Tests of Bias in Log-Periodogram Regression

James Davidson,
Department of Economics, University of Exeter, Exter EX4 4PU, UK

Philipp Sibbertsen,
Department of Economics, University of Hannover, D-30167 Hannover, Germany

Abstract

This paper proposes simple Hausman-type tests to check for bias in the log-periodogram regression of a time series believed to be long memory. The statistics are asymptotically standard normal on the null hypothesis that no bias is present, and the tests are consistent.

JEL Classification: C12, C22

Keywords: Long memory, log periodogram regression, Hausman test.

1 Introduction

Long memory models, and specifically fractionally integrated models, are a popular method of representing the persistence characteristics of time series. Inference about long memory is often based on the results of the narrow band log-periodogram regression, first introduced by Geweke and Porter-Hudak (1983) (henceforth, GPH). This method has applications both to testing for the presence of long memory in individual time series and to testing for fractional cointegration (Hassler et al., 2006). However, as shown by Agiakloglou et al. (1993) inter alia, the presence of short-run dynamic components can still severely bias the GPH estimator in finite samples, falsely indicating the existence of long memory. A stable autoregressive root is nominally a 'short run' component, yet if close to unity it approaches observational equivalence with long memory. There is, moreover, a trade-off that investigators must make in the matter of bandwidth choice, between excessive bias and lost efficiency.

In this context, it appears desirable to have a means of testing whether a significant bias is present. In this paper we propose simple tests of the Hausman (1978) type. Numerous variants of the basic log-periodogram estimator have been proposed, notably by Phillips and Shimotsu (2002) who advocate a frequency-grouping approach with a fixed-effects treatment of the intercept, Andrews and Guggenberger (2003) who include polynomial terms in the frequencies in the narrow-band regression, and Moulines and Soulier (1999) (henceforth MS) who advocate a broad-band regression with dummy variables alone controlling for the short-run effects. In this note we confine attention to two cases, although our idea could in principle be generalized. Section 2 derives the test in the context of GPH, showing consistency and deriving the null distribution. Some simulations are reported in Section 3. Section 4 extends to idea to the MS estimator, and Section 5 contains concluding remarks.

2 The Test

Consider the class of covariance stationary Gaussian processes $\{Y_t\}$ whose spectrum takes the form

$$f(\lambda) = |1 - \exp(-i\lambda)|^{-2d} f^*(\lambda), \quad -\pi \leq \lambda \leq \pi \tag{2.1}$$

where $|d| < 0.5$ is the fractional integration parameter, and $f^*(\lambda)$ represents the short-term correlation structure of the model. We assume the regularity conditions of Hurvich et al. (1998) (henceforth HDB). Given a sample $Y_1, \ldots, Y_T$, GPH suggested estimating d in (2.1) from the regression

$$\log I(\lambda_k) = c + dX_k + \varepsilon_k, \ k = 1, \ldots, m \tag{2.2}$$

for some $m \leq [T/2]$ where $X_k = -2\log(\sin \lambda_k)$, $\lambda_k = 2\pi k/T$ denotes the kth Fourier frequency, and

$$I(\lambda) := \frac{1}{2\pi T} \left| \sum_{t=1}^{T} Y_t \exp(-it\lambda) \right|^2.$$

To minimize bias due to the omission of the unknown function f^*, GPH suggested setting $m = O(T^{1/2})$, in the hope that in this narrow band of low frequencies, the variations in f^* are small. Let $\hat{d}$ denote this estimator. Defining $a_j = X_j - \bar{X}$ where $\bar{X} = m^{-1} \sum_{j=1}^{m} X_j$, and $S_{XX} = \sum_{k=1}^{m} a_j^2$, HDB derive the bias expression

$$\begin{aligned}
E(\hat{d} - d) &= \frac{1}{S_{XX}} \sum_{k=1}^{m} a_k \log f_k^* + \frac{1}{S_{XX}} \sum_{k=1}^{m} a_k E(\varepsilon_k) \\
&= -\frac{2\pi^2}{9} \frac{f^{*''}(0)}{f^*(0)} \frac{m^2}{T^2} + o\left(\frac{m^2}{T^2}\right) + O\left(\frac{\log^3 m}{m}\right).
\end{aligned} \tag{2.3}$$

They further show that for $m = o(T^{4/5})$,

$$\sqrt{m}(\hat{d} - d) \xrightarrow{d} \mathrm{N}\left(0, \frac{\pi^2}{24}\right).$$

There is evidently a trade-off between efficiency and bias in the choice of m. HDB show that the mean squared error is minimized by setting $m = CT^{4/5}$ where, the constant C depends inversely on $f^{*''}(0)/f^*(0)$. Hurvich and Deo (1999) propose a plug-in estimator of C, but their Monte Carlo evidence shows that the asymptotic optimality criteria are successful only when there is a limited amount of short-run dependence. When the bias component is large enough to dominate the MSE, the GPH narrow-bandwidth proposal proves more effective.

Knowledge of the magnitude of the bias term would obviously be desirable in guiding the choice of bandwidth. Our suggestion is, in essence, to compare the results of setting two different bandwidths. Let the null hypothesis have the form

$$H_0 : f^*(\lambda) = \text{constant}, 0 \leq \lambda \leq \Lambda \leq \pi \tag{2.4}$$

where Λ is a bound to be specified. The test statistics we consider are of the form

$$\mathrm{TS} = \frac{\hat{d}_1 - \hat{d}_2}{\mathrm{SE}(\hat{d}_1 - \hat{d}_2)} \tag{2.5}$$

where $\hat{d}_1$ and $\hat{d}_2$ are estimators of d constructed with m_1 and $m_2 < m_1$ periodogram ordinates, respectively, and $\mathrm{SE}(\cdot)$ denotes a suitable estimator of the standard error. Letting $S_i = \sum_{j=1}^{m_i} a_{ij}^2$

for $i = 1$ and 2 respectively, we can adapt HDB's formula (5) to give

$$\text{Var}(\hat{d}_1 - \hat{d}_2) = E \left(\frac{1}{S_1} \sum_{j=1}^{m_1} a_{1j}\varepsilon_j - \frac{1}{S_2} \sum_{j=1}^{m_2} a_{2j}\varepsilon_j \right)^2$$

$$= E \left(\sum_{j=1}^{m_2} \left(\frac{a_{1j}}{S_1} - \frac{a_{2j}}{S_2} \right) \varepsilon_j + \frac{1}{S_1} \sum_{j=m_2+1}^{m1} a_{1j}\varepsilon_j \right)^2$$

$$= \frac{\pi^2}{6} \left(\sum_{j=1}^{m_2} \left(\frac{a_{1j}}{S_1} - \frac{a_{2j}}{S_2} \right)^2 + \frac{1}{S_1^2} \sum_{j=m_2+1}^{m1} a_{1j}^2 \right) + o \left(\frac{1}{m_2} \right). \quad (2.6)$$

The last member of (2.6), less the small-order term, is used to provide the standard error in (2.5).
For a test of hypothesis (2.4) we need to set m_1 such that

$$\lambda_{m_1} = 2\pi m_1 / T \ \leq \Lambda.$$

and then to choose m_2 so as to optimize the power if the test. Applying expression (2.3) yields

$$E(\hat{d}_1 - \hat{d}_2) = \frac{2\pi^2}{9} \frac{f^{*''}(0)}{f^*(0)} \frac{m_1^2}{T^2} \left(1 - \frac{m_2^2}{m_1^2} \right) + o \left(\frac{m_1^2}{T^2} \right) + O \left(\frac{\log^3 m_2}{m_2} \right). \quad (2.7)$$

Suppose we set $m_1 = CT^\beta$ in this expression for some $\beta \leq 1$ and also set $m_2 = Km_1$, where $K \in (0, 1)$ is another constant to be chosen. Substituting these choices into (2.7) yields

$$E(\hat{d}_1 - \hat{d}_2) = \frac{2\pi^2}{9} \frac{f^{*''}(0)}{f^*(0)} T^{2\beta-2} \left[1 - K^2 \right] + o(T^{2\beta-2}) + O \left(\frac{\log^3 T^\beta}{T^\beta} \right). \quad (2.8)$$

Also observe that (2.6) implies

$$\text{Var}(\hat{d}_1 - \hat{d}_2) = \frac{\pi^2}{6} \left[\left(\frac{1}{m_1} - \frac{1}{m_2} \right)^2 m_2 + \frac{m_1 - m_2}{m_1^2} \right] + o \left(\frac{1}{m_2} \right)$$

$$= \frac{\pi^2}{6T^\beta} \left[\frac{1}{K} - 1 \right] + o(T^{-\beta}). \quad (2.9)$$

Compare the leading terms in these expressions. For a consistent test, it is necessary that $|E(\hat{d}_1 - \hat{d}_2)|/\sqrt{\text{Var}(\hat{d}_1 - \hat{d}_2)} \to \infty$ when $f^{*''}(0) \neq 0$, and according to (2.8) and (2.9), this requires $2\beta - 2 > -\beta/2$ or, in other words, $\beta > 4/5$. Subject to this condition, we should then choose K to maximize the expression $(1 - K^2)/\left(K^{-1} - 1 \right)^{1/2}$, so as to make $|E(\hat{d}_1 - \hat{d}_2)|/\sqrt{\text{Var}(\hat{d}_1 - \hat{d}_2)}$ as large as possible with any given T. The solution to this problem on $[0, 1]$ can be verified numerically to be about $K = 0.64$. The following theorem establishes sufficient conditions for the statistic to have a normal limiting distribution under the null hypothesis. (See the Appendix for proof.)

Theorem 2.1 *If Y_t is a Gaussian process satisfying (2.1), hypothesis (2.4) holds, and $m_1 \leq [T\Lambda/2\pi]$, then*

$$\text{TS} \xrightarrow{d} N(0, 1).$$

Note that the Gaussianity is required since the CLT is established by the method of moments, using the approach pioneered by Robinson (1995).

Provided $4/5 < \beta \leq 1$, the test is consistent for the null hypothesis in (2.4) for any $\Lambda > 0$. Setting $\beta = 1$ corresponds to testing the case $\Lambda = \pi$. In other words $\hat{d}_1$ is obtained from the broad band regression including all the periodogram points. In this form, the null hypothesis takes the strict form that the spectrum depends on the single dynamic parameter d, such that the fractional differences are white noise. On the other hand, by choosing $\beta < 1$ we obtain a test of the hypothesis simply that $f^{*''}(0) = 0$, and hence that f^* is constant in an arbitrary neighbourhood of the origin. To see how this works in practice, note that the width of the interval of $[0, \pi]$ from which periodogram points for the estimators $\hat{d}_1$ and $\hat{d}_2$ are drawn is of order $m_1/T = O(T^{\beta-1})$, and hence shrinks to zero as $T \to \infty$. Heuristically, we can say that in a sample of size T, f^* needs to be constant in a region of this width to approximate the limiting null distribution of the statistic. The smaller β is set, the more rapid is this convergence and hence the greater potential generality of the null, although the power of the test grows correspondingly more slowly. Note the interesting fact that the test depends on comparing two estimators which are both consistent, under both null and alternative. Their t ratios nonetheless diverge under the alternative (even when $d = 0$, note) with β in the $(4/5, 1)$ range. By taking the difference of two such estimators we remove the dependence on d, and so obtain the null distribution. However, the bias term still converges to 0 more slowly than its standard deviation, so that statistic TS yields a consistent test. We note that in the class of ARFIMA(p, d, q) models only the case $p = q = 0$ yields a restriction corresponding to a case of H_0. The possibility of short-run autocorrelations existing, compatibly with the null distribution, can be realized only by invoking a wider class of DGPs. Nonetheless, we note that it is straightforward to construct such processes in the frequency domain, and to simulate them, by applying an inverse Fourier transform to a sample of suitably heteroscedastic normal innovations. The generality in the choice of nulls afforded by (2.4) is therefore not inconsiderable.

3 Monte Carlo Evidence

Extensive simulation results (available on request from the authors) suggest that our test has ample power to detect the existence of bias, at least in the context of linear Gaussian models. However, we note the unusual feature of our test, that in many applications of interest the null hypothesis is known to be technically false, and the test will therefore always reject when the sample is large enough. We therefore need to emphasize that the question of interest is not "Is there bias?" but rather "Is there important bias?", where 'important' is to be interpreted, typically, in the context of bandwidth choice in relation to sample size. Ideally, failure to reject a false null should imply that the bias is small enough to be innocuous from the viewpoint of standard inference on d, while a rejection should indicate a need for corrective action. The choice of significance level needs to be geared to these implicit trade-offs.

With these issues in mind, we present evidence of the performance of the test in the context of implementing a pretest estimator. Figures 1-3 show the results in 30,000 replications of GPH estimations in samples of 500 observations. The data generation processes have the ARFIMA$(1, d, 0)$ form

$$(1 - \phi L)(1 - L)^d Y_t = u_t \sim \mathrm{NI}(0, 1) \tag{3.1}$$

with the AR coefficient ϕ set to values 0, 0.25, 0.5 and 0.75, and either $d = 0$ or $d = 0.25$. Note that in this case $f^*(\lambda) = (1 - 2\phi \cos \lambda + \phi^2)^{-1}$ and

$$\frac{f^{*''}(0)}{f^*(0)} = \frac{-2\phi}{(1 - \phi)^2} \tag{3.2}$$

4

so that the bias is related directly to the magnitude of ϕ.

In addition to GPH estimators with bandwidths $m = 22 \approx [T^{0.5}]$ and $m = 78 \approx [T^{0.7}]$, we construct a composite estimator defined to equal the second case if the bias test rejects at the nominal 20% level (asymptotic criterion) and otherwise the first case. This conservative significance level was chosen on the basis of some preliminary experiments. Figure 1 shows the RMSE of the three estimators for the four cases of ϕ and $d = 0$. The same simulations, performed with $d = 0.25$, yielded effectively identical results, and are not shown. Figure 2 shows the rejection rates using the standard asymptotic criterion of the conventional Wald significance tests on d when the null hypothesis is true, with $d = 0$. Figure 3 shows a case of the alternative, the proportion of rejections in the corresponding cases with $d = 0.25$.

4 Broadband Estimator

We have adopted the GPH estimator for our exposition as the best-known case of log-periodogram regression, but analogous tests could be constructed for other variants. The MS estimator uses the whole range of frequencies. MS expand $\log f^*$ in Fourier series to pth order, yielding terms of the form h_j, $j = 1, ..., p$ where $h_j(\lambda_k) = \cos(j\lambda_k)/\sqrt{\pi}$. Assuming the expansion

$$\log f^*(\lambda) = \sum_{j=1}^{\infty} \theta_j h_j(\lambda) \tag{4.1}$$

they show that if the coefficients decline exponentially, such that $|\theta_j| = O(\rho^j)$ for $0 < \rho < 1$, then if $p = p_T = O(\log T)$, the mean squared error of the estimator of d is of $O(\log T/T)$. MS show that

$$\sqrt{T/p_T}(\hat{d}_{MS} - d) \xrightarrow{d} N\left(0, \frac{\pi^2}{6}\right).$$

A Hausman-type test can be constructed for this estimator similarly, although only to test the strict version of the null hypothesis. The cosine expansion of f^* corresponding to (2.4) with $\Lambda < \pi$ cannot be constructed with only a finite number of terms. To obtain a statistic having the form (2.5) for this case, choose $\hat{d}_1$ to be the estimator computed in the regression of the log periodogram points $\log I(\lambda_k)$ onto $(X_k, 1)$, and $\hat{d}_2$ to be the estimator computed from the regression of $I(\lambda_k)$ onto $(X_k, 1, h_1(\lambda_k), \ldots, h_{p_T}(\lambda_k))$. The asymptotic variance of $\hat{d}_1 - \hat{d}_2$ under H_0 is straightforwardly shown to be the difference of their asymptotic variances, by the usual Hausman (1978) variance formula. The counterpart of Theorem 2.1 for this case follows directly from MS Theorem 1.

5 Concluding Remarks

This paper proposes simple diagnostic procedures to detect bias in log periodogram regressions. The tests are very easy to implement as routine outputs of a log-periodogram regression package.[1] In particular, being nonparametric, they are easier to apply than goodness-of-fit tests based on a specified spectral functional form (e.g. Chen and Deo 2004, Paparoditis 2000). We have focused in our simulation experiments on linear Gaussian (ARFIMA) alternatives, in which an autoregressive root may 'mimic' long memory, but recent research (Davidson and Sibbertsen 2005, Hsieh et al. 2007) has considered nonlinear processes giving rise to long memory. The properties of long-periodogram regression in this context are still poorly understood, but the availability of diagnostic techniques assume special importance in the context of model uncertainty.

---

[1] They are implemented in the package Time Series Modelling 4.26 (Davidson 2008)

Appendix: Proof of Theorem 2.1

We adapt the proof of MS Theorem 1 to show that both $\sqrt{m_1}(\hat{d}_1 - d)$ and $\sqrt{m_2}(\hat{d}_2 - d)$ are asymptotically normal under H_0. Applying the continuous mapping theorem, these results extend directly to yield the stated limiting distribution of the TS statistic.

The MS theorem is given for the case of a broadband log-periodogram regression containing p_T dummy variables representing the leading terms of the expansion in (4.1). The MS estimator allows local smoothing of the periodogram by using averages of $m \geq 1$ consecutive periodogram points (their notation) to construct the regressand, and includes the maximum number of $K_T = [T/2m]$ periodogram points in the regression.[2] For simplicity, MS's smoothing factor m can validly be set to 1. We then modify the conditions by setting, respectively, $K_T = m_1$ and $K_T = m_2$. We also set $p_T = 0$, although note that in reading the MS proof, p_T (or p_n in MS's notation) should be replaced by 1 whenever it appears in orders-of-magnitude expressions.

Since, by hypothesis H_0, the $f^*(\lambda_j)$ are constant for $\lambda_j < \Lambda = 2\pi m_1/T$, the bias terms containing l_p^* in MS equation (C.2) can be replaced by 0, and the tail sums in MS equations (C.6) treated similarly. With these modifications, MS Theorem 1 holds for $\sqrt{m_1}(\hat{d}_1 - d)$ and $\sqrt{m_2}(\hat{d}_2 - d)$ under the stated assumptions, noting that MS Assumption 2 is equivalent to (2.1) under hypothesis H_0, and likewise MS Assumption 3 (summability of the cosine weights) is subsumed under our assumptions. ∎

References

Agiakloglou, C., P. Newbold and M. Wohar (1993) Bias in an estimator of the fractional difference parameter, *Journal of Time Series Analysis* 14, 235–246

Andrews D. W. K. and P. Guggenberger, A Bias-reduced log-periodogram regression estimator for the long-memory parameter, *Econometrica* 71, 675-712.

Chen, W. S. and R. S. Deo (2004) A Generalized Portmanteau Goodness-of-Fit Test for Time Series Models. *Econometric Theory* 20, 382-416.

Davidson, J. (2007) *Time Series Modelling 4.26* at http://www.timeseriesmodelling.com

Davidson, J. and P. Sibbertsen (2005) Generating schemes for long memory processes: regimes, aggregation and linearity" *Journal of Econometrics* 128 (2) 253-282

Geweke, J. and S. Porter-Hudak (1983) The estimation and application of long-memory time series models. *Journal of Time Series Analysis* 4, 221-237

Hassler, U., Marmol, F. and C. Velasco (2006) Residual log-periodogram inference for long-run relationships. *Journal of Econometrics* 130, 165-207

Hausman, J. A. (1978) Specification Tests in Econometrics, *Econometrica* 46, 1251-1272.

Hsieh, M.-C., Hurvich, C. M. and P. Soulier (2007) Asymptotics for duration-driven long range dependent processes. *Journal of Econometrics*141, 913-949

Hurvich, C. M. and R. Deo (1999) Plug-in selection of the number of frequencies in regression estimates of the memory parameter of a long memory time series, *Journal of Time Series Analysis* 20, 331–341

Hurvich, C. M., R. Deo and J. Brodsky (1998) The mean squared error of Geweke and Porter-Hudak's estimator of a long memory time series, *Journal of Time Series Analysis* 19, 19-46

---

[2]For clarity, we substitute here our notation T for sample size, in place of MS's n.

Moulines, E. and P. Soulier (1999) Broad ban log-periodogram estimation of time series with long-range dependence. *Annals of Statistics* 27, 1415-1439

Paparoditis, E. (2000) Spectral Density Based Goodness-of-Fit Tests for Time Series Models. *Scandinavian Journal of Statistics* 27(1), 143-176

Phillips, P. C. B. and K. Shimotsu (2002) *Journal of Time Series Analysis* 23, 57-93.

Robinson, P. (1995) Log-periodogram regression of time series with long-range dependence, *Annals of Statistics* 23, 1048-1072.
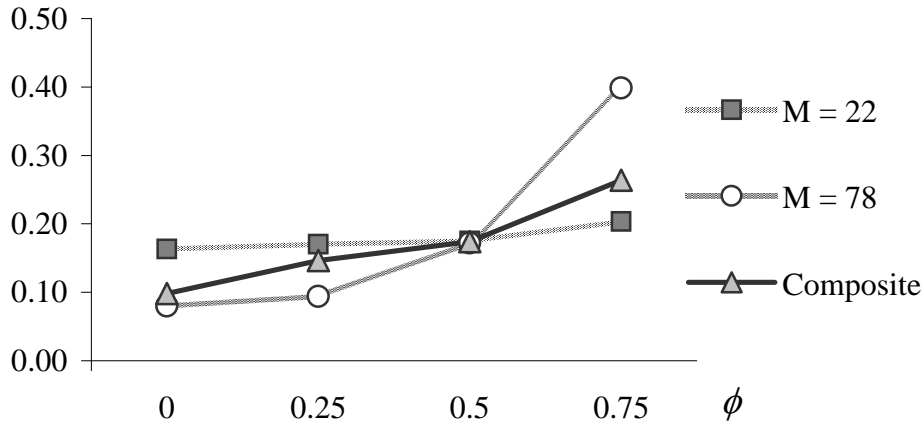
Figure 1. Root Mean Squared Error of GPH Estimates, $d = 0$, $T = 500$
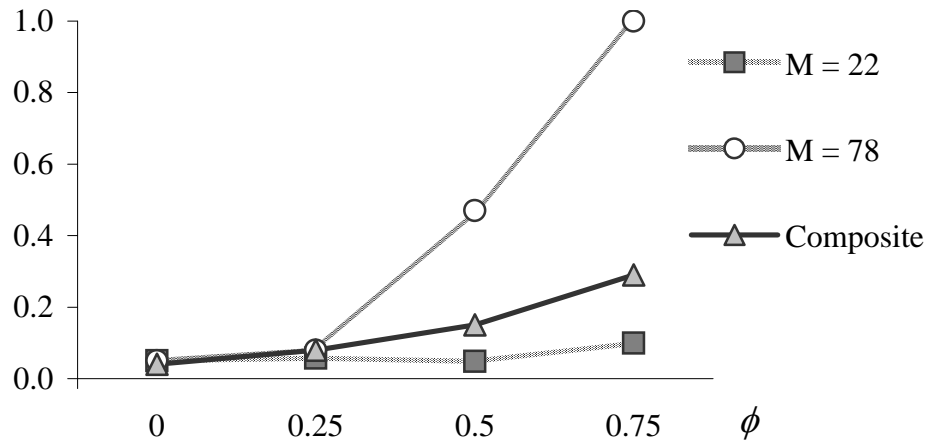

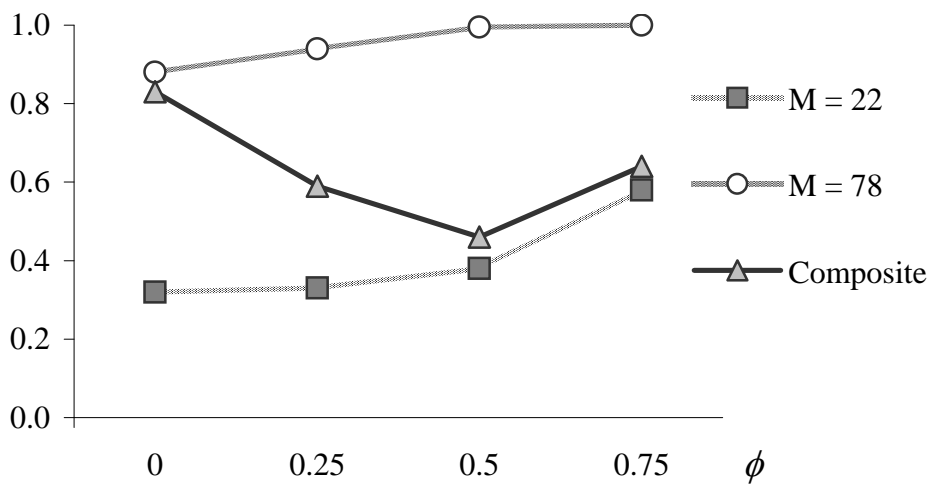Figure 2. Rejection Frequencies of 5% Significance Test, $d = 0$, $T = 500$

Figure 3. Rejection Frequencies of 5% Significance Test, $d = 0.25$, $T = 500$