JETEM

Journal of Economic Theory and Econometrics

The Journal of the Korean Econometric Society

(계　량　경　제　학　보)

March 2011

Volume 22. Issue 1. pp.1-27

Testing the Hypothesis on the Cointegrating Vector When Data May Have a Near Unit Root *

Masako Miyanishi†

Abstract This paper considers the size distortion problem when testing the hypothesis on the cointegrating vector. The test on the cointegrating vector tends to be erroneously rejected. We propose constructing a valid confidence interval by application of the Bonferroni's inequality, and compare the performance to another alternative procedure.

Keywords Cointegration, Local-to-unity, Bonferroni procedure

JEL Classification C12, C32

* I am grateful to Graham Elliott for his advice. I also thank two anonymous referees for helpful comments. All mistakes are mine.

† Assistant Professor, College of Business Administration, Seoul National University, E-mail address: mmiyanis@snu.ac.kr

Received July 7, 2010, Revised November 8, 2010, Accepted November 29, 2010

1 Introduction

This paper considers the size distortion problem in testing the hypothesis about the cointegrating vector when data may not have an exact unit root, but have a near unit root.

Cointegration is precisely defined by Engle and Granger (1987). Assuming variables have a unit root or are exactly integrated (which we denote as $I(1)$), researchers are often interested in testing a cointegrating vector, which captures the relationship among variables. The cointegrating vector is consistently estimable by least squares (Engle and Granger (1987)). However it is well known that OLS estimators could be severely biased in a finite sample due to correlations between variables and the short-run dynamics (see Saikkonen (1992), for example). As a solution, numerous studies have proposed a variety of efficient classes of estimators, including Johansen (1988, 1991), Phillips and Hansen (1991), Saikkonen (1991, 1992), and Stock and Watson (1993).

Those efficient estimators all assume that variables have an exact unit root. In practice, researchers conduct a unit root test such as Dickey-Fuller (1979) test. If the test does not reject the null hypothesis of a unit root, they assume that the data has a unit root. However, it is impossible to determine whether an economic variable has an exact unit root in a finite sample (and the data are always finite in practice). The result of a unit root test simply tells us that the largest root in the data is close to unity. Elliott (1998) points out that even if the largest root is only slightly less than unity, the test on the cointegrating vector could have a severe size distortion. Numerical analysis indicates that if the largest root is near but not exact unity, the actual size of a test could be as high as 40% for the nominal size of 10% (Elliott (1998)).

This paper makes two contributions. First, this paper proposes a valid testing procedure in the presence of a near unit root. We apply Bonferroni's inequality to take the sample information on the largest root into account. Application of Bonferroni's inequality was first proposed by Cavanagh, Elliott and Stock (1995) for a predictive regression. Torous et al. (2004) and Campbell and Yogo (2006) applied the method to test stock return predictability. The test of predictability is rejected too often when a predictor variable such as the dividend-price ratio is highly persistent. Rossi (2007) applies Bonferroni's inequality to a long-horizon predictability parameter. We apply

Bonferroni's inequality to Phillips and Hansen's (1990) fully modified estimator, which is an efficient estimator. We construct a valid confidence interval for the cointegration coefficient from a valid confidence interval of the largest root in the regressor. The method is simple and applicable to any efficient estimator.

Secondly, we evaluate the finite sample performance of the proposed test. Since the data is always finite in an empirical study, we examine how the test is improved in a finite sample rather than looking at asymptotics. Through numerical work, we compare the size and power of our test to an efficient test that does not take the presence of a near unit root into account and another valid test proposed by Wright (2000). Compared to the fully modified estimator, our test greatly improves size. It is obvious that an efficient class of estimator leads to invalid inference when the data does not have a unit root. The performance of our test is comparable to that of Wright (2000).

The rest of paper is organized as follows. The second section presents the model under examination and two parameters that bias the test statistics. We discuss the estimation method as well. In the third section, we propose a valid test, and also discuss another testing method in the presence of a near unit root. The fourth section evaluates the finite sample performance of our test by simulations. The fifth section gives a brief conclusion.

2 Test Statistics in the Presence of Near Unit Root

2.1 Model

We consider the following simple triangle model.

$$x_t = \alpha_1 + \rho x_{t-1} + u_{1t} \tag{1}$$

$$y_t = \alpha_2 + \beta x_t + u_{2t} \tag{2}$$

where $t = 1, .., T$, α_1 and α_2 are constant terms, y_t and x_t are both univariate.

The second equation is the cointegration equation. The first equation is the innovation of the regressor of the cointegrating equation. It is modeled as the first-order autoregressive process. In our study, x_t may or may not have an exact unit root, that is, ρ may be one or less than one, but close to one. When we consider a variable whose largest root is near unity, we typically employ the local-to-asymptotic theory developed by Bobkoski (1983), Cavanagh (1985), Chan and Wei (1987), and Phillips (1987). Under this theory, we reparameterize ρ as $\rho = 1+\frac{c}{T}$, where c is a local-to-unity parameter, $c \leq 0$, and T is the sample size.

Residuals $u_t = [u_{1t}, u_{2t}]'$ can be serially correlated and $\Phi(L)\ u_t = \varepsilon_t$, where $\varepsilon_t = [\varepsilon_{1t}, \varepsilon_{2t}]'$ and $E[\varepsilon_t \varepsilon_t'] = \Sigma$. $\Phi(L)$ is a lag polynomial of known order with all roots outside the unit circle, that is, u_t is stationary. 2π times spectral density at frequency zero of u_t is $\Omega = \Phi(1)^{-1}\Sigma\Phi(1)^{-1\prime} = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{12}' & \Omega_{22} \end{pmatrix}$, where $\Phi(1) = \sum_i \Phi_i$, and the zero frequency correlation between residuals is given as $\delta = \frac{\Omega_{12}}{\Omega_{11}^{1/2}\Omega_{22}^{1/2}}$.

We are interested in testing the null hypothesis about the cointegrating coefficient β as follows:

$$H_o : \beta = \beta_o \quad vs. \quad H_1 : \beta \neq \beta_o.$$

2.2 Distribution of the test statistic in the presence of a near unit root

A consistent estimator for β can be obtained by the usual ordinary least squares (OLS) (Engle and Granger, 1987). However, the OLS estimate of β tends to be severely biased in finite samples due to correlation and short-run dynamics. As a result, a variety of 'efficient' classes of estimators are introduced. Examples are Saikkonen's (1992) VAR, the full information maximum likelihood (FIML) method of Johansen (1988, 1991) and Ahn and Reinsel (1990), the fully modified estimator of Phillips and Hansen (1990), the dynamic OLS (DOLS) of Saikkonen (1991), Phillips and Loretan (1991) and Stock and Watson (1993). However, when the regressor is not exactly integrated, a test statistic such as t-statistic or Wald statistic of the efficient estimator will be still biased (Elliott, 1998). Elliott (1998) shows that the Wald statistic of Saikkonen's (1992) estimator, denoted as $\widehat{\beta}$, converges to the usual chi-square distribution plus a bias term which depends on the local-to-unity parameter c and zero frequency correlation δ. The results hold for Phillips and Hansen's (1995) fully modfied estimator and Stock and Watson's (1993) dynamic OLS procedure in the bivariate model (Elliott, 1998).

Here we explain bias in t-statistics on β in a univariate case. By Theorem 1 in Elliott (1998),

$$\widehat{V}^{-1/2}\{T(\widehat{\beta}-\beta)-B\} \sim_a N(0,1)$$

under his regularity conditions and the null, $\sim_a$ indicates 'asymptotically distributed' and

$$B = -\Omega_{21}\Omega_{11}^{-1}c; \tag{3}$$

$$\widehat{V} \Rightarrow V = \left[\Omega_{11}^{1/2}\left(\int (J_c^d(\lambda))^2 d\lambda\right)\Omega_{11}^{1/2\prime}\right]^{-1}\Omega_{2.1} \tag{4}$$

$$\Omega_{2.1} = \Omega_{22} - \Omega_{21}\Omega_{11}^{-1}\Omega_{12}; \tag{5}$$

$\Rightarrow$ indicates 'weakly converges,' and $dJ_c^d(\lambda)$ is the detrended Ornstein Uhlenbeck process, i.e.,

$$dJ_c(\lambda) = cJ_c(\lambda)d\lambda + dW_1(\lambda) \tag{6}$$

such that $J_c(0) = 0$, $W_1(\lambda)$ is a standard Brownian Motion associated with x_t,and $J_c^d(\lambda) = J_c(\lambda) - \int_0^1 J_c(s)ds$. Thus, the $t-$statistic has a bias term $\widehat{V}^{-1/2}B$, and,

$$\widehat{V}^{-1/2}B \Rightarrow -c\delta\sqrt{\frac{\int (J_c^d)^2 d\lambda}{\left(1-\delta^2\right)}}, \tag{7}$$

implying that if $\delta = 0$, there is no bias even if x_t does not have an exact unit root, although δ is highly unlikely to be zero. Alternatively, if x_t does have an exact unit root (i.e., $c = 0$), there is no bias, even if $\delta \neq 0$. On the contrary, as δ approaches one and $c < 0$, bias will go to infinity; if $c < 0$, the sign of δ determines the direction of bias. For exmaple, if $\delta > 0$, t-statistic will be negatively biased.

2.3 Two nuisance parameters: c and δ

When the largest root of x_t does not have an exact unit root, the test statistic has the bias given in (7), which depends on two nuisance parameters, c and δ. Thus, to conduct a correct statistical inference, it is essential to have information about these parameters. We are going to discuss how to obtain the information about c and δ.

2.3.1 The largest root (c)

First, when ρ is near one, we reparameterize ρ as $\rho = 1 + \frac{c}{T}$ by local-to-unity asymptotic theory, by which we treat c as fixed. Notice that for a fixed sample size T (which is always the case in empirical applications), there is a one-to-one correspondence between the values of ρ and c. In this sense we refer to them interchangeably even though c is not a parameter in the model. This local to unity parameter makes it difficult to conduct a correct statististical inference, since it is not consistently estimable. The estimator of c has a random distribution and is strongly biased and skewed.[1] Although the point estimate of c may give some idea about the true c, it will generally underestimate the true parameter.

An alternative approach is to place a confidence interval on the value for ρ (or c interchangeably in a finite sample). Since the estimator of ρ does not have asymptotically a symmetric distribution, we cannot obtain equally tailed confidence intervals by adding and subtracting a proportion of the estimated standard error of the estimator (Stock (1991)). Instead, a valid confidence interval can be constructed by collecting the set of values for c that cannot be rejected in the hypothesis test that $c = c_o$.

We explain how an equally tailed confidence interval for ρ (and c) is constructed. Stock (1991) first provided such a method through the inversion of the augmented Dickey-Fuller (ADF) test for a unit root of a first or higher order autoregressive (AR) process. Consider a higher-order AR

[1] Note that if u_{1t} is independent and identically distributed, then as $T \to \infty$, the estimator of c, $\widehat{c} = T(\widehat{\rho} - 1) \Rightarrow c + \frac{\int_0^1 J_c(r)dW(r)}{\int_0^1 \{J_c(r)\}^2 dr}$, where $J_c(r)$ is Ornstein-Uhlenbeck process with $dJ_c(r) = cJ_c(r)\,dr + dW(r)$, $0 \leq r \leq 1$, and $W(r)$ is a standard Brownian motion. See Phillips (1987) for detailed discussion. Furthermore, $T(\widehat{\rho} - \rho) = O_p(T)$ if $\widehat{\rho} = 1 + \frac{\widehat{c}}{T}$, then, $\widehat{c} - c = O_p(1)$. Hence, c is not consistently estimable.

model,

$$x_t = \mu_1 + v_t$$

where $a(L)v_t = \varepsilon_t$, $a(L) = b(L)(1-\rho L)$, $b(L) = \sum_{i=0}^{k} b_i L^i$, $b_0 = 1$ and L is the lag operator; $b(1) \neq 0$ is assumed, and ε_t is a martingale difference sequence with a finite fourth moment. The parameter ρ is the largest root factored out from $a(L)$ and assumed to be equal to or near one. Local-to-unity asymptotic theory is applied. The ADF regression model is

$$\Delta x_t = \widetilde{\mu}_1 + (\alpha(1)-1)x_{t-1} + \sum_{j=1}^{k} \alpha_{j-1}^{*}\Delta x_{t-1} + \varepsilon_t,$$

where $\widetilde{\mu}_1 = -\frac{cb(1)\mu_1}{T}$, $\alpha(1) = L^{-1}(1-a(L))$, so that $\alpha(1) = 1 + \frac{cb(1)}{T}$, and $\alpha_j^{*} = -\sum_{j=1}^{k}\alpha_j$. The ADF t-test is a test statistic testing the null hypothesis that $a(1) - 1 = 0$.

As shown in Stock (1991), the ADF t-statistic, which we denote as $\widehat{\tau}$ has an asymptotic distribution depending only on the local-to-unity parameter c and is continuous in c;

$$\widehat{\tau} \Rightarrow \left(\int_0^1 J(s)^2\, ds\right)^{1/2} \left\{ \left(\int_0^1 J(s)^2\, ds\right)^{-1} \int_0^1 J(s)\, dW(s) + c \right\}$$

where $J(.)$ is a diffusion process such that $dJ(s) = cJ(s)ds + dW(s)$, and $W(.)$ is a Brownian motion. If x_t is exactly a unit root process, $J(.)$ is a Brownian motion.

For each value of c, there is a distribution of $\widehat{\tau}$, and the $100(1-\alpha)\%$ confidence set of $\widehat{\tau}$ can be obtained from this. Suppose $A_\alpha(c)$ is the asymptotic $100(1-\alpha)\%$ acceptance region of c. Then, for each c in $A_\alpha(c)$, there is the corresponding distribution of $\widehat{\tau}$. The closed confidence set of c, $S(\widehat{\tau})$ can be described as;

$$S(\widehat{\tau}) = \left\{ c : f_{l;\frac{1}{2}\alpha}(c) \leq \widehat{\tau} \leq f_{u;\frac{1}{2}\alpha}(c) \right\},$$

where $f_{l;\frac{1}{2}\alpha}(c)$ and $f_{u;\frac{1}{2}\alpha}(c)$ are respectively the lower and upper $\frac{1}{2}\alpha$ percentiles of $\widehat{\tau}$ as a function of c. Because the function $f_{.}(.)$ is strictly monotonically increasing in c, by inverting the equally tailed confidence set of $\widehat{\tau}$, it is possible to obtain an equally tailed confidence interval for c, so that for each $\widehat{\tau}$,

$$S(\widehat{\tau}) = \left\{ c : f^{-1}_{u;\frac{1}{2}\alpha}(c) \leq c \leq f^{-1}_{l;\frac{1}{2}\alpha}(c) \right\}$$

is a closed confidence interval of c, and thus an equally tailed confidence interval for ρ can be constructed in a finite sample.

2.3.2 Zero frequency correlation (δ)

Unless long-run correlation δ is zero, the test statistic of β will be biased in the presence of a near unit root. If economic variables are closely related and moving together, δ is likely to be non zero. To determine whether the test statistics are biased, therefore it is necessary to estimate δ, estimating the long-run variance Ω. For the long-run variance, a variety of consistent estimators is available. The estimators can be categorized into two approaches: the sum of autocovariances approach (nonparametric approach) and the autoregressive approximation approach (parametric approach). In both approaches, they first estimate the simple regression models in equations (1) and (2) ignoring any serial correlation and then to construct the OLS residuals from these regressions. Denote these estimated residuals as $\widehat{u}_t = [\widehat{u}_1, \widehat{u}_{2t}]'$ for $t = 1, ..., T$.

The sum of autocovariances approach arises naturally out of the definition of the zero frequency and includes popular methods such as Newey and West's (1987) estimator. We define the $j-$th lag autocovariance matrix as $\Gamma(j) = E\left[\widehat{u}_t \widehat{u}'_{t+j}\right]$. It is estimated as

$$\widehat{\Gamma}(j) = T^{-1} \sum_{t=|j|+1}^{T} \widehat{u}_t \widehat{u}'_{t+j}.$$

Then, a finite sum of $\widehat{\Gamma}(j)$ is a consistent estimator of Ω. The problem is that the finite sum of autocovariances is not necessarily positive definite. In order to make the estimated long-run variance positive definite, we add weight ω_j for each estimated autocovariance.

$$\widehat{\Omega} = \sum_{j=-j_{\max}}^{j_{\max}} \omega_j \widehat{\Gamma}(j),$$

where usually $\omega_0 = 1$, and $j_{\max}$ is called the lag truncation or bandwidth parameter because the autocovariance with the lag larger than $j_{\max}$ has zero weight. Different weights result in different estimators. Newey and West's (1987) estimator uses the Bartlett kernel (weight), and Andrews (1991) examined various estimators with different choices of weights and window sizes.

Autoregressive approximations take a different approach, based on the results of Berk (1974). In the case of stationary u_t, the Wold representation theorem states that there is a moving average representation $u_t = C(L)\varepsilon_t$, where $E[\varepsilon_t\varepsilon_t'] = \Sigma$, implying that we have a vector autoregression $A(L)u_t = \varepsilon_t$, which can be inverted as $u_t = A(L)^{-1}\varepsilon_t$ and $\Omega = A(1)^{-1}\Sigma A(1)^{-1\prime}$. In this sense a two step approach can be considered. First, run a vector autoregression on the estimated residuals, that is, run $\widehat{u}_t = A^*(L)\widehat{u}_{t-1} + \varepsilon_t$ and save the estimates $\widehat{A}_i, \widehat{\varepsilon}_t$. We then have the estimator

$$\widehat{\Omega} = \left(I - \sum_{i=1}^{k_{\max}} \widehat{A}_i^*\right)^{-1} \left(T^{-1}\sum_{t=1}^{T} \widehat{\varepsilon}_t\widehat{\varepsilon}_t'\right) \left(I - \sum_{i=1}^{k_{\max}} \widehat{A}_i^*\right)^{-1\prime}.$$

In principle, we should have an infinite sequence of A_i, but they are impossible to compute with finite T observations. Hence as above we need to choose a 'window' $k_{\max}$. Here the window is just the lag order in the VAR run on the estimated residuals. There is no need for considering weights as the formula is a quadratic, which must be positive definite by construction.

The estimate of the long-run correlation is obtained as $\widehat{\delta} = \frac{\widehat{\Omega}_{12}}{\widehat{\Omega}_{11}^{1/2}\widehat{\Omega}_{22}^{1/2}}$.

3 Alternative Procedures

We propose the application of Bonferroni's method to Phillips and Hansen's (1990) fully modified estimator. We also discuss another testing procedure.

3.1 Bonferroni test

3.1.1 Bonferroni's inequality

In the presence of a near unit root, t-statistics will not follow the standard normal distribution and the conventional critical values under the normal assumption lead to an invalid statistical inference.

Supposing that the nominal size is 10%; the conventional critical values for a two-sided test are ± 1.645. Whereas the actual asymptotic distribution may be different from the standard normal distribution, the size will exceed 10%. As the value of c moves further away from zero while δ is nonzero, the size distortion will become larger. If $c \ll 0$ and $\delta < 0$, the actual t-statistics distribution will be skewed to the right, and the actual size of the right tail above 1.645 will exceed 5%.

To conduct a correct statistical inference, we construct a valid confidence interval for the true value of β. In this section, we obtain it by Bonferroni's inequality:

$$\Pr\left(A_1^c \cap A_2^c\right) \geqq 1 - \left(\Pr\left(A_1\right) + \Pr\left(A_2\right)\right).$$

In this inference problem, event A_1 is that c (or ρ equivalently in finite sample) is outside of a confidence region and event A_2 is that β is outside of a confidence region. Let $\Pr\left(A_1\right) = \eta_1$ and $\Pr\left(A_2\right) = \eta_2$. Let $C_c\left(\eta_1\right)$ denote a $100\left(1-\eta_1\right)\%$ confidence region for c, and $C_{\beta|c}\left(\eta_2\right)$ denote $100\left(1-\eta_2\right)\%$ confidence region for β that depends on c. Let us denote $\eta = \eta_1 + \eta_2$. Then, if we obtain a confidence region for any possible c, a $100\left(1-\eta\right)\%$ valid confidence region for β, which is independent of c, can be constructed as

$$C_\beta^B\left(\eta\right) = \cup_{c \in C_c(\eta_1)} C_{\beta|c}\left(\eta_2\right)$$

By Bonferroni's inequality, the confidence region $C_\beta^B\left(\eta\right)$ has a confidence level of at least $100(1-\eta)\%$,. Thus, by applying Bonferroni's inequality we can obtain a valid confidence region for β and can effectively control the size.

3.1.2 The fully modified estimator

We propose to apply the above idea to Phillips and Hansen's (1990) fully modified estimator. First we briefly explain the fully modified estimator. The fully modified estimator is an efficient class estimator. This method 'modifies' the cointegration equation (2) so that the cointegration residual term becomes orthogonal to the regressor. The fully modified cointegration equation is:

$$y_t^* = \alpha_2 + \beta x_t + u_{2t}^*, \tag{8}$$

where

$$\begin{aligned} y_t^* &= y_t - \frac{\Omega_{21}}{\Omega_{11}} u_{1t} \\ u_{2t}^* &= u_{2t} - \frac{\Omega_{21}}{\Omega_{11}} u_{1t}. \end{aligned}$$

Here, we assume that there is no constant term in the innovation of x_t for a simple presentation. By running regression with the modified cointegration equation, we obtain the fully modified estimate of β, denoted as $\widehat{\beta}_{FM}$. This method will effectively eliminate an endogeneity bias, because the cointegration residual term u_{2t} can be described as $u_{2t} = \frac{\Omega_{21}}{\Omega_{11}} u_{1t} + u_{2.1t}$, where $u_{2.1t}$ is orthogonal to x_t. By transforming y_t into $y_t^* = y_t - \frac{\Omega_{21}}{\Omega_{11}} u_{1t}$, the cointegration equation residual becomes orthogonal to x_t. Hence, the test statistic on β will have the standard asymptotic distribution.

In practice, we estimate y_t^* as follows:

$$\widehat{y}_t^* = y_t - \frac{\widehat{\Omega}_{21}}{\widehat{\Omega}_{11}} \Delta x_t, \tag{9}$$

where $\widehat{\Omega}$ is any consistent estimator of Ω. We estimate Ω by the non-parametric method with the Bartlett kernel as in Newey and West (1987).

Notice that $\Delta x_t = u_{1t}$ if $\rho = 1$. However, if x_t is not exactly integrated or $\rho \neq 1$, then, $\Delta x_t \neq u_{1t}$. Consequently, the test statistics testing on β will have size distortion as shown in section 4.

3.1.3 Application of Bonferroni's procedure

Bonferroni's inequality has been applied to a near unit root problem in a predictive regression setting. Cavanagh, Elliott, and Stock (1995) first suggested applying the procedure to the one-period-ahead predictive regression to obtain a valid critical value. Torous, Valkanov, and Yan

(2004) and Campbell and Yogo (2006) applied the procedure to testing the predictability of stock returns. In testing predictability, the dividend-price ratio or the earning-price ratio is often used as a predictor variable and it is often found to have a near unit root. In a predictive regression, the regressor which is supposed to be stationary is found to have a near unit root. In the cointegration setting which we are now considering, the regressor that is supposed to have a unit root may have a near but not exact unit root.

Campbell and Yogo (2006) offered constructing valid confidence intervals for the regression coefficient. We also consider the construction of valid confidence intervals for the cointegration coefficient.

We slightly change the equation (9), replacing the first difference term Δx_t with the quasi-difference term $(1-\rho L)x_t$:

$$\widetilde{y}_t^* = y_t - \frac{\widehat{\Omega}_{21}}{\widehat{\Omega}_{11}}(1-\rho L)x_t. \tag{10}$$

Notice that $(1-\rho L)x_t = u_{1t}$. That is, as long as ρ is known (though it may be different from unity), the test statistics with this equation will have the standard distribution. In reality, however, we do not have the true value of ρ. Thus, as suggested in Stock (1991), we construct a confidence interval of ρ by inverting the ADF t-statistic and, then apply Bonferroni's method to have a valid confidence interval for β.

We apply Bonferroni's procedure as follows. First, consider a $100\,(1-\eta_1)\,\%$ confidence interval for ρ. Any value of ρ within the interval can be a true value with the confidence level of $100(1-\eta_1)\%$. Second, a confidence interval for β that depends on ρ is as follows:

$$C_{\beta|\rho}\,(\eta_2) = \left[\underline{\beta}, \overline{\beta}\right],$$

where

$$\begin{aligned}\underline{\beta} &= \widehat{\beta}_{FM} - z_{\eta/2} SE(\widehat{\beta}_{FM}) \\ \overline{\beta} &= \widehat{\beta}_{FM} + z_{\eta/2} SE(\widehat{\beta}_{FM}),\end{aligned}$$

where $z_{\eta/2}$ is $\frac{\eta}{2}$th percentile of standard normal distribution, $SE(\widehat{\beta}_{FM})$ is standard error of $\widehat{\beta}_{FM}$.

A valid confidence interval for β can be obtained as:

$$C_\beta(\eta) = \left[\inf \underline{\beta}, \sup \overline{\beta}\right],$$

where $\inf \underline{\beta}$ is the minimum of $\underline{\beta}'s$ and $\sup \overline{\beta}$ is the maximum of $\overline{\beta}'s$.

Alternatively, we could obtain valid critical values through simulations, as suggested by Cavanagh et al. (1995). In order to obtain critical values, we first construct $100\,(1-\eta_1)\,\%$ confidence interval for ρ and estimate δ, on which the distribution of test statistics depends. With the knowledge of ρ and δ, we simulate the distribution of test statistics and obtain the critical values with the $100\eta\%$ confidence level. That is, in order to obtain critical values, we need to estimate an additional nuisance parameter δ. It is consistently estimable because the consistent estimator of Ω exists. However, it may not be an accurate estimate in a small sample. Furthermore, to obtain critical values, we must perform simulation for each data set, which is computationally demanding. Thus, we choose to construct confidence intervals rather than to obtain critical values.

It is known that tests based on Bonferroni's inequality tend to be conservative. That is, the actual rejection rate is below 10% for the nominal size 10%, especially when δ is small. We may correct the Bonferroni test so that the size is close enough to 10%. Basically, we refine the test by tightening the confidence interval for ρ depending on δ. Cavanagh et al. (1995) recommend η_1 equal to 30% for $\delta = 0$ or 0.5 and η_1 equal to 24% for $\delta = 0.7$. Other discussions are found in Campbell and Yogo (2006).

3.2 Inference based on stationarity tests

We would like to compare Bonferroni's method with another testing procedure. Wright (2000) suggested an inference method based on the stationarity test proposed by Nyblom (1989), and Kwitkowski, Phillips, Schmidt, and Shin (1992). The stationarity test is a unit root test, testing the null hypothesis of stationarity against the alternative hypothesis of a unit root. Wright (2000) applied the test to testing for stationarity of the residuals of a cointegration equation. Let us denote ξ_t as the residual term of the cointegration equation under the null hypothesis of $\beta = \beta_o$ when there is no deterministic term:

$$\xi_t = y_t - \beta_o x_t.$$

Under the null hypothesis of $\beta = \beta_o$, ξ_t will be stationary ($I(0)$), that is, the variables y_t and x_t are cointegrated with the hypothesized cointegration coefficient, β_o. On the other hand, if $\beta \neq \beta_o$, ξ_t will have a (local to) unit root, if the variables have an exact unit root. Thus, the stationarity test is applicable to the residuals from the cointegration equation. Additionally, the stationarity test is the locally most powerful invariant (LMPI) test, which is free of nuisance parameters under the null hypothesis. Thus, the test can effectively avoid a size distortion problem in the presence of nuisance parameters. Under the null hypothesis of $\beta = \beta_o$, the test statistic will not be biased even if x_t does not have an exact unit root ($c < 0$) and there is endogeneity ($\delta \neq 0$). Under the alternative hypothesis of $\beta \neq \beta_o$, the test will depend on the nuisance parameters.

The test statistic is a Lagrange multiplier (LM) test, denoted as $L(.)$. Let us denote $\widehat{\sigma}_2^2$[2] as the estimated zero frequency variance of the residuals from the cointegration regression. Then, the test statistic has three forms:

$$\begin{aligned}
L_1(\beta_o) &= \frac{T^{-1}\left(\sum_{t=1}^{T}\xi_t\right)^2}{\widehat{\sigma}_2^2} \\
L_2(\beta_o) &= \frac{T^{-2}\sum_{t=1}^{T}\left(\sum_{s=1}^{t}\xi_s^{\mu}\right)^2}{\widehat{\sigma}_2^2} \\
L_3(\beta_o) &= \frac{T^{-2}\sum_{t=1}^{T}\left(\sum_{s=1}^{t}\xi_s^{\tau}\right)^2}{\widehat{\sigma}_2^2},
\end{aligned}$$

where ξ_t^{μ} and ξ_t^{τ} are residuals of ξ_t after projection on a constant and a constant and after projection on a trend, respectively.

To compute $\widehat{\sigma}_2^2$, we obtain $\widehat{u}_{2t} = y_t - \widehat{\alpha} - \widehat{\beta} x_t$ by regression, where $\widehat{\alpha}$ and $\widehat{\beta}$ can be any consistent estimates. Under the null hypothesis, the test statistic has the chi-square distribution asymptot-

[2] Although σ_2^2 is equivalent to Ω_{22}, we use the notation $\widehat{\sigma}_2^2$ instead of $\widehat{\Omega}_{22}$, because we compute $\widehat{\sigma}_2^2$ by imposing the fully modified estimator that requires $\widehat{\Omega}_{22}$.

ically. Under the alternative hypothesis of $\beta \neq \beta_o$, the test statistic will depend on the nuisance parameters, c and δ. Thus, there would be a substantial power loss if $\beta \neq \beta_o$. As $c \ll 0$ (or $\rho \ll 1$) in x_t, any linear combination of y_t and x_t will be closer to being stationary. It must be emphasized that this test actually tests the joint hypothesis of stationarity of the cointegration equation error and $\beta \neq \beta_o$. Consequently, even if $\beta \neq \beta_o$, the stationarity test tends to fail to reject the null hypothesis of cointegration under $\beta = \beta_o$.

4 Finite Sample Performance

This section evaluates finite sample performance. Size and power for testing the null of $\beta = \beta_0$ are reported. Data are generated from equations (1) and (2), constant terms are set as $\alpha_1 = \alpha_2 = 0.2$, and the initial value x_0 is drawn from the standard normal distribution. A mild serial correlation is allowed in u_t as the VAR (1) process. Each simulation has 1000 replications. Nominal size is 10%.

The notation "Fully Modified" denotes Phillips and Hansen's (1990) method that assumes the data has an exact unit root even in the presence of a near unit root. We compute a Wald statistic. "Bonferroni" indicates the application of Bonferroni's procedure. We apply this method to the fully modified estimator, and construct valid confidence intervals for β. "Wright" indicates the stationarity test procedure proposed by Wright (2000). As our model includes a constant term in the cointegration equation, we use $L_2(\beta_0)$. We obtain $\widehat{u}_{2t}$ by the fully modified estimator. Then, we compute $\widehat{\sigma}_2^2$ with the Bartlett kernel as in Newey and West (1987).

Table 1 presents the size for the three procedures as a function of the true values of c and δ with the nominal size 10%. They are computed for the sample sizes of $T = 100, 150, 300$, and 500. The size of "Fully Modified" is computed as $\Pr\left[W_\beta > \chi^2_{0.9}(1)\,|H_0 \text{ is true}\right]$, where W_β is the Wald statistic and $\chi^2_{0.9}(1)$ is the 90th percentile of a chi-square distribution with the one degree of freedom. The size of "Bonferroni" is $\Pr\left[\beta_0 \notin C_\beta(\eta = 0.1)\,|H_0 \text{ is true}\right]$, where $C_\beta(\eta = 10)$ is the confidence interval of β by the Bonferroni's method with the level of 10%. The size of "Wright" is computed as $\Pr\left[L_2(\beta_0) > \chi^2_{0.9}(1)\,|H_0 \text{ is true}\right]$.

"Fully Modified" tells us how large the actual size can be when the presence of a near unit root is ignored. When $T = 100$, $\delta = 0.5$, and $c = -5$, the actual size is 29.5% against the nominal size of 10%, and for $c = -10$, it is as large as 44%. Increasing the sample size does not eliminate this size distortion problem. "Wright" handles the size relatively well. Because this procedure is based on the LMPI test, it is stable for any value of c and δ when the null hypothesis holds. "Bonferroni" also controls size well when the sample size is relatively large. As the sample size increases, the actual size becomes relatively stable around 10%. For example, when $T = 100$ and $c = -10$, the actual size is 0.291 for $\delta = 0.5$ and 0.464 for $\delta = 0.7$. As T increases, say, $T = 300$, sizes are 0.132 and 0.172, respectively.

Figures 1a and 1b present the local alternative power for the 10% level test of the null hypothesis of $\beta = \beta_0$ against the local alternative of $\beta = \beta_0 + g/T$, where $g = -20, -19, ..., 19, 20$. The local alternative power is computed for "Wright" and "Bonferroni." Power is evaluated for the sample sizes of $T = 100$ and 500, the local to unity parameter c of 0, -5, and -10, and zero frequency correlation δ of 0.3 and 0.7. Figure 1a presents the power for $T = 100$. Once the largest root deviates from unity, both procedures lose power. For "Wright", the loss of power is especially substantial. This relatively large loss of power is attributable to the fact the null hypothesis of the LMPI test is existence of stationarity. Once c falls below zero (or ρ is less than unity) and variables are in the direction of being stationary, the test tends to lose power to reject the alternatives of $\beta \neq \beta_0$. As for "Bonferroni," loss of power is relatively smaller when the long-run correlatioin is high ($\delta = 0.7$).

Figure 1b presents the local alternative power for $T = 500$. As the sample size increases, power improves for both "Wright" and "Bonferroni". Wright's power improvement for is especially large. For example, at $g = -20$, power increases from 40% to 80% for $c = -10$ and $\delta = 0.7$. Overall, however, "Bonferroni" outperforms "Wright" in power comparisons.

5 Conclusion

This paper has proposed a valid testing procedure on the cointegrating vector in the presence of a near unit root, applying Bonferroni's procedure. We first constructed a confidence interval for the largest root in the regressor of the cointegration equation, and then, constructed a valid confidence interval for the cointegration coefficient. Our simulation study showed that Bonferroni's procedure is helpful in reducing the size distortion problem when data may not have an exact unit root. The size distortion will be effectively eliminated in a larger sample size. Local alternative power is comparable with, or higher than the existing alternative procedure. However, this procedure is not an optimal test, because we do not know about the true value of the largest root in the regressor. Instead, we construct confidence intervals.

One way to improve the performance of Bonferroni's procedure is to construct more accurate confidence intervals for the largest root by using more powerful unit root test. In this paper, we have used the ADF test. Alternatively, one may use the Dickey-Fuller GLS (DF-GLS) test of Elliott, Rothenberg, and Stock (1996). The DF-GLS test is more powerful than the ADF test in the presence of deterministic terms because the DF-GLS test estimates deterministic terms more efficiently. By constructing a confidence interval of the largest root by DF-GLS, our procedure may improve.

References

Ahn, S. K. and G. C. Reinsel, 1990, Estimation for partially nonstationary autoregressive models, Journal of the American Statistical Association 85, 813-823.

Anderson, T. W., 1971, The statistical analysis oftime series, New York, Wiley.

Andrews, D. W. K., 1991, Heteroskedasticity and autocorrelation consistent covariance matrix estimation, Econometrica 59, 817-858.

Berk, K. N., 1974, Consistent autoregressive spectral estimates, The Annals of Statistics XLIII, 661-677.

Bobkoski, M. J. 1983. Hypothesis Testing in Nonstationary Time Series. Unpublished PhD thesis (Department of Statistics, University of Wisconsin, Madison, WI).

Campbell, J. Y. and M. Yogo, 2006, Efficient tests of stock return predictability, Journal of Financial Economics 81, 27-60.

Cavanagh, C., 1985, Roots local to unity, manuscript (Department of Economics, Harvard University, Cambridge, MA).

Cavanagh, C., G. Elliott, and J. H. Stock, 1995, Inference in models with nearly integrated regressors, Econometric Theory 11, 1131-1147.

Chan, N. H. and C. Z. Wei, 1987, Asymptotic inference for nearly nonstationary AR (1) processes, Annals of Statistics 15, 1050-1063.

Dickey, D. A. and W. A. Fuller, 1979, Distribution of the estimators for autoregressive time series with a unit root, Journal of the American Statistical Association 74, 427-31.

Elliott, G., 1998, On the robustness of cointegration methods when regressors almost have unit roots, Econometrica 66, 149-158.

Elliott, G., T. J. Rothenberg, and J. H. Stock, 1996, Efficient tests for an autoregressive unit root, Econometrica 64, 813-836.

Engle, R. F. and C. W. J. Granger, 1987, Cointegration and error correction: representation, estimation, and testing, Econometrica 55, 251-276.

Johansen, S., 1988, Statistical analysis of cointegrating vectors, Journal of Economic Dynamics and Control 12, 231-254.

Johansen, S. 1991, Estimation and hypothesis testing for cointegrating vectors in Gaussian vector autoregression models, Econometrica 59, 1551-1580.

Kwitokowski, D., P. C. B. Phillips, P. Schmidt, and Y. Shin, 1992, Testing the null hypothesis of stationarity against the alternative hypothesis of a unit root: how sure are we that economic time series have a unit root? Journal of Econometrics 54, 158-179.

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

Nyblom, J., 1989, Testing for the constancy ofparameters over time, Journal of the American Statistical Association 84, 223-230.

Park, J. Y., and P.C. B. Phillips, 1988, Statistical inference in regressions with integrated processes: I, Econometric Theory 4, 467-497.

Phillips, P. C. B., 1987, Towards a unified asymptotic theory for autoregression, Biometrika 65, 533-547.

Phillips, P. C. B. and B. E. Hansen, 1990, Statistical inference in instrumental variables regression with I(1) processes, Review of Economic Studies 59, 99-125.

Phillips, P. C. B. and M. Loretan, 1991, Estimating long-run economic equilibria, Review of Economic Studies 59, 407-436.

Rossi, B., 2007, Expectations hypothesis at long horizons, Econometrics Journal 10, 554-579.

Saikkonen, P.,1991, Asymptotically efficient estimation of cointegration regressions, Econometric Theory 7, 1-21.

Saikkonen, P., 1992, Estimation and testing of cointegrated systems by an autoregressive approximation, Econometric Theory 8, 1-27.

Stock, J. H., 1991, Confidence intervals for the largest autoregressive roots in macroeconomic time series, Journal of Monetary Economics 28, 435-460.

Stock, J. H. and M. W. Watson, 1993, A simple estimator of cointegrating vectors in higher order integrated systems, Econometrica 61, 783-820.

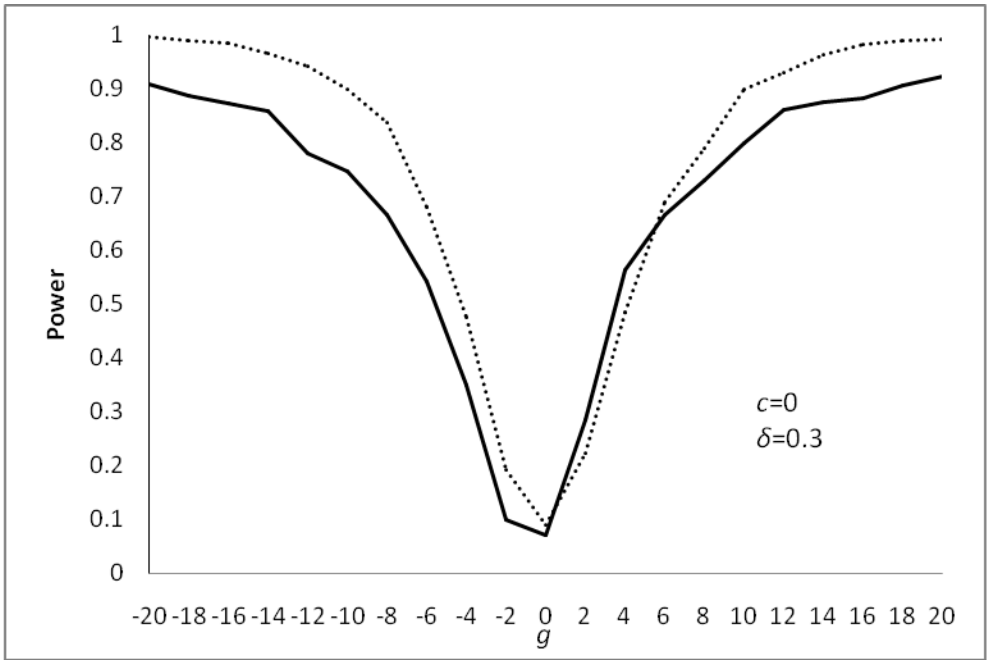
Torous, W., R. Valkanov, and S. Yan, 2004, On predicting stock returns with nearly integrated explanatory variables, Journal of Business 77, 937-965.

Wright, J. H, 2000, Confidence sets for cointegrating coefficients based on stationarity tests, Journal of Business and Economic Statistics 18, 211-222.

Table 1. Size. Rejection rates of test of $\beta = \beta_o$ in the finite sample with nominal size 10%

	Fully Modified			Wright			Bonferroni		
δ / c	0	0.5	0.7	0	0.5	0.7	0	0.5	0.7
T=100									
0	0.154	0.194	0.124	0.092	0.085	0.073	0.102	0.102	0.112
-5	0.158	0.295	0.440	0.089	0.083	0.103	0.143	0.187	0.228
-10	0.185	0.438	0.779	0.085	0.094	0.110	0.141	0.291	0.464
-20	0.193	0.653	0.935	0.094	0.112	0.147	0.144	0.583	0.849
T=150									
0	0.121	0.148	0.085	0.087	0.098	0.080	0.092	0.104	0.113
-5	0.133	0.219	0.380	0.095	0.099	0.098	0.103	0.119	0.143
-10	0.134	0.384	0.631	0.086	0.093	0.115	0.112	0.222	0.343
-20	0.174	0.671	0.919	0.105	0.121	0.149	0.139	0.498	0.771
T=300									
0	0.111	0.093	0.103	0.125	0.124	0.136	0.078	0.111	0.148
-5	0.123	0.188	0.513	0.144	0.136	0.139	0.090	0.089	0.105
-10	0.133	0.353	0.659	0.117	0.138	0.155	0.103	0.132	0.172
-20	0.155	0.588	0.899	0.137	0.136	0.167	0.118	0.288	0.482
T=500									
0	0.103	0.130	0.079	0.206	0.175	0.191	0.075	0.119	0.208
-5	0.118	0.203	0.363	0.188	0.184	0.213	0.088	0.056	0.089
-10	0.117	0.301	0.637	0.175	0.194	0.207	0.100	0.117	0.134
-20	0.106	0.572	0.901	0.194	0.197	0.205	0.106	0.217	0.316

Figure 1a: Local Alternative Power of 10 % level testing $\beta = \beta_o + g/T$, T=100, c=0. Solid line, Wright; Dots, Bonferroni.



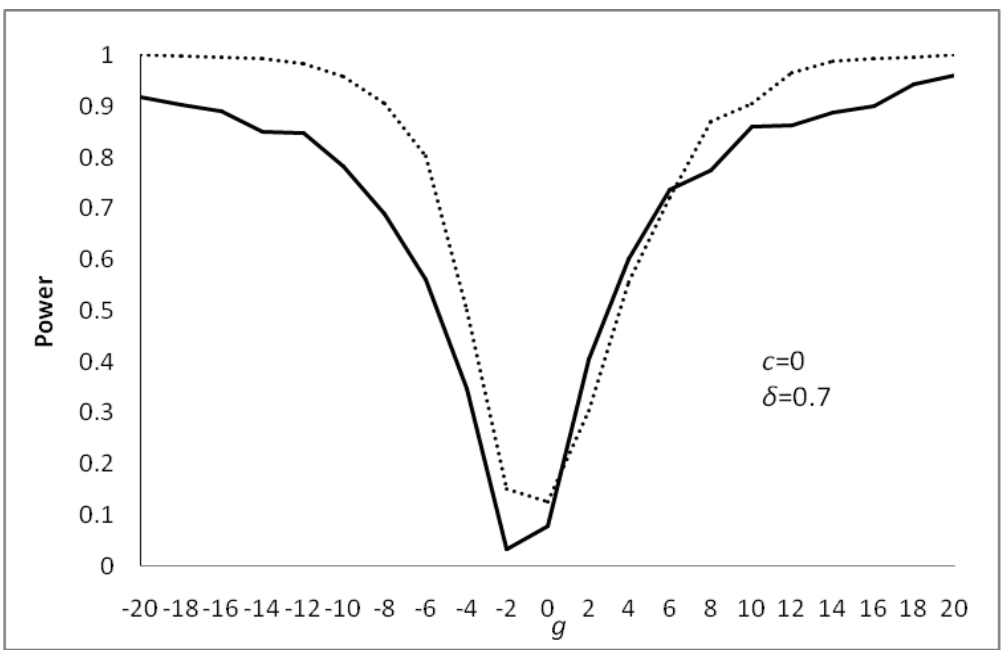
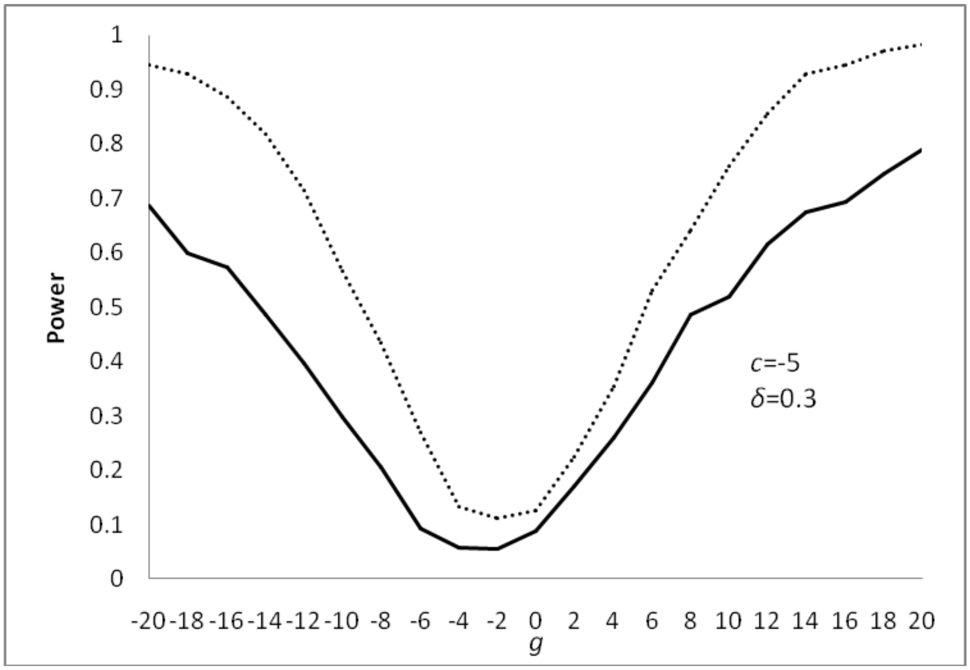


Figure 1a (continued): Local Alternative Power of 10 % level testing $\beta = \beta_o + g/T$, T=100, c=-5. Solid line, Wright; Dots, Bonferroni.



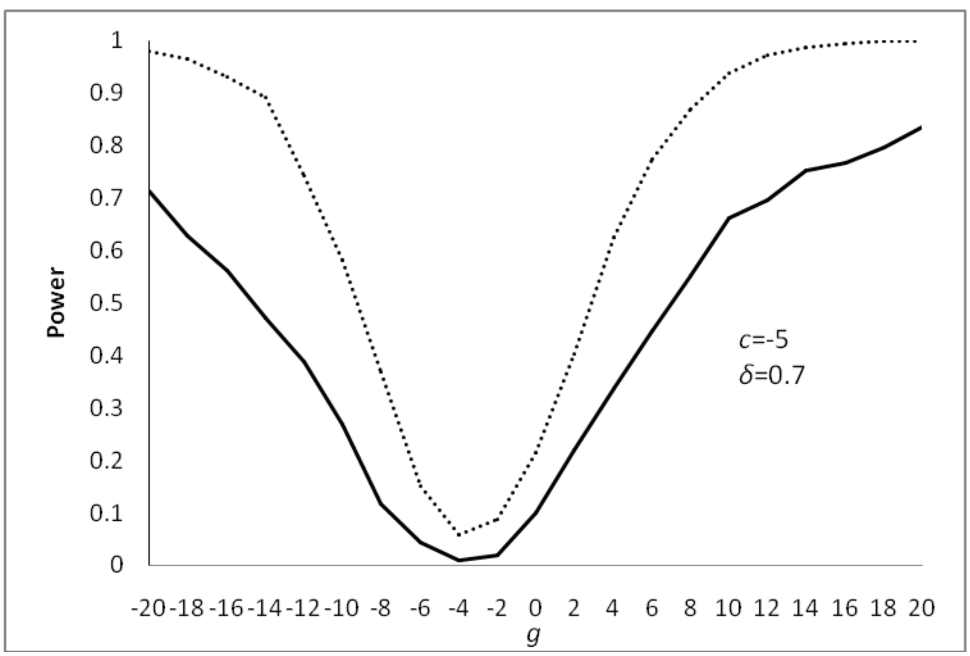
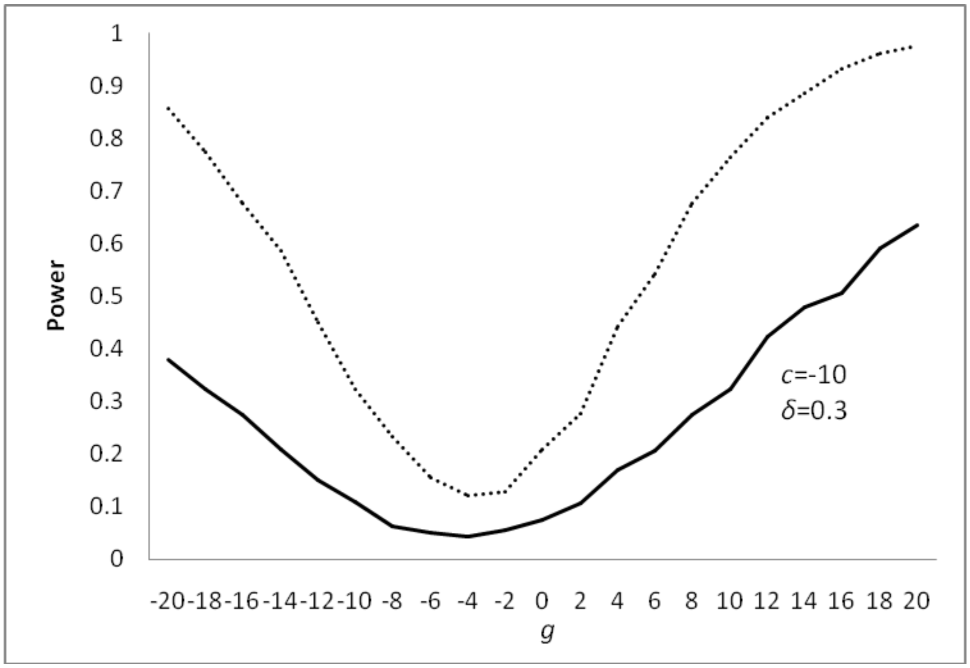


Figure 1a (continued): Local Alternative Power of 10 % level testing $\beta = \beta_o + g/T$, T=100, c=-10. Solid line, Wright; Dots, Bonferroni.



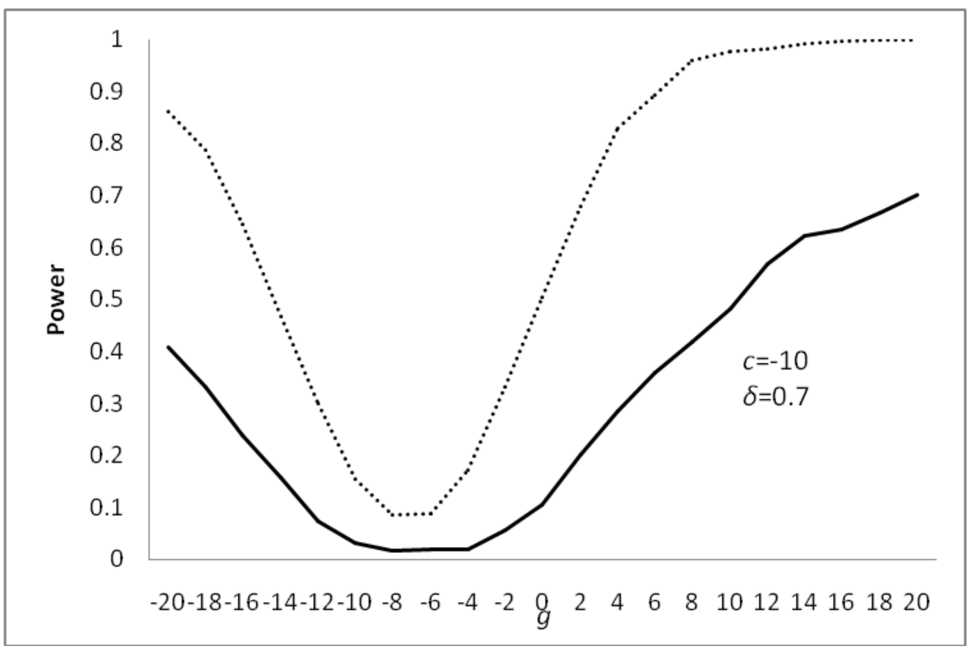


Figure 1b: Local Alternative Power of 10 % level testing $\beta = \beta_o + g/T$, T=500, c=0. Solid line, Wright; Dots, Bonferroni.

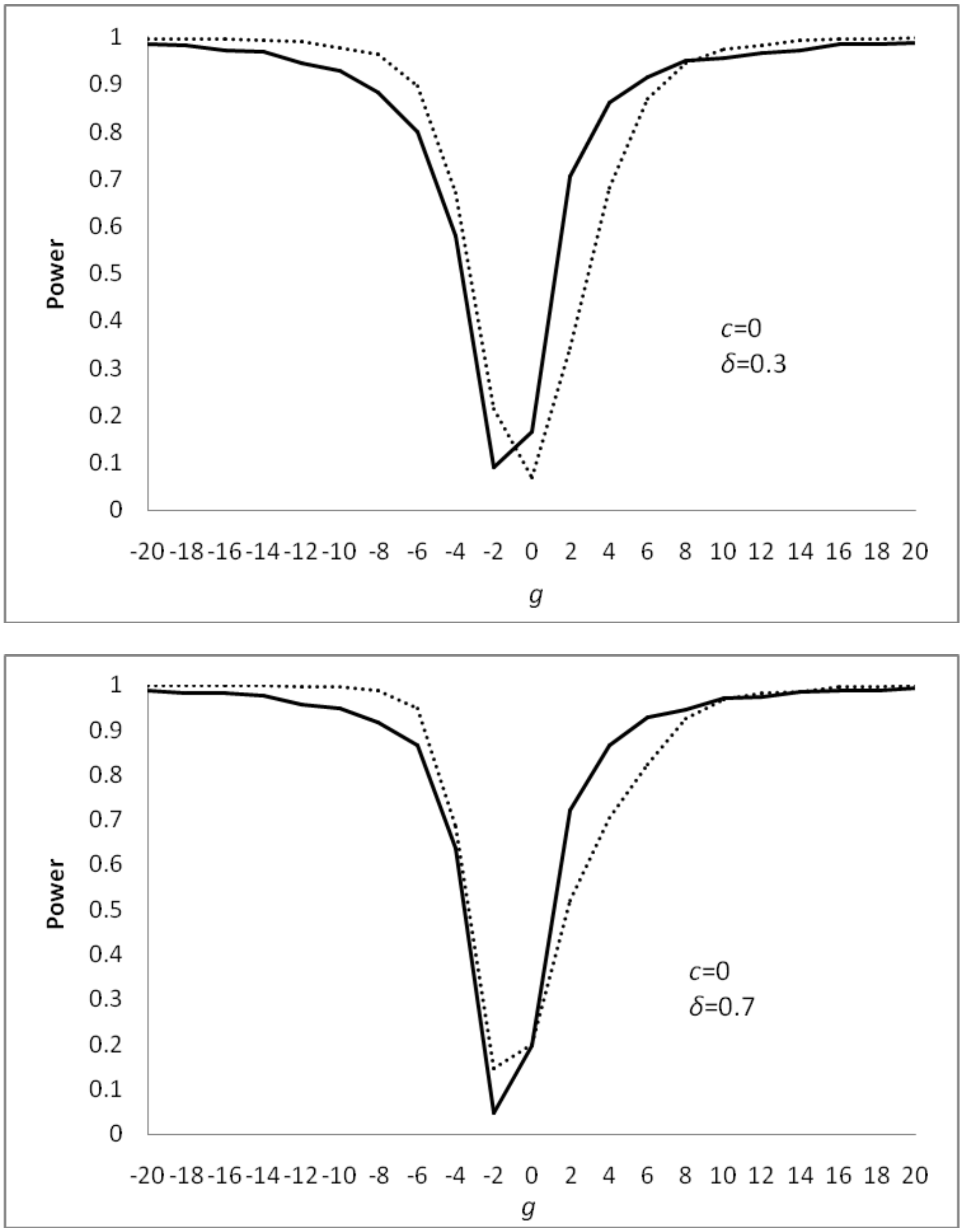


Figure 1b (continued): Local Alternative Power of 10 % level testing $\beta = \beta_o + g/T$, T=500, c=-5. Solid line, Wright; Dots, Bonferroni.

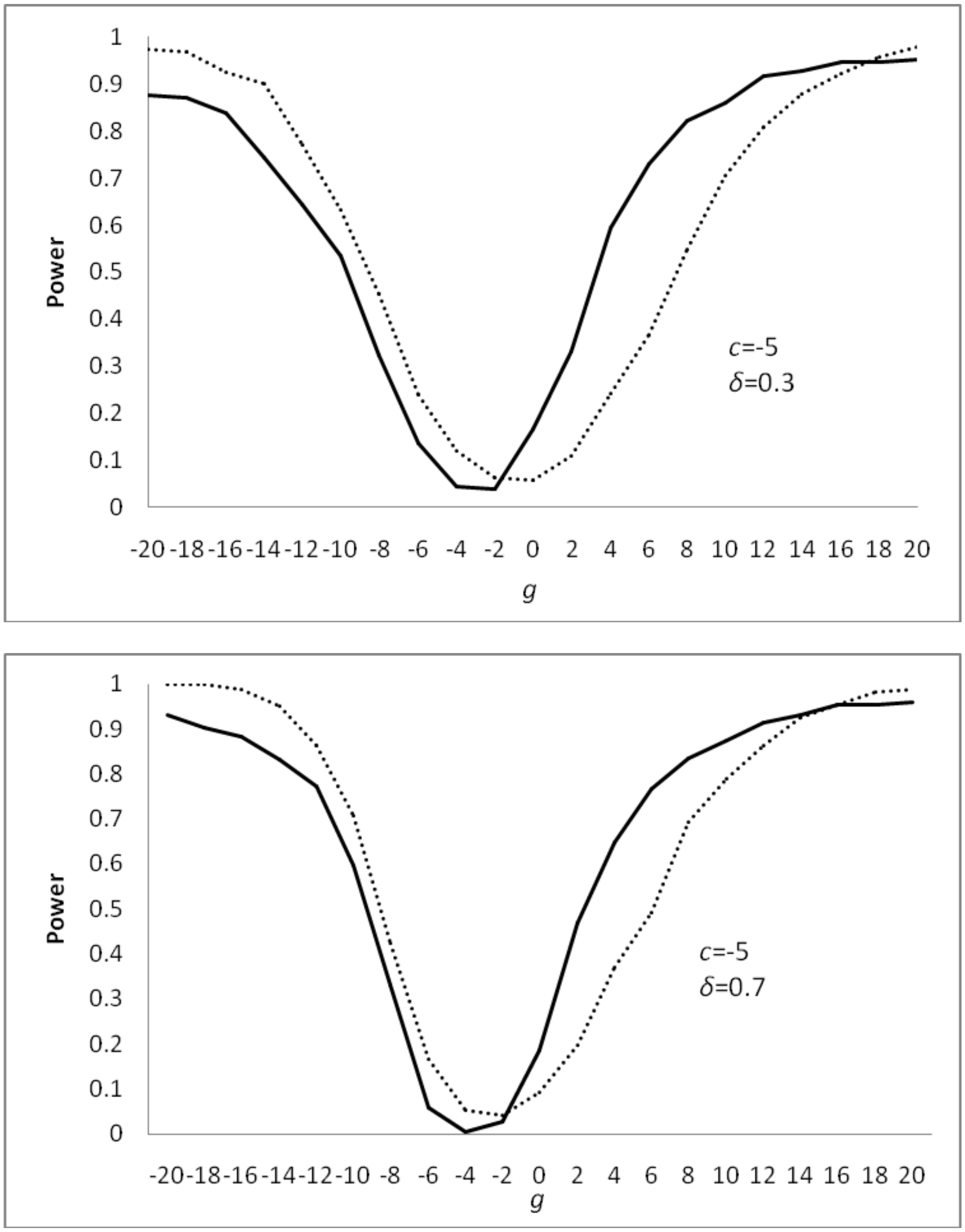


Figure 1b (continued): Local Alternative Power of 10 % level testing $\beta = \beta_o + g/T$, T=500, c=-10. Solid line, Wright; Dots, Bonferroni.