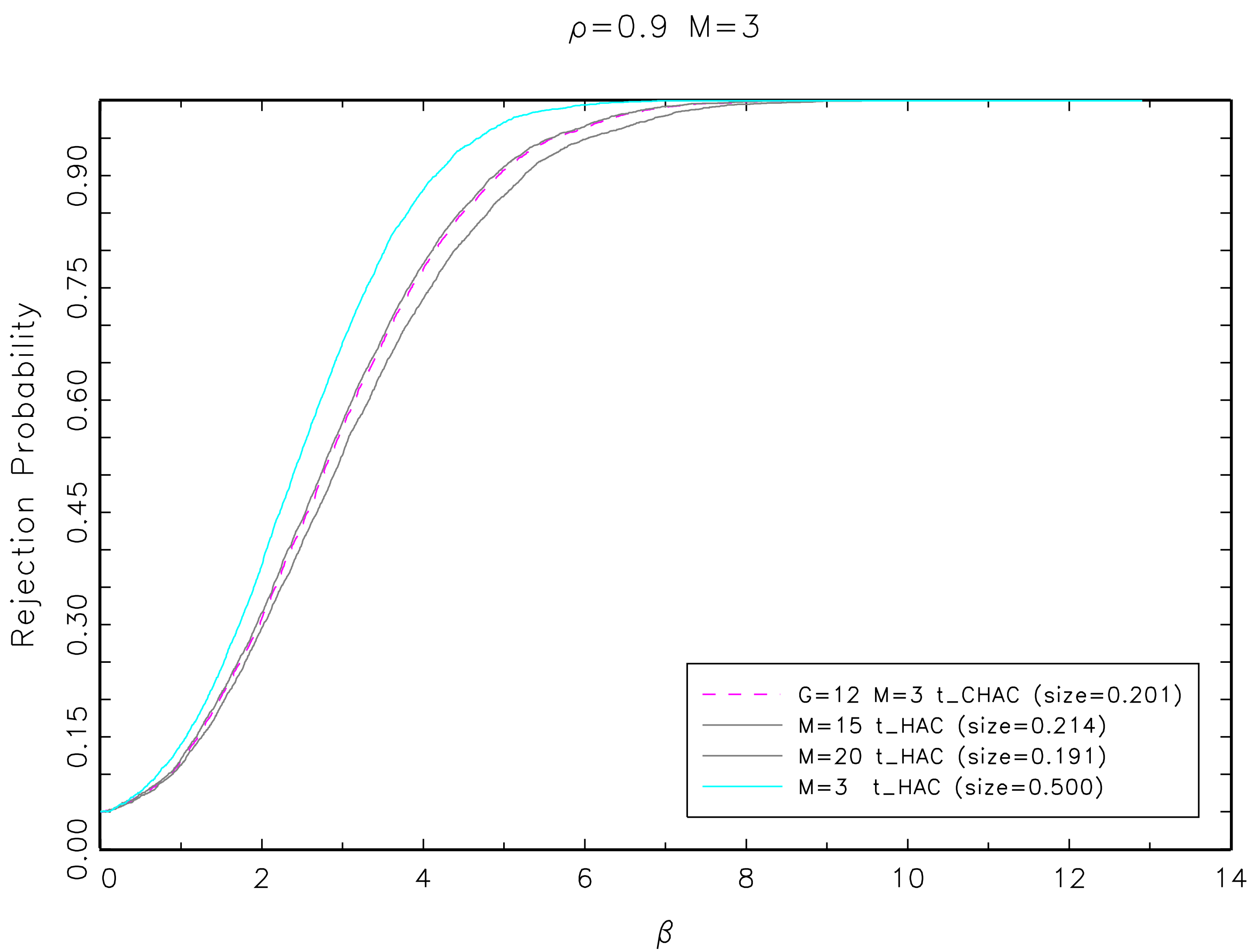


Figure 3.22: Size Adjusted Power Comparision- Clustering V.S. Smoothing, $\rho = 0.9$, $M = 4$

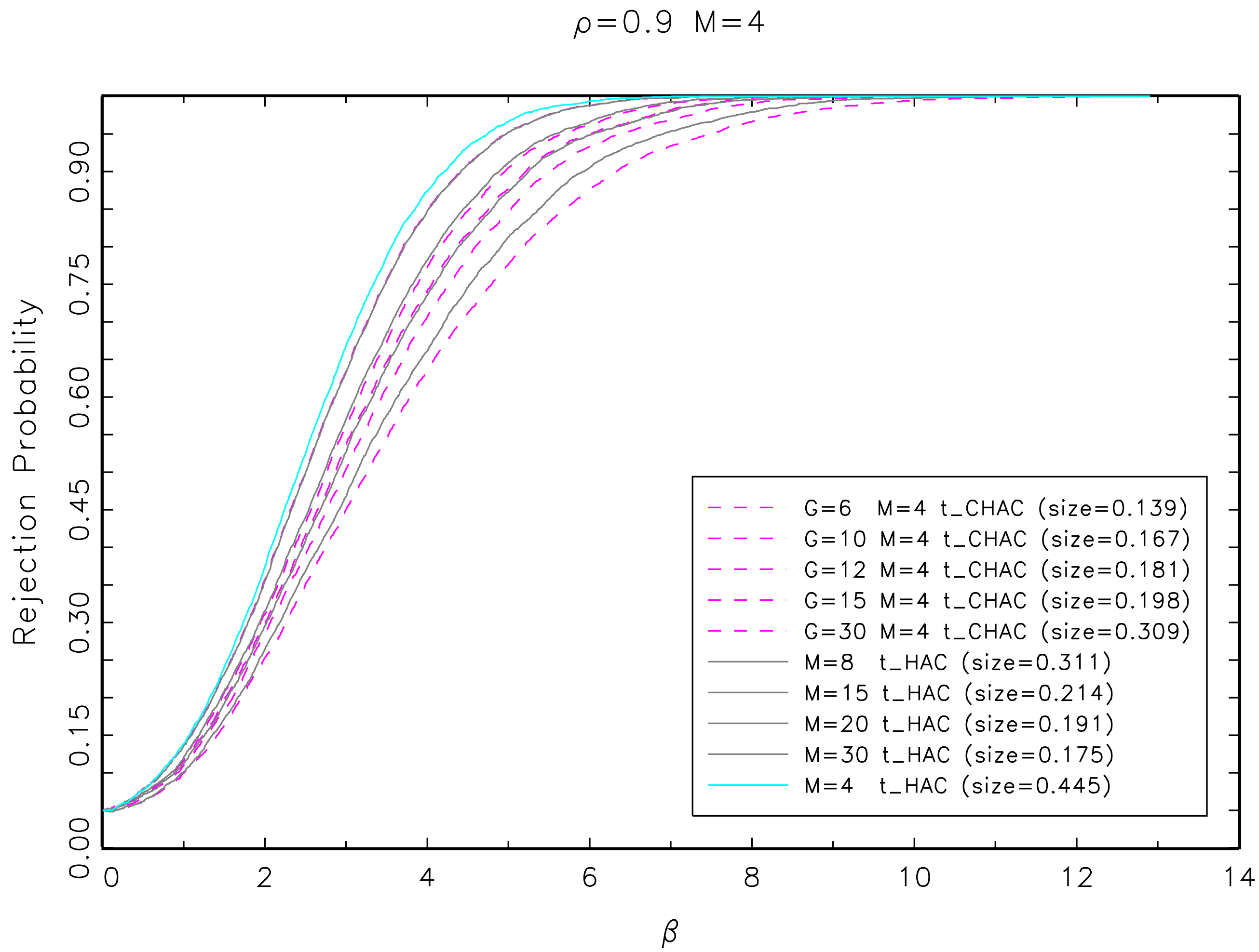


Figure 3.23: Size Adjusted Power Comparision- Clustering V.S. Smoothing, $\rho = 0.9$, $M = 5$

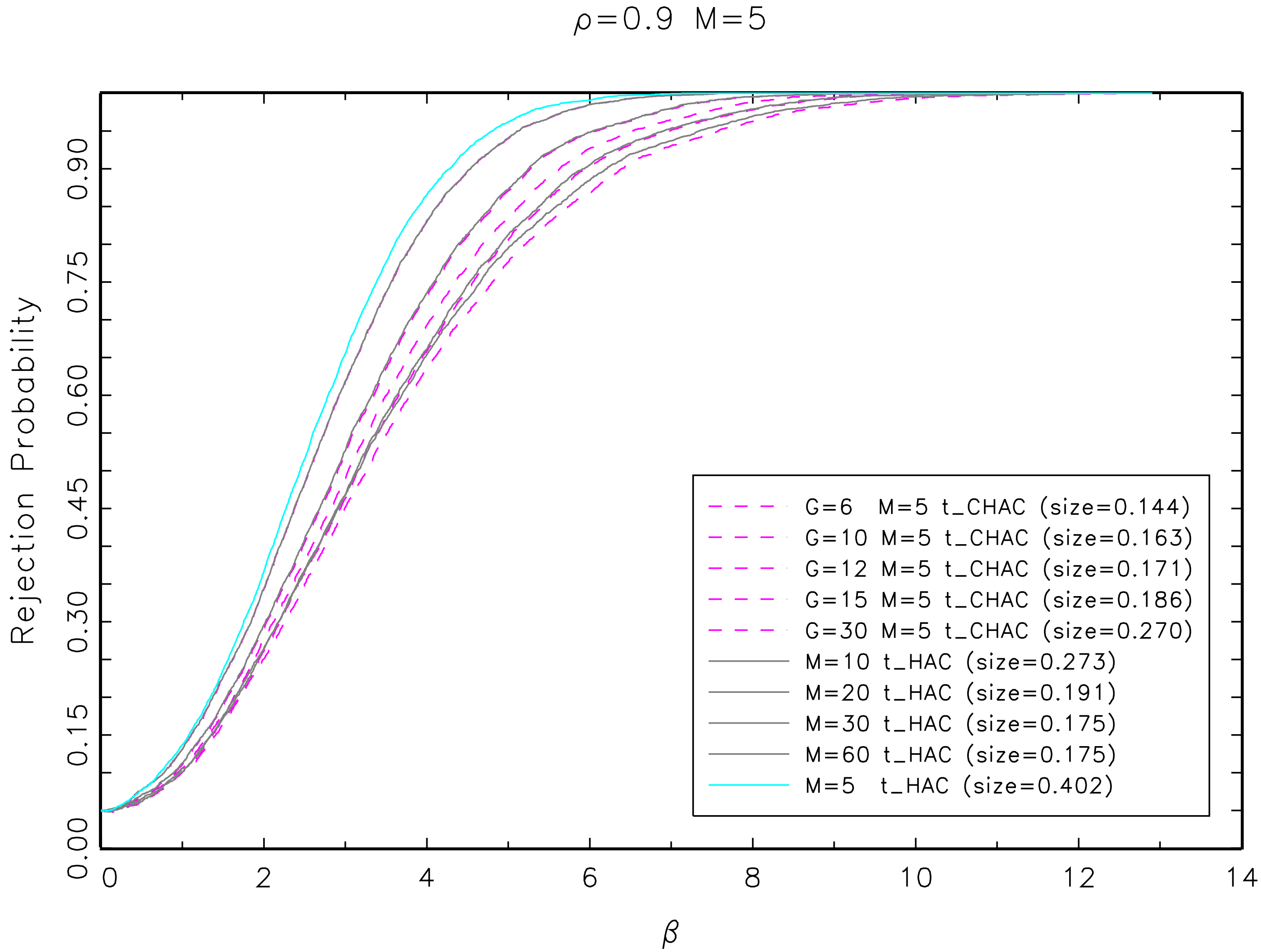


Figure 3.24: Size Adjusted Power Comparision- Clustering V.S. Smoothing, $\rho = 0.9$, $M = 6$

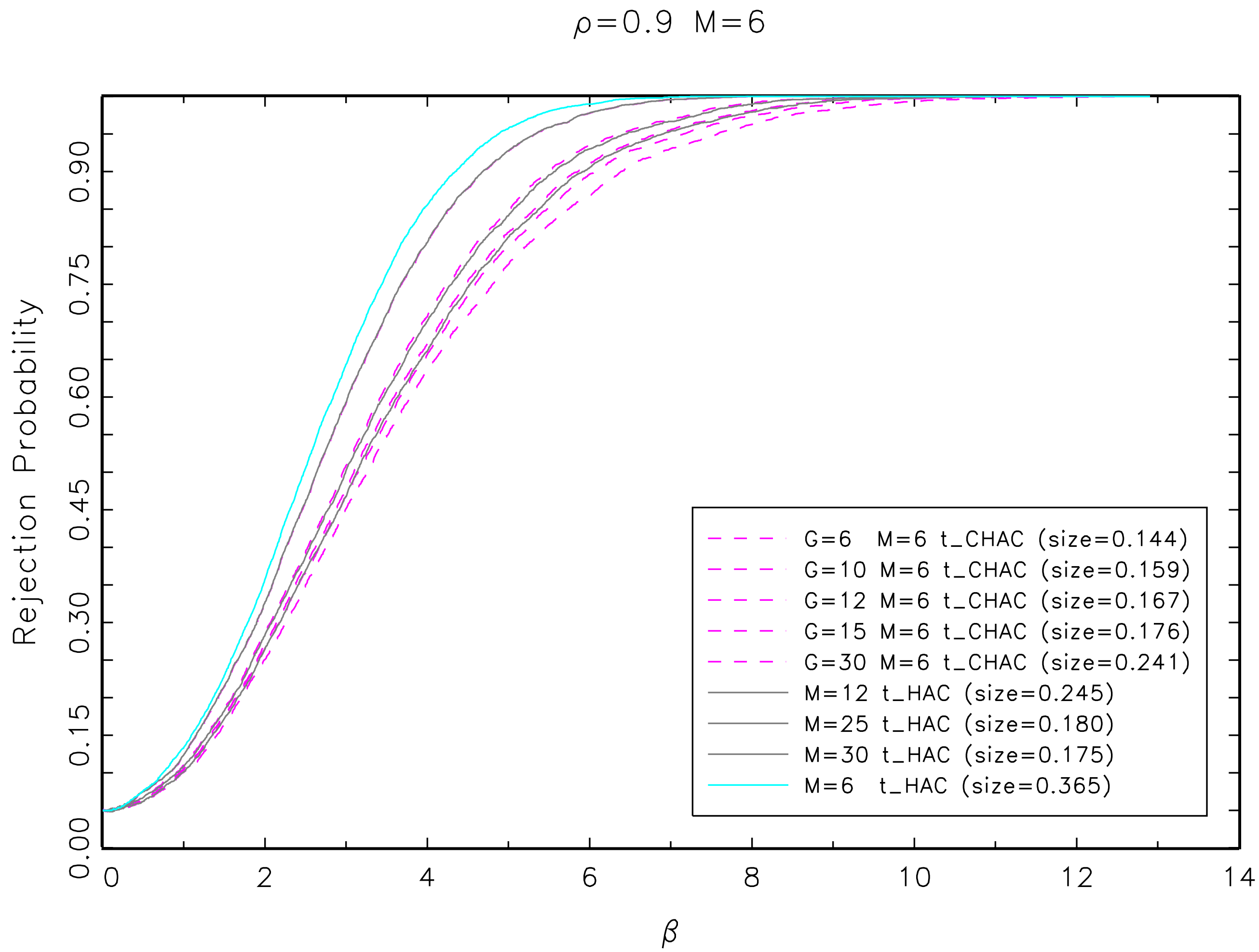


Figure 3.25: Size Adjusted Power Comparision- Clustering V.S. Smoothing, $\rho = 0.9$, $M = 7$

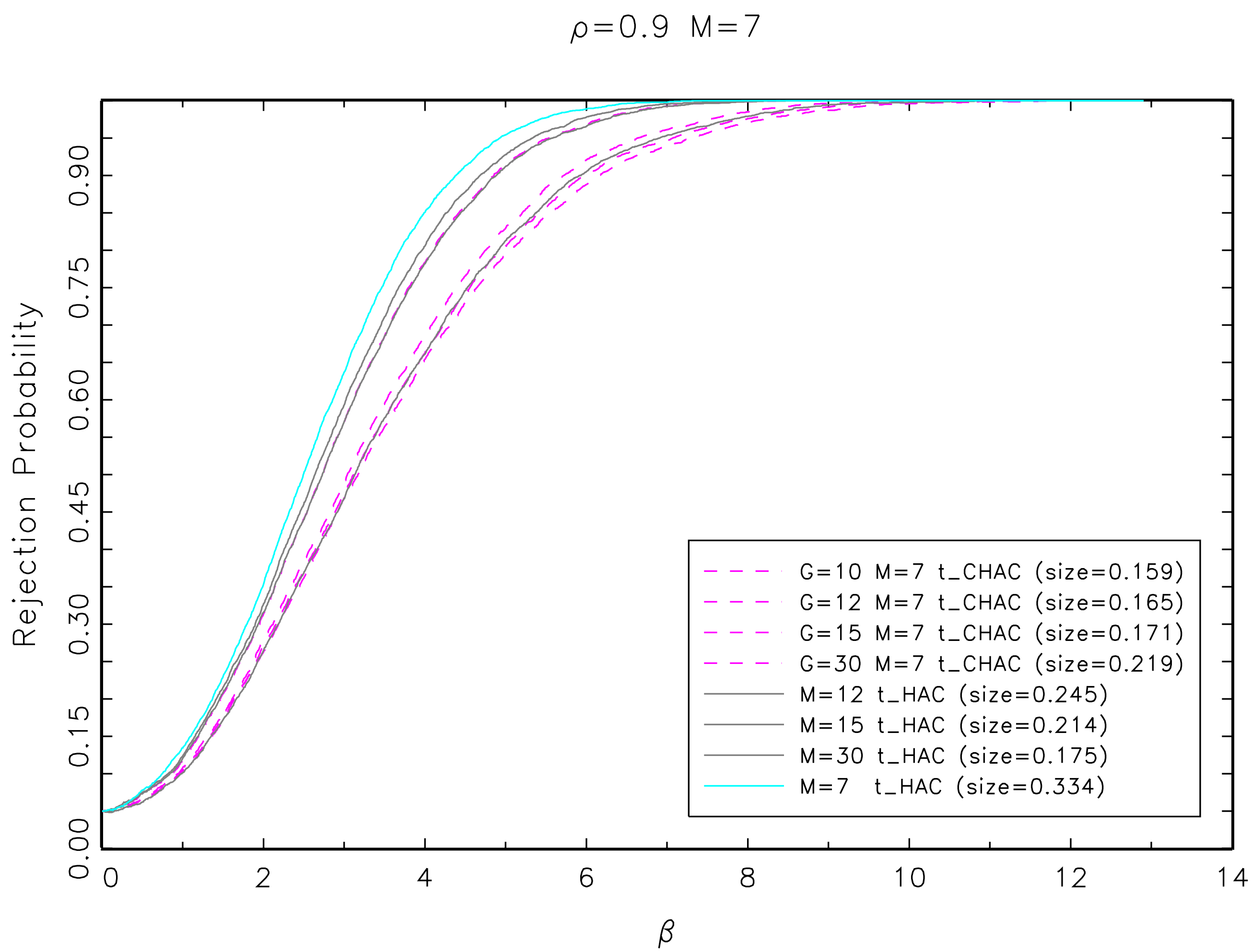


Figure 3.26: Size Adjusted Power Comparision- Clustering V.S. Smoothing, $\rho = 0.9$, $M = 8$

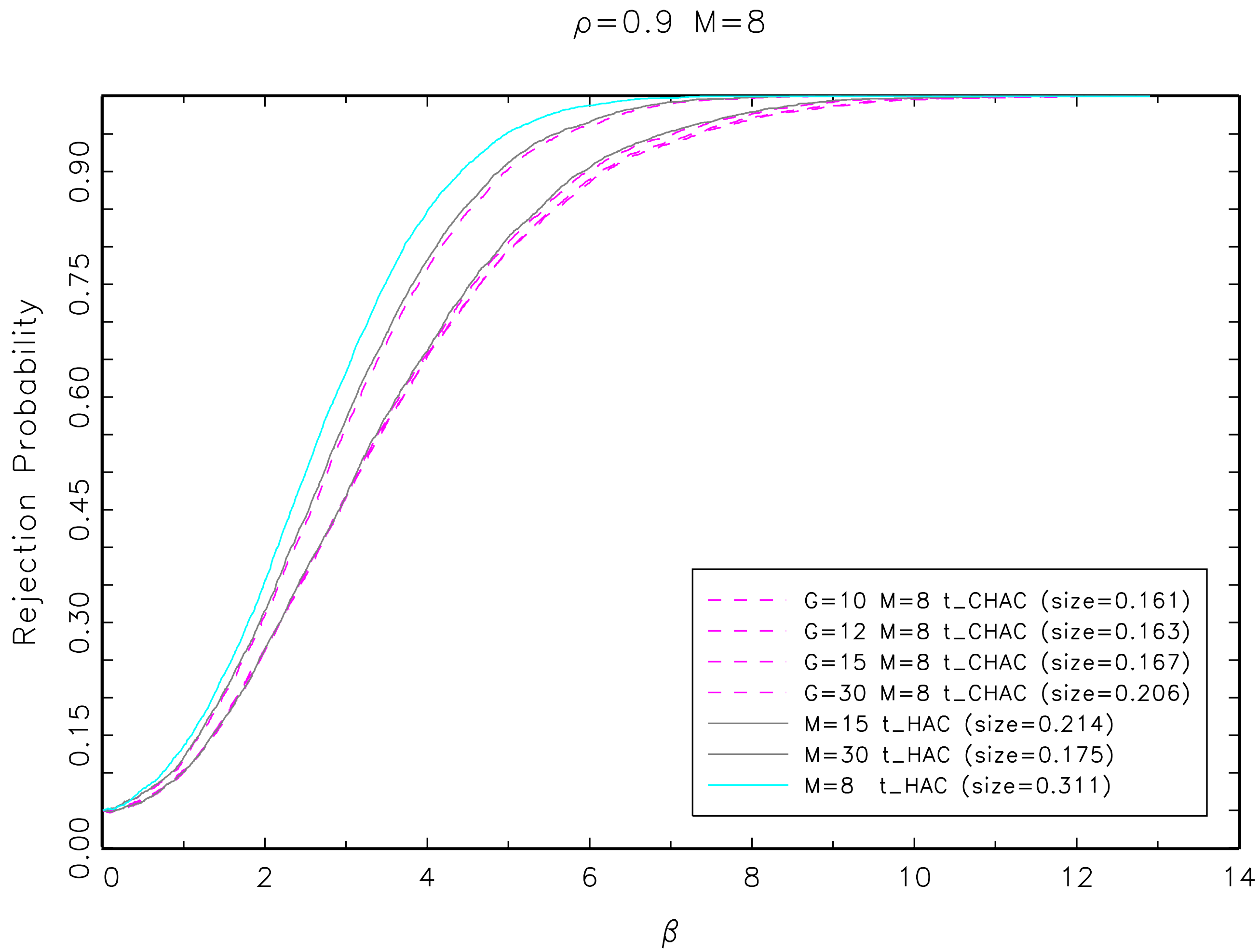
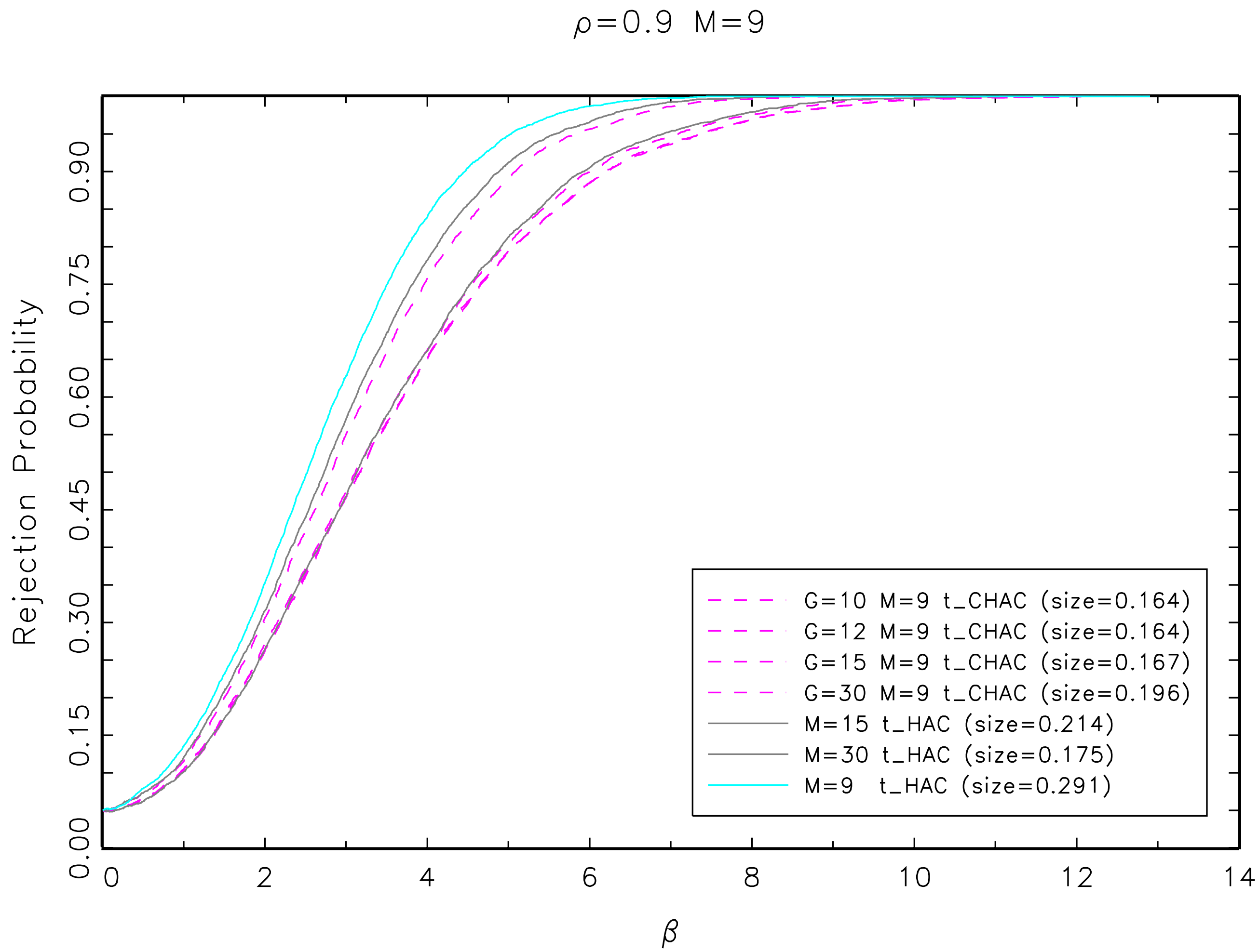


Figure 3.27: Size Adjusted Power Comparision- Clustering V.S. Smoothing, $\rho = 0.9$, $M = 9$



APPENDICES

Appendix A

PROOFS FOR CHAPTER 1

We will use the following notation. Let $f_v(\varepsilon_i) = \sqrt{\frac{1+\lambda^2}{\sigma^2}}\phi\left(\varepsilon_i\sqrt{\frac{1+\lambda^2}{\sigma^2}}\right)$, $f_\varepsilon(\varepsilon_i) = \frac{2}{\sigma}\phi\left(\frac{\varepsilon_i}{\sigma}\right)\left(1-\Phi\left(\frac{\varepsilon_i\lambda}{\sigma}\right)\right)$, $f_p(\varepsilon_i) = pf_v(\varepsilon_i) + (1-p)f_\varepsilon(\varepsilon_i)$, $\ln L = \sum \ln f_p(\varepsilon_i)$, $m_i = \frac{\phi\left(\frac{\varepsilon_i\lambda}{\sigma}\right)}{1-\Phi\left(\frac{\varepsilon_i\lambda}{\sigma}\right)}$, $\theta = \left(\beta', \lambda, \sigma^2, p\right)'$, $\beta = k \times 1$ vector, $\theta^{**} = \left(\hat{\beta}', \hat{\lambda}, \hat{\sigma}^2, \hat{p}\right)'$, where $\hat{\beta}$ =OLS, $\hat{\lambda} = 0$, $\hat{\sigma}^2 = \frac{1}{n}\sum\hat{\varepsilon}_i^2$, $\hat{\varepsilon}_i = y_i - \mathbf{x}_i'\hat{\beta}$, $\hat{p} \in [0,1]$. $\vee$ indicates maximum.

Result 1. *θ^{**} is a stationary point of the log likelihood function.*

Proof. The first derivative of $\ln L$ is :

$$S(\theta) = \begin{pmatrix} \frac{\partial \ln L}{\partial \beta} \\ \frac{\partial \ln L}{\partial \lambda} \\ \frac{\partial \ln L}{\partial \sigma^2} \\ \frac{\partial \ln L}{\partial p} \end{pmatrix}$$

$$= \begin{pmatrix} \sum_{i=1}^{n} \frac{pf_v(\varepsilon_i)\left(\frac{1+\lambda^2}{\sigma^2}\varepsilon_i\mathbf{x}_i\right) + (1-p)f_\varepsilon(\varepsilon_i)\left(\frac{\varepsilon_i\mathbf{x}_i}{\sigma^2} + \frac{m_i\mathbf{x}_i\lambda}{\sigma}\right)}{f_p(\varepsilon_i)} \\ \sum_{i=1}^{n} \frac{pf_v(\varepsilon_i)\left(\frac{\lambda}{1+\lambda^2} - \frac{\lambda}{\sigma^2}\varepsilon_i^2\right) - (1-p)f_\varepsilon(\varepsilon_i)\left(\frac{1}{\sigma}m_i\varepsilon_i\right)}{f_p(\varepsilon_i)} \\ \sum_{i=1}^{n} \frac{pf_v(\varepsilon_i)\left(-\frac{1}{2\sigma^2} + \frac{1+\lambda^2}{2\sigma^4}\varepsilon_i^2\right) + (1-p)f_\varepsilon(\varepsilon_i)\left(-\frac{1}{2\sigma^2} + \frac{1}{2\sigma^4}\varepsilon_i^2 + \frac{\lambda}{2\sigma^3}m_i\varepsilon_i\right)}{f_p(\varepsilon_i)} \\ \sum_{i=1}^{n} \frac{f_v(\varepsilon_i) - f_\varepsilon(\varepsilon_i)}{f_p(\varepsilon_i)} \end{pmatrix}.$$

When $\lambda = 0$,

$$S(\theta)|_{\lambda=0} = \begin{pmatrix} \frac{1}{\sigma^2} \sum_{i=1}^{n} \varepsilon_i \mathbf{x}_i \\ -(1-p)\sqrt{\frac{2}{\pi}} \frac{1}{\sigma} \sum_{i=1}^{n} \varepsilon_i \\ -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^{n} \varepsilon_i^2 \\ 0 \end{pmatrix}.$$

It is straightforward that $S(\theta^{**}) = 0$, since, with $\varepsilon_i = \hat{\varepsilon}_i$, $\sum_{i=1}^{n} \hat{\varepsilon}_i = 0$ and $\sum_{i=1}^{n} \hat{\varepsilon}_i \mathbf{x}_i = 0$. Therefore, θ^{**} is a stationary point.

□

Result 2. *Evaluated at the stationary point, θ^{**}, the Hessian of the log likelihood is negative semi-definite with two zero eigenvalues.*

Proof. The Hessian evaluated at the stationary point θ^{**} is

$$H(\theta^{**}) = \begin{pmatrix} -\frac{1}{\hat{\sigma}^2} \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i' & (1-\hat{p})\sqrt{\frac{2}{\pi}} \frac{1}{\hat{\sigma}} \sum_{i=1}^{n} \mathbf{x}_i & 0 & 0 \\ (1-\hat{p})\sqrt{\frac{2}{\pi}} \frac{1}{\hat{\sigma}} \sum_{i=1}^{n} \mathbf{x}_i' & -(1-\hat{p})^2 \frac{2n}{\pi} & 0 & 0 \\ 0 & 0 & -\frac{n}{2\hat{\sigma}^4} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

When $\hat{p} = 1$,

$$H(\theta^{**}) = \begin{pmatrix} -\frac{1}{\hat{\sigma}^2} \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i' & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -\frac{n}{2\hat{\sigma}^4} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Because $-\frac{1}{\hat{\sigma}^2} \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i'$ is a negative definite matrix, $H(\theta^{**})$ is a negative semi-definite matrix with two zero eigenvalues.

Now suppose that $\hat{p} \neq 1$. Note that the first row of $H\left(\theta^{**}\right)$, $\left(-\frac{1}{\hat{\sigma}^2}\sum_{i=1}^{n}\mathbf{x}_i', (1-\hat{p})\sqrt{\frac{2}{\pi}}\frac{n}{\hat{\sigma}}, 0, 0\right)$, is linearly dependent with the $(k+1)^{th}$ row of $H\left(\theta^{**}\right)$. Multiplying the first row by $(1-\hat{p})\sqrt{\frac{2}{\pi}}\hat{\sigma}$ and adding to the $(k+1)^{th}$row results in a row vector of zeros. Hence,

$$H\left(\theta^{**}\right) \sim \begin{pmatrix} -\frac{1}{\hat{\sigma}^2}\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i' & (1-\hat{p})\sqrt{\frac{2}{\pi}}\frac{1}{\hat{\sigma}}\sum_{i=1}^{n}\mathbf{x}_i & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -\frac{n}{2\hat{\sigma}^4} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

where $\sim$ stands for an elementary row operation. Again, the first column and the $(k+1)^{th}$ column of the transferred matrix are linearly dependent. Similarly, multiplying the first column by $(1-\hat{p})\sqrt{\frac{2}{\pi}}\hat{\sigma}$ and adding to the $(k+1)^{th}$ column results in a column vector of zeros. In other words,

$$H\left(\theta^{**}\right) \sim \begin{pmatrix} -\frac{1}{\hat{\sigma}^2}\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i' & (1-\hat{p})\sqrt{\frac{2}{\pi}}\frac{1}{\hat{\sigma}}\sum_{i=1}^{n}\mathbf{x}_i & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -\frac{n}{2\hat{\sigma}^4} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\sim \begin{pmatrix} -\frac{1}{\hat{\sigma}^2}\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i' & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -\frac{n}{2\hat{\sigma}^4} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Elementary operations preserve the rank of a matrix. Hence, the rank of $H\left(\theta^{**}\right)$ is $k+1$, i.e., $H\left(\theta^{**}\right)$ has two zero eigenvalues.

Now we will show that $H(\theta^{**})$ is negative semi-definite. Let $\alpha = \left(\alpha_1', \alpha_2, \alpha_3, \alpha_4\right)'$ be an arbitrary non-zero $(k+3)\times 1$ vector, where α_1 is a $k\times 1$ vector, and $\alpha_2, \alpha_3, \alpha_4$ are scalars. Then,

$$
\begin{aligned}
&\alpha' H\left(\theta^{**}\right)\alpha \\
=&-\left(\frac{1}{\hat{\sigma}}\frac{1}{\sqrt{n}}\alpha_1'\sum_{i=1}^{n}\mathbf{x}_i-\alpha_2\left(1-p\right)\sqrt{\frac{2}{\pi}}\sqrt{n}\right)^2-\frac{1}{\hat{\sigma}^2}\alpha_1'\left(\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i'-\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i\sum_{i=1}^{n}\mathbf{x}_i'\right)\alpha_1 \\
&-\frac{n}{2\hat{\sigma}^4}\alpha_3^2 \\
\leq&0,
\end{aligned}
$$

because

$$
\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i'-\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i\sum_{i=1}^{n}\mathbf{x}_i'=\sum_{i=1}^{n}\left(\mathbf{x}_i-\frac{1}{n}\sum_{j=1}^{n}\mathbf{x}_j\right)\left(\mathbf{x}_i-\frac{1}{n}\sum_{j=1}^{n}\mathbf{x}_j\right)'
$$

is positive semi-definite. Therefore $H\left(\theta^{**}\right)$ is negative semi-definite.

□

Result 3. *θ^{**} with $\hat{p}\in[0,1)$ is a local maximizer of the log likelihood function if and only if $\sum_{i=1}^{n}\hat{\varepsilon}_i^3>0$.*

Proof. From Result 2, we know that the Hessian evaluated at θ^{**} is negative semi-definite. Therefore, if the log likelihood decreases in the direction of the two eigenvectors associated with zero eigenvalues, θ^{**} is a local maximizer of the log likelihood. The two eigenvectors that are associated with the two zero eigenvalue are

$$
\begin{pmatrix} (1-\hat{p})\sqrt{\frac{2}{\pi}}\hat{\sigma} \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}.
$$

Let

$$\Delta\theta = \mu \begin{pmatrix} (1-\hat{p})\sqrt{\frac{2}{\pi}}\hat{\sigma} \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} + \phi \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} (1-\hat{p})\sqrt{\frac{2}{\pi}}\hat{\sigma}\mu \\ 0 \\ \mu \\ 0 \\ \phi \end{pmatrix}.$$

Because $\lambda \geq 0,\ \mu > 0$. $\Delta\theta$ has only three non-zero arguments. Thus, relevant parameters would be β_0, λ, and p. By Taylor's expansion,

$$\begin{aligned}
&L\left(\theta^* + \Delta\theta\right) - L\left(\theta^*\right) \\
&= \frac{1}{6}\Bigg[L_{\beta_0\beta_0\beta_0}\left((1-\hat{p})\sqrt{\frac{2}{\pi}}\hat{\sigma}\mu\right)^3 + 3L_{\beta_0\beta_0\lambda}\left((1-\hat{p})\sqrt{\frac{2}{\pi}}\hat{\sigma}\mu\right)^2\mu \\
&+ 3L_{\beta_0\lambda\lambda}\left((1-\hat{p})\sqrt{\frac{2}{\pi}}\hat{\sigma}\mu\right)\mu^2 + 3L_{\beta_0\beta_0 p}\left((1-\hat{p})\sqrt{\frac{2}{\pi}}\hat{\sigma}\mu\right)^2\phi \\
&+ 3L_{\beta_0 pp}\left((1-\hat{p})\sqrt{\frac{2}{\pi}}\hat{\sigma}\mu\right)\phi^2 + L_{\lambda\lambda\lambda}\,\mu^3 + 3L_{\lambda\lambda p}\mu^2\phi + 3L_{\lambda pp}\,\mu\phi^2 + L_{ppp}\,\phi^3 \\
&+ 6L_{\beta_0 p\lambda}\left((1-\hat{p})\sqrt{\frac{2}{\pi}}\hat{\sigma}\mu\right)\mu\phi\Bigg] + o\left((\mu\vee\phi)^4\right) \\
&= (1-\hat{p})\frac{1}{6\pi}\sqrt{\frac{2}{\pi}}\frac{1}{\hat{\sigma}^3}\left(-4\hat{p}^2 + \hat{p}(8-3\pi) + \pi - 4\right)\sum_{i=1}^{n}\hat{\varepsilon}_i^3\mu^3 + o\left((\mu\vee\phi)^4\right).
\end{aligned}$$

The 1st order term is zero because θ^{**} is a stationary point (Result 1). The 2nd order term is zero by the definition of the eigenvector. Note that $\left(-4\hat{p}^2 + \hat{p}(8-3\pi) + \pi - 4\right)$ has its maximum, $\pi - 4 < 0$, when $\hat{p} = 0$. Since $\mu > 0$, $L\left(\theta^* + \Delta\theta\right) - L\left(\theta^*\right) < 0$ if and only if $\sum\hat{\varepsilon}_i^3 > 0$. Therefore, θ^{**} with $\hat{p} \in [0,1)$ is a local maximizer if and only if $\sum\hat{\varepsilon}_i^3 > 0$. When $\hat{p} = 0$, the expression goes back to the one in Waldman (1982). □

Result 4. *θ^{**} with $\hat{p} = 1$ is a local maximizer of the likelihood function if $\sum_{i=1}^{n}\hat{\varepsilon}_i^3 > 0$.*

Proof. The two eigenvectors associated with the zero eigenvalues are

$$\begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}.$$

Let

$$\Delta\theta = \mu \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} + \phi \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ \mu \\ 0 \\ \phi \end{pmatrix}.$$

Because $\lambda \geq 0$ and $p \leq 1$, $\mu > 0$ and $\phi < 0$. $\Delta\theta$ has only two non-zero arguments. Thus, the relevant parameters would be λ and p. By Taylor's expansion,

$$\begin{aligned}
&L\left(\theta^{*}+\Delta\theta\right)-L\left(\theta^{*}\right) \\
&=\frac{1}{24}\left[L_{\lambda\lambda\lambda\lambda}\mu^{4}+4L_{\lambda\lambda\lambda p}\mu^{3}\phi+6L_{\lambda\lambda pp}\,\mu^{2}\phi^{2}+4L_{\lambda ppp}\mu\phi^{3}+L_{pppp}\,\phi^{4}\right] \\
&+o\left((\mu\vee\phi)^{5}\right) \\
&=-\frac{1}{4}\mu^{4}+\frac{1}{3\hat{\sigma}^{3}}\sqrt{\frac{2}{\pi}}\sum_{i=1}^{n}\hat{\varepsilon}_{i}^{3}\mu^{3}\phi-\frac{n}{\pi}\mu^{2}\phi^{2}+o\left((\mu\vee\phi)^{5}\right)
\end{aligned}$$

The 1st order term is zero because θ^{**} is a stationary point (Result 1). The 2nd order term is zero by the definition of the eigenvector. The third order term is zero because $L\left(\theta^{*}+\Delta\theta\right)-L\left(\theta^{*}\right)$ in Result 3 is zero when $\hat{p}=1$. Since $\phi<0$ and $\mu>0$, $\frac{1}{3\hat{\sigma}^{3}}\sqrt{\frac{2}{\pi}}\sum_{i=1}^{n}\hat{\varepsilon}_{i}^{3}\mu^{3}\phi<0$ when $\sum\hat{\varepsilon}_{i}^{3}>0$. Therefore, if $\sum\hat{\varepsilon}_{i}^{3}>0$, $L\left(\theta^{*}+\Delta\theta\right)-L\left(\theta^{*}\right)<0$ and θ^{**} with $\hat{p}=1$ is a local maximizer.

□

Appendix B

PROOFS FOR CHAPTER 2

In this Section, we prove that $\hat{\Omega} \overset{p}{\to} \Omega$ under Assumption R$''$ (footnote on p10) and Theorem 2.2. We use the following notation. Given block resample $\omega_t^\bullet = (y_t^\bullet, x_t^{\bullet\prime})'$ in Section 2.3.1, we let $v_{0t}^\bullet = x_t^\bullet(y_t^\bullet - x_t^{\bullet\prime}\beta) \equiv x_t^\bullet u_{0t}^\bullet$ and $v_t^\bullet = x_t^\bullet(y_t^\bullet - x_t^{\bullet\prime}\hat{\beta}) \equiv x_t^\bullet u_t^\bullet$. Following the notation in Gonçalves and Vogelsang (2011), $p^\bullet$ denotes the probability measure induced by the bootstrap resampling, conditional on a realization of the original time series. Let $Z_T^\bullet$ be bootstrap statistics. Then, we write $Z_T^\bullet = o_{p^\bullet}(1)$ in probability or $Z_T^\bullet \overset{p^\bullet}{\to} 0$ if for any $\varepsilon > 0$, $\delta > 0$, $\lim_{T\to\infty} p[p^\bullet(|Z_T^\bullet| > \delta) > \varepsilon] = 0$. Similarly we say that $Z_T^\bullet = O_{p^\bullet}(1)$ in probability if for all $\varepsilon > 0$ there exists an $M_\varepsilon < \infty$ such that $\lim_{T\to\infty} p[p^\bullet(|Z_T^\bullet| > M_\varepsilon) > \varepsilon] = 0$. Finally, we write $Z_T^\bullet \overset{p^\bullet}{\Rightarrow} Z$ in probability if conditional on the sample, if $Z_T^\bullet$ weakly converges to Z under $p^\bullet$, for all samples contained in a set with probability converging to one. Specifically, we write $Z_T^\bullet \overset{p^\bullet}{\Rightarrow} Z$ in probability if and only if $E^\bullet[f(Z_T^\bullet)] \to E[f(Z)]$ in probability for any bounded and uniformly continuous function f.

Lemma B1. *Let $r \geq p \geq 1$. Suppose $\|w_t\|_r \leq \Delta < \infty$. Let $\{a_t\}$ be a random sequence which takes values either 0 or 1. If $\{(a_t, \varepsilon_t)\}$ is a α-mixing sequence with α_m of size $-a$ and $\{w_t\}$ is L_p-NED on $\{\varepsilon_t\}$ with ν_m of size $-b$, then $\{a_t w_t - E(a_t w_t), \mathcal{F}^t\}$ is L_p-mixingale of size $-\min\{b, a\frac{r-2}{2r}\}$ with uniformly bounded mixingale constants where $\mathcal{F}^t$ is a nondecreasing sequence of σ-fields, $\sigma\left(\boldsymbol{X_t}, \boldsymbol{X_{t-1}}, \dots\right)$, $\boldsymbol{X_t} = (a_t, \varepsilon_t)$.*

Proof. We start by defining the following notation. Let $\boldsymbol{X_t} = (a_t, \varepsilon_t)$,
$\mathcal{F}_s^t = \sigma\left(\boldsymbol{X_s}, \boldsymbol{X_{s+1}}, \dots, \boldsymbol{X_t}\right)$, $\mathcal{G}_s^t = \sigma\left(\varepsilon_s, \varepsilon_{s+1}, \dots, \varepsilon_t\right)$. First we prove that $\{a_t w_t - E(a_t w_t)\}$ is L_p-mixingale. Note that

$$
\begin{aligned}
&\left\| E\left[a_t w_t - E(a_t w_t) | \mathcal{F}_{-\infty}^{t-m}\right] \right\|_p \\
&= \left\| E\left[a_t w_t - a_t E\left[w_t | \mathcal{G}_{t-k}^{t+k}\right] + a_t E\left[w_t | \mathcal{G}_{t-k}^{t+k}\right] - E\left(a_t E\left[w_t | \mathcal{G}_{t-k}^{t+k}\right]\right)\right.\right. \\
&\left.\left. +E\left(a_t E\left[w_t | \mathcal{G}_{t-k}^{t+k}\right]\right) - E(a_t w_t) \right| \mathcal{F}_{-\infty}^{t-m}\right] \Big\|_p \\
&\leq \left\| E\left[a_t \left(w_t - E\left[w_t | \mathcal{G}_{t-k}^{t+k}\right]\right) \middle| \mathcal{F}_{-\infty}^{t-m}\right] \right\|_p \\
&+ \left\| E\left[a_t E\left[w_t | \mathcal{G}_{t-k}^{t+k}\right] \middle| \mathcal{F}_{-\infty}^{t-m}\right] - E\left(a_t E\left[w_t | \mathcal{G}_{t-k}^{t+k}\right]\right) \right\|_p \\
&+ \left\| E\left(a_t E\left[w_t | \mathcal{G}_{t-k}^{t+k}\right]\right) - E(a_t w_t) \right\|_p \because \text{Minkowski inequality} \\
&\leq \left\| a_t \left(w_t - E\left[w_t | \mathcal{G}_{t-k}^{t+k}\right]\right) \right\|_p \\
&+ \left\| E\left[a_t E\left[w_t | \mathcal{G}_{t-k}^{t+k}\right] \middle| \mathcal{F}_{-\infty}^{t-m}\right] - E\left(a_t E\left[w_t | \mathcal{G}_{t-k}^{t+k}\right]\right) \right\|_p \\
&+ \left\| a_t \left(w_t - E\left[w_t | \mathcal{G}_{t-k}^{t+k}\right]\right) \right\|_p \because \text{Conditional Jensen's inequality} \\
&\leq 2 \left\| w_t - E\left[w_t | \mathcal{G}_{t-k}^{t+k}\right] \right\|_p + \left\| E\left[a_t E\left[w_t | \mathcal{G}_{t-k}^{t+k}\right] \middle| \mathcal{F}_{-\infty}^{t-m}\right] - E\left(a_t E\left[w_t | \mathcal{G}_{t-k}^{t+k}\right]\right) \right\|_p \\
&\leq 2 d_t \nu_k + 6 \alpha_m^{\frac{1}{p} - \frac{1}{r}} \left\| a_t E_{t-k}^{t+k} w_t \right\|_r \\
&\because \{w_t\} \text{ is } L_p\text{-NED on } \{\varepsilon_t\} \text{ with } \nu_m \text{ of size } -b \\
&\text{and } \left\{a_t E\left[w_t | \mathcal{G}_{t-k}^{t+k}\right]\right\} \text{ is } \alpha\text{-mixing with } \alpha_m \text{ of size } -a \\
&\leq 2 d_t \nu_k + 6 \alpha_m^{\frac{1}{p} - \frac{1}{r}} \| w_t \|_r \\
&\leq \max\left\{d_t, \|w_t\|_r\right\} \left(2\nu_k + 6\alpha_m^{\frac{1}{p} - \frac{1}{r}}\right) \equiv c_t \psi_m
\end{aligned}
$$

Also note that

$$
\begin{aligned}
&\left\|(a_t w_t - E(a_t w_t)) - E\left[a_t w_t - E(a_t w_t)|\, \mathcal{F}_{-\infty}^{t+m}\right]\right\|_p \\
&= \left\|a_t w_t - E\left[a_t w_t|\, \mathcal{F}_{-\infty}^{t+m}\right]\right\|_p \\
&\leq 2\left\|a_t w_t - E\left[a_t w_t|\, \mathcal{F}_{t-m}^{t+m}\right]\right\|_p \qquad \because \text{Davidson (2002, Theorem 10.28)} \\
&= 2\left\|a_t w_t - a_t E\left[w_t|\, \mathcal{F}_{t-m}^{t+m}\right]\right\|_p \qquad \because a_t \text{ is } \mathcal{F}_{t-m}^{t+m} - \text{measurable} \\
&\leq 2\left\|w_t - E\left[w_t|\, \mathcal{F}_{t-m}^{t+m}\right]\right\|_p \\
&\leq 2 d_t v_m \leq c_t \psi_{m+1}
\end{aligned}
$$

Therefore $\{a_t w_t - E(a_t w_t)\}$ is L_p-mixingale with ψ_m of size $-\min\left\{b, a\frac{r-p}{pr}\right\}$ with $c_t << \max\left\{d_t, \|w_t\|_r\right\}$. Next we show that mixingale constants are uniformly bounded. According to the Minkowski and conditional modulus inequalities,

$$
\begin{aligned}
\left\|w_t - E\left[w_t|\mathcal{G}_{t-m}^{t+m}\right]\right\|_p &\leq \|w_t\|_p + \left\|E\left[w_t|\mathcal{G}_{t-k}^{t+k}\right]\right\|_p \\
&\leq \|w_t\|_p + \|w_t\|_p \\
&= 2\|w_t\|_p
\end{aligned}
$$

Since $\|w_t\|_p \leq \|w_t\|_r$ is uniformly bounded by assumption, we can set d_t equal to a finite constant for all t. Furthermore, by imposing $d_t = 2\|w_t\|_p$, we can set $v_m \leq 1$ without loss of generality. Thus, mixingale constant, $c_t << \max\left\{d_t, \|w_t\|_r\right\} \leq \max\left\{2\|w_t\|_p, \|w_t\|_r\right\}$, is uniformly bounded under the assumed moment conditions. □

Lemma B2. *Let x_t and w_t be L_p-NED on ε_t with v_m^x and v_m^w of respective sizes $-\phi_x$ and $-\phi_w$. Then $\{x_t w_t\}$ is $L_{p/2}$-NED of size $-\min\{\phi_x, \phi_w\}$.*

Proof. We follow the proof similar to that of Davidson (2002, Theorem 17.9). Define $\mathcal{F}_s^t =$

$\sigma\left(\varepsilon_s, \varepsilon_{s+1}, \ldots, \varepsilon_t\right)$. Note that

$$\left\|x_t w_t - E\left[x_t w_t \middle| \mathcal{F}_{t-m}^{t+m}\right]\right\|_{\frac{p}{2}}$$

$$= \left\|x_t w_t - x_t E\left[w_t \middle| \mathcal{F}_{t-m}^{t+m}\right] + x_t E\left[w_t \middle| \mathcal{F}_{t-m}^{t+m}\right] - E\left[x_t \middle| \mathcal{F}_{t-m}^{t+m}\right] E\left[w_t \middle| \mathcal{F}_{t-m}^{t+m}\right]\right.$$

$$\left. + E\left[x_t \middle| \mathcal{F}_{t-m}^{t+m}\right] E\left[w_t \middle| \mathcal{F}_{t-m}^{t+m}\right] - E\left[x_t w_t \middle| \mathcal{F}_{t-m}^{t+m}\right]\right\|_{\frac{p}{2}}$$

$$\leq \left\|x_t w_t - x_t E\left[w_t \middle| \mathcal{F}_{t-m}^{t+m}\right]\right\|_{\frac{p}{2}} + \left\|x_t E\left[w_t \middle| \mathcal{F}_{t-m}^{t+m}\right] - E\left[x_t \middle| \mathcal{F}_{t-m}^{t+m}\right] E\left[w_t \middle| \mathcal{F}_{t-m}^{t+m}\right]\right\|_{\frac{p}{2}}$$

$$+ \left\|E\left[x_t \middle| \mathcal{F}_{t-m}^{t+m}\right] E\left[w_t \middle| \mathcal{F}_{t-m}^{t+m}\right] - E\left[x_t w_t \middle| \mathcal{F}_{t-m}^{t+m}\right]\right\|_{\frac{p}{2}}$$

$\because$ Minkowski's Inequality

$$= \left\|x_t \left(w_t - E\left[w_t \middle| \mathcal{F}_{t-m}^{t+m}\right]\right)\right\|_{\frac{p}{2}} + \left\|\left(x_t - E\left[x_t \middle| \mathcal{F}_{t-m}^{t+m}\right]\right) E\left[w_t \middle| \mathcal{F}_{t-m}^{t+m}\right]\right\|_{\frac{p}{2}}$$

$$+ \left\|E\left[\left(x_t - E\left[x_t \middle| \mathcal{F}_{t-m}^{t+m}\right]\right)\left(w_t - E\left[w_t \middle| \mathcal{F}_{t-m}^{t+m}\right]\right) \middle| \mathcal{F}_{t-m}^{t+m}\right]\right\|_{\frac{p}{2}}$$

$$\leq \|x_t\|_p \left\|w_t - E\left[w_t \middle| \mathcal{F}_{t-m}^{t+m}\right]\right\|_p + \left\|x_t - E\left[x_t \middle| \mathcal{F}_{t-m}^{t+m}\right]\right\|_p \|w_t\|_p$$

$$+ \left\|x_t - E\left[x_t \middle| \mathcal{F}_{t-m}^{t+m}\right]\right\|_p \left\|w_t - E\left[w_t \middle| \mathcal{F}_{t-m}^{t+m}\right]\right\|_p$$

$\because$ Hölder's inequality and Conditional Jensen's inequality

$$\leq \|x_t\|_p d_t^w \nu_m^w + d_t^x \nu_m^x \|w_t\|_p + d_t^x \nu_m^x d_t^w \nu_m^w$$

$$\leq \max\left\{\|x_t\|_p d_t^w, \|w_t\|_p d_t^x, d_t^x d_t^w\right\} \left(\nu_m^w + \nu_m^x + \nu_m^x \nu_m^w\right) \equiv d_t \nu_m$$

In other words, $d_t = \max\left\{\|x_t\|_p d_t^w, \|w_t\|_p d_t^x, d_t^x d_t^w\right\}$ and $\nu_m = \nu_m^w + \nu_m^x + \nu_m^x \nu_m^w = O\left(m^{-\min\{\phi x, \phi w\}}\right)$.

□

Lemma B3. *Define $f(w) : \mathbb{T} \mapsto \mathbb{R}, \mathbb{T} \subset \mathbb{R}^k$, a function of k real variables, and $\rho(w^1, w^2) = \sum_{i=1}^k \left|w_i^1 - w_i^2\right|$ that measures the distance between points w^1 and w^2. Let $\{w_t\}$ be a k dimensional random sequence, of which each element is $L_2 - NED$ of size $-b$ on $\{\varepsilon_t\}$. Suppose that $f(w_t)$ is L_2-bounded. Further assume that $\left|f(w_t^1) - f(w_t^2)\right| \leq B_t(w_t^1, w_t^2)\rho(w_t^1, w_t^2)$*

a.s. where $\rho()$ and $B_t()$ satisfy the following conditions: $B_t(w_t^1, w_t^2) : \mathbb{T} \times \mathbb{T} \mapsto \mathbb{R}^+$for $1 \leq q \leq 2$, $\left\| \rho \left(w_t, E \left[w_t | \mathcal{G}_{t-m}^{t+m} \right] \right) \right\|_q < \infty$, $\left\| B \left(w_t, E \left[w_t | \mathcal{G}_{t-m}^{t+m} \right] \right) \right\|_{q/(q-1)} < \infty$, and for $r > 2$, $\left\| B \left(w_t, E \left[w_t | \mathcal{G}_{t-m}^{t+m} \right] \right) \rho \left(w_t, E \left[w_t | \mathcal{G}_{t-m}^{t+m} \right] \right) \right\|_r < \infty$. Then, $\{f(w_t)\}$ is $L_2 - NED$ on $\{\varepsilon_t\}$ of size $-b(r-2)/(2(r-1))$.

Proof. See Davidson (2002, Theorem 7.16).

□

Lemma B4. *For some nondecreasing sequence of σ-fields $\{\mathcal{F}^t\}$ and for some $p > 1$, let $\left\{ w_t \mathcal{F}^t \right\}$ be an L_p-mixingale with mixingale coefficients $\psi_m$and mixingale constants c_t. Then letting $S_j = \sum_{t=1}^{j} w_t$ and $\Psi = \sum_{m=1}^{\infty} \psi_m$, it follows that*

$$\left\| \max_{j \leq T} \left| S_j \right| \right\|_p \leq K\Psi \left(\sum_{t=1}^{T} c_t^\beta \right)^{\frac{1}{\beta}}, \quad \beta = \min\{p, 2\}$$

for some generic constant K.

Proof. See Hansen (1991),Hansen (1992). □

Proof of footnote on page 10: First we show that $\left\{ v_t v_{t+j}' - E\left(v_t v_{t+j}' \right) \right\}$ is $L_{(2+\delta)/2}$-mixingale of size -1 with uniformly bounded mixingale constants. Note that under the Assumption R''-4, $\{v_t^*\}$ is $L_{2+\delta}$-NED on $\{\varepsilon_t\}$ of size -1, which implies that $\{v_{t+j}^*\}$ is $L_{2+\delta}$-NED on $\{\varepsilon_t\}$ of size -1 as well. See Davidson (2002, Theorem 17.10). Then $\left\{ v_t^* v_{t+j}^{*\prime} \right\}$ is $L_{(2+\delta)/2}$-NED on $\{\varepsilon_t\}$ of size -1 by Lemma B2. Also note that under the Assumption R''-5, $\{(a_t, \varepsilon_t)\}$ is α-mixing of size $-(2+\delta)(r+\delta)/(r-2)$ and

$$\left\| v_t^* v_{t+j}^{*\prime} \right\|_{\frac{2+\delta}{2}} \leq \|v_t^*\|_{2+\delta} \left\| v_{t+j}^* \right\|_{2+\delta} \leq \Delta^2 < \infty.$$

Using Lemma B1, this implies that $\{a_t v_t^* v_{t+j}^* - E(a_t v_t^* v_{t+j}^*)\}$ is $L_{(2+\delta)/2}$-mixingale of size -1 with uniformly bounded mixingale constants. In other words, $\{v_t v_{t+j} - E(v_t v_{t+j})\}$ is $L_{(2+\delta)/2}$-mixingale of size -1. Secondly, we show that $\tilde{\Omega} \xrightarrow{p} E\tilde{\Omega}$. Using Lemma B4, we can write

$$\left\| \frac{1}{T} \sum_{t=1}^{T-j} \left(v_t v'_{t+j} - E(v_t v'_{t+j}) \right) \right\|_{(2+\delta)/2} \leq \frac{1}{T} K \Psi \left(\sum_{t=1}^{T-j} c_t^{\min\{\frac{2+\delta}{2}, 2\}} \right)^{\max\{\frac{2}{2+\delta}, \frac{1}{2}\}}$$
$$\leq K' T^{\left(-1+\max\{\frac{2}{2+\delta}, \frac{1}{2}\}\right)}$$

uniformly in T for some finite constant K'. The last inequality follows by the fact that c_t is uniformly bounded constants and $\Psi < \infty$ which is due to ψ_m being of size -1. Hence,

$$\frac{1}{M} T^{\left(1-\max\left\{\frac{2}{2+\delta}, \frac{1}{2}\right\}\right)} \|\tilde{\Omega} - E\tilde{\Omega}\|_{\frac{2+\delta}{2}}$$
$$= \frac{1}{M} T^{\left(1-\max\left\{\frac{2}{2+\delta}, \frac{1}{2}\right\}\right)} \left\| \sum_{j=-T}^{T} k\left(\frac{j}{M}\right) \frac{1}{T} \sum_{t=1}^{T-j} \left(v_t v'_{t+j} - E\left(v_t v'_{t+j}\right) \right) \right\|_{\frac{2+\delta}{2}}$$
$$\leq \frac{1}{M} T^{\left(1-\max\left\{\frac{2}{2+\delta}, \frac{1}{2}\right\}\right)} \sum_{j=-T}^{T} \left| k\left(\frac{j}{M}\right) \right| \left\| \frac{1}{T} \sum_{t=1}^{T-j} \left(v_t v'_{t+j} - E\left(v_t v'_{t+j}\right) \right) \right\|_{\frac{2+\delta}{2}}$$
$$\because \text{Minkowski's inequality}$$
$$\leq K' \int_R |k(x)|\, dx < \infty,$$

uniformly in T, where

$$M T^{\left(\max\left\{\frac{2}{2+\delta}, \frac{1}{2}\right\}\right)-1} = M T^{-\frac{1}{2q+1}} T^{\frac{1}{2q+1} + \max\{-\frac{\delta}{2+\delta}, -\frac{1}{2}\}}$$
$$= O(1) T^{\max\left\{\frac{2(1-q\delta)}{(2q+1)(2+\delta)}, \frac{1-2q}{(2q+1)2}\right\}} = O(1)o(1) = o(1)$$

since $\max\{1/2, 1/\delta\} < q$. Therefore $\|\tilde{\Omega} - E\tilde{\Omega}\|_{\frac{2+\delta}{2}} \to 0$, and thus $\tilde{\Omega} \xrightarrow{p} E\tilde{\Omega}$ by Markov's inequality. Thirdly, we prove that $E\tilde{\Omega} \xrightarrow{p} \Omega$ so that combining with above result, $\tilde{\Omega} \xrightarrow{p} \Omega$.

By definition we can write

$$
\begin{aligned}
&\Omega - E\tilde{\Omega} \\
&= (1-k(0))\frac{1}{T}\sum_{t=1}^{T} E\left(v_t v_t'\right) + \frac{1}{T}\sum_{j=1}^{M}\left(1-k\left(\frac{j}{M}\right)\right)\sum_{t=1}^{T-j}\left(E\left(v_t v_{t+j}'\right) + E\left(v_{t+j}v_t'\right)\right) \\
&+ \frac{1}{T}\sum_{j=M+1}^{T-1}\sum_{t=1}^{T-j}\left(E\left(v_t v_{t+j}'\right) + E\left(v_{t+j}v_t'\right)\right)
\end{aligned}
$$

Note that

$$
\left\|v_t v_t'\right\|_{\frac{2+\delta}{2}} \leq \left\|v_t^* v_t^{*\prime}\right\|_{\frac{2+\delta}{2}} \leq \|v_t^*\|_{2+\delta}\,\|v_t^*\|_{2+\delta} \leq \Delta^2 < \infty,
$$

which implies that ${}^{1}/{}_{T}\sum_{t=1}^{T} E\left(v_t v_t'\right) = O_P(1)$. Therefore, $k(0) \to 1$ implies that the first term vanishes as $T \to \infty$. Showing the second term being $o_P(1)$ is the same as showing that the equation below is $o_P(1)$.

$$
\left|\sum_{j=1}^{M}\left[1-k\left(\frac{j}{M}\right)\right]\frac{1}{T}\sum_{t=1}^{T-j}E\left(v_t v_{t+l}'\right)\right| \leq \sum_{j=1}^{M}\left|1-k\left(\frac{j}{M}\right)\right|\frac{1}{T}\sum_{t=1}^{T-j}\left|E\left(v_t v_{t+l}'\right)\right|
$$

Using Lemma B1, $\left\{v_t, \mathcal{F}^t\right\}$ is $L_{2+\delta}$-mixingale of size -1 with uniformly bounded mixingale constants, where $\mathcal{F}_s^t = \sigma\left(\boldsymbol{X_t}, \boldsymbol{X_{t-1}}, \ldots, \boldsymbol{X_s}\right)$, $\boldsymbol{X_t} = (a_t, \varepsilon_t)$. Then, we can write

$$
\begin{aligned}
\left|E\left(v_t v_{t+l}'\right)\right| &= \left|E\left(E\left[v_t v_{t+j}' \middle| \mathcal{F}_{-\infty}^{t+j-[j/2]}\right]\right)\right| \\
&= \left|E\left(v_t E\left[v_{t+j}' \middle| \mathcal{F}_{-\infty}^{t+j-[j/2]}\right]\right)\right| \\
&\leq \|v_t\|_2 \left\|E\left[v_{t+j}' \middle| \mathcal{F}_{-\infty}^{t+j-[j/2]}\right]\right\|_2 \\
&\leq \Delta d_t \nu_{[j/2]} \\
&\leq K\nu_{[j/2]}. \qquad \text{(eqB.1)}
\end{aligned}
$$

Hence,

$$\left| \sum_{j=1}^{M} \left[1 - k\left(\frac{j}{M}\right) \right] \frac{1}{T} \sum_{t=1}^{T-j} E\left(v_t v_{t+l}'\right) \right| \leq \sum_{j=1}^{M} \left| 1 - k\left(\frac{j}{M}\right) \right| \frac{1}{T} \sum_{t=1}^{T-j} \left| E\left(v_t v_{t+l}'\right) \right|$$

$$\leq \sum_{j=1}^{M} \left| 1 - k\left(\frac{j}{M}\right) \right| \frac{1}{T} \sum_{t=1}^{T-j} K\nu_{[j/2]}$$

$$= \frac{T-j}{T} K \sum_{j=1}^{M} \left| 1 - k\left(\frac{j}{M}\right) \right| \nu_{[j/2]} \qquad \text{(eqB.2)}$$

If we show that (eqB.2) converges to zero then we are done with the second term. We use the same approach as done in the proof of Gallant and White (1995, Lemma 6.6). First define μ to be a counting measure on the positive integers. Then, we can write

$$\sum_{i=1}^{M} \left| 1 - k\left(\frac{j}{M}\right) \right| \nu_{[j/2]} = \int_0^{\infty} \mathbb{1}\{j \leq M\} \left| 1 - k\left(\frac{j}{M}\right) \right| \nu_{[j/2]}. \qquad \text{(eqB.3)}$$

Note that for each $j \in \mathbb{N}$, $\lim_{T\to\infty} k\left(j/M\right) \to 1$ implies

$$\lim_{T\to\infty} \mathbb{1}\{j \leq M\} \left| 1 - k\left(\frac{j}{M}\right) \right| \nu_{[j/2]} d\mu(j) \to 0. \qquad \text{(eqB.4)}$$

Also note that since $|1 - k(j/M)|$ is bounded,

$$\mathbb{1}\{j \leq M\} \left| 1 - k\left(\frac{j}{M}\right) \right| \nu_{[j/2]} \leq K\nu_{[j/2]}$$

for some finite constant K. $K\nu_{[j/2]}$ is integrable because ν_m is of size -1. Therefore by the dominated convergence theorem, (eqB.4) implies that (eqB.3) converges to zero as well. This in turn implies that (eqB.2) converges to zero as $T \to \infty$. Hence the second term vanishes as $T \to \infty$. Now consider the third term. It is sufficient to show that

$$\left| \frac{1}{T} \sum_{j=M+1}^{T-1} \sum_{t=1}^{T-j} E\left(v_t v_{t+j}'\right) \right| \to 0 \text{ as } T \to \infty.$$

Using (eqB.1),

$$\left|\frac{1}{T}\sum_{j=M+1}^{T-1}\sum_{t=1}^{T-j}E\left(v_t v_{t+j}'\right)\right| \leq \frac{1}{T}\sum_{j=M+1}^{T-1}\sum_{t=1}^{T-j}\left|E\left(v_t v_{t+j}'\right)\right|$$

$$\leq \frac{1}{T}\sum_{j=M+1}^{T-1}\sum_{t=1}^{T-j}K\nu_{[j/2]}$$

$$= \frac{1}{T}\sum_{j=1}^{T-1}\sum_{t=1}^{T-j}K\nu_{[j/2]} - \frac{1}{T}\sum_{j=1}^{M}\sum_{t=1}^{T-j}K\nu_{[j/2]}.$$

The two terms above converge to the same limit as $T \to \infty$ by the similar argument as above. Hence, the third term converges to zero as well. Therefore we've shown that $E\tilde{\Omega} \xrightarrow{p} \Omega$. Combining with above result, $\tilde{\Omega} \xrightarrow{p} \Omega$. Lastly, note that given assumptions are sufficient for Andrews (1991, Assumption B). Hence, $\sqrt{T}/M\left(\hat{\Omega} - \tilde{\Omega}\right) = O_p(1)$. See Andrews (1991, Proof of Theorem 1(1)). Therefore $\hat{\Omega} - \tilde{\Omega} = o_p(1)$ because $M^{1+2q}/T = O(1), q \in (\max\{1/2, 1/\delta\}, \infty)$. Therefore, $\hat{\Omega} \xrightarrow{p} \Omega$. □

Lemma B5. *Suppose that $\{w_t - E(w_t)\}$ is a weakly stationary L_2-mixingale with $\|w_t\|_p \leq \Delta < \infty$ for some $p > 2$ such that its mixingale coefficients ψ_m satisfy $\sum_1^\infty \psi_m < \infty$ and its mixingale constants are uniformly bounded. Let $\{w_t^\bullet : t = 1, \ldots, T\}$ denote an MBB resample of $\{w_t : t = 1, \ldots, T\}$ with block size l satisfying either of the two following conditions: (a) l is fixed as $T \to \infty$, or (b) $l \to \infty$ as $T \to \infty$ with $l = o(T)$. Then, for any $\eta > 0$, as $T \to \infty$,*

$$p^\bullet\left(\sup_{r\in[0,1]}\left|T^{-1}\sum_{t=1}^{[rT]}\left(w_t^\bullet - E^\bullet\left(w_t^\bullet\right)\right)\right| > \eta\right) = o_P(1).$$

Proof. See Gonçalves and Vogelsang (2011, Proof of Lemma A.4). □

Lemma B6. *Under Assumption R',*

(a) For any fixed l such that $1 \leq l < T$, $T \to \infty$,

$$p\lim_{T\to\infty}\Omega_T^\bullet = \Gamma_0 + \sum_{j=1}^{l}\left(1 - \frac{j}{l}\right)\left(\Gamma_j + \Gamma_j'\right) \equiv \Omega_l,$$

where $\Gamma_j = E\left(v_t v_{t-j}'\right)$.

(b) *Let $l = l_T \to \infty$ as $T \to \infty$ such that $l^2/T \to 0$. Then*

$$p \lim_{T\to\infty} \Omega_T^{\bullet} = \Gamma_0 + \sum_{j=1}^{\infty} \left(\Gamma_j + \Gamma_j'\right) \equiv \Omega,$$

Proof. See Gonçalves and Vogelsang (2011, Proof of Lemma A.2). □

Lemma B7. *Suppose Assumption R'' holds and let Ω_l and Ω as defined in Lemma B6 be positive definite matrices. It follows that*

(a) *For any fixed l such that $1 \leq l < T$, $T \to \infty$,*

$$Z_T^{\bullet}(r) \Rightarrow^{p^{\bullet}} \Lambda_l W_k(r),$$

in probability where Λ_l is the square root matrix of Ω_l.

(b) *Let $l = l_T \to \infty$ as $T \to \infty$ such that $l^2/T \to 0$. Then*

$$Z_T^{\bullet}(r) \Rightarrow^{p^{\bullet}} \Lambda_l W_k(r),$$

in probability where Λ is the square root matrix of Ω.

Proof. See the proof of Gonçalves and Vogelsang (2011, Lemma A.3). The sufficient condition for the proof is that $\{v_t\}$ is $L_{2+\delta} - mixingale$ with size -1 with uniformly bounded mixing coefficients, which is implied by Assumption R'' and Lemma B1. □

***Proof of Theorem 2.2**:* Define the vector $\omega_t = (y_t, x_t')'$ that collects dependent and explanatory variables. Let $l \in \mathbb{N}(1 \leq l \leq T)$ be a 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. Draw $k_0 = T/l$ blocks randomly with replacement from the set of overlapping blocks $\{B_{1,l}, \ldots, B_{T-l+1,l}\}$ to obtain a bootstrap resample denoted as $\omega_t^{\bullet} = (y_t^{\bullet}, x_t^{\bullet\prime})'$, $t = 1, \ldots, T$. We want to show

1. $T^{-1} \sum_{t=1}^{[rT]} x_t^{\bullet} x_t^{\bullet\prime} \Rightarrow^{p^{\bullet}} rQ^{\bullet}$ for some $Q^{\bullet}$
2. $T^{-1/2} \sum_{t=1}^{[rT]} v_t^{\bullet} \Rightarrow^{p^{\bullet}} \Lambda^{\bullet} \mathcal{W}_k(r)$ for some $\Lambda^{\bullet}$

is true under Assumption R$'$ with Assumption R$'$ 3-5 strengthened to Assumption R$''$ 3-5. Let $p^{\bullet}$ denote the probability measure induced by the bootstrap resampling conditional on a realization of the original time series. Assumption R$'$ 1-2 and Lemma B3 implies that $\{x_t^* x_t^{*\prime}\}$ is $L_2 - NED$ of size -1. Then from Lemma B1, $\{x_t x_t' - Q\}$ is L_2-mixingale of size -1 with uniformly bounded mixingale constants. Also Assumption R$'$ 1 implies that $\left\| x_t x_t' \right\|_r \leq \Delta, r > 2$. Therefore Lemma B5 applies and the first condition follows straightforwardly. Now we prove the second condition. Given our definitions $v_{0t}^{\bullet}$ and $v_t^{\bullet}$, we can write

$$v_t^{\bullet} = v_{0t}^{\bullet} - x_t^{\bullet} x_t^{\bullet\prime} \left(\hat{\beta} - \beta\right),$$

which implies that

$$\begin{aligned}
& T^{-1/2} \sum_{t=1}^{[rT]} v_t^{\bullet} \\
&= T^{-1/2} \sum_{t=1}^{[rT]} \left(v_{0t}^{\bullet} - E^{\bullet}\left(v_{0t}^{\bullet}\right)\right) + T^{-1/2} \sum_{t=1}^{[rT]} E^{\bullet}\left(v_{0t}^{\bullet}\right) - T^{-1/2} \sum_{t=1}^{[rT]} x_t^{\bullet} x_t^{\bullet\prime} \left(\hat{\beta} - \beta\right) \\
&\equiv Z_T^{\bullet}(r) + A_{1T}^{\bullet}(r) - A_{2T}^{\bullet}(r).
\end{aligned}$$

We show the second condition in the following two steps.

Step 1. We show that $Z_T^{\bullet}(1) \Rightarrow^{\bullet} \Lambda^{\bullet} W_k(r)$.

Proof of Step 1. Straightforward from Lemmas B6-B7 and Assumption R''.

Step 2. We show that $\sup_{r\in[0,1]} \left| A_{1T}^{\bullet}(r) - A_{2T}^{\bullet}(r) \right| = o_{p^{\bullet}}(1)$ in probability.

Proof of Step 2. Note that

$$
\begin{aligned}
&A^{\bullet}_{1T}(r) - A^{\bullet}_{2T}(r) \\
&= T^{-1/2}\sum_{t=1}^{[rT]} E^{\bullet}\left(x^{\bullet}_t\left(y^{\bullet}_t - x^{\bullet\prime}_t\hat{\beta} + x^{\bullet\prime}_t\hat{\beta} - x^{\bullet\prime}_t\beta\right)\right) - T^{-1/2}\sum_{t=1}^{[rT]} x^{\bullet}_t x^{\bullet\prime}_t\left(\hat{\beta}-\beta\right) \\
&= T^{-1/2}\sum_{t=1}^{[rT]} E^{\bullet}\left(x^{\bullet}_t\left(y^{\bullet}_t - x^{\bullet\prime}_t\hat{\beta}\right)\right) + T^{-1/2}\sum_{t=1}^{[rT]} E^{\bullet}\left(x^{\bullet}_t x^{\bullet\prime}_t\hat{\beta} - x^{\bullet}_t x^{\bullet\prime}_t\beta\right) \\
&- T^{-1/2}\sum_{t=1}^{[rT]} x^{\bullet}_t x^{\bullet\prime}_t\left(\hat{\beta}-\beta\right) \\
&= T^{-1/2}\sum_{t=1}^{[rT]} E^{\bullet}\left(v^{\bullet}_t\right) - T^{-1/2}\sum_{t=1}^{[rT]}\left(x^{\bullet}_t x^{\bullet\prime}_t - E^{\bullet}\left(x^{\bullet}_t x^{\bullet\prime}_t\right)\right)\left(\hat{\beta}-\beta\right) \\
&\equiv B^{\bullet}_{1T}(r) - B^{\bullet}_{2T}(r).
\end{aligned}
$$

It is sufficient to show that $\sup_{r\in[0,1]}\left|B^{\bullet}_{1T}(r)\right| = o_{p^{\bullet}}(1)$ and $\sup_{r\in[0,1]}\left|B^{\bullet}_{2T}(r)\right| = o_{p^{\bullet}}(1)$, in probability.

Step 2-1. We prove that $\sup_{r\in[0,1]}\left|B^{\bullet}_{1T}(r)\right| = o_{p^{\bullet}}(1)$.

$$
\begin{aligned}
B^{\bullet}_{1T}(r) &= T^{-1/2}\sum_{t=1}^{[rT]} E^{\bullet}\left(v^{\bullet}_t\right) \\
&= T^{-1/2}\sum_{m=1}^{M_r}\sum_{s=1}^{\bullet} E^{\bullet}\left(\hat{v}_{I_m+s}\right) \\
&= T^{-1/2}\sum_{m=1}^{M_r}\sum_{s=1}^{l} E^{\bullet}\left(\hat{v}_{I_m+s}\right) - T^{-1/2}\sum_{s=B+1}^{l} E^{\bullet}\left(\hat{v}_{I_{M_r}+s}\right) \\
&\equiv b^{\bullet}_{1T} - b^{\bullet}_{2T},
\end{aligned}
$$

where $M_r = [([rT]-1)/l] + 1$ and $B = \min\{l, [rT]-(m-1)l\}$. Note that $M_r \in \{1,\ldots,k_0\}$, $B \in \{1,\ldots,l\}$, and $I_1,\ldots,I_{k_0}$ are *i.i.d.* uniformly distributed on $\{0,1,\ldots,T-l\}$ (See

Paparoditis and Politis (2003)).

$$
\begin{aligned}
&\sup_{r\in[0,1]} \left| b^{\bullet}_{1T} \right| \\
&= \sup_{r\in[0,1]} \left| T^{-1/2} \sum_{m=1}^{M_r} \sum_{s=1}^{l} E^{\bullet}\left(\hat{v}_{I_m+s}\right) \right| \\
&= \sup_{r\in[0,1]} \left| T^{-1/2} \sum_{m=1}^{M_r} \sum_{s=1}^{l} \frac{1}{T-l+1} \sum_{j=0}^{T-l} \hat{v}_{j+s} \right| \\
&= \sup_{r\in[0,1]} \left| T^{-1/2} M_r \sum_{s=1}^{l} \frac{1}{T-l+1} \sum_{j=0}^{T-l} \hat{v}_{j+s} \right| \\
&\le T^{-1/2} k_0 \left| \frac{1}{T-l+1} \sum_{s=1}^{l} \sum_{j=0}^{T-l} \hat{v}_{j+s} \right| \\
&\le T^{-1/2} k_0 \left| \frac{1}{T-l+1} \left(l \sum_{t=1}^{T} \hat{v}_t - \Big[(l-1)\hat{v}_1 + (l-2)\hat{v}_2 + \cdots + \hat{v}_{l-1} \right.\right. \\
&\left.\left. + (l-1)\hat{v}_T + (l-2)\hat{v}_{T-1} + \cdots + \hat{v}_{T-l+2} \Big] \right) \right| \\
&= T^{-1/2} k_0 \left| \frac{1}{T-l+1} \Big((l-1)\hat{v}_1 + (l-2)\hat{v}_2 + \cdots + \hat{v}_{l-1} + (l-1)\hat{v}_T \right. \\
&\left. + (l-2)\hat{v}_{T-1} + \cdots + \hat{v}_{T-l+2} \Big) \right| \quad \because \text{OLS FOC} \\
&= T^{-1/2} k_0 O_p\left(\frac{l^2}{T}\right) \quad \because \hat{v}_t \text{ is uniformly bounded in probability (See below)} \\
&= O_p\left(\frac{l}{\sqrt{T}}\right) \\
&= o_p(1) \quad \because l \text{ is fixed or } \frac{l^2}{T} \to 0.
\end{aligned}
$$

We show that $\hat{v}_t$ is uniformly bounded in probability. Given our definitions,

$$\hat{v}_t = x_t(y_t - x_t'\hat{\beta}) = v_t - x_t x_t'(\hat{\beta} - \beta).$$

First note that v_t and $x_t x_t'$ are uniformly $L_{\frac{q}{2}}$−bounded, which implies that both are uniformly bounded in probability.

$$\|v_t\|_q = \|a_t v_t^*\|_q \leq \|v_t^*\|_q \leq \Delta < \infty$$
$$\left\|x_t x_t'\right\|_{\frac{q}{2}} = \left\|a_t x_t^* x_t^{*\prime}\right\|_{\frac{q}{2}} \leq \left\|x_t^* x_t^{*\prime}\right\|_{\frac{q}{2}} \leq \|x_t^*\|_q \left\|x_t^{*\prime}\right\|_q \leq \Delta^2 < \infty$$

Also, we know that $\left|\hat{\beta} - \beta\right| = o_P(1)$. Hence $\hat{\beta} - \beta$ is uniformly bounded in probability. Therefore $\hat{v}_t$ is uniformly bounded in probability.

Finally, note that

$$
\begin{aligned}
&\sup_{r\in[0,1]} \left| b^{\bullet}_{2T} \right| \\
&= \sup_{r\in[0,1]} \left| T^{-1/2} \sum_{s=B+1}^{l} E^{\bullet}\left(\hat{v}_{I_{M_r}+s} \right) \right| \\
&= \sup_{r\in[0,1]} \left| T^{-1/2} \sum_{s=B+1}^{l} E^{\bullet}\left(v_{I_{M_r}+s} - x_{I_{M_r}+s} x_{I_{M_r}+s'} (\hat{\beta}-\beta) \right) \right| \\
&= \sup_{r\in[0,1]} \left| T^{-1/2} \sum_{s=B+1}^{l} E^{\bullet}\left(v_{I_{M_r}+s} \right) - T^{-1/2} \sum_{s=B+1}^{l} E^{\bullet}\left(x_{I_{M_r}+s} x_{I_{M_r}+s'} (\hat{\beta}-\beta) \right) \right| \\
&\leq \sup_{r\in[0,1]} \left| T^{-1/2} \sum_{s=B+1}^{l} E^{\bullet}\left(v_{I_{M_r}+s} \right) \right| \\
&+ \sup_{r\in[0,1]} \left| T^{-1/2} \sum_{s=B+1}^{l} E^{\bullet}\left(x_{I_{M_r}+s} x_{I_{M_r}+s'} (\hat{\beta}-\beta) \right) \right| \\
&= \sup_{r\in[0,1]} \left| T^{-1/2} \sum_{s=B+1}^{l} E^{\bullet}\left(v_{I_{M_r}+s} \right) \right| \\
&+ \sup_{r\in[0,1]} \left| T^{-1} \sum_{s=B+1}^{l} E^{\bullet}\left(x_{I_{M_r}+s} x_{I_{M_r}+s'} \right) \right| \left| \sqrt{T}(\hat{\beta}-\beta) \right| \\
&= \sup_{r\in[0,1]} \left| T^{-1/2} \sum_{s=B+1}^{l} \frac{1}{T-l+1} \sum_{j=0}^{T-l} v_{j+s} \right| \\
&+ \sup_{r\in[0,1]} \left| T^{-1} \sum_{s=B+1}^{l} \frac{1}{T-l+1} \sum_{j=0}^{T-l} \left(x_{j+s} x_{j+s'} \right) \right| \left| \sqrt{T}(\hat{\beta}-\beta) \right| \\
&\leq \frac{T^{-1/2}}{T-l+1} \sum_{j=0}^{T-l} \sup_{r\in[0,1]} \left| \sum_{s=B+1}^{l} v_{j+s} \right| \\
&+ \frac{T^{-1/2}}{(T-l+1)} \sum_{j=0}^{T-l+1} \sup_{r\in[0,1]} \left| \sum_{s=B+1}^{l} x_{j+s} x'_{j+s} \right| \left| (\hat{\beta}-\beta) \right|
\end{aligned}
$$

In terms of the first term,

$$
\begin{aligned}
& E\left(k_0^{\frac{1}{2}}\frac{T^{-1/2}}{T-l+1}\sum_{j=0}^{T-l}\sup_{r\in[0,1]}\left|\sum_{s=B+1}^{l}v_{j+s}\right|\right) \\
& \leq E\left(\frac{k_0^{\frac{1}{2}}T^{-1/2}}{T-l+1}\sum_{j=0}^{T-l}\max_{1\leq i\leq l}\left|\sum_{s=j+i}^{j+l}v_s\right|\right) \\
& \leq \frac{k_0^{\frac{1}{2}}T^{-1/2}}{T-l+1}\sum_{j=0}^{T-l}\left\|\max_{1\leq i\leq l}\left|\sum_{s=j+i}^{j+l}v_s\right|\right\|_{2+\delta} \\
& \leq \frac{k_0^{\frac{1}{2}}T^{-1/2}}{T-l+1}\sum_{j=0}^{T-l}K'l^{\frac{1}{2}} \\
& = O(1) \quad \because \frac{k_0 l}{T} \to 1
\end{aligned}
$$

The first inequality is obvious because for $r \in [0,1]$, $B \in \{1,\ldots,l\}$. Using Lemma B4 and the fact that $\{v_t\}$ is $L_{2+\delta}$-mixingale of size -1 with uniformly bounded mixingale constants due to Lemma B1, the third inequality is also straightforward. Therefore by the Markov inequality,

$$
\frac{T^{-1/2}}{T-l+1}\sum_{j=0}^{T-l}\sup_{r\in[0,1]}\left|T^{-1/2}\sum_{s=B+1}^{l}v_{j+s}\right| = O_p(k_0^{-\frac{1}{2}}) = o_p(1).
$$

Now we consider the second term.

$$
\begin{aligned}
& \frac{T^{-1/2}}{(T-l+1)}\sum_{j=0}^{T-l}\sup_{r\in[0,1]}\left|\sum_{s=B+1}^{l}x_{j+s}x'_{j+s}\right|\left|\hat{\beta}-\beta\right| \\
& = \frac{T^{-1/2}}{(T-l+1)}\sum_{j=0}^{T-l}\sup_{r\in[0,1]}\left|\sum_{s=B+1}^{l}\left\{\left(x_{j+s}x'_{j+s}-Q\right)+Q\right\}\right|\left|\hat{\beta}-\beta\right| \\
& \leq \frac{T^{-1/2}}{(T-l+1)}\sum_{j=0}^{T-l}\sup_{r\in[0,1]}\left|\sum_{s=B+1}^{l}\left(x_{j+s}x'_{j+s}-Q\right)\right| + T^{-1/2}Q\left|\hat{\beta}-\beta\right|,
\end{aligned}
$$

where the second term is $Op\left(T^{-1}\right) = o_P(1)$ because $\hat{\beta} - \beta = O_P(T^{-1/2})$, and the first term is $O_P(k_0^{-1/2}) = o_P(1)$ because of the following.

$$\begin{aligned}
&E\left(k_0^{\frac{1}{2}}\frac{T^{-1/2}}{(T-l+1)}\sum_{j=0}^{T-l}\sup_{r\in[0,1]}\left|\sum_{s=B+1}^{l}\left(x_{j+s}x'_{j+s}-Q\right)\right|\right)\\
&\leq E\left(\frac{k_0^{\frac{1}{2}}T^{-1/2}}{(T-l+1)}\sum_{j=0}^{T-l}\max_{1\leq i\leq l}\left|\sum_{s=j+i}^{j+l}\left(x_{j+s}x'_{j+s}-Q\right)\right|\right)\\
&\leq \frac{k_0^{\frac{1}{2}}T^{-1/2}}{(T-l+1)}\sum_{j=0}^{T-l}\left\|\max_{1\leq i\leq l}\left|\sum_{s=j+i}^{j+l}\left(x_{j+s}x'_{j+s}-Q\right)\right|\right\|_2\\
&\leq \frac{k_0^{\frac{1}{2}}T^{-1/2}}{(T-l+1)}\sum_{j=0}^{T-l}K'l^{\frac{1}{2}}\\
&= O(1) \quad \because \frac{k_0 l}{T}\to 1.
\end{aligned}$$

The first inequality is obvious because for $r \in [0,1]$, $B \in \{1,\ldots,l\}$. Note that $\{x_t x'_t\}$ is $L_2 - NED$ with given assumptions due to Lemma B3 and combining Lemma B1, $\{x_t x'_t - Q\}$ is L_2-mixingale of size -1 with uniformly bounded mixingale constants. Then using Lemma B4, the third inequality is straightforward. Therefore by the Markov inequality,

$$\frac{T^{-1/2}}{(T-l+1)}\sum_{j=0}^{T-l}\sup_{r\in[0,1]}\left|\sum_{s=B+1}^{l}\left(x_{j+s}x'_{j+s}-Q\right)\right| = O_P\left(k_0^{-\frac{1}{2}}\right) = o_P(1).$$

Hence we have $\sup_{r\in[0,1]}\left|b^{\bullet}_{1T}\right| = o_{p^{\bullet}}(1)$ and $\sup_{r\in[0,1]}\left|b^{\bullet}_{2T}\right| = o_{p^{\bullet}}(1)$, which implies that $\sup_{r\in[0,1]}\left|B^{\bullet}_{1T}\right| = o_{p^{\bullet}}(1)$ in probability.

Step 2-2. We prove that $\sup_{r\in[0,1]}\left|B^{\bullet}_{2T}(r)\right| = o_{p^{\bullet}}(1)$. We can write

$$\begin{aligned}
\sup_{r\in[0,1]}\left|B^{\bullet}_{2T}(r)\right| &= \sup_{r\in[0,1]}\left|T^{-1/2}\sum_{t=1}^{[rT]}\left(x^{\bullet}_t x^{\bullet\prime}_t - E^{\bullet}\left(x^{\bullet}_t x^{\bullet\prime}_t\right)\right)\right|\left|\hat{\beta}-\beta\right|\\
&= \sup_{r\in[0,1]}\left|T^{-1}\sum_{t=1}^{[rT]}\left(x^{\bullet}_t x^{\bullet\prime}_t - E^{\bullet}\left(x^{\bullet}_t x^{\bullet\prime}_t\right)\right)\right|\left|\sqrt{T}(\hat{\beta}-\beta)\right|\\
&= o_{p^{\bullet}}(1)
\end{aligned}$$

We know that $|\left|\sqrt{T}(\hat{\beta}-\beta)\right| = O_p(1)$. From Lemma B5,

$$\sup_{r\in[0,1]}\left|T^{-1}\sum_{t=1}^{[rT]}\left(x_t^{\bullet}x_t^{\bullet\prime} - E^{\bullet}\left(x_t^{\bullet}x_t^{\bullet\prime}\right)\right)\right| = o_{p^{\bullet}}(1).$$

Hence we have the third equality. $\square$

Appendix C

PROOFS FOR AMPLITUDE MODULATED STATISTIC, NON-RANDOM MISSING DATA

We define λ to be the total fraction of observed data points as

$$\lambda_{2C+1} - \lambda_{2C} + \lambda_{2C-1} - \cdots + \lambda_1 = \sum_{j=1}^{2C+1} (-1)^{j+1} \lambda_j \equiv \lambda.$$

Also define

$$\overline{\mathcal{W}}_k = \sum_{j=1}^{2C+1} (-1)^{j+1} \mathcal{W}_k \left(\lambda_j\right)$$

We first prove a lemma that shows that Assumptions NR' imply that Assumptions NR hold.

Lemma C1. *Assumption NR′ is sufficient for Assumption NR.*

Proof: Under Assumption NR′ 6 the locations of missing observations are fixed. Hence,

$$\lim_{T\to\infty} \frac{T_n}{T} = \lambda_n, \text{ for } n = 0, \ldots, 2C+1,$$

where it holds trivially that $\lambda_0 = 0$ and $\lambda_{2C+1} = 1$. Assumptions NR′ 1,2, and 5, imply that for all $r \in (0, 1]$,

$$T^{-1} \sum_{t=1}^{[rT]} x_t^* x_t^{*\prime} \Rightarrow rQ^*.$$

Assumptions NR′ 3, 4, and 5, imply that for all $r \in (0, 1]$,

$$T^{-1/2} \sum_{t=1}^{[rT]} v_t^* \Rightarrow \Lambda^* \mathcal{W}_k(r).$$

Also note that Assumption NR′ 7 implies that there exists Λ^* such that $\Lambda^* \Lambda' = \Omega^*$.

The next two lemmas establish the limits of scaled sums of the amplitude modified processes.

Lemma C2. *Under Assumption NR$'$ 1,2, 5, and 6, for all $r \in (0,1]$,*

$$T^{-1}\sum_{t=1}^{[rT]} x_t x_t' \Rightarrow \sum_{n=0}^{C} \mathbb{1}\left\{\lambda_{2n} < r \leq \lambda_{2(n+1)}\right\} \sum_{j=1}^{2n+1} (-1)^{j+1}\left(r \wedge \lambda_j\right) Q^*,$$

where $\lambda_0 = 0, \lambda_{2C+1} = 1$, and C is the total number of missing clusters.

Proof: Assumptions NR$'$ 1, 2, 5, and 6 are sufficient for Assumptions NR 1 and 2 by Lemma C1. Hence we can write for $r \in (\lambda_{2n}, \lambda_{2n+1}]$,

$$\begin{aligned}
&T^{-1}\sum_{t=1}^{[rT]} x_t x_t' \\
&= T^{-1}\sum_{t=1}^{[rT]} x_t^* x_t^{*\prime} - T^{-1}\sum_{t=1}^{[\lambda_{2n}T]} x_t^* x_t^{*\prime} + T^{-1}\sum_{t=1}^{\left[\lambda_{2n-1}T\right]} x_t^* x_t^{*\prime} - \ldots + T^{-1}\sum_{t=1}^{[\lambda_1 T]} x_t^* x_t^{*\prime} \\
&\Rightarrow \left(r - \lambda_{2n} + \lambda_{2n-1} - \ldots + \lambda_1\right) Q^*,
\end{aligned}$$

whereas for $r \in (\lambda_{2n+1}, \lambda_{2n+2}]$ we have

$$\begin{aligned}
&T^{-1}\sum_{t=1}^{[rT]} x_t x_t' \\
&= T^{-1}\sum_{t=1}^{\left[\lambda_{2n+1}T\right]} x_t^* x_t^{*\prime} - T^{-1}\sum_{t=1}^{[\lambda_{2n}T]} x_t^* x_t^{*\prime} + T^{-1}\sum_{t=1}^{\left[\lambda_{2n-1}T\right]} x_t^* x_t^{*\prime} - \ldots \\
&+ T^{-1}\sum_{t=1}^{[\lambda_1 T]} x_t^* x_t^{*\prime} \\
&\Rightarrow \left(\lambda_{2n+1} - \lambda_{2n} + \lambda_{2n-1} - \ldots + \lambda_1\right) Q^*.
\end{aligned}$$

Combining these two results we have for $r \in \left(\lambda_{2n}, \lambda_{2n+2}\right]$,

$$T^{-1}\sum_{t=1}^{[rT]} x_t x_t' \Rightarrow \sum_{j=1}^{2n+1} (-1)^{j+1}\left(r \wedge \lambda_j\right) Q^*.$$

It immediately follows for $r \in [0,1]$ that

$$T^{-1}\sum_{t=1}^{[rT]} x_t x_t' \Rightarrow \sum_{n=0}^{C} \mathbb{1}\left\{\lambda_{2n} < r \leq \lambda_{2n+2}\right\} \sum_{j=1}^{2n+1} (-1)^{j+1}\left(r \wedge \lambda_j\right) Q^*.$$

Lemma C3. *Under Assumption NR′ 3, 4, 5, 6, and 7, for* $r \in (0,1]$,

$$T^{-1/2}\sum_{t=1}^{[rT]} v_t \Rightarrow \Lambda^* \sum_{n=0}^{C} \mathbb{1}\left\{\lambda_{2n} < r \leq \lambda_{2n+2}\right\} \sum_{j=1}^{2n+1} (-1)^{j+1}\, \mathcal{W}_k\left(r \wedge \lambda_j\right).$$

Proof: Assumptions NR′ 3, 4, 5 and 6 are sufficient for Assumptions NR 1 and 3 by Lemma C1. Using similar algebra as the proof in Lemma C2, we have for $r \in \left(\lambda_{2n}, \lambda_{2n+1}\right]$,

$$\begin{aligned}
&T^{-1/2}\sum_{t=1}^{[rT]} v_t \\
&= T^{-1/2}\sum_{t=1}^{[rT]} a_t v_t^* \\
&= T^{-1/2}\sum_{t=1}^{[rT]} v_t^* - T^{-1/2}\sum_{t=1}^{[\lambda_{2n}T]} v_t^* + T^{-1/2}\sum_{t=1}^{\left[\lambda_{2n-1}T\right]} v_t^* - \ldots + T^{-1/2}\sum_{t=1}^{[\lambda_1 T]} v_t^* \\
&\Rightarrow \Lambda^* \left[\mathcal{W}_k(r) - \mathcal{W}_k(\lambda_{2n}) + \mathcal{W}_k\left(\lambda_{2n-1}\right) - \ldots + \mathcal{W}_k(\lambda_1)\right],
\end{aligned}$$

and for $r \in \left(\lambda_{2n+1}, \lambda_{2n+2}\right]$,

$$\begin{aligned}
&T^{-1/2}\sum_{t=1}^{[rT]} v_t \\
&= T^{-1/2}\sum_{t=1}^{[rT]} a_t v_t^* \\
&= T^{-1/2} v_t^* - T^{-1/2}\sum_{t=1}^{[\lambda_{2n}T]} v_t^* + T^{-1/2}\sum_{t=1}^{\left[\lambda_{2n-1}T\right]} v_t^* - \ldots + T^{-1/2}\sum_{t=1}^{[\lambda_1 T]} v_t^* \\
&\Rightarrow \Lambda^* \left[\mathcal{W}_k\left(\lambda_{2n+1}\right) - \mathcal{W}_k(\lambda_{2n}) + \mathcal{W}_k\left(\lambda_{2n-1}\right) - \ldots + \mathcal{W}_k(\lambda_1)\right].
\end{aligned}$$

Therefore, for $r \in \left(\lambda_{2n}, \lambda_{2n+2}\right]$,

$$T^{-1/2}\sum_{t=1}^{[rT]} v_t \Rightarrow \Lambda \sum_{j=1}^{2n+1} (-1)^{j+1}\, \mathcal{W}_k\left(r \wedge \lambda_j\right),$$

and it immediately follows for $r \in (0,1]$,

$$T^{-1/2}\sum_{t=1}^{[rT]} v_t \Rightarrow \Lambda^* \sum_{n=0}^{C} \mathbb{1}\left\{\lambda_{2n} < r \leq \lambda_{2n+2}\right\} \sum_{j=1}^{2n+1} (-1)^{j+1}\, \mathcal{W}_k\left(r \wedge \lambda_j\right).$$

Proof of Theorem 2.3 (a). Using Lemmas C2 and C3 it follows that

$$T^{-1}\sum_{t=1}^{T} x_t x_t' \Rightarrow \sum_{j=1}^{2C+1} (-1)^{j+1}\left(1\wedge\lambda_j\right) Q^* = \lambda Q^*,$$

$$T^{-1/2}\sum_{t=1}^{T} v_t \Rightarrow \Lambda^* \sum_{j=1}^{2C+1} (-1)^{j+1}\mathcal{W}_k\left(1\wedge\lambda_j\right) = \Lambda^*\overline{\mathcal{W}}_k,$$

which imply that

$$\sqrt{T}\left(\hat{\beta}-\beta\right) = \left(T^{-1}\sum_{t=1}^{T} x_t x_t'\right)^{-1} T^{-1/2}\sum_{t=1}^{T} v_t \Rightarrow \lambda^{-1}Q^{*-1}\Lambda^*\overline{\mathcal{W}}_k.$$

Lemma C4. *Let $T^{-1/2}\hat{S}_{[rT]} = T^{-1/2}\sum_{t=1}^{[rT]}\hat{v}_t$. Let $\{\lambda_i\}$ denote the set $\{\lambda_1,\lambda_2,...,\lambda_{2C}\}$. Under Assumption NR$'$, for $r\in(0,1]$, as $T\to\infty$,*

$$T^{-1/2}\hat{S}_{[rT]} \Rightarrow \Lambda^*\breve{B}_k\left(r,\{\lambda_i\}\right),$$

where

$$\begin{aligned}
&\breve{B}_k\left(r,\{\lambda_i\}\right)\\
&= \sum_{n=0}^{C}\mathbb{1}\left\{\lambda_{2n} < r \leq \lambda_{2(n+1)}\right\}\sum_{j=1}^{2n+1}(-1)^{j+1}\left(\mathcal{W}_k\left(r\wedge\lambda_j\right) - \left(r\wedge\lambda_j\right)\lambda^{-1}\overline{\mathcal{W}}_{k\cdot}\right).
\end{aligned}$$

Proof: For $r\in(0,1]$ we can write

$$\begin{aligned}
&T^{-1/2}\hat{S}_{[rT]}\\
&= T^{-1/2}\sum_{t=1}^{[rT]}\hat{v}_t = T^{-1/2}\sum_{t=1}^{[rT]} v_t - T^{-1}\sum_{t=1}^{[rT]} x_t x_t'\sqrt{T}\left(\hat{\beta}-\beta\right)\\
&\Rightarrow \Lambda^*\sum_{n=0}^{C}\mathbb{1}\left\{\lambda_{2n} < r \leq \lambda_{2n+2}\right\}\sum_{j=1}^{2n+1}(-1)^{j+1}\mathcal{W}_k\left(r\wedge\lambda_j\right)\\
&-\sum_{n=0}^{C}\mathbb{1}\left\{\lambda_{2n} < r \leq \lambda_{2n+2}\right\}\sum_{j=1}^{2n+1}(-1)^{j+1}\left(r\wedge\lambda_j\right)Q^*\left(\lambda Q^*\right)^{-1}\Lambda^*\overline{\mathcal{W}}_k\\
&= \Lambda^*\sum_{n=0}^{C}\mathbb{1}\left\{\lambda_{2n} < r \leq \lambda_{2n+2}\right\}\sum_{j=1}^{2n+1}(-1)^{j+1}\left(\mathcal{W}_k\left(r\wedge\lambda_j\right) - \left(r\wedge\lambda_j\right)\lambda^{-1}\overline{\mathcal{W}}_k\right)\\
&\equiv \Lambda^*\breve{B}(r,\{\lambda_i\})
\end{aligned}$$

where the weak convergence, $\Rightarrow$, is straightforward given Lemmas C2-C3 and Theorem 2.3 (a).

Proof of Theorem 2.3 (b): We can write

$$\begin{aligned}
&\hat{\Omega}^{AM} \\
&= T^{-1} \sum_{t=1}^{T} \sum_{s=1}^{T} k\left(\frac{t-s}{bT}\right) \hat{v}_t \hat{v}_s' \\
&= T^{-1} \sum_{t=1}^{T-1} T^{-1} \sum_{s=1}^{T-1} T^{-1/2} \hat{S}_t \\
&\times T^2 \left[k\left(\frac{t-s}{bT}\right) - k\left(\frac{t-s-1}{bT}\right) - k\left(\frac{t-s+1}{bT}\right) + k\left(\frac{t-s}{bT}\right)\right] T^{-1/2} \hat{S}_s',
\end{aligned}$$

where the second line follows by application of summation by parts to each sum. By Lemma C4 and Kiefer and Vogelsang (2005) it follows that

$$\hat{\Omega}^{AM} \Rightarrow \Lambda^* P\left(b, \breve{B}_k(r, \{\lambda_i\}\right) \Lambda^{*\prime}.$$

Given the lemmas, we can now sketch the proof of Theorem 2.3 (c).

Proof of Theorem 2.3 (c): Using Theorem 2.3 (a) and the delta method, it directly follows that

$$\sqrt{T} r(\hat{\beta}) \Rightarrow R\left(\beta_0\right) \left(\lambda Q^*\right)^{-1} \Lambda^* \overline{\mathcal{W}}_k$$

where $R(\beta_0) = {}^{\partial r(\beta)}/{}_{\partial \beta'}|_{\beta=\beta_0}$. Note that the limit is q linear combinations of k independent Wiener processes. Because Wiener processes are Gaussian, linear combinations of Wiener processes are also Gaussian. Thus, we can rewrite the q linear combinations of k independent Wiener processes as q linear combinations of q independent Wiener processes. Define the $q \times q$ matrix Δ^* such that

$$\Delta^* \Delta^{*\prime} = R\left(\beta_0\right) \left(\lambda Q^*\right)^{-1} \Omega^* \left(\lambda Q^*\right)^{-1} R\left(\beta_0\right)'.$$

An equivalent representation for $R\left(\beta_0\right) \left(\lambda Q^*\right)^{-1} \Lambda^* \overline{\mathcal{W}}_k$ is given by

$$R\left(\beta_0\right) \left(\lambda Q^*\right)^{-1} \Lambda^* \overline{\mathcal{W}}_k = \Delta^* \overline{\mathcal{W}}_q.$$

Using Lemmas C2 and Theorem 2.3 (b), it follows that

$$
\begin{aligned}
& W_T \\
& = \sqrt{T} r(\hat{\beta})' \left[r(\hat{\beta}) \left(T^{-1} \sum_{t=1}^{T} x_t x_t' \right)^{-1} \hat{\Omega}^{AM} \left(T^{-1} \sum_{t=1}^{T} x_t x_t' \right)^{-1} r(\hat{\beta})' \right]^{-1} \sqrt{T} r(\hat{\beta}) \\
& \Rightarrow \left[R(\beta_0) \left(\lambda Q^* \right)^{-1} \Lambda^* \overline{\mathcal{W}}_k \right]' \\
& \times \left[R\left(\beta_0\right) \left(\lambda Q^* \right)^{-1} \Lambda^* P\left(b, \breve{B}_k(r, \{\lambda_i\}) \right) \Lambda^{*\prime} \left(\lambda Q^* \right)^{-1} R\left(\beta_0\right)' \right]^{-1} \\
& \times \left[R(\beta_0) \left(\lambda Q^* \right)^{-1} \Lambda^* \overline{\mathcal{W}}_k \right] \\
& = \left(\Delta^* \overline{\mathcal{W}}_q \right)' \left[\Delta^* P \left(b, \breve{B}_q(r, \{\lambda_i\} \right) \Delta^{*\prime} \right]^{-1} \Delta^* \overline{\mathcal{W}}_q \\
& = \overline{\mathcal{W}}_q' \left[P \left(b, \breve{B}_q(r, \{\lambda_i\}) \right) \right]^{-1} \overline{\mathcal{W}}_q,
\end{aligned}
$$

and the proof is complete. Note that for the case of one restriction, $q = 1$, it follows for the t-statistic that,

$$
t_T \Rightarrow \frac{\overline{\mathcal{W}}_1}{\sqrt{P\left(b, \breve{B}_1(r, \{\lambda_i\})\right)}}.
$$

Appendix D

PROOFS FOR EQUAL SPACED STATISTIC, NON-RANDOM MISSING DATA

We first define some relevant functions similar to those defined in Kiefer and Vogelsang (2005). Define the functions

$$k_b(x) = k\left(\frac{x}{b}\right),$$

$$\Delta^2 K^a_{ts}$$

$$= k\left(\frac{\sum_{i=1}^{t} a_i - \sum_{i=1}^{s} a_i}{bT_{ES}}\right) - k\left(\frac{\sum_{i=1}^{t} a_i - \sum_{i=1}^{s} a_i - 1}{bT_{ES}}\right) - k\left(\frac{\sum_{i=1}^{t} a_i - \sum_{i=1}^{s} a_i + 1}{bT_{ES}}\right)$$

$$+ k\left(\frac{\sum_{i=1}^{t} a_i - \sum_{i=1}^{s} a_i}{bT_{ES}}\right),$$

$$D_{T_{ES}}(r)$$

$$= T_{ES}^2\left[\left(k_b\left(\frac{[rT_{ES}]+1}{T_{ES}}\right) - k_b\left(\frac{[rT_{ES}]}{T_{ES}}\right)\right) - \left(k_b\left(\frac{[rT_{ES}]}{T_{ES}}\right) - k_b\left(\frac{[rT_{ES}]-1}{T_{ES}}\right)\right)\right].$$

When $k(x)$ is twice continuously differentiable,

$$\lim_{T_{ES}\to\infty} D_{T_{ES}}(r) = k_b''(r) = \frac{1}{b^2}k\left(\frac{r}{b}\right), \tag{eqD.1}$$

where the limit holds uniformly in r by the definition of the second derivative and the continuity of $k''(x)$. Let $k_{b-}'(b)$ denote the first derivative of $k_b(x)$ from the left at $x = b$. Then, by definition,

$$\begin{aligned} b^{-1}k'(1)_- = k_{b-}'(b) &= \lim_{T_{ES}\to\infty} T_{ES}\left[k_b(b) - k_b\left(b - \frac{1}{T_{ES}}\right)\right] \qquad \text{(eqD.2)} \\ &= \lim_{T_{ES}\to\infty} T_{ES}\left[k_b\left(\frac{[bT_{ES}]}{T_{ES}}\right) - k_b\left(\frac{[bT_{ES}]}{T_{ES}} - \frac{1}{T_{ES}}\right)\right]. \end{aligned}$$

Throughout this section we assume that $M_{ES} = bT_{ES}$ where $b \in (0,1]$ is fixed. For notational purposes we let the summation be zero whenever the starting value is larger than the final value. For example, for a sequence $\{a_k\}$, then we have $\sum_{k=1}^{0} a_k = 0$.

Lemma D1. *An equivalent expression for $\hat{\Omega}^{ES}$ is given by*

$$\hat{\Omega}^{ES} = \frac{1}{T_{ES}} \sum_{t=1}^{T-1} \sum_{s=1}^{T-1} a_{t+1} a_{s+1} \hat{S}_t \Delta^2 K_{ts}^{a} \hat{S}_s'.$$

Proof: First rewrite $\hat{\Omega}^{ES}$ using summation by parts (see Kiefer and Vogelsang (2005) for details):

$$\begin{aligned}
\hat{\Omega}^{ES} &= \frac{1}{T_{ES}} \sum_{t=1}^{T} \sum_{s=1}^{T} k \left(\frac{\sum\limits_{i=1}^{t} a_i - \sum\limits_{i=1}^{s} a_i}{bT_{ES}} \right) \hat{v}_t \hat{v}_s', \\
&= \frac{1}{T_{ES}} \sum_{t=1}^{T-1} \sum_{s=1}^{T-1} \hat{S}_t \left[k \left(\frac{\sum\limits_{i=1}^{t} a_i - \sum\limits_{i=1}^{s} a_i}{bT_{ES}} \right) - k \left(\frac{\sum\limits_{i=1}^{t} a_i - \sum\limits_{i=1}^{s+1} a_i}{bT_{ES}} \right) \right. \\
&\quad \left. -k \left(\frac{\sum\limits_{i=1}^{t+1} a_i - \sum\limits_{i=1}^{s} a_i}{bT_{ES}} \right) + k \left(\frac{\sum\limits_{i=1}^{t+1} a_i - \sum\limits_{i=1}^{s+1} a_i}{bT_{ES}} \right) \right] \hat{S}_s'
\end{aligned}$$

Note that if $a_{t+1} = a_{s+1} = 1$, it follows that

$$k\left(\frac{\sum_{i=1}^{t} a_i - \sum_{i=1}^{s} a_i}{bT_{ES}}\right) - k\left(\frac{\sum_{i=1}^{t} a_i - \sum_{i=1}^{s+1} a_i}{bT_{ES}}\right) - k\left(\frac{\sum_{i=1}^{t+1} a_i - \sum_{i=1}^{s} a_i}{bT_{ES}}\right)$$
$$+ k\left(\frac{\sum_{i=1}^{t+1} a_i - \sum_{i=1}^{s+1} a_i}{bT_{ES}}\right)$$
$$= k\left(\frac{\sum_{i=1}^{t} a_i - \sum_{i=1}^{s} a_i}{M}\right) - k\left(\frac{\sum_{i=1}^{t} a_i - \sum_{i=1}^{s} a_i - 1}{M}\right) - k\left(\frac{\sum_{i=1}^{t} a_i - \sum_{i=1}^{s} a_i + 1}{M}\right)$$
$$+ k\left(\frac{\sum_{i=1}^{t} a_i - \sum_{i=1}^{s} a_i}{M}\right) \equiv \Delta^2 K_{ts}^{a}.$$

However when $a_{t+1} = 0$ and/or $a_{s+1} = 0$, it follows that

$$k\left(\frac{\sum_{i=1}^{t} a_i - \sum_{i=1}^{s} a_i}{bT_{ES}}\right) - k\left(\frac{\sum_{i=1}^{t} a_i - \sum_{i=1}^{s+1} a_i}{bT_{ES}}\right) - k\left(\frac{\sum_{i=1}^{t+1} a_i - \sum_{i=1}^{s} a_i}{bT_{ES}}\right)$$
$$+ k\left(\frac{\sum_{i=1}^{t+1} a_i - \sum_{i=1}^{s+1} a_i}{bT_{ES}}\right) = 0 \neq \Delta^2 K_{ts}^{a}.$$

This holds because if $a_{t+1} = 0$, then the first term cancels out the third term and the second term cancels out the fourth term, and if $a_{s+1} = 0$ then the first two terms cancel each other out and the last two terms cancel each other out. Whenever $a_{t+1} = 0$ and/or $a_{s+1} = 0$, we require the argument of the sum to be zero. This can be accomplished by scaling the argument of the sum by $a_{t+1}a_{s+1}$. Using this device it follows that

$$\hat{\Omega}^{ES} = \frac{1}{T_{ES}} \sum_{t=1}^{T-1} \sum_{s=1}^{T-1} a_{t+1}a_{s+1}\hat{S}_t \Delta^2 K^a_{ts} \hat{S}'_s,$$

completing the proof.

We next prove a collection of lemmas used to establish the limit of $\hat{\Omega}^{ES}$. The first set of lemmas are algebraic and mechanical whereas the second set of lemmas work out the limits of components of $\hat{\Omega}^{ES}$.

Lemma D2. *Under Assumption NR$'$ 6, , it follows as $T \to \infty$,*

$$T_{ES}^{-1}\left(\sum_{t=1}^{[rT]} a_t - \sum_{t=1}^{[uT]} a_t\right) \to \lambda^{-1}\left(\sum_{n=0}^{C} \mathbb{1}\left\{\lambda_{2n} < r \leq \lambda_{2(n+1)}\right\} \sum_{j=1}^{2n+1} (-1)^{j+1}\left(r \wedge \lambda_j\right) \right.$$
$$\left. - \sum_{l=0}^{C} \mathbb{1}\left\{\lambda_{2l} < u \leq \lambda_{2(l+1)}\right\} \sum_{j=1}^{2l+1} (-1)^{j+1}\left(u \wedge \lambda_j\right)\right).$$

Proof: It is sufficient to establish the limit of $T_{ES}^{-1} \sum_{t=1}^{[rT]} a_t$. First consider the behavior of $T^{-1} \sum_{t=1}^{[rT]} a_t$. There are two possibilities depending on the value of r. The first is when r is in an interval such that data are observed at $t = [rT]$ and the second is such that data is missing at $t = [rT]$. Note that data are observed at $t = [rT]$ whenever $r \in \left(\lambda_{2n}, \lambda_{2n+1}\right]$, $n = 0, \ldots, C$. Therefore, when $r \in \left(\lambda_{2n}, \lambda_{2n+1}\right]$, we can write

$$\begin{aligned} T^{-1}\sum_{t=1}^{[rT]} a_t &= T^{-1}\sum_{t=1}^{[rT]} 1 - T^{-1}\sum_{t=1}^{[\lambda_{2n}T]} 1 + T^{-1}\sum_{t=1}^{[\lambda_{2n-1}T]} 1 - \ldots + T^{-1}1, \\ &\to r - \lambda_{2n} + \lambda_{2n-1} - \ldots + \lambda_1, \\ &= r + \sum_{j=1}^{2n} (-1)^{j+1}\lambda_j \qquad \text{(eqD.3)} \end{aligned}$$

where the term $\sum_{j=1}^{2n}(-1)^{j+1}\lambda_j$ is removing the missing portions from r. In contrast, when $r \in \left(\lambda_{2n+1}, \lambda_{2n+2}\right]$, data is missing at $t = [rT]$. Hence when $r \in \left(\lambda_{2n+1}, \lambda_{2n+2)}\right]$, we can write

$$T^{-1}\sum_{t=1}^{[rT]} a_t = T^{-1}\sum_{t=1}^{\left[\lambda_{2n+1}T\right]} 1 - T^{-1}\sum_{t=1}^{\left[\lambda_{2n}T\right]} 1 + T^{-1}\sum_{t=1}^{\left[\lambda_{2n-1}T\right]} 1 - \ldots + T^{-1}\sum_{t=1}^{\left[\lambda_{1}T\right]} 1,$$

$$\Rightarrow \lambda_{2n+1} - \lambda_{2n} + \lambda_{2n-1} - \ldots + \lambda_1,$$

$$= \lambda_{2n+1} + \sum_{j=1}^{2n} (-1)^{j+1}\lambda_j \qquad \text{(eqD.4)}$$

The reason that we have a different expression compared to (eqD.3) is because r is now located where the observations are missing and we have to remove the portion of missing observations from λ_{2n+1} rather than from r. From λ_{2n+1} to r, there is no observed data and thus $a_t = 0$ for t in the range $[\lambda_{2n+1}T] < t \leq [rT]$. Combining (eqD.3) and (eqD.4), the following holds for $r \in (\lambda_{2n}, \lambda_{2n+2}]$:

$$T^{-1}\sum_{t=1}^{[rT]} a_t = \mathbb{1}\left\{\lambda_{2n} < r \leq \lambda_{2n+1}\right\} T^{-1}\sum_{t=1}^{[rT]} a_t + \mathbb{1}\left\{\lambda_{2n+1} < r \leq \lambda_{2n+2}\right\} T^{-1}\sum_{t=1}^{[rT]} a_t$$

$$\to \mathbb{1}\left\{\lambda_{2n} < r \leq \lambda_{2n+1}\right\}\left(r + \sum_{j=1}^{2n} (-1)^{j+1}\lambda_j\right)$$

$$+ \mathbb{1}\left\{\lambda_{2n+1} < r \leq \lambda_{2n+2}\right\}\left(\lambda_{2n+1} + \sum_{j=1}^{2n} (-1)^{j+1}\lambda_j\right)$$

$$= \mathbb{1}\left\{\lambda_{2n} < r \leq \lambda_{2n+2}\right\}\sum_{j=1}^{2n+1} (-1)^{j+1}\left(r \wedge \lambda_j\right).$$

It immediately follows for a general value of $r \in (0, 1]$:

$$T^{-1}\sum_{t=1}^{[rT]} a_t \to \sum_{n=0}^{C} \mathbb{1}\{\lambda_{2n} < r \leq \lambda_{2n+2}\}\sum_{j=1}^{2n+1} (-1)^{j+1}\left(r \wedge \lambda_j\right). \qquad \text{(eqD.5)}$$

Applying the result given by (eqD.5) for the case of $r = 1$ gives

$$\frac{T_{ES}}{T} = T^{-1}\sum_{t=1}^{T} a_t \to \sum_{j=1}^{2C+1} \lambda_j (-1)^{j+1} \equiv \lambda. \qquad \text{(eqD.6)}$$

Using (eqD.5) and (eqD.6) it follows that

$$T_{ES}^{-1} \sum_{t=1}^{[rT]} a_t = \left(\frac{T_{ES}}{T} \right)^{-1} T^{-1} \sum_{t=1}^{[rT]} a_t$$
$$\to \frac{1}{\lambda} \sum_{n=0}^{C} \mathbb{1} \left\{ \lambda_{2n} < r \leq \lambda_{2(n+1)} \right\} \sum_{j=1}^{2n+1} (-1)^{j+1} \left(r \wedge \lambda_j \right),$$

and the lemma is established.

Lemma D3. *The following algebraic relationship holds:*

$$a_{t+1} a_{s+1} = \sum_{n=0}^{C} \sum_{l=0}^{C} \mathbb{1} \left\{ T_{2n} \leq t \leq T_{2n+1} - 1 \right\} \mathbb{1} \left\{ T_{2l} \leq s \leq T_{2l+1} - 1 \right\}.$$

Proof: Recall that data are observed when there exists a value of n such that $T_{2n} + 1 \leq t \leq T_{2n+1}$ (see Definition 1). Therefore, $a_{t+1} = 1$ implies that there is a value of n such that

$$T_{2n} \leq t + 1 \leq T_{2n+1},$$

or equivalently

$$T_{2n} \leq t \leq T_{2n+1} - 1.$$

If $a_{t+1} = 0$, then t does not satisfy this inequality for any value of n. Therefore, we may write

$$a_{t+1} = \sum_{n=0}^{C} \mathbb{1} \left\{ T_{2n} \leq t \leq T_{2n+1} - 1 \right\},$$

and it directly follows that

$$a_{t+1} a_{s+1} = \sum_{n=0}^{C} \mathbb{1} \left\{ T_{2n} \leq t \leq T_{2n+1} - 1 \right\} \sum_{l=0}^{C} \mathbb{1} \left\{ T_{2l} \leq s \leq T_{2l+1} - 1 \right\}$$
$$= \sum_{n=0}^{C} \sum_{l=0}^{C} \mathbb{1} \left\{ T_{2n} \leq t \leq T_{2n+1} - 1 \right\} \mathbb{1} \left\{ T_{2l} \leq s \leq T_{2l+1} - 1 \right\}.$$

Lemma D4. *The following algebraic relationship holds:*

$$a_{t+1}a_{s+1}\mathbb{1}\left\{\left|\sum_{i=1}^{t}a_i-\sum_{i=1}^{s}a_i\right|<\left[bT_{ES}\right]\right\}=\sum_{n=0}^{C}\sum_{l=0}^{C}\mathbb{1}\left\{T_{2n}\leq t\leq T_{2n+1}-1\right\}$$
$$\times\,\mathbb{1}\left\{T_{2l}\leq s\leq T_{2l+1}-1\right\}\mathbb{1}\left\{|t-s|<\left[bT_{ES}\right]+\sum_{k=2(n\wedge l)+1}^{2(n\vee l)}(-1)^kT_k\right\}$$

Proof: Note that the number of missing observations in the first missing cluster is (T_2-T_1), the number of missing observations in the second missing cluster is (T_4-T_3), and so forth. Hence the n^{th} missing cluster has $(T_{2n}-T_{2n-1})$ missing observations. Therefore, the total number of missing observations in the first n missing clusters is

$$\sum_{k=1}^{n}\left(T_{2k}-T_{2k-1}\right). \tag{eqD.7}$$

Suppose that t is in the range $T_{2n}\leq t\leq T_{2n+1}-1$. We want to count the number of observed data points up to time t. We further divide this interval for t into two parts because when t is in the range $T_{2n}<t\leq T_{2n+1}-1$ data is observed at time t while for $t=T_{2n}$ data is missing. First consider the case $T_{2n}<t\leq T_{2n+1}-1$. In this case there are n missing clusters before time t. Hence the number of missing observations up to time t is $\sum_{k=1}^{n}\left(T_{2k}-T_{2k-1}\right)$ from (eqD.7). Subtracting this number of missing observations from t, we obtain the number of observed data points up to time t. Therefore it follows that

$$\begin{aligned}\sum_{i=1}^{t}a_i&=t-\sum_{k=1}^{n}\left(T_{2k}-T_{2k-1}\right)\\&=t-\sum_{k=1}^{2n}(-1)^kT_k.\end{aligned} \tag{eqD.8}$$

Next, consider the case of $t=T_{2n}$. Because data is not observed at $t=T_{2n}$, instead of counting all the way up to time t, we only count up to time T_{2n-1}, which is the last time period where the data is available. There are $(n-1)$ missing clusters up to time T_{2n-1}. Then using (eqD.7), the number of missing observations in those $(n-1)$ clusters is $\sum_{k=1}^{n-1}(T_{2k}-T_{2k-1})$. Hence the number of observed data points up to time T_{2n} is

$$\sum_{i=1}^{T_{2n}} a_i = T_{2n-1} - \sum_{k=1}^{n-1}(T_{2k}-T_{2k-1}),$$

which can be re-expressed as

$$\begin{aligned}
T_{2n-1} - \sum_{k=1}^{n-1}\left(T_{2k}-T_{2k-1}\right) &= T_{2n} - (T_{2n}-T_{2n-1}) - \sum_{k=1}^{n-1}\left(T_{2k}-T_{2k-1}\right) \\
&= T_{2n} - \sum_{k=1}^{n}\left(T_{2k}-T_{2k-1}\right) \\
&= T_{2n} - \sum_{k=1}^{2n}(-1)^k T_k,
\end{aligned}$$

showing that the $t = T_{2n}$ case can also be expressed as (eqD.8). Therefore, when t falls in the range $T_{2n} \leq t \leq T_{2n+1}-1$, it follows that

$$\sum_{i=1}^{t} a_i = t - \sum_{k=1}^{2n}(-1)^k T_k. \tag{eqD.9}$$

Now consider values of t and s with $t \geq s$ such that $T_{2n} \leq t \leq T_{2n+1}-1$, $T_{2l} \leq s \leq T_{2l+1}-1$. Note that because $t \geq s$ it follows that $n \geq l$. Using (eqD.9) gives,

$$\begin{aligned}
\sum_{i=1}^{t} a_i - \sum_{i=1}^{s} a_i &= \left(t - \sum_{k=1}^{2n}(-1)^k T_k\right) - \left(s - \sum_{k=1}^{2l}(-1)^k T_k\right) \\
&= (t-s) - \sum_{k=2l+1}^{2n}(-1)^k T_k. \qquad \text{(eqD.10)}
\end{aligned}$$

Similarly, when $t \leq s$,

$$\sum_{i=1}^{s} a_i - \sum_{i=1}^{t} a_i = (s-t) - \sum_{k=2n+1}^{2l}(-1)^k T_k.$$

Therefore, we can write

$$\left|\sum_{i=1}^{t} a_i - \sum_{i=1}^{s} a_i\right| = |t-s| - \sum_{k=2(n\wedge l)+1}^{2(n\vee l)}(-1)^k T_k.$$

Using this expression we have

$$\left| \sum_{i=1}^{t} a_i - \sum_{i=1}^{s} a_i \right| < [bT_{ES}]$$

is equivalent to

$$|t-s| < [bT_{ES}] + \sum_{k=2(n\wedge l)+1}^{2(n\vee l)} (-1)^k T_k.$$

From the proof of Lemma D3 we know that $a_{t+1} = 1$ and $a_{s+1} = 1$ if and only if there is a value of n and a value of l such that

$$T_{2n} \leq t \leq T_{2n+1} - 1, \quad T_{2l} \leq s \leq T_{2l+1} - 1,$$

and when this is the case, $\left|\sum_{i=1}^{t} a_i - \sum_{i=1}^{s} a_i\right| < [bT_{ES}]$ if and only if

$$|t-s| < [bT_{ES}] + \sum_{k=2(n\wedge l)+1}^{2(n\vee l)} (-1)^k T_k,$$

and it immediately follows for this case that

$$a_{t+1}a_{s+1}\mathbb{1}\left\{\left|\sum_{i=1}^{t} a_i - \sum_{i=1}^{s} a_i\right| < [bT_{ES}]\right\} = \mathbb{1}\left\{T_{2n} \leq t \leq T_{2n+1} - 1\right\}$$
$$\times \mathbb{1}\left\{T_{2l} \leq s \leq T_{2l+1} - 1\right\} \mathbb{1}\left\{|t-s| < [bT_{ES}] + \sum_{k=2(n\wedge l)+1}^{2(n\vee l)} (-1)^k T_k\right\}.$$

As was done in the proof of Lemma D3 we can write for general values of t and s:

$$a_{t+1}a_{s+1}\mathbb{1}\left\{\left|\sum_{i=1}^{t} a_i - \sum_{i=1}^{s} a_i\right| < [bT_{ES}]\right\} = \sum_{n=0}^{C}\sum_{l=0}^{C} \mathbb{1}\left\{T_{2n} \leq t \leq T_{2n+1} - 1\right\}$$
$$\times \mathbb{1}\left\{T_{2l} \leq s \leq T_{2l+1} - 1\right\} \mathbb{1}\left\{|t-s| < [bT_{ES}] + \sum_{k=2(n\wedge l)+1}^{2(n\vee l)} (-1)^k T_k\right\}.$$

This completes the proof.

Lemma D5. *Suppose that $t > s$. Then the following algebraic result holds:*

$$a_{t+1}a_{s+1}\mathbb{1}\left\{\sum_{i=1}^{t} a_i - \sum_{i=1}^{s} a_i = [bT_{ES}]\right\} = \sum_{n=0}^{C}\sum_{l=0}^{n}$$
$$\mathbb{1}\left\{T_{2n} - [bT_{ES}] - \sum_{k=2l+1}^{2n}(-1)^k T_k \leq s \leq T_{2n+1} - 1 - [bT_{ES}] - \sum_{k=2l+1}^{2n}(-1)^k T_k\right\}$$
$$\times\, \mathbb{1}\left\{T_{2l} \leq s \leq T_{2l+1} - 1\right\}\mathbb{1}\left\{t = s + [bT_{ES}] + \sum_{k=2l+1}^{2n}(-1)^k T_k\right\}$$

Proof: From the proof of Lemma D3 we know that $a_{t+1} = 1$ and $a_{s+1} = 1$ if and only if there is a value of n and a value of l such that

$$T_{2n} \leq t \leq T_{2n+1} - 1, \quad T_{2l} \leq s \leq T_{2l+1} - 1.$$

From (eqD.10) in Lemma D4, we also know that when

$$\sum_{i=1}^{t} a_i - \sum_{i=1}^{s} a_i = [bT_{ES}]$$

it follows that

$$t = s + [bT_{ES}] + \sum_{k=2l+1}^{2n}(-1)^k T_k$$

because $t > s$. Plugging this formula for t into the inequality $T_{2n} \leq t \leq T_{2n+1} - 1$ gives

$$T_{2n} \leq s + [bT_{ES}] + \sum_{k=2l+1}^{2n}(-1)^k T_k \leq T_{2n+1} - 1,$$

which can be rearranged as

$$T_{2n} - [bT_{ES}] - \sum_{k=2l+1}^{2n}(-1)^k T_k \leq s \leq T_{2n+1} - 1 - [bT_{ES}] - \sum_{k=2l+1}^{2n}(-1)^k T_k.$$

Hence, given $t > s$, the conditions: $a_{t+1} = 1$, $a_{s+1} = 1$ and $\sum_{i=1}^{t} a_i - \sum_{i=1}^{s} a_i = [bT_{ES}]$ hold if and only if the following three conditions are satisfied for some value of n and some value of l:

$$T_{2n} - [bT_{ES}] - \sum_{k=2l+1}^{2n} (-1)^k T_k \le s \le T_{2n+1} - 1 - [bT_{ES}] - \sum_{k=2l+1}^{2n} (-1)^k T_k,$$

$$T_{2l} \le s \le T_{2l+1} - 1,$$

$$t = s + [bT_{ES}] + \sum_{k=2l+1}^{2n} (-1)^k T_k.$$

In terms of indicator functions we express this equivalence as

$$\begin{aligned}
&a_{t+1}a_{s+1}\mathbb{1}\left\{\sum_{i=1}^{t} a_i - \sum_{i=1}^{s} a_i = [bT_{ES}]\right\} \\
=&\mathbb{1}\left\{T_{2n} - [bT_{ES}] - \sum_{k=2l+1}^{2n} (-1)^k T_k \le s \le T_{2n+1} - 1 - [bT_{ES}] - \sum_{k=2l+1}^{2n} (-1)^k T_k\right\} \\
&\times \mathbb{1}\left\{T_{2l} \le s \le T_{2l+1} - 1\right\} \mathbb{1}\left\{t = s + [bT_{ES}] + \sum_{k=2l+1}^{2n} (-1)^k T_k\right\}.
\end{aligned}$$

Writing more generally as done in the proof of Lemma D3 by combining the above expression for all possible values of n and l with $n \ge l$ gives the desired relationship:

$$\begin{aligned}
&a_{t+1}a_{s+1}\mathbb{1}\left\{\sum_{i=1}^{t} a_i - \sum_{i=1}^{s} a_i = [bT_{ES}]\right\} = \sum_{n=0}^{C}\sum_{l=0}^{n} \\
&\mathbb{1}\left\{T_{2n} - [bT_{ES}] - \sum_{k=2l+1}^{2n} (-1)^k T_k \le s \le T_{2n+1} - 1 - [bT_{ES}] - \sum_{k=2l+1}^{2n} (-1)^k T_k\right\} \\
&\times \mathbb{1}\left\{T_{2l} \le s \le T_{2l+1} - 1\right\} \mathbb{1}\left\{t = s + [bT_{ES}] + \sum_{k=2l+1}^{2n} (-1)^k T_k\right\}.
\end{aligned}$$

Lemma D6. *Suppose that $t > s$. Then the following algebraic result holds:*

$$\begin{aligned}
&a_{t+1}a_{s+1}\mathbb{1}\left\{\sum_{i=1}^{t} a_i - \sum_{i=1}^{s} a_i = [bT_{ES}] + 1\right\} = \sum_{n=0}^{C}\sum_{l=0}^{n} \\
&\mathbb{1}\left\{T_{2n} - [bT_{ES}] - 1 - \sum_{k=2l+1}^{2n} (-1)^k T_k \le s \le T_{2n+1} - 2 - [bT_{ES}] - \sum_{k=2l+1}^{2n} (-1)^k T_k\right\} \\
&\times \mathbb{1}\left\{T_{2l} \le s \le T_{2l+1} - 1\right\} \mathbb{1}\left\{t = s + [bT_{ES}] + 1 + \sum_{k=2l+1}^{2n} (-1)^k T_k\right\}
\end{aligned}$$

The proof is essentially the same as the proof of Lemma D5.

Lemma D7. *The following algebraic result holds:*

$$a_{t+1}a_{s+1}\mathbb{1}\left\{\sum_{i=1}^{t} a_i - \sum_{i=1}^{s} a_i = 0\right\} = \sum_{n=0}^{C} \mathbb{1}\left\{T_{2n} \leq t \leq T_{2n+1} - 1\right\}\mathbb{1}\left\{t = s\right\}$$

Proof: Note that when $a_{t+1} = 1$ and $a_{s+1} = 1$, it follows that

$$\sum_{i=1}^{t} a_i - \sum_{i=1}^{s} a_i = 0$$

if and only if $t = s$. It is obvious that the difference in sums is zero when $t = s$. The difference in sums cannot be zero if t and s are different. Suppose that $t > s$. Then we have

$$\sum_{i=1}^{t} a_i - \sum_{i=1}^{s} a_i = a_t + a_{t-1} + \cdots + a_{s+2} + a_{s+1} \neq 0,$$

because $a_{s+1} = 1$. We have the same conclusion when $t < s$ due to the fact that $a_{t+1} = 1$. Hence, $a_{t+1} = 1$, $a_{s+1} = 1$ and $\sum_{i=1}^{t} a_i - \sum_{i=1}^{s} a_i = 0$ are satisfied if and only if $t = s$ and there is a value of n such that

$$T_{2n} \leq t \leq T_{2n+1} - 1.$$

In terms of indicator functions we can write these conditions as

$$a_{t+1}a_{s+1}\mathbb{1}\left\{\sum_{i=1}^{t} a_i - \sum_{i=1}^{s} a_i = 0\right\} = \mathbb{1}\left\{T_{2n} \leq t \leq T_{2n+1} - 1\right\}\mathbb{1}\left\{t = s\right\}.$$

Writing more generally as done in the proof of Lemma D3 we have the desired result:

$$a_{t+1}a_{s+1}\mathbb{1}\left\{\sum_{i=1}^{t} a_i - \sum_{i=1}^{s} a_i = 0\right\} = \sum_{n=0}^{C} \mathbb{1}\left\{T_{2n} \leq t \leq T_{2n+1} - 1\right\}\mathbb{1}\left\{t = s\right\}.$$

The next collection of lemmas establish the limit of $\hat{\Omega}^{ES}$.

Lemma D8. *Let $M_{ES} = bT_{ES}$ where b is a fixed constant with $b \in [0, 1]$. When $k(x)$ is twice continuously differentiable, under Assumptions NR$'$, as $T \to \infty$*

$$\hat{\Omega}^{ES} \Rightarrow \Lambda^* P_1^{ES}\left(b, \breve{B}_k(\{\lambda\}_1^{2C})\right)\Lambda^{*\prime}$$

where,

$$P_1^{ES}\left(b,\breve{B}_k(\{\lambda\}_1^{2C})\right) = -\frac{1}{b^2\lambda^3}\sum_{n=0}^{C}\sum_{l=0}^{C}\int_{\lambda_{2n}}^{\lambda_{2n+1}}\int_{\lambda_{2l}}^{\lambda_{2l+1}}$$
$$k''\left((\lambda b)^{-1}\left[\sum_{j=1}^{2n+1}(-1)^{j+1}(r\wedge\lambda_j) - \sum_{j=1}^{2l+1}(-1)^{j+1}(u\wedge\lambda_j)\right]\right)$$
$$\times\,\breve{B}_k(r,\{\lambda_i\})\breve{B}_k(u,\{\lambda_i\})' dudr.$$

Proof: Using the definitions at the beginning of this appendix, it is straightforward to show that

$$T_{ES}^2\Delta^2 K_{ts}^a = -D_{T_{ES}}\left(T_{ES}^{-1}\left(\sum_{i=1}^{t}a_i - \sum_{i=1}^{s}a_i\right)\right).$$

From Lemma D1 we know that

$$\hat{\Omega}^{ES} = \frac{1}{T_{ES}}\sum_{t=1}^{T-1}\sum_{s=1}^{T-1}a_{t+1}a_{s+1}\hat{S}_t\Delta^2 K_{ts}^a\hat{S}_s'.$$

Re-expressing $\hat{\Omega}^{ES}$ in terms of $D_{T_{ES}}(r)$ gives

$$\hat{\Omega}^{ES} = -\frac{1}{T_{ES}}\sum_{t=1}^{T-1}\sum_{s=1}^{T-1}a_{t+1}a_{s+1}\hat{S}_t T_{ES}^{-2}D_{T_{ES}}\left(T_{ES}^{-1}\left(\sum_{i=1}^{t}a_i - \sum_{i=1}^{s}a_i\right)\right)\hat{S}_s'.$$

Plugging in the expression from Lemma D3 gives,

$$\begin{aligned}
\hat{\Omega}^{ES} &= -\frac{1}{T_{ES}}\sum_{t=1}^{T-1}\sum_{s=1}^{T-1}\sum_{n=0}^{C}\sum_{l=0}^{C}\mathbb{1}\{T_{2n}\le t\le T_{2n+1}-1\}\mathbb{1}\{T_{2l}\le s\le T_{2l+1}-1\}\\
&\quad\times\hat{S}_t T_{ES}^{-2}D_{T_{ES}}\left(T_{ES}^{-1}\left(\sum_{i=1}^{t}a_i - \sum_{i=1}^{s}a_i\right)\right)\hat{S}_s'\\
&= -\left(\frac{T}{T_{ES}}\right)^3 T^{-1}\sum_{t=1}^{T-1}T^{-1}\sum_{s=1}^{T-1}\Bigg[\sum_{n=0}^{C}\sum_{l=0}^{C}\mathbb{1}\{T_{2n}\le t\le T_{2n+1}-1\}\\
&\quad\times\mathbb{1}\{T_{2l}\le s\le T_{2l+1}-1\}T^{-1/2}\hat{S}_t D_{T_{ES}}\left(T_{ES}^{-1}\left(\sum_{i=1}^{t}a_i - \sum_{i=1}^{s}a_i\right)\right)T^{-1/2}\hat{S}_s'\Bigg].
\end{aligned}$$

Mapping t to $[rT]$ and s to $[uT]$ we can write

$$\hat{\Omega}_{ES} = -\left(\frac{T}{T_{ES}}\right)^3 \int_0^1 \int_0^1$$
$$\left[\sum_{n=0}^{C} \sum_{l=0}^{C} \mathbb{1}\left\{\lambda_{2n} \le r < \lambda_{2n+1}\right\} \mathbb{1}\left\{\lambda_{2l} \le u < \lambda_{2l+1}\right\} T^{-1/2} \hat{S}_{[rT]}\right.$$
$$\left. \times D_{T_{ES}}\left(T_{ES}^{-1}\left(\sum_{i=1}^{[rT]} a_i - \sum_{i=1}^{[uT]} a_i\right)\right) T^{-1/2} \hat{S}'_{[uT]}\right] dudr.$$

Using Lemma D2, Lemma C4, (eqD.1), (eqD.6) and the continuous mapping theorem it follows that

$$\hat{\Omega}_{ES} \Rightarrow \lambda^{-3} \int_0^1 \int_0^1 \left[\sum_{n=0}^{C} \sum_{l=0}^{C} \mathbb{1}\left\{\lambda_{2n} \le r < \lambda_{2n+1}\right\} \mathbb{1}\left\{\lambda_{2l} \le u < \lambda_{2l+1}\right\}\right.$$
$$\times k''\left(b^{-1}\lambda^{-1}\left(\sum_{n=0}^{C} \mathbb{1}\left\{\lambda_{2n} < r \le \lambda_{2(n+1)}\right\} \sum_{j=1}^{2n+1} (-1)^{j+1}\left(r \wedge \lambda_j\right)\right.\right.$$
$$\left.\left. - \sum_{l=0}^{C} \mathbb{1}\left\{\lambda_{2l} < u \le \lambda_{2(l+1)}\right\} \sum_{j=1}^{2l+1} (-1)^{j+1}\left(u \wedge \lambda_j\right)\right)\right)$$
$$\left. \times \Lambda^* \breve{B}_k(r, \{\lambda_i\}) \breve{B}_k(u, \{\lambda_i\})' \Lambda^{*\prime} \right] dudr.$$

The limiting expression can be simplified by breaking up the integrals into the ranges indicated by the indicator functions and using the fact that when $\lambda_{2n} \le r < \lambda_{2n+1}$:

$$\sum_{n=0}^{C} \mathbb{1}\{\lambda_{2n} < r \le \lambda_{2(n+1)}\} \sum_{j=1}^{2n+1} (-1)^{j+1}\left(r \wedge \lambda_j\right) = \sum_{j=1}^{2n+1} (-1)^{j+1}\left(r \wedge \lambda_j\right).$$

Therefore we have

$$\hat{\Omega}_{ES} \Rightarrow -\Lambda^* \frac{1}{b^2\lambda^3} \sum_{n=0}^{C} \sum_{l=0}^{C} \int_{\lambda_{2n}}^{\lambda_{2n+1}} \int_{\lambda_{2l}}^{\lambda_{2l+1}}$$
$$\left[k''\left((\lambda b)^{-1}\left[\sum_{j=1}^{2n+1} (-1)^{j+1}(r \wedge \lambda_j) - \sum_{j=1}^{2l+1} (-1)^{j+1}(u \wedge \lambda_j)\right]\right)\right.$$
$$\left. \times \breve{B}_k(r, \{\lambda_i\}) \breve{B}_k(u, \{\lambda_i\})' \right] dudr \Lambda^{*\prime},$$
$$= \Lambda^* P_1^{ES}\left(b, \breve{B}_k(\{\lambda\}_1^{2C})\right) \Lambda^{*\prime},$$

completing the proof.

Lemma D9. *Let $M_{ES} = bT_{ES}$ where b is a fixed constant with $b \in [0,1]$. Under Assumption NR$'$, when $k(x)$ is continuous, $k(x) = 0$ for $|x| \leq 1$, and twice continuously differentiable everywhere except for $|x| = 1$, as $T \to \infty$,*

$$\hat{\Omega}^{ES} \Rightarrow \Lambda^* P_2^{ES}\left(b, \breve{B}_k(\{\lambda\}_1^{2C})\right)\Lambda^{*\prime}$$

where,

$$\begin{aligned}
&P_2^{ES}\left(b, \breve{B}_k(\{\lambda\}_1^{2C})\right) \equiv \\
&-\frac{1}{b^2\lambda^3}\sum_{n=0}^{C}\sum_{l=0}^{C}\int_{\lambda_{2n}}^{\lambda_{2n+1}}\int_{\lambda_{2l}}^{\lambda_{2l+1}}\left[\mathbb{1}\left\{|r-u| < b\lambda + \sum_{j=2(n\wedge l)+1}^{2(n\vee l)}(-1)^j\lambda_j\right\}\right.\\
&\times k''\left((\lambda b)^{-1}\left[\sum_{j=1}^{2n+1}(-1)^{j+1}(r\wedge\lambda_j) - \sum_{j=1}^{2l+1}(-1)^{j+1}(u\wedge\lambda_j)\right]\right)\\
&\quad\left.\times \breve{B}_k(r,\{\lambda_i\})\breve{B}_k(u,\{\lambda_i\})'\right] drdu\\
&+\frac{k'(1)_-}{b\lambda^2}\sum_{n=0}^{C}\sum_{l=0}^{n}\int_{\lambda_{2l}}^{\lambda_{2l+1}}\\
&\left[\mathbb{1}\left\{\lambda_{2n} - b\lambda - \sum_{j=2l+1}^{2n}(-1)^j\lambda_j < u \leq \lambda_{2n+1} - b\lambda - \sum_{j=2l+1}^{2n}(-1)^j\lambda_j\right\}\right.\\
&\times\left\{\breve{B}_k\left(u + b\lambda + \sum_{j=2l+1}^{2n}(-1)^j\lambda_j, \{\lambda_i\}\right)\breve{B}_k(u,\{\lambda_i\})'\right.\\
&\left.\left.+\breve{B}_k(u,\{\lambda\}_1^{2C})\breve{B}_k\left(u + b\lambda + \sum_{j=2l+1}^{2n}(-1)^j\lambda_j, \{\lambda_i\}\right)'\right\}\right] du.
\end{aligned}$$

Proof: Straightforward calculations give

$$\Delta^2 K^a_{ts} =$$
$$\begin{cases} -\frac{1}{T_{ES}^2} D_{T_{ES}}\left(T_{ES}^{-1}\left(\sum\limits_{i=1}^{t} a_i - \sum\limits_{i=1}^{s} a_i\right)\right) & \left|\sum^t a_i - \sum^s a_i\right| < [bT_{ES}] \\ \left[k\left(\frac{[bT_{ES}]}{bT_{ES}}\right) - k\left(\frac{[bT_{ES}]-1}{bT_{ES}}\right)\right] + k\left(\frac{[bT_{ES}]}{bT_{ES}}\right) & \left|\sum^t a_i - \sum^s a_i\right| = [bT_{ES}] \\ -k\left(\frac{[bT_{ES}]}{bT_{ES}}\right) & \left|\sum^t a_i - \sum^s a_i\right| = [bT_{ES}] + 1 \\ 0 & \text{otherwise} \end{cases}.$$

We rewrite $\hat{\Omega}^{ES}$ using Lemma D1 and dividing it into the three nonzero cases as determined by $\Delta^2 K^a_{ts}$:

$$\begin{aligned}
&\hat{\Omega}^{ES} \\
&= \frac{1}{T_{ES}} \sum_{t=1}^{T-1} \sum_{s=1}^{T-1} a_{t+1} a_{s+1} \hat{S}_t \Delta^2 K^a_{ts} \hat{S}_s' \\
&= -\frac{1}{T_{ES}} \sum_{t=1}^{T-1} \sum_{s=1}^{T-1} a_{t+1} a_{s+1} \mathbb{1}\left\{\left|\sum_{i=1}^{t} a_i - \sum_{i=1}^{s} a_i\right| < [bT_{ES}]\right\} \\
&\times \frac{1}{T_{ES}^2} D_{T_{ES}}\left(T_{ES}^{-1}\left(\sum_{i=1}^{t} a_i - \sum_{i=1}^{s} a_i\right)\right) \hat{S}_t \hat{S}_s' \\
&+ \frac{1}{T_{ES}} \sum_{t=1}^{T-1} \sum_{s=1}^{T-1} a_{t+1} a_{s+1} \mathbb{1}\left\{\left|\sum_{i=1}^{t} a_i - \sum_{i=1}^{s} a_i\right| = [bT_{ES}]\right\} \\
&\times \left[k\left(\frac{[bT_{ES}]}{bT_{ES}}\right) - k\left(\frac{[bT_{ES}]-1}{bT_{ES}}\right) + k\left(\frac{[bT_{ES}]}{bT_{ES}}\right)\right] \hat{S}_t \hat{S}_s' \\
&- \frac{1}{T_{ES}} \sum_{t=1}^{T-1} \sum_{s=1}^{T-1} a_{t+1} a_{s+1} \mathbb{1}\left\{\left|\sum_{i=1}^{t} a_i - \sum_{i=1}^{s} a_i\right| = [bT_{ES}] + 1\right\} k\left(\frac{[bT_{ES}]}{bT_{ES}}\right) \hat{S}_t \hat{S}_s'.
\end{aligned}$$

Expanding the second term gives

$$
\begin{aligned}
&\hat{\Omega}^{ES} \\
&= -\frac{1}{T_{ES}} \sum_{t=1}^{T-1} \sum_{s=1}^{T-1} a_{t+1} a_{s+1} \mathbb{1}\left\{\left|\sum_{i=1}^{t} a_i - \sum_{i=1}^{s} a_i\right| < [bT_{ES}]\right\} \\
&\times \frac{1}{T_{ES}^2} D_{T_{ES}}\left(T_{ES}^{-1}\left(\sum_{i=1}^{t} a_i - \sum_{i=1}^{s} a_i\right)\right) \hat{S}_t \hat{S}_s' \\
&+ \frac{1}{T_{ES}} \sum_{t=1}^{T-1} \sum_{s=1}^{T-1} a_{t+1} a_{s+1} \mathbb{1}\left\{\left|\sum_{i=1}^{t} a_i - \sum_{i=1}^{s} a_i\right| = [bT_{ES}]\right\} \\
&\times \left[k\left(\frac{[bT_{ES}]}{bT_{ES}}\right) - k\left(\frac{[bT_{ES}]-1}{bT_{ES}}\right)\right] \hat{S}_t \hat{S}_s' \\
&+ \frac{1}{T_{ES}} \sum_{t=1}^{T-1} \sum_{s=1}^{T-1} a_{t+1} a_{s+1} \mathbb{1}\left\{\left|\sum_{i=1}^{t} a_i - \sum_{i=1}^{s} a_i\right| = [bT_{ES}]\right\} k\left(\frac{[bT_{ES}]}{bT_{ES}}\right) \hat{S}_t \hat{S}_s' \\
&- \frac{1}{T_{ES}} \sum_{t=1}^{T-1} \sum_{s=1}^{T-1} a_{t+1} a_{s+1} \mathbb{1}\left\{\left|\sum_{i=1}^{t} a_i - \sum_{i=1}^{s} a_i\right| = [bT_{ES}]+1\right\} k\left(\frac{[bT_{ES}]}{bT_{ES}}\right) \hat{S}_t \hat{S}_s' \\
&= \zeta_1 + \zeta_2 + \zeta_3 + \zeta_4.
\end{aligned}
$$

First consider ζ_1. Plugging in the expression from Lemma D4 gives,

$$
\begin{aligned}
\zeta_1 &= -\frac{1}{T_{ES}} \sum_{t=1}^{T-1} \sum_{s=1}^{T-1} \sum_{n=0}^{C} \sum_{l=0}^{C} \Bigg[\mathbb{1}\left\{T_{2n} \le t \le T_{2n+1} - 1\right\} \mathbb{1}\left\{T_{2l} \le s \le T_{2l+1} - 1\right\} \\
&\times \mathbb{1}\left\{|t-s| < [bT_{ES}] + \sum_{k=2(n\wedge l)+1}^{2(n\vee l)} (-1)^k T_k\right\} \\
&\times \frac{1}{T_{ES}^2} D_{T_{ES}}\left(T_{ES}^{-1}\left(\sum_{i=1}^{t} a_i - \sum_{i=1}^{s} a_i\right)\right) \hat{S}_t \hat{S}_s'\Bigg] \\
&= -\left(\frac{T}{T_{ES}}\right)^3 \frac{1}{T} \sum_{t=1}^{T-1} \frac{1}{T} \sum_{s=1}^{T-1} \sum_{n=0}^{C} \sum_{l=0}^{C} \Bigg[\mathbb{1}\left\{T_{2n} \le t \le T_{2n+1} - 1\right\} \\
&\times \mathbb{1}\left\{T_{2l} \le s \le T_{2l+1} - 1\right\} \mathbb{1}\left\{|t-s| < [bT_{ES}] + \sum_{k=2(n\wedge l)+1}^{2(n\vee l)} (-1)^k T_k\right\} \\
&\times D_{T_{ES}}\left(T_{ES}^{-1}\left(\sum_{i=1}^{t} a_i - \sum_{i=1}^{s} a_i\right)\right) T^{-1/2} \hat{S}_t T^{-1/2} \hat{S}_s'\Bigg]
\end{aligned}
$$

where the second equality holds by rescaling. Next, consider ζ_2. We use the expression from Lemma D5 when $t > s$. When $s > t$, the expression is the same with t and s interchanged. When $t = s$, $\zeta_2 = 0$. Therefore we have

$$
\begin{aligned}
&\zeta_2 = \frac{1}{T_{ES}} \sum_{t=1}^{T-1} \sum_{s=1}^{T-1} \sum_{n=0}^{C} \sum_{l=0}^{n} \\
&\mathbb{1}\left\{ T_{2n} - [bT_{ES}] - \sum_{k=2l+1}^{2n} (-1)^k T_k \leq s \leq T_{2n+1} - 1 - [bT_{ES}] - \sum_{k=2l+1}^{2n} (-1)^k T_k \right\} \\
&\times \mathbb{1}\left\{ T_{2l} \leq s \leq T_{2l+1} - 1 \right\} \mathbb{1}\left\{ t = s + [bT_{ES}] + \sum_{k=2l+1}^{2n} (-1)^k T_k \right\} \\
&\times \left[k\left(\frac{[bT_{ES}]}{bT_{ES}} \right) - k\left(\frac{[bT_{ES}] - 1}{bT_{ES}} \right) \right] \left(\hat{S}_t \hat{S}_s' + \hat{S}_s \hat{S}_t' \right).
\end{aligned}
$$

We further simplify ζ_2 by plugging in $t = s + [bT_{ES}] + \sum_{k=2l+1}^{2n} (-1)^k T_k$ directly rather than denoting it as an indicator function. The double sum collapses to a single sum giving

$$\zeta_2 = \frac{1}{T_{ES}} \sum_{s=1}^{T-1} \sum_{n=0}^{C} \sum_{l=0}^{n}$$
$$\mathbb{1}\left\{ T_{2n} - [bT_{ES}] - \sum_{k=2l+1}^{2n} (-1)^k T_k \leq s \leq T_{2n+1} - 1 - [bT_{ES}] - \sum_{k=2l+1}^{2n} (-1)^k T_k \right\}$$
$$\times \mathbb{1}\left\{ T_{2l} \leq s \leq T_{2l+1} - 1 \right\} \left[k\left(\frac{[bT_{ES}]}{bT_{ES}}\right) - k\left(\frac{[bT_{ES}] - 1}{bT_{ES}}\right) \right]$$
$$\times \left(\hat{S}_{s+[bT_{ES}]+\sum_{k=2l+1}^{2n}(-1)^k T_k} \hat{S}_s' + \hat{S}_s \hat{S}'_{s+[bT_{ES}]+\sum_{k=2l+1}^{2n}(-1)^k T_k} \right)$$
$$= \left(\frac{T}{T_{ES}}\right)^2 \frac{1}{T} \sum_{s=1}^{T-1} \sum_{n=0}^{C} \sum_{l=0}^{n}$$
$$\mathbb{1}\left\{ T_{2n} - [bT_{ES}] - \sum_{k=2l+1}^{2n} (-1)^k T_k \leq s \leq T_{2n+1} - 1 - [bT_{ES}] - \sum_{k=2l+1}^{2n} (-1)^k T_k \right\}$$
$$\times \mathbb{1}\left\{ T_{2l} \leq s \leq T_{2l+1} - 1 \right\} T_{ES} \left[k\left(\frac{[bT_{ES}]}{bT_{ES}}\right) - k\left(\frac{[bT_{ES}] - 1}{bT_{ES}}\right) \right]$$
$$\times \left(T^{-1/2} \hat{S}_{s+[bT_{ES}]+\sum_{k=2l+1}^{2n}(-1)^k T_k} T^{-1/2} \hat{S}_s' \right.$$
$$\left. + T^{-1/2} \hat{S}_s T^{-1/2} \hat{S}'_{s+[bT_{ES}]+\sum_{k=2l+1}^{2n}(-1)^k T_k} \right).$$

Finally consider ζ_3 and ζ_4. Because $\frac{[bT_{ES}]+1}{bT_{ES}}$ is beyond the truncation point, it follows that $k\left(\frac{[bT_{ES}]+1}{bT_{ES}}\right) = 0$. Therefore, we have

$$k\left(\frac{[bT_{ES}]}{bT_{ES}}\right) = k\left(\frac{[bT_{ES}]}{bT_{ES}}\right) - k\left(\frac{[bT_{ES}] + 1}{bT_{ES}}\right),$$

and notice that

$$(bT_{ES}) k\left(\frac{[bT_{ES}]}{bT_{ES}}\right) = (bT_{ES}) \left[k\left(\frac{[bT_{ES}]}{bT_{ES}}\right) - k\left(\frac{[bT_{ES}] + 1}{bT_{ES}}\right) \right] \to k'_+(1) = 0.$$

We obtain zero because $k'_+(1)$ is the derivative from the right of the truncation point. Using similar arguments as used for ζ_2, it follows that $\zeta_3 = o_p(1)$ and $\zeta_4 = o_p(1)$ because

$k'_+(1) = 0$. Combining the results for $\zeta_1, \zeta_2, \zeta_3, \zeta_4$ allows us to write

$$\hat{\Omega}^{ES} = -\left(\frac{T}{T_{ES}}\right)^3 T^{-1} \sum_{t=1}^{T-1} T^{-1} \sum_{s=1}^{T-1} \sum_{n=0}^{C} \sum_{l=0}^{C}$$

$$\mathbb{1}\left\{T_{2n} \leq t \leq T_{2n+1} - 1\right\} \mathbb{1}\left\{T_{2l} \leq s \leq T_{2l+1} - 1\right\}$$

$$\times \mathbb{1}\left\{|t - s| < [bT_{ES}] + \sum_{k=2(l\wedge n)+1}^{2(n\vee l)} (-1)^k T_k\right\} D_{T_{ES}}\left(T_{ES}^{-1}\left(\sum_{i=1}^{t} a_i - \sum_{i=1}^{s} a_i\right)\right)$$

$$\times T^{-\frac{1}{2}} \hat{S}_t T^{-\frac{1}{2}} \hat{S}'_s$$

$$+ \left(\frac{T}{T_{ES}}\right)^2 \frac{1}{T} \sum_{s=1}^{T-1} \sum_{n=0}^{C} \sum_{l=0}^{n}$$

$$\mathbb{1}\left\{T_{2n} - [bT_{ES}] - \sum_{k=2l+1}^{2n} (-1)^k T_k \leq s \leq T_{2n+1} - 1 - [bT_{ES}] - \sum_{k=2l+1}^{2n} (-1)^k T_k\right\}$$

$$\times \mathbb{1}\left\{T_{2l} \leq s \leq T_{2l+1} - 1\right\} T_{ES}\left[k\left(\frac{[bT_{ES}]}{bT_{ES}}\right) - k\left(\frac{[bT_{ES}] - 1}{bT_{ES}}\right)\right]$$

$$\times \left(T^{-1/2} \hat{S}_{s+[bT_{ES}]+\sum_{k=2l+1}^{2n}(-1)^k T_k} T^{-1/2} \hat{S}_s'^{-\frac{1}{2}}\right.$$

$$\left. + T^{-1/2} \hat{S}_s T^{-1/2} \hat{S}'_{s+[bT_{ES}]+\sum_{k=2l+1}^{2n}(-1)^k T_k}\right) + o_p(1).$$

Using similar arguments as in the proof of Lemma D8 it follows that

$$\hat{\Omega}^{ES} = -\frac{1}{\lambda^3}\int_0^1\int_0^1 \sum_{l=0}^{C}\sum_{n=0}^{C} \mathbb{1}\left\{\lambda_{2n} < r < \lambda_{2n+1}\right\}\mathbb{1}\left\{\lambda_{2l} < u < \lambda_{2l+1}\right\}$$

$$\times \mathbb{1}\left\{|r-u| < b\lambda + \sum_{k=2(l\wedge n)+1}^{2(n\vee l)} (-1)^k \lambda_k\right\}$$

$$\times D_{T_{ES}}\left(T_{ES}^{-1}\left(\sum_{i=1}^{t} a_i - \sum_{i=1}^{s} a_i\right)\right) T^{-\frac{1}{2}}\hat{S}_{[rT]} T^{-\frac{1}{2}}\hat{S}'_{[uT]} du dr$$

$$+ \frac{1}{\lambda^2}\int_0^1 \sum_{n=0}^{C}\sum_{l=0}^{n} \mathbb{1}\left\{\lambda_{2l} < u < \lambda_{2l+1}\right\}$$

$$\times \mathbb{1}\left\{\lambda_{2n} - b\lambda - \sum_{k=2l+1}^{2n}(-1)^k\lambda_k < u < \lambda_{2n+1} - b\lambda - \sum_{k=2l+1}^{2n}(-1)^k\lambda_k\right\}$$

$$\times T_{ES}\left[k_b(b) - k_b\left(b - \frac{1}{T_{ES}}\right)\right]\left(T^{-1/2}\hat{S}_{\left[\left(u+b\lambda+\sum_{k=2l+1}^{2n}(-1)^k\lambda_k\right)T\right]} T^{-1/2}\hat{S}'_{[uT]}\right.$$

$$\left. + T^{-1/2}\hat{S}_{[uT]} T^{-1/2}\hat{S}'_{\left[\left(u+b\lambda+\sum_{k=2l+1}^{2n}(-1)^k\lambda_k\right)T\right]}\right) du$$

$$+ o_p(1).$$

Using $\lim_{T_{ES}\to\infty} T_{ES}\left[k_b(b) - k_b(b - 1/T_{ES})\right] = b^{-1}k'(1)_-$, Lemma D2, Lemma C4, (eqD.1), (eqD.6), the continuous mapping theorem, and the simplifications used in the proof of Lemma D8, it follows that

$$
\begin{aligned}
&\hat{\Omega}^{ES} \\
&\Rightarrow \Lambda^* \left[-\frac{1}{b^2\lambda^3} \sum_{l=0}^{C} \sum_{n=0}^{C} \int_{\lambda_{2n}}^{\lambda_{2n+1}} \int_{\lambda_{2l}}^{\lambda_{2l+1}} \right. \\
&\mathbb{1}\left\{ |r-u| < \left(b \sum_{l=0}^{2C+1} \lambda_l (-1)^{l+1} + \sum_{k=2(l\wedge n)+1}^{2(n\vee l)} (-1)^k \lambda_k \right) \right\} \\
&\times k'' \left((\lambda b)^{-1} \left(\sum_{j=1}^{2n+1} (-1)^{j+1} (r \wedge \lambda_j) - \sum_{j=1}^{2l+1} (-1)^{j+1} (u \wedge \lambda_j) \right) \right) \\
&\times \breve{B}_k(r, \{\lambda_i\}) \breve{B}_k(u, \{\lambda_i\})' dudr \\
&+ \frac{k'(1)_-}{b\lambda^2} \sum_{n=0}^{C} \sum_{l=0}^{n} \int_{\lambda_{2l}}^{\lambda_{2l+1}} \\
&\mathbb{1}\left\{ \lambda_{2n} - b\lambda - \sum_{k=2l+1}^{2n} (-1)^k \lambda_k < u < \lambda_{2n+1} - b\lambda - \sum_{k=2l+1}^{2n} (-1)^k \lambda_k \right\} \\
&\times \left\{ \breve{B}_k \left(u + b\lambda + \sum_{k=2l+1}^{2n} (-1)^k \lambda_k, \{\lambda_i\} \right) \breve{B}_k(u, \{\lambda_i\})' \right. \\
&\left. \left. + \breve{B}_k(u, \{\lambda_i\}) \breve{B}_k \left(u + b\lambda + \sum_{k=2l+1}^{2n} (-1)^k \lambda_k, \{\lambda_i\} \right)' \right\} du \right] \Lambda^{*\prime} \\
&\equiv \Lambda^* P_2^{ES} \left(b, \breve{B}_k(\{\lambda_i\}) \right) \Lambda^{*\prime}.
\end{aligned}
$$

Lemma D10. *Let $M_{ES} = bT_{ES}$ where b is a fixed constant with $b \in [0,1]$. Under Assumptions NR$'$, when $k(x)$ is the Bartlett kernel, as $T \to \infty$,*

$$
\hat{\Omega}^{ES} \Rightarrow \Lambda^* P_3^{ES} \left(b, \breve{B}_k(\{\lambda_i\}) \right) \Lambda^{*\prime}
$$

where,

$$P_3^{ES}\left(b,\breve{B}_k(\{\lambda_i\})\right)=\frac{2}{b}\frac{1}{\lambda^2}\sum_{n=0}^{C}\int_{\lambda_{2n}}^{\lambda_{2n+1}}\breve{B}_k\left(r,\{\lambda_i\}\right)\breve{B}_k\left(r,\{\lambda_i\}\right)'dr$$

$$-\frac{1}{b}\frac{1}{\lambda^2}\sum_{n=0}^{C}\sum_{l=0}^{n}\int_{\lambda_{2l}}^{\lambda_{2l+1}}$$

$$\left[\mathbb{1}\left\{\lambda_{2n}-b\lambda-\sum_{k=2l+1}^{2n}(-1)^k\lambda_k\le u\le\lambda_{2n+1}-b\lambda-\sum_{k=2l+1}^{2n}(-1)^k\lambda_k\right\}\right.$$

$$\left\{\breve{B}_k\left(u,\{\lambda_i\}\right)\breve{B}_k\left(u+b\lambda+\sum_{k=2l+1}^{2n}\lambda_k(-1)^k,\{\lambda_i\}\right)'\right.$$

$$\left.\left.+\breve{B}_k\left(u+b\lambda+\sum_{k=2l+1}^{2n}\lambda_k(-1)^k,\{\lambda_i\}\right)\breve{B}_k\left(r,\{\lambda_i\}\right)'\right\}\right]du.$$

Proof: Using straightforward algebra we have

$$\Delta^2K_{ts}^a=\begin{cases}\frac{2}{bT_{ES}} & \left|\sum^t a_i-\sum^s a_i\right|=0\\ -\frac{1}{bT_{ES}}+1-\frac{[bT_{ES}]}{bT_{ES}} & \left|\sum^t a_i-\sum^s a_i\right|=[bT_{ES}]\\ -\left(1-\frac{[bT_{ES}]}{bT_{ES}}\right) & \left|\sum^t a_i-\sum^s a_i\right|=[bT_{ES}]+1\\ 0 & \text{otherwise}\end{cases}.$$

We rewrite $\hat{\Omega}^{ES}$ using Lemma D1 and dividing it into the three nonzero cases as determined by $\Delta^2K_{ts}^a$ while expanding the second term into two parts giving

$$\hat{\Omega}^{ES}=\frac{1}{T_{ES}}\sum_{t=1}^{T-1}\sum_{s=1}^{T-1}a_{t+1}a_{s+1}\mathbb{1}\left\{\left|\sum^t a_i-\sum^s a_i\right|=0\right\}\frac{2}{bT_{ES}}\hat{S}_t\hat{S}_s'$$

$$-\frac{1}{T_{ES}}\sum_{t=1}^{T-1}\sum_{s=1}^{T-1}a_{t+1}a_{s+1}\mathbb{1}\left\{\left|\sum^t a_i-\sum^s a_i\right|=[bT_{ES}]\right\}\frac{1}{bT_{ES}}\hat{S}_t\hat{S}_s'$$

$$+\frac{1}{T_{ES}}\sum_{t=1}^{T-1}\sum_{s=1}^{T-1}a_{t+1}a_{s+1}\mathbb{1}\left\{\left|\sum^t a_i-\sum^s a_i\right|=[bT_{ES}]\right\}\left(1-\frac{[bT_{ES}]}{bT_{ES}}\right)\hat{S}_t\hat{S}_s'$$

$$-\frac{1}{T_{ES}}\sum_{t=1}^{T-1}\sum_{s=1}^{T-1}a_{t+1}a_{s+1}\mathbb{1}\left\{\left|\sum^t a_i-\sum^s a_i\right|=[bT_{ES}]+1\right\}\left(1-\frac{[bT_{ES}]}{bT_{ES}}\right)\hat{S}_t\hat{S}_s'.$$

Using similar arguments as used in the proof of Lemma D9 for ζ_3 and ζ_4, it is easy to show that the third and fourth terms are $o_p(1)$. Therefore, we have

$$\begin{aligned}\hat{\Omega}^{ES} =& \frac{1}{T_{ES}} \sum_{t=1}^{T-1} \sum_{s=1}^{T-1} a_{t+1} a_{s+1} \mathbb{1} \left\{ \left| \sum^{t} a_i - \sum^{s} a_i \right| = 0 \right\} \frac{2}{bT_{ES}} \hat{S}_t \hat{S}_s' \\ &- \frac{1}{T_{ES}} \sum_{t=1}^{T-1} \sum_{s=1}^{T-1} a_{t+1} a_{s+1} \mathbb{1} \left\{ \left| \sum^{t} a_i - \sum^{s} a_i \right| = [bT_{ES}] \right\} \frac{1}{bT_{ES}} \hat{S}_t \hat{S}_s' + o_p(1)\end{aligned}$$

Using Lemmas D5 and D7 we can write

$$\begin{aligned}\hat{\Omega}^{ES} =& \frac{1}{T_{ES}} \sum_{t=1}^{T-1} \sum_{s=1}^{T-1} \sum_{n=0}^{C} \mathbb{1} \left\{ T_{2n} \leq t \leq T_{2n+1} - 1 \right\} \mathbb{1} \left\{ t = s \right\} \frac{2}{bT_{ES}} \hat{S}_t \hat{S}_s' \\ &- \frac{1}{T_{ES}} \sum_{t=1}^{T-1} \sum_{s=1}^{T-1} \sum_{n=0}^{C} \sum_{l=0}^{n} \\ &\mathbb{1} \left\{ T_{2n} - [bT_{ES}] - \sum_{k=2l+1}^{2n} (-1)^k T_k \leq s \leq T_{2n+1} - 1 - [bT_{ES}] - \sum_{k=2l+1}^{2n} (-1)^k T_k \right\} \\ &\times \mathbb{1} \left\{ T_{2l} \leq s \leq T_{2l+1} - 1 \right\} \mathbb{1} \left\{ t = s + [bT_{ES}] + \sum_{k=2l+1}^{2n} (-1)^k T_k \right\} \\ &\times \frac{1}{bT_{ES}} \left(\hat{S}_t \hat{S}_s' + \hat{S}_s \hat{S}_t' \right) + o_p(1).\end{aligned}$$

We can simplify $\hat{\Omega}^{ES}$ by plugging in $t = s$ and $t = s + [bT_{ES}] + \sum_{k=2l+1}^{2n} (-1)^k T_k$ into the first and second terms respectively instead of using the indicator functions to give

$$\begin{aligned}\hat{\Omega}^{ES} =& \frac{1}{T_{ES}} \sum_{t=1}^{T-1} \sum_{n=0}^{C} \mathbb{1} \left\{ T_{2n} \leq t \leq T_{2n+1} - 1 \right\} \frac{2}{bT_{ES}} \hat{S}_t \hat{S}_t' \\ &- \frac{1}{T_{ES}} \sum_{s=1}^{T-1} \sum_{n=0}^{C} \sum_{l=0}^{n} \\ &\left[\mathbb{1} \left\{ T_{2n} - [bT_{ES}] - \sum_{k=2l+1}^{2n} (-1)^k T_k \leq s \leq T_{2n+1} - 1 - [bT_{ES}] - \sum_{k=2l+1}^{2n} (-1)^k T_k \right\} \right. \\ &\times \mathbb{1} \left\{ T_{2l} \leq s \leq T_{2l+1} - 1 \right\} \frac{1}{bT_{ES}} \left(\hat{S}_{s+[bT_{ES}]+\sum_{k=2l+1}^{2n}(-1)^k T_k} \hat{S}_s' \right. \\ &\left. \left. + \hat{S}_s \hat{S}'_{s+[bT_{ES}]+\sum_{k=2l+1}^{2n}(-1)^k T_k} \right) \right] + o_p(1)\end{aligned}$$

$$
\begin{aligned}
&= \frac{2}{b}\left(\frac{T}{T_{ES}}\right)^2 T^{-1}\sum_{t=1}^{T-1}\sum_{n=0}^{C}\mathbb{1}\left\{T_{2n}\le t\le T_{2n+1}-1\right\}T^{-\frac{1}{2}}\hat{S}_t\hat{S}_t'\\
&-\frac{1}{b}\left(\frac{T}{T_{ES}}\right)^2 T^{-1}\sum_{s=1}^{T-1}\sum_{n=0}^{C}\sum_{l=0}^{n}\\
&\left[\mathbb{1}\left\{T_{2n}-[bT_{ES}]-\sum_{k=2l+1}^{2n}(-1)^k T_k\le s\le T_{2n+1}-1-[bT_{ES}]-\sum_{k=2l+1}^{2n}(-1)^k T_k\right\}\right.\\
&\times\mathbb{1}\left\{T_{2l}\le s\le T_{2l+1}-1\right\}\left(T^{-\frac{1}{2}}\hat{S}_{s+[bT_{ES}]+\sum_{k=2l+1}^{2n}(-1)^k T_k}T^{-\frac{1}{2}}\hat{S}_s'\right.\\
&\left.\left.+T^{-\frac{1}{2}}\hat{S}_sT^{-\frac{1}{2}}\hat{S}'_{s+[bT_{ES}]+\sum_{k=2l+1}^{2n}(-1)^k T_k}\right)\right]+op(1),\\
&= \frac{2}{b}\left(\frac{T}{T_{ES}}\right)^2\int_0^1\sum_{n=0}^{C}\mathbb{1}\left\{\lambda_{2n}\le r<\lambda_{2n+1}\right\}T^{-\frac{1}{2}}\hat{S}_{[rT]}T^{-\frac{1}{2}}\hat{S}'_{[rT]}dr\\
&-\frac{1}{b}\left(\frac{T}{T_{ES}}\right)^2\int_0^1\sum_{n=0}^{C}\sum_{l=0}^{n}\\
&\left[\mathbb{1}\left\{\lambda_{2n}-b\lambda-\sum_{k=2l+1}^{2n}(-1)^k\lambda_k\le u\le\lambda_{2n+1}-b\lambda-\sum_{k=2l+1}^{2n}(-1)^k\lambda_k\right\}\right.\\
&\times\mathbb{1}\left\{\lambda_{2l}\le u<\lambda_{2l+1}\right\}\left(T^{-\frac{1}{2}}\hat{S}_{[(u+b\lambda+\sum_{k=2l+1}^{2n}(-1)^k\lambda_k)T]}T^{-\frac{1}{2}}\hat{S}'_{[uT]}\right.\\
&\left.\left.+T^{-\frac{1}{2}}\hat{S}_{[uT]}T^{-\frac{1}{2}}\hat{S}'_{[(u+b\lambda+\sum_{k=2l+1}^{2n}(-1)^k\lambda_k)T]}\right)\right]du
\end{aligned}
$$

Further simplifications can be obtained by denoting the indicator functions as the ranges of the integrals:

$$
\begin{aligned}
\hat{\Omega}^{ES} &= \frac{2}{b}\left(\frac{T}{T_{ES}}\right)^2 \sum_{n=0}^{C} \int_{\lambda_{2n}}^{\lambda_{2n+1}} T^{-\frac{1}{2}} \hat{S}_{[rT]} T^{-\frac{1}{2}} \hat{S}'_{[rT]} dr \\
&- \frac{1}{b}\left(\frac{T}{T_{ES}}\right)^2 \sum_{n=0}^{C} \sum_{l=0}^{n} \int_{\lambda_{2l}}^{\lambda_{2l+1}} \Bigg[\\
&\mathbb{1}\left\{\lambda_{2n} - b\lambda - \sum_{k=2l+1}^{2n} (-1)^k \lambda_k \le u \le \lambda_{2n+1} - b\lambda - \sum_{k=2l+1}^{2n} (-1)^k \lambda_k\right\} \\
&\times \Bigg(T^{-\frac{1}{2}} \hat{S}_{[(u+b\lambda+\sum_{k=2l+1}^{2n}(-1)^k\lambda_k)T]} T^{-\frac{1}{2}} \hat{S}'_{[uT]} \\
&+ T^{-\frac{1}{2}} \hat{S}_{[uT]} T^{-\frac{1}{2}} \hat{S}'_{[(u+b\lambda+\sum_{k=2l+1}^{2n}(-1)^k\lambda_k)T]} \Bigg) \Bigg] du \\
&+ o_p(1)
\end{aligned}
$$

Then, by Lemma C4 and the continuous mapping theorem,

$$
\begin{aligned}
&\hat{\Omega}^{ES} \Rightarrow \Lambda^* \Bigg\{ \\
&\frac{2}{b}\frac{1}{\lambda^2} \sum_{n=0}^{C} \int_{\lambda_{2n}}^{\lambda_{2n+1}} \breve{B}_k\left(r, \{\lambda_i\}\right) \breve{B}_k\left(r, \{\lambda_i\}\right)' dr - \frac{1}{b}\frac{1}{\lambda^2} \sum_{n=0}^{C} \sum_{l=0}^{n} \int_{\lambda_{2l}}^{\lambda_{2l+1}} \Bigg[\\
&\mathbb{1}\left\{\lambda_{2n} - b\lambda - \sum_{k=2l+1}^{2n} (-1)^k \lambda_k \le u \le \lambda_{2n+1} - b\lambda - \sum_{k=2l+1}^{2n} (-1)^k \lambda_k\right\} \\
&\times \Bigg\{ \breve{B}_k\left(u, \{\lambda_i\}\right) \breve{B}_k\left(u + b\lambda + \sum_{k=2l+1}^{2n} \lambda_k (-1)^k, \{\lambda_i\}\right)' \\
&+ \breve{B}_k\left(u + b\lambda + \sum_{k=2l+1}^{2n} \lambda_k (-1)^k, \{\lambda_i\}\right) \breve{B}_k\left(u, \{\lambda_i\}\right)' \Bigg\} \Bigg] du \Bigg\} \Lambda^{*\prime} \\
&= \Lambda^* P_3^{ES}\left(b, \breve{B}_k(\{\lambda_i\})\right) \Lambda^{*\prime}.
\end{aligned}
$$

Proof of Theorem 2.5 (a): Theorem 2.5 (a) directly follows from Lemmas D8-D10.

Proof of Theorem 2.5 (b): Recall that the null hypothesis is $H_0 : r(\beta_0) = 0$ with q

restrictions. The Wald statistic is defined as

$$W_T^{ES} = r\left(\hat{\beta}^{ES}\right)' \left[R\left(\hat{\beta}\right)\hat{V}_{ES}R\left(\hat{\beta}^{ES}\right)'\right]^{-1} r\left(\hat{\beta}^{ES}\right),$$

where

$$\hat{V}_{ES} = T_{ES}\left(\sum_{t=1}^{T_{ES}} x_t^{ES}x_t^{ES\prime}\right)^{-1}\hat{\Omega}^{ES}\left(\sum_{t=1}^{T_{ES}} x_t^{ES}x_t^{ES\prime}\right)^{-1}.$$

Using $\hat{\beta}^{ES} = \widehat{\beta}$ and $\sum_{t=1}^{T_{ES}} x_t^{ES}x_t^{ES\prime} = \sum_{t=1}^{T} x_t x_t'$, we can write

$$\begin{aligned}
&W_T^{ES} \\
&= r\left(\hat{\beta}^{ES}\right)'\left[R\left(\hat{\beta}^{ES}\right)\hat{V}_{ES}R\left(\hat{\beta}^{ES}\right)'\right]^{-1} r\left(\hat{\beta}^{ES}\right) \\
&= r\left(\hat{\beta}^{ES}\right)'\left[R\left(\hat{\beta}^{ES}\right)T_{ES}\left(\sum_{t=1}^{T_{ES}} x_t^{ES}x_t^{ES\prime}\right)^{-1}\hat{\Omega}^{ES}\left(\sum_{t=1}^{T_{ES}} x_t^{ES}x_t^{ES\prime}\right)^{-1} R\left(\hat{\beta}^{ES}\right)'\right]^{-1} \\
&\times r\left(\hat{\beta}^{ES}\right) \\
&= \sqrt{T}r\left(\hat{\beta}\right)'\left[\frac{T_{ES}}{T}R\left(\hat{\beta}\right)\left(T^{-1}\sum_{t=1}^{T} x_t x_t'\right)^{-1}\hat{\Omega}^{ES}\left(T^{-1}\sum_{t=1}^{T} x_t x_t'\right)^{-1} R\left(\hat{\beta}\right)'\right]^{-1} \\
&\times \sqrt{T}r\left(\hat{\beta}\right).
\end{aligned}$$

From the proof of Theorem 2.3 (a), we know that

$$\sqrt{T}r(\hat{\beta}) \Rightarrow R\left(\beta_0\right)\left(\lambda Q^*\right)^{-1}\Lambda^*\overline{\mathcal{W}}_k.$$

There exists a $q \times q$ matrix Δ^* such that

$$\Delta^*\Delta^{*\prime} = R\left(\beta_0\right)\left(\lambda Q^*\right)^{-1}\Omega^*\left(\lambda Q^*\right)^{-1}R\left(\beta_0\right)',$$

and it follows that

$$R\left(\beta_0\right)\left(\lambda Q^*\right)^{-1}\Lambda^*\overline{\mathcal{W}}_k = \Delta^*\overline{\mathcal{W}}_k.$$

Using this result and Lemmas D8- D10 (depending on the type of kernel) it follows that

$$
\begin{aligned}
W_T \Rightarrow & \left[R(\beta_0)\left(\lambda_j Q^*\right)^{-1}\Lambda^*\overline{\mathcal{W}}_k\right]' \\
& \times \left[\lambda R\left(\beta_0\right)\left(\lambda Q^*\right)^{-1}\Lambda^* P^{ES}\left(b, \breve{B}_k(\{\lambda_i\})\right)\Lambda^{*\prime}\left(\lambda Q^*\right)^{-1} R\left(\beta_0\right)'\right]^{-1} \\
& \times \left[R(\beta_0)\left(\lambda Q^*\right)^{-1}\Lambda^*\overline{\mathcal{W}}_k\right] \\
& = \left(\Delta^*\overline{\mathcal{W}}_q\right)'\left[\lambda\Delta^* P^{ES}\left(b, \breve{B}_q(\{\lambda_i\})\right)\Delta^{*\prime}\right]^{-1}\left(\Delta^*\overline{\mathcal{W}}_q\right) \\
& = \overline{\mathcal{W}}_q'\left[\lambda P^{ES}\left(b, \breve{B}_q(\{\lambda_i\})\right)\right]^{-1}\overline{\mathcal{W}}_q.
\end{aligned}
$$

For the special case of $q = 1$, we have the following limit for the t-statistic:

$$
t_T \Rightarrow \frac{\overline{\mathcal{W}}_1}{\sqrt{\lambda P^{ES}\left(b, \breve{B}_1(\{\lambda_i\})\right)}}.
$$

Note that the particular form of $P^{ES}\left(b, \breve{B}_q(\{\lambda_i\})\right)$ is given by Lemmas D8- D10 depending on the form of kernel.

Appendix E

PROOFS FOR FIXED-G, LARGE-n_G CASE WHEN G EVENLY DIVIDES T

Proof of Theorem 3.2 (a) (Asymptotic Limit of OLS): Plugging in $n_G = T/G$ to (3.2) in Section 3.2, we can write

$$\hat{\beta} - \beta = \left(\sum_{g=1}^{G} \sum_{t=(g-1)n_G+1}^{gn_G} x_t x_t' \right)^{-1} \sum_{g=1}^{G} \sum_{t=(g-1)n_G+1}^{gn_G} v_t$$

$$= \left(\sum_{g=1}^{G} \sum_{t=(\frac{g-1}{G})T+1}^{(\frac{g}{G})T} x_t x_t' \right)^{-1} \sum_{g=1}^{G} \sum_{t=(\frac{g-1}{G})T+1}^{(\frac{g}{G})T} v_t.$$

It directly follows that

$$\sqrt{T}\left(\hat{\beta} - \beta\right) = \left(\sum_{g=1}^{G} T^{-1} \sum_{t=(\frac{g-1}{G})T+1}^{(\frac{g}{G})T} x_t x_t' \right)^{-1} \sum_{g=1}^{G} T^{-\frac{1}{2}} \sum_{t=(\frac{g-1}{G})T+1}^{(\frac{g}{G})T} v_t. \quad \text{(eqE.1)}$$

Assumption B implies

$$T^{-1} \sum_{t=1}^{(\frac{g}{G}T)} x_t x_t' \Rightarrow \frac{g}{G} Q,$$

and

$$T^{-1/2} \sum_{t=1}^{(\frac{g}{G}T)} v_t \Rightarrow \Lambda \mathcal{W}_k(\tfrac{g}{G}).$$

Therefore from (eqE.1) it follows that

$$\sqrt{T}\left(\hat{\beta} - \beta\right) \Rightarrow \left(\sum_{g=1}^{G} \left(\frac{g}{G} - \frac{g-1}{G} \right) Q \right)^{-1} \sum_{g=1}^{G} \Lambda \left(\mathcal{W}_k\left(\frac{g}{G}\right) - \mathcal{W}_k\left(\frac{g-1}{G}\right) \right)$$

$$= \left(\sum_{g=1}^{G} \frac{Q}{G} \right)^{-1} \Lambda \mathcal{W}_k(1)$$

$$= Q^{-1} \Lambda \mathcal{W}_k(1).$$

Define $\widehat{\widehat{S}}_g = \sum_{j=1}^{g} \widehat{\widehat{v}}_j$. We now establish the following lemma about $\widehat{\widehat{S}}_g$.

Lemma E2. *Let $\widetilde{\mathcal{W}}_k(r) = \mathcal{W}_k(r) - r\mathcal{W}_k(1)$. Under Assumption B as $T \to \infty$,*

$$T^{-1/2}\widehat{\widehat{S}}_g \Rightarrow \Lambda \widetilde{\mathcal{W}}_k\left(\frac{g}{G}\right).$$

Proof: Plugging in $n_G = {}^{T}\!/_{G}$, gives

$$\begin{aligned} T^{-1/2}\widehat{\widehat{S}}_g &= T^{-1/2}\sum_{j=1}^{g}\widehat{\widehat{v}}_j = T^{-1/2}\sum_{j=1}^{g}\sum_{t=(j-1)n_G+1}^{jn_G}\hat{v}_t \\ &= T^{-1/2}\sum_{j=1}^{g}\sum_{t=(\frac{j-1}{G})T+1}^{(\frac{j}{G})T}\hat{v}_t \end{aligned} \tag{eqE.2}$$

Note that $\hat{v}_t = x_t(y_t - x_t'\hat{\beta}) = v_t - x_t x_t'(\hat{\beta} - \beta)$. Thus, we can write

$$\begin{aligned} T^{-1/2}\widehat{\widehat{S}}_g &= T^{-1/2}\sum_{j=1}^{g}\sum_{t=(\frac{j-1}{G})T+1}^{(\frac{j}{G})T}\left(v_t - x_t x_t'\left(\hat{\beta} - \beta\right)\right) \\ &= \sum_{j=1}^{g}\left(T^{-1/2}\sum_{t=(\frac{j-1}{G})T+1}^{(\frac{j}{G})T}v_t - T^{-1}\sum_{t=(\frac{j-1}{G})T+1}^{(\frac{j}{G})T}x_t x_t'\sqrt{T}\left(\hat{\beta} - \beta\right)\right). \end{aligned}$$

Assumption B implies that

$$T^{-1/2}\sum_{t=\left(\frac{j-1}{G}\right)T+1}^{\left(\frac{j}{G}\right)T} v_t \Rightarrow \Lambda\left(\mathcal{W}_k\left(\frac{j}{G}\right) - \mathcal{W}_k\left(\frac{j-1}{G}\right)\right),$$

and

$$T^{-1}\sum_{t=\left(\frac{j-1}{G}\right)T+1}^{\left(\frac{j}{G}\right)T} x_t x_t' \Rightarrow \left(\frac{j}{G} - \frac{j-1}{G}\right)Q.$$

From Theorem 3.2 (a), we know that

$$\sqrt{T}(\hat{\beta}-\beta) \Rightarrow Q^{-1}\Lambda \mathcal{W}_k(1).$$

Therefore from (eqE.2),

$$\begin{aligned}
T^{-1/2}\widehat{\widehat{S}}_g &\Rightarrow \sum_{j=1}^{g}\left\{\Lambda\left(\mathcal{W}_k\left(\frac{j}{G}\right)-\mathcal{W}_k\left(\frac{j-1}{G}\right)\right)-\left(\frac{j}{G}-\frac{j-1}{G}\right)QQ^{-1}\Lambda\mathcal{W}_k(1)\right\}, \\
&= \Lambda\left(\mathcal{W}_k\left(\frac{g}{G}\right)-\frac{g}{G}\mathcal{W}_k(1)\right), \\
&\equiv \Lambda\widetilde{\mathcal{W}}_k\left(\frac{g}{G}\right).
\end{aligned}$$

Proof of Theorem 3.2 (b) (Asymptotic Distribution of CHAC): Define

$$\begin{aligned}
&P(G,M,\widetilde{\mathcal{W}}_k) = \\
&\sum_{g=1}^{G-1}\sum_{h=1}^{G-1}\widetilde{\mathcal{W}}_k\left(\frac{g}{G}\right)\left(2k\left(\frac{|g-h|}{M}\right)-k\left(\frac{|g-h+1|}{M}\right)-k\left(\frac{|g-h-1|}{M}\right)\right)\widetilde{\mathcal{W}}_k\left(\frac{h}{G}\right)'.
\end{aligned}$$

Using summation by parts we can rewrite $\frac{G}{T}\widehat{\widehat{\Omega}}$ as

$$\begin{aligned}
&\frac{G}{T}\widehat{\widehat{\Omega}} \\
&= \sum_{g=1}^{G-1}\sum_{h=1}^{G-1}T^{-1/2}\widehat{\widehat{S}}_g\left[2k\left(\frac{|g-h|}{M}\right)-k\left(\frac{|g-h+1|}{M}\right)-k\left(\frac{|g-h-1|}{M}\right)\right]T^{-1/2}\widehat{\widehat{S}}_h' \\
&\Rightarrow \Lambda\left[\sum_{g=1}^{G-1}\sum_{h=1}^{G-1}\widetilde{\mathcal{W}}_k\left(\frac{g}{G}\right)\left(2k\left(\frac{|g-h|}{M}\right)-k\left(\frac{|g-h+1|}{M}\right)-k\left(\frac{|g-h-1|}{M}\right)\right)\right. \\
&\left.\times\widetilde{\mathcal{W}}_k\left(\frac{h}{G}\right)'\right]\Lambda' \\
&\equiv \Lambda P(G,M,\widetilde{\mathcal{W}}_k)\Lambda'
\end{aligned}$$

Weak convergence follows from Lemma E2. The above expression is valid for any kernel but in the case of the Bartlett kernel we can further simplify $P(G,M,\widetilde{\mathcal{W}}_k)$ as follows. Note that for the Bartlett kernel we have

$$2k\left(\frac{|g-h|}{M}\right)-k\left(\frac{|g-h+1|}{M}\right)-k\left(\frac{|g-h-1|}{M}\right)=\begin{cases}\frac{2}{M} & \text{when } g=h\\ -\frac{1}{M} & \text{when } |g-h|=M\\ 0 & \text{otherwise.}\end{cases} \cdot$$

Therefore

$$P(G,M,\widetilde{\mathcal{W}}_k)=\frac{2}{M}\sum_{g=1}^{G-1}\widetilde{\mathcal{W}}_k\left(\frac{g}{G}\right)\widetilde{\mathcal{W}}_k\left(\frac{g}{G}\right)'$$
$$-\frac{1}{M}\sum_{g=1}^{G-M-1}\left(\widetilde{\mathcal{W}}_k\left(\frac{g}{G}\right)\widetilde{\mathcal{W}}_k\left(\frac{g+M}{G}\right)'+\widetilde{\mathcal{W}}_k\left(\frac{g+M}{G}\right)\widetilde{\mathcal{W}}_k\left(\frac{g}{G}\right)'\right).$$

Proof of Theorem 3.2 (c) (Asymptotic Distribution of W_{CHAC} and t_{CHAC}) :Using the definition of W_{CHAC} and Theorem 3.2 (a,b) it follows from standard calculations that

$$\begin{aligned}W_{CHAC}&=(R\hat{\beta}-r)'\left[R\widehat{V}_{CHAC}R'\right]^{-1}(R\hat{\beta}-r)\\
&=\sqrt{T}\,(R\hat{\beta}-r)'\left[RT\widehat{V}_{CHAC}R'\right]^{-1}\sqrt{T}\,(R\hat{\beta}-r)\\
&=\left[R\sqrt{T}\,(\hat{\beta}-\beta)\right]'\left[R\left(\frac{1}{T}\sum_{t=1}^{T}x_tx_t'\right)^{-1}\frac{G}{T}\widehat{\widehat{\Omega}}\left(\frac{1}{T}\sum_{t=1}^{T}x_tx_t'\right)^{-1}R'\right]^{-1}\\
&\times R\sqrt{T}\,(\hat{\beta}-\beta)\\
&\Rightarrow\left[RQ^{-1}\Lambda\mathcal{W}_k(1)\right]'\left[RQ^{-1}\Lambda P(G,M,\widetilde{\mathcal{W}}_k)\Lambda'^{-1}R\right]^{-1}RQ^{-1}\Lambda\mathcal{W}_k(1).\end{aligned}$$

(eqE.3)

There exists a $q\times q$ matrix Δ^* such that

$$\Delta^*\Delta^{*\prime}=RQ^{-1}\Omega Q^{-1}R',$$

and it follows that

$$RQ^{-1}\Lambda\mathcal{W}_k=\Delta^*\mathcal{W}_q.$$

Then from (eqE.3),

$$
\begin{aligned}
W_{CHAC} &\Rightarrow \left[RQ^{-1}\Lambda\mathcal{W}_k(1)\right]' \left[RQ^{-1}\Lambda P(G,M,\widetilde{\mathcal{W}}_k)\Lambda'^{-1}R\right]^{-1} RQ^{-1}\Lambda\mathcal{W}_k(1) \\
&= \mathcal{W}_q'(1)\Delta^{*\prime}[\Delta^* P(G,M,\widetilde{\mathcal{W}}_q)\Delta'^*]^{-1}\Delta^*\mathcal{W}_q(1) \\
&= \mathcal{W}_q'(1)P(G,M,\widetilde{\mathcal{W}}_q)^{-1}\mathcal{W}_q(1)
\end{aligned}
$$

When $q = 1$,

$$
t_{CHAC} \Rightarrow \frac{\mathcal{W}_1(1)}{\sqrt{P(G,M,\widetilde{\mathcal{W}}_1)}}.
$$

Appendix F

PROOFS FOR FIXED-G, LARGE-n_G CASE WHEN THE NUMBER OF OBSERVATIONS ARE NOT THE EXACT MULTIPLE OF G

In this appendix we obtain the fixed-G limits for the case where the number of clusters does not evenly divide the sample. Suppose that there are n_G observations in first $G-1$ clusters and $n_l \leq n_G$ observations in the last cluster. Hence it follows that $T = n_G(G-1) + n_l$. Assume that $\frac{n_l}{T} \to \lambda_l$ as $T \to \infty$.

Asymptotic Limit of OLS: First notice that $\hat{\beta}$ can be rewritten as

$$\hat{\beta} = \left(\sum_{g=1}^{G-1} \sum_{t=(g-1)n_G+1}^{gn_G} x_t x_t' + \sum_{t=T-n_l+1}^{T} x_t x_t' \right)^{-1} \times \left(\sum_{g=1}^{G-1} \sum_{t=(g-1)n_G+1}^{gn_G} x_t y_t + \sum_{t=T-n_l+1}^{T} x_t y_t \right).$$

Also, note that

$$T = n_G(G-1) + n_l \Leftrightarrow n_G = \frac{T - n_l}{G-1} \Leftrightarrow \frac{1}{G-1}\left(1 - \frac{n_l}{T}\right) T. \qquad \text{(eqF.1)}$$

Then with (eqF.1) and Assumption B,

$$
\begin{aligned}
&\sqrt{T}\left(\hat{\beta}-\beta\right) \\
&= \left(\sum_{g=1}^{G-1} T^{-1} \sum_{t=(g-1)n_G+1}^{gn_G} x_t x_t' + T^{-1} \sum_{t=T-n_l+1}^{T} x_t x_t'\right)^{-1} \\
&\times \left(\sum_{g=1}^{G-1} T^{-1/2} \sum_{t=(g-1)n_G+1}^{gn_G} v_t + T^{-1/2} \sum_{t=T-n_l+1}^{T} v_t\right) \\
&= \left(\sum_{g=1}^{G-1} T^{-1} \sum_{t=\frac{g-1}{G-1}(1-\frac{n_l}{T})T+1}^{\frac{g}{G-1}(1-\frac{n_l}{T})T} x_t x_t' + T^{-1} \sum_{t=(1-\frac{n_l}{T})T+1}^{T} x_t x_t'\right)^{-1} \\
&\times \left(\sum_{g=1}^{G-1} T^{-1/2} \sum_{t=\frac{g-1}{G-1}(1-\frac{n_l}{T})T+1}^{\frac{g}{G-1}(1-\frac{n_l}{T})T} v_t + T^{-1/2} \sum_{t=(1-\frac{n_l}{T})T+1}^{T} v_t\right) \\
&\Rightarrow \left(\sum_{g=1}^{G-1}\left(\frac{g}{G-1}-\frac{g-1}{G-1}\right)(1-\lambda_l)Q + \left(1-(1-\lambda_l)\right)Q\right)^{-1} \\
&\times \Lambda\left[\sum_{g=1}^{G-1}\left(\mathcal{W}_k\left(\frac{g(1-\lambda_l)}{G-1}\right) - \mathcal{W}_k\left(\frac{(g-1)(1-\lambda_l)}{G-1}\right)\right) + \mathcal{W}_k(1) - \mathcal{W}_k\left(1-\lambda_l\right)\right] \\
&= Q^{-1}\Lambda\mathcal{W}_k\left(1\right)
\end{aligned}
$$

Lemma F2. *Let* $\widetilde{\mathcal{W}}_k(r) = \mathcal{W}_k(r) - r\mathcal{W}_k(1)$*. Under Assumption B as* $T \to \infty$*, when* $g \leq G-1$*,*

$$T^{-1/2}\widehat{\widehat{S}}_g \Rightarrow \Lambda\widetilde{\mathcal{W}}_k\left(\frac{g(1-\lambda_l)}{G-1}\right).$$

When $g = G$*,*

$$T^{-\frac{1}{2}}\widehat{\widehat{S}}_g = 0.$$

Proof: When $g \leq G-1$, it follows by simple algebra

$$
\begin{aligned}
& T^{-1/2}\widehat{\hat{S}}_g \\
&= T^{-1/2}\sum_{j=1}^{g}\sum_{t=(j-1)n_G+1}^{jn_G}\hat{v}_t \\
&= \sum_{j=1}^{g} T^{-1/2}\sum_{t=(j-1)n_G+1}^{jn_G} v_t - \sum_{j=1}^{g} T^{-1}\sum_{t=(j-1)n_G+1}^{jn_G}\left(x_t x_t'\right)\sqrt{T}\left(\hat{\beta}-\beta\right) \\
&= \sum_{j=1}^{g} T^{-1/2}\sum_{t=\frac{j-1}{G-1}(1-\frac{n_l}{T})T+1}^{\frac{j}{G-1}(1-\frac{n_l}{T})T} v_t \\
&- \sum_{j=1}^{g} T^{-1}\sum_{t=\frac{j-1}{G-1}(1-\frac{n_l}{T})T+1}^{\frac{j}{G-1}(1-\frac{n_l}{T})T}\left(x_t x_t'\right)\sqrt{T}\left(\hat{\beta}-\beta\right) \quad \because n_g = \frac{T-n_l}{G-1} \text{ (eqF.1)} \\
&\Rightarrow \sum_{j=1}^{g}\Lambda\left[\mathcal{W}_k\left(\frac{j(1-\lambda_l)}{G-1}\right) - \mathcal{W}_k\left(\frac{(j-1)(1-\lambda_l)}{G-1}\right)\right] \\
&- \sum_{j=1}^{g}\left[\frac{j(1-\lambda_l)}{G-1}Q - \frac{(j-1)(1-\lambda_l)}{G-1}Q\right]Q^{-1}\Lambda\mathcal{W}_k(1) \\
&= \Lambda\left[\mathcal{W}_k\left(\frac{g(1-\lambda_l)}{G-1}\right) - \frac{g(1-\lambda_l)}{G-1}\mathcal{W}_k(1)\right] \equiv \Lambda\widetilde{\mathcal{W}}_k\left(\frac{g(1-\lambda_l)}{G-1}\right)
\end{aligned}
$$

When $g = G$,

$$T^{-\frac{1}{2}}\widehat{\hat{S}}_G = 0$$

because it is the first order condition for the OLS estimator. Note that when $\lambda_l = 0$, we obtain the same result as in Lemma E2 as expected.

Asymptotic Limit of $\frac{G}{T}\widehat{\widehat{\Omega}}$: Recalling the algebra using the proof of Theorem 3.2 (b):

$$
\begin{aligned}
&\frac{G}{T}\widehat{\widehat{\Omega}} \\
&= \sum_{g=1}^{G-1}\sum_{h=1}^{G-1} T^{-1/2}\widehat{\widehat{S}}_g \left[2k\left(\frac{|g-h|}{M}\right) - k\left(\frac{|g-h+1|}{M}\right) - k\left(\frac{|g-h-1|}{M}\right)\right] T^{-1/2}\widehat{\widehat{S}}_h' \\
&\Rightarrow \Lambda \left[\sum_{g=1}^{G-1}\sum_{h=1}^{G-1} \widetilde{\mathcal{W}}_k\left(\frac{g(1-\lambda_l)}{G-1}\right)\left(2k\left(\frac{|g-h|}{M}\right) - k\left(\frac{|g-h+1|}{M}\right) - k\left(\frac{|g-h-1|}{M}\right)\right)\right. \\
&\left.\widetilde{\mathcal{W}}_k\left(\frac{h(1-\lambda_l)}{G-1}\right)'\right]\Lambda' \\
&\equiv \Lambda P^l\left(G, M, \widetilde{\mathcal{W}}_k, \lambda_l\right)\Lambda',
\end{aligned}
$$

which follows from Lemma F2. For the Bartlett kernel we have

$$
\begin{aligned}
&\frac{G}{T}\widehat{\widehat{\Omega}} \\
&= \sum_{g=1}^{G-1}\sum_{h=1}^{G-1} T^{-1/2}\widehat{\widehat{S}}_g \left[2k\left(\frac{|g-h|}{M}\right) - k\left(\frac{|g-h+1|}{M}\right) - k\left(\frac{|g-h-1|}{M}\right)\right] T^{-1/2}\widehat{\widehat{S}}_h' \\
&= \frac{2}{M}\sum_{g=1}^{G-1} T^{-1/2}\widehat{\widehat{S}}_g T^{-1/2}\widehat{\widehat{S}}_g' \\
&- \frac{1}{M}\sum_{g=1}^{G-M-1}\left(T^{-1/2}\widehat{\widehat{S}}_g T^{-1/2}\widehat{\widehat{S}}_{g+M}' + T^{-1/2}\widehat{\widehat{S}}_{g+M} T^{-1/2}\widehat{\widehat{S}}_g'\right) \\
&\Rightarrow \Lambda\left[\frac{2}{M}\sum_{g=1}^{G-1}\widetilde{\mathcal{W}}_k\left(\frac{g(1-\lambda_l)}{G-1}\right)\widetilde{\mathcal{W}}_k\left(\frac{g(1-\lambda_l)}{G-1}\right)' - \right. \\
&\frac{1}{M}\sum_{g=1}^{G-M-1}\left(\widetilde{\mathcal{W}}_k\left(\frac{g(1-\lambda_l)}{G-1}\right)\widetilde{\mathcal{W}}_k\left(\frac{(g+M)(1-\lambda_l)}{G-1}\right)'\right. \\
&\left.\left.+\widetilde{\mathcal{W}}_k\left(\frac{(g+M)(1-\lambda_l)}{G-1}\right)\widetilde{\mathcal{W}}_k\left(\frac{g(1-\lambda_l)}{G-1}\right)'\right)\right]\Lambda' \\
&\equiv \Lambda P^l(G, M, \widetilde{\mathcal{W}}_k, \lambda_l)\Lambda'
\end{aligned}
$$

Note that when $\lambda_l = 0$, the asymptotic approximation is the same as in Theorem 3.2 (b).

Asymptotic Limit of W_{CHAC}: Using similar arguments as in the proof of Theorem 3.2

(c), it follows from the previous results in this appendix that

$$
\begin{aligned}
& W_{CHAC} \\
&= (R\hat{\beta} - r)' \left[R\widehat{V}_{CHAC}R' \right]^{-1} (R\hat{\beta} - r) \\
&= \sqrt{T}\,(R\hat{\beta} - r)' \left[RT\widehat{V}_{CHAC}R' \right]^{-1} \sqrt{T}\,(R\hat{\beta} - r) \\
&= \left[R\sqrt{T}\,(\hat{\beta} - \beta) \right]' \left[R \left(T^{-1} \sum_{t=1}^{T} x_t x_t' \right)^{-1} T^{-1} G\widehat{\widehat{\Omega}} \left(T^{-1} \sum_{t=1}^{T} x_t x_t' \right)^{-1} R' \right]^{-1} \\
&\times R\sqrt{T}\,(\hat{\beta} - \beta) \\
&\Rightarrow \left[RQ^{-1}\Lambda \mathcal{W}_k(1) \right]' \left[RQ^{-1}\Lambda P^{l}(G, M, \widetilde{\mathcal{W}}_k, \lambda_l)\Lambda'^{-1}R \right]^{-1} RQ^{-1}\Lambda \mathcal{W}_k(1) \\
&= \mathcal{W}_q(1)' P^{l}(G, M, \widetilde{\mathcal{W}}_q, \lambda_l)^{-1} \mathcal{W}_q(1).
\end{aligned}
$$

When $q = 1$,

$$
\begin{aligned}
t_{CHAC} &= \frac{R\hat{\beta} - r}{\sqrt{R\hat{V}_{CHAC}R'}} \\
&\Rightarrow \frac{\mathcal{W}_1(1)}{\sqrt{P^{l}(G, M, \widetilde{\mathcal{W}}_1, \lambda_l)}}.
\end{aligned}
$$

BIBLIOGRAPHY

BIBLIOGRAPHY

Dennis Aigner, C.A. Knox Lovell, and Peter Schmidt. Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics*, 6:21–37, 1977.

Hirotugu Akaike. A new look at the statistical model identication. *I.E.E.E. Transactions on Automatic Control*, 19:716–723, 1974.

Antonio Alvarez, Christine Amsler, Luis Orea, and Peter Schmidt. Interpreting and testing the scaling property in models where inefficiency depends on firm characteristics. *Journal of Productivity Analysis*, 25:201–212, 2006.

Donald W. K. Andrews. Testing when a parameter is on the boundary of the maintained hypothesis. *Econometrica*, 69:683–734, 2001.

Donald W.K. Andrews. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59:817–858, 1991.

George E. Battese and Tim J. Coelli. Prediction of firm-level technical efficiencies with a generalized frontier production function and panel data. *Journal of Econometrics*, 38: 387–399, 1988.

George E. Battese and Tim J. Coelli. A model for technical inefficiency effects in a stochastic frontier production function for panel data. *Empirical Economics*, 20:325–332, 1995.

C. Alan Bester, Timothy G. Conley, and Christian B. Hansen. Inference with dependent data using cluster covariance estimators. *Journal of Econometrics*, 165:137–151, 2011.

Peter Bloomfield. Spectral analysis with randomly missing observations. *Journal of the Royal Statistical Society*, 32:369–380, 1970.

Steven B. Caudill. Estimating a mixture of stochastic frontier regression models via the em algorithm: A multiproduct cost function application. *Empirical Econometrics*, 28: 581–598, 2003.

Steven B. Caudill and Jon M. Ford. Biases in frontier estimation due to heteroskedasticity. *Economics Letters*, 41:17–20, 1993.

Steven B. Caudill, Jon M. Ford, and Daniel M. Gropper. Frontier estimation and firm-specific inefficiency measures in the presence of heteroscedasticity. *Journal of Business and Economic Statistics*, 13:105–111, 1995.

Yong Chen and Kung-Yee Liang. On the asymptotic behaviour of the pseudolikelihood ratio test statistic with boundary problems. *Biometrika*, 97:603–620, 2010.

Timothy J. Coelli, D.S. Pradada Rao, Christopher J. O'Donnell, and George E. Battese. *An introduction to efficiency and productivity analysis*. Springer, New York, 2005.

Deepa Dhume Datta and Wenxin Du. Nonparametric HAC estimation for time series data with missing observations. *Working Paper*, 2012.

James Davidson. *Stochastic Limit Theory*. Advanced Texts in Econometrics. Oxford University Press, 2002.

John C. Driscoll and Aart C. Kraay. Consistent covariance matrix estimation with spatially dependent panel data. *The Review of Economics and Statistics*, 80:549–560, 1998.

William Dunsmuir and Peter Robinson. Asymptotic theory for time series containing missing and amplitude modulated observations. *The Indian Journal of Statistics*, 43:260–281, 1981.

A.Ronald Gallant and Halbert White. *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*. Basil Blackwell, 1995.

Sílvia Gonçalves and Timothy J. Vogelsang. Block bootstrap HAC robust tests: the sophistication of the naive bootstrap. *Econometric Theory*, 27:745–791, 2011.

Christian Gouriéroux and Alain Monfort. *Statistics and Econometric Models*. Number v. 2 in Statistics and Econometric Models: Testing, Confidence Regions, Model Selection, and Asymptotic Theory. Cambridge University Press, New York, 1995.

Christian Gouriéroux, Alberto Holly, and Alain Monfort. Likelihood ratio test, Wald test, and Kuhn-Tucker test in linear models with inequality constraints on the regression parameters. *Econometrica*, 50:63–80, 1982.

Luca Grassetti. A novel mixture based stochastic frontier model with application to hospital efficiency. *unpublished manuscript, University of Udine*, 2011.

William Greene. Reconsidering heterogeneity in panel data estimators of the stochastic frontier model. *Journal of Econometrics*, 126:269–303, 2005.

E. J. Hannan and B. G. Quinn. The determination of the order of an autoregression. *Journal of the Royal Statistical Society. Series B (Methodological)*, 41:190–195, 1979.

Bruce E. Hansen. Strong laws for dependent heterogeneous processes. *Econometric Theory*, 7:213–221, 1991.

Bruce E. Hansen. Erratum. *Econometric Theory*, 8:421–422, 1992.

Clief J. Huang and Jin-Tan Liu. Estimation of a non-neutral stochastic frontier production function. *Journal of Productivity Analysis*, 5:171–180, 1994.

Michael Jansson. Consistent covariance matrix estimation for linear processes. *Econometric Theory*, 18:1449–1459, 2002.

James Jondrow, C.A. Knox Lovell, Ivan S. Materov, and Peter Schmidt. On the estimation of technical inefficiency in the stochastic frontier production function model. *Journal of Econometrics*, 19:233–238, 1982.

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

Subal C. Kumbhakar, Soumendra Ghosh, and J. Thomas McGuckin. A generalized production frontier approach for estimating determinants of inefficiency in U.S. dairy farms. *Journal of Business and Economic Statistics*, 9:279–286, 1991.

Subal C. Kumbhakar, Christopher F. Parmeter, and Efthymios G. Tsionas. A zero inefficiency stochastic frontier model. *Journal of Econometrics*, 172:66–76, 2013.

Wim Meeusen and Julien van den Broeck. Efficiency estimation from Cobb-Douglas production function with composed error. *International Economic Review*, 18:435–444, 1977.

Henry R. Neave. Spectral analysis of a stationary time series using initially scarce data. *Biometrika*, 57:111–122, 1970.

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

Luis Orea and Subal C. Kumbhakar. Efficiency measurement using a latent class stochastic frontier model. *Empirical Econometrics*, 29:169–183, 2004.

Efstathios Paparoditis and Dimitris N. Politis. Residual-based block bootstrap for unit root testing. *Econometrica*, 71:813–855, 2003.

Emanuel Parzen. On spectral analysis with missing observations and amplitude modulation. *The Indian Journal of Statistics*, 25, 1963.

Peter C.B. Phillips and Steven N. Durlauf. Multiple time series regression with integrated processes. *Review of Economic Studies*, 53(4):473–95, 1986.

David Reifschneider and Rodney Stevenson. Systematic departures from the frontier: a framework for the analysis of firm inefficiency. *International Economic Review*, 32:715–723, 1991.

Alan J. Rogers. Modified lagrange multiplier tests for problems with one-sided alternatives. *Journal of Econometrics*, 31:341–361, 1986.

Perry A. Scheinok. Spectral analysis with randomly missed observations: The binomial case. *The Annals of Mathematical Statistics*, 36(3):971–977, 1965.

Gideon Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6:461–464, 1978.

Steven G. Self and Kung-Yee Liang. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*, 82:605–610, 1987.

Yixiao Sun, Peter C. B. Phillips, and Sainan Jin. Optimal bandwidth selection in heteroskedasticity-autocorrelation robust testing. *Econometrica*, 76:175–194, 2008.

Timothy J. Vogelsang. Heteroskedasticity, autocorrelation, and spatial correlation robust inference in linear panel models with fixed-effects. *Journal of Econometrics*, 166:303–319, 2012.

Donald M. Waldman. A stationary point for the stochastic frontier likelihood. *Journal of Econometrics*, 18:275–279, 1982.

Hung-Jen Wang. Heteroscedasticity and non-monotonic efficiency effects of a stochastic frontier model. *Journal of Productivity Analysis*, 18:241–253, 2002.

Halbert White, editor. *Asymptotic Theory for Econometricians*. Academic Press, 2001. Revised Edition.