Testing Parameter Constancy When the Regressor May Have a Near Unit Root

Masako Miyanishi*†

Department of Economics, University of Georgia

February, 2009

Abstract

This paper considers testing a long run stable relationship among variables in the presence of a near unit root. Since each variable may have a unit root or near unit root, the the conventional asymptotic theory such as Andrews (1993, Econometrica) is not applicable. As an alternative, we propose modified fixed regressor bootstrap. Fixed regressor bootstrap is originally proposed by Hansen (2000, Journal of Econometrics), and it assumes strict exogeneity. We modify the test so that it is applicable in the presence of endogeneity and serial correlation in the cointegration equation error. In empirical study, we test the present value model and the expectations hypothesis. For the present value model, we detect change in the relationship in the early 1970's. For the expectations hypothesis, we fail to detect a change in the late 1970's when the Fed changed the operating procedure for the relatively short yields.

Keywords: Parameter instability; Bootstrap; Cointegration; Near unit root

JEL Classification: C22

*I am grateful to Graham Elliott for his advice. I also thank Elena Pesavento and Guido Kuersteiner for helpful comments

†Corresponding auther: Masako Miyanishi, Dept of Economics, University of Georgia, Athens GA 30602 U.S.A., Phone: 1(706)542-3689, E-mail: masako@terry.uga.edu

1 Introduction

Engle and Granger (1987) defined cointegration as a long run equilibrium relationship among unit root variables. The cointegrating vector that captures the long run relationship should be constant over time.

Some authors have relaxed this strict definition of cointegration and investigate the long run relationship among variables, allowing a structural change in the model. For example, Hansen (1992) proposed a procedure for testing the null hypothesis of a constant cointegration coefficient (cointegration) against the alternative hypothesis of a random walk coefficient (that is, no cointegration) by extending $supF$ test, mean F test, and LM test to an $I(1)$ regressor. Gregory and Hansen (1996) proposed the residual based tests such as the modified augmenting Dickey Fuller (ADF) and Ztests that allow the possibility of a regime shift (change in a cointegrating vector) under the alternative hypothesis. They found the supporting evidence for money demand equation when they allow a regime shift. Inoue (1999) designed a test so that it has power against cointegration with a broken trend. He tested money demand equation and found a supporting result for the finding in Campbell and Perron (1991). He found money, output and interest rate are cointegrated with a broken deterministic trend. Seo (1998) applied LM test and found the supporting result for money demand equation. Hansen (2003) generalized Johansen's (1989) VAR test allowing structural changes in various parameters in the model. He tested the expectations hypothesis and found the cointegrating vector and common factors are constant in the long run while there are structural changes in covariance and term premium. Their findings are quite interesting. If they allow structural change in the model, they find a long run stable economic relationship. Kejriwal and Perron (2008 a and b) extend Bai and Perron's (1998) work to test multiple structural breaks when $I(1)$ regressors are included in the model. They propose tests for the number of structural changes and offer the construction of joint confidence intervals for break dates.

This paper also relaxes the strict concept of cointegration by Engle and Granger's (1987). Our aim is to test the long-run stable relationship among variables, allowing the presence of a near unit root in data or allowing data to switch between $I(1)$ and near $I(1)$. It is important to consider the possibility that the largest root is close to but not exactly unity especially for empirical application. Because we have the limited number of observations in practice, it is difficult to tell from the sample if the data diverge away in the long run. The behavior of data may be different from an exact unit root process. When testing the cointegrating vector, the limit distribution of the test statistics will be nonstandard if the variable is not exactly integrated (Elliott (1998)). It is shown that size distortion is quite large even if the efficient class estimator is employed. If the largest root experiences a structural change or deviates away from unity, constancy of a cointegrating vector might be rejected even if there actually exists a stable long-run relationship among variables. The presence of a near unit root is a plausible concern in empirical application. Furthermore, because we examine a long span of data for testing cointegration, there may be the possibility that the marginal distribution of data experiences structural change, switching between $I(1)$ (or near $I(1)$) and $I(0)$. Consider the expectations hypothesis of the term structure. It is theoretically hard to believe that interest rates have a unit root; it is difficult to expect they diverge away in the long run. Nevertheless, the hypothesis is often tested in a cointegration framework, since a unit root test tends to fail to reject the presence of a unit root. And the hypothesis is often rejected in empirical study. If we construct confidence intervals for the largest autoregressive root in data, there seems to be a structural change in the interest rate around the end of 1979 when the Federal Reserve ceased targeting. We can observe that, while the interval tends to be narrow around unity during the targeting period, it becomes wider, that is, the largest root is smaller than unity after the targeting period ends. Hansen (2003) indeed found a structural change in term premium and variance around the end of 1979. In order to take into account the presence of a near unit root, we utilize the local-to-unity asymptotic theory (Bobkoski (1983), Cavanagh (1985), and Phillips (1987)).

One of major contributions in this paper is that we propose a test for a structural change in the long-run relationship among variables when the largest root is close to, but may not be exact unity. We show the asymptotic distribution of $\sup F$ statistic depends on two nuisance parameters and break date in a complicated manner, and the asymptotic theory in Andrews (1993) is not applicable. As an alternative testing procedure, we propose modified fixed regressor bootstrap. Fixed regressor bootstrap is originally proposed by Hansen (2000), assuming strict exogeneity. We extend the procedure so that it is robust to endogeneity and serial correlation in a cointegration error and consider application of Bonferroni test to allow the presence of a near unit root. We investigate finite sample performance. We show that there is a large power gain by adding quasi-difference term to a cointegration equation especially when endogeneity is large. Size distortion problem is also reduced by our modification. When a serial correlation is allowed, there is a power loss especially when endogeneity is small. We find size of test stays around the nominal size even if there is a serial correlation.

Another contribution is empirical application. We test the present value model in the stock market and the expectations hypothesis. For the present value model, we detect a structural change in the relationship in the early 1970's. For the expectations hypothesis, we fail to detect a change in the late 1970's when the Fed changed the operating procedure but a structural change in the late 1980's.

The rest of this paper is organized as follows; section 2 presents the model we consider, section 3 presents the asymptotic distribution of $\sup F$, section 4 considers application of fixed regressor bootstrap proposed by Hansen (2000), section 5 small sample performance, section 6 applies the fixed regressor bootstrap empirically, and section 7 concludes.

2 The Model

We consider a simple triangular model;

$$x_t = \alpha_1 + \rho x_{t-1} + u_{1t} \quad (1)$$

$$y_t = \alpha_2 + \beta x_t + u_{2t} \quad (2)$$

for $t = 1, .., T$. It is a simple bivariate model. The first equation is an innovation of the regressor, x_t. The largest autoregressive root in x_t is ρ. The second equation is a cointegration equation. The deterministic terms are constants $\alpha_1$and α. The error terms $u_t = [u_{1t}, u_{2t}]'$ are stationary, and $\Phi(L)u_t = \varepsilon_t$. where $\varepsilon_t = [\varepsilon'_{1t}, \varepsilon'_{2t}]'$ and $\Phi(L)$ is a lag polynomial in the lag operator L. We have the following assumptions throughout the paper

A1: $|\Phi(L)| = 0$ has roots outside the unit circle.

A2: $E_{t-1}(\varepsilon_t) = 0, E(\varepsilon_t \varepsilon'_t) = \Sigma$, and $\sup_t ||\varepsilon_t||^{4+\gamma} < \infty$ (a.s.) for some $\gamma > 0$, where Σ is positive definite and $E_{t-1}(.)$ denotes conditional expectation with respect to $\{\varepsilon_{t-1}, \varepsilon_{t-2}, ...\}$

A3: $\max_{-k \leq t \leq 0} || [u_{1t}, u_{2t}]' || = O_p(1)$, where $||.||$ is the Euclidean norm.

The assumptions are fairly standard. A1 is stationarity condition. A2 implies that $\{\varepsilon_t\}$ follows the martingale difference sequence. A3 ensures that the initial values are asymptotically negligible.

Define 2π times spectral density of u_t at frequency zero as $\Omega = \Phi(1)^{-1}\Sigma\Phi(1)^{-1\prime}$, where $\Phi(1) = \sum_i \Phi_i$. The scaled long run variance-covariance matrix can be denoted as $\Omega = \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix} = \begin{bmatrix} \sigma_{11}^2 & \delta\sigma_{11}\sigma_{22} \\ \delta\sigma_{11}\sigma_{22} & \sigma_{22}^2 \end{bmatrix}$, where $\frac{\Omega_{12}}{\Omega_{11}^{1/2}\Omega_{22}^{1/2}} = \delta$ is the zero frequency correlation.

If we allow the regressor to have a near unit root, two nuisance parameters will have a crucial role for the distribution of test for parameter constancy as we see in the following section. First one is a local to unity parameter under the local to unity approximation. Since it is possible in our setting that the largest autoregressive root is equal to or less than unity, we employ the local-to-unity asymptotic theory (Bobkoski (1983), Cavanagh (1985), Phillips (1987)), with which we reparameterize ρ as $\rho = 1 + \frac{c}{T}$, where c is a local to unity parameter, which can be zero or negative. The local-to-unity approximation is useful because we can avoid discontinuity between a unit root and stationarity. We may regard this reparameterization as local alternative to a unit root hypothesis. A problem is that c is not consistently estimable. If c is estimated as $\widehat{c} = T(\widehat{\rho} - 1)$, where $\widehat{\rho}$ is OLS estimate, then, it tends to be underestimated, which may lead us to an incorrect inference.

Another nuisance parameter is the long-run correlation δ. The long run variance covariance matrix Ω is consistently estimable by using methods such as the sum of autocovariances such as Newey and West (1987), the inversion of vector autoregressive coefficients by Berk (1974), and pre-whitening methods by Andrews and Monahan (1991). Because δ tends to be high in a cointegration mode, the regressor x_t may not be orthogonal to u_{2t}.

We are interested in constancy of the cointegrating coefficient, β, over the entire sample. We would like to test;

$$H_0: \quad \beta = \beta_o \quad \text{for all } t \qquad vs. \qquad H_1: \quad \left\{ \begin{array}{ll} \beta = \beta_o & \text{for } t < \tau_o \\ \beta = \beta_o + \theta_T & \text{for } t \geqslant \tau_o \end{array} \right. \tag{3}$$

where τ_o is the unknown true timing of a structural change and θ is the magnitude of the change. Thus we test constancy of β, and are interested in testing $H_0 : \theta = 0$ for all t against $H_1 : \theta_T \neq 0$ for $t \geqslant \tau_o$. H_1 is local alternative and we assume $\theta_T = \zeta\sigma_2/T$, where constant ζ indicates magnitude of a structural change. Specification of $\theta_T = \zeta\sigma_2/T$ is necessary to derive asymptotic distribution of the test statistic under the alternative hypothesis.

Testing constancy of a parameter has been extensively studied. Among the contributions, for testing a structural change with unknown timing, Quandt (1960) proposed taking the supremum of $F - statistc$ over all the possible break dates, and the asymptotic theory for this Quandt (1960) test is provided by Andrews (1993). The exponentially weighted tests are considered by Andrews and Ploberger (1994).

This paper considers $\sup F$ tests. However, the asymptotic theory given by Andrews (1993) is not applicable here. Andrews (1993) assumes stationarity of the regressor. To apply Andrews' (1993) asymptotic theory, the second moments of x_t have to grow linearly, that is, if we denote $M = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} E(x_t^2)$, then, $\frac{1}{T}\sum_{t=1}^{Tr} x_t^2$ should weakly converge to rM for $0 < r < 1$. In this paper, we assume the regressor has a unit root or a near unit root. Thus, $\frac{1}{T^2}\sum_{t=1}^{Tr} x_t^2$ weakly converges to $\int_0^r J^{\mu}(a)^2 da$, where $J^{\mu}(.)$ is a demeaned diffusion process that satisfies $dJ(a) = cJ(a)da + dW(a)$ and $W_1(.)$ is a Brownian motion associated with u_{1t}, and $J^{\mu} = J(\lambda) - \int_0^r J(a)da$. The second moments of the regressor do not grow linearly.

3 Asymptotic Distribution When the Regressor is I(1) or near I(1)

As mentioned in the previous section, when x_t is integrated or nearly integrated, the asymptotic (unconditional) distribution of $\sup F$ will be different from one derived by

Andrews (1993). This section derives the asymptotic distribution of *supF* test when the regressor has a near unit root. It depends on two nuisance parameters c and δ in a complicated manner.

When the true timing of a structural change is unknown, a typical test for a parameter constancy is to take the Wald statistic,

$$F_t = \frac{(T-m)\widehat{\sigma}_2^2 - (T-2m)\widehat{\sigma}_{2t}^2}{\widehat{\sigma}_{2t}^2} \tag{4}$$

for all the possible break dates, t, and take the supremum of the statistics, $\sup F$, as suggested by Quandt (1960). $\widehat{\sigma}_2^2$ is estimated variance of u_{2t} under the null hypothesis, while $\widehat{\sigma}_{2t}^2$ is estimated variance of u_{2t} under the local alternative $H_1 : \theta_T \neq 0$.

Lemma 1 *We have the following convergence results.*[1] *As* $T \to \infty$,

$$\begin{array}{ll} i) & \frac{1}{T^{1/2}} x_t \Rightarrow \sigma_{11} J_c(r) \\ ii) & \frac{1}{T^2} \sum_{t=1}^{[Tr]} x_t^2 \Rightarrow M(r) = \sigma_{11}^2 \int_0^r J_c(s)^2 ds \\ iii) & \frac{1}{\sigma_{22} T} \sum_{t=1}^{[Tr]} x_t u_{2t} \Rightarrow N(r) = \sigma_{11} \int_0^r J_c(\lambda) dW_{2.1} + \frac{\delta \sigma_{2.1}}{\sqrt{(1-\delta^2)}} \left(\int_0^r J_c(\lambda) dW_1 + r \frac{S_{11}^2}{\sigma_{11}^2} \right) \end{array} \tag{5}$$

where $\Rightarrow$ *indicates weak convergence, for* $0 < r < 1$, $t = 1, ..., T$. $S_{11}^2 = \lim_{T\to\infty} \frac{1}{[Tr]} \sum_1^{[Tr]} u_{1t}^2$. $\sigma_{2.1}$ *is the long run variance of* $u_{2.1t}$ *that is orthogonal to* u_{1t-j} *for all* $j = 0, \pm 1, \pm 2, ...$ *We have* $W_2(.) = W_{2.1}(.) + \delta \frac{\sigma_{22}}{\sigma_{11}} W_1(.)$ $W_{2.1}(.)$ *is independent of* $W_1(.)$. $S_{11}^2 = \lim_{T\to\infty} \frac{1}{[Tr]} \sum_1^{[Tr]} u_{1t}^2$. *Proof is given in appendix.*

Lemma 1 shows that, if $c \neq 0$, the second moment of x_t does not converge to a Brownian motion. And unless x_t is strictly exogenous, $N(r)$ depends on δ, and c. Note that if $\delta = 0$, the second term in $N(r)$ disappears.

[1] The results are derived ignoring the deterministic term. We can get the same results for $x_t^\mu = x_t - \alpha_1$

Theorem 2 *$F-$statistic will be;*

$$F_{[Tr]} \Rightarrow (N^*(r) - Q(r)\zeta)' M^*(r)^{-1}(N^*(r) - Q(r)\zeta) \tag{6}$$
$$\equiv F(r|\zeta) \tag{7}$$

where $N^*(r) = N(r) - M(r)M(1)^{-1}N(1)$, $Q(r) = (M(r) - M(r)M(1)^{-1}M(\tau_0)) - (M(r) - M(\tau_0))I(r \geq \tau_0)$, and $M^*(r) = M(r) - M(r)M(1)^{-1}M(r)$, where $\tau_0 = \frac{t_0}{T}$. Then, $\sup F$ *will converge to* $\sup F(r|\zeta)$ *in distribution.*

Theorem above tells us that the asymptotic distribution $\sup F(r|\zeta)$ has a complicated form that depends on two nuisance parameters and timing of a structural change. Under the null hypothesis $F_{[Tr]}$ weakly converges to $N^*(r)'M^*(r)^{-1}N^*(r)$. Under the alternative, τ_0, true timing of a structural change also appears as a parameter.

Because the asymptotic distribution under the null hypothesis is now known, it is possible to generate the distribution by simulation and tabulate critical values for any plausible combination of δ and c. However, it might be difficult to use them in practice. The reason is that a local to unity parameter is not consistently estimable. Thus, we need an alternative testing method which is robust to c.

We would like to test a structural change in the relationship among variables with unknown timing. Thus the test we are interested in is similar to the one in Hansen (1992) but more generalized one in the sense that we allow the presence of a near unit root.

4 Fixed Regressor Bootstrap

We have learned that if the regressor does not have an exact unit root and there is endogeneity in the cointegration equation, the asymptotic null distribution depends on nuisance parameters in a complicated manner. It might not be very useful to generate the (unconditional) null distribution by simulation and tabulate critical values. Alternatively, we consider Hansen's (2000) fixed regressor bootstrap.

4.1 Fixed Regressor Bootstrap

Because the asymptotic distribution depends on parameters in a complicated manner, the bootstrap will be a useful method (see Hall (1994)). Although the bootstrap method is often found invalid when the data is $I(1)$ (see Basawa et al. (1991)), Hansen's (2000) fixed regressor bootstrap is applicable to nonstationary data.

This bootstrap treats the regressor as fixed (exogenous). Thus, this method assumes a strict exogeneity. We replicate the bootstrap distribution of $\sup F$ denoted as $\sup F(b)$ conditional on data. Then,we count how many percent of the replicated distribution exceeds the sample test statistic $\sup F_T$, that is, $p-$value is computed (or we have critical values from the replicated bootstrap distribution). I briefly describe the how to compute homoskedastic fixed regressor bootstrap statistic. First, a random sample $\{y_t\,(b) : t = 1, ..., T\}$ is generated from the $N(0, 1)$ distribution. By regressing $y_t\,(b)$ on x_t to get residual variance $\widehat{\sigma}^2(b)$, and regress $y_t\,(b)$ on x_t and $x_t I(t \geqq \tau)$ to get the residual variance $\widehat{\sigma}^2_\tau(b)$ and the sequence of Wald statistics;

$$F_t\,(b) = \frac{(T-m)\widehat{\sigma}^2(b) - (T-2m)\widehat{\sigma}^2_\tau(b)}{\widehat{\sigma}^2_\tau(b)} \tag{8}$$

The bootstrap test statistic is $\sup F_T(b) = \sup_{\tau_1 \leq \tau \leq \tau_2} F_\tau\,(b)$. While the true distribution of $\sup F_T(b)$ is unknown, we may obtain the distribution by simulation. Theorem 5 in Hansen (2000) ensures the replicated bootstrap distribution $\sup F_T(b)$ converges to the null distribution in probability. Because $y_t\,(b)$ is generated independently from data, x_t is treated as exogenous. Since the fixed regressor bootstrap method conditions on x_t, it generates a conditional distribution, while the sample statistic $\sup F_T(b)$ is a draw from unconditional distribution.

We would like to apply the fixed regressor bootstrap to testing parameter constancy in the cointegration relationship. Since this method treats x_t as fixed, the magnitude of c will not affect the inference. Furthermore, it should be robust to a structural change in the marginal distribution of x_t. A problem is strict exogeneity assumption. Suppose we define the regression error $e_t(b) = y_t(b) - \beta x_t$, then, $e_t(b)$ is orthogonal to x_t because $y_t(b)$ is generated independently. Hence, $\frac{1}{T}\sum_1^{[Tr]} x_t e_t(b)$ will converge to $\int_0^r J(a)dW_e$, where W_e is the standard Brownian motion associated with $e_t(b)$ and it can be different from $N(r)$ for the unconditional distribution of $\sup F$ derived in the previous section. Because we test parameter constancy in a cointegration framework, the long run correlation δ is likely to be nonzero, which implies endogeneity in the cointegration equation. Thus, fixed regressor bootstrap is not directly applicable to our model.

4.2 Application of Fixed Regressor Bootstrap

The distribution of $supF$ depends on two nuisance parameters c and δ under the null hypothesis. Local to unity parameter c is not consistently estimable. Nonzero long run correlation δ implies endogeneity in the cointegration equation. We apply fixed regressor bootstrap, hoping to make a correct inference. The method would automatically resolve the first issue. Since fixed regressor bootstrap treats as if x_t were fixed, it will correctly replicate the null distribution of $\sup F_T(b)$ even though the regressor is nonstationary or near nonstationary. However, the method assumes strict exogeneity that requires that the cointegration equation error be orthogonal to regressors. Hence, if we stick with the original cointegration equation (2), where usually $\delta \neq 0$ and $E(u_{2t}|x_t) \neq 0$, the fixed regressor bootstrap will not correctly yield the null distribution.

First we assume no serial correlation in u_t, for a simple illustration purpose. We modify the cointegration equation (2) to make the fixed regressor bootstrap applicable. For simplicity we ignore deterministic terms and assume that u_t is white noise and there is only contemporaneous correlation between u_{1t} and u_{2t}.In order to orthogonalize the error term to the regressor, we add a quasi-differnce term $(1-\rho L)x_t$ to the equation $(2)^2$. Now the equation is;

$$y_t = \beta x_t + \delta \frac{\sigma_{22}}{\sigma_{11}}(1-\rho L)x_t + u_{2.1t} \tag{9}$$

To see why $u_{2.1t}$ is orthogonal to x_t, consider rotating error terms $u_t = [u_{1t}, u_{2t}]'$ so that $Ru_t = \begin{pmatrix} u_{1t} \\ u_{2.1t} \end{pmatrix}$, where $R = \begin{pmatrix} 1 & 0 \\ -\delta\frac{\sigma_{22}}{\sigma_{11}} & 1 \end{pmatrix}$. Then $E[Ru_t u_t' R'] = \begin{pmatrix} \sigma_{11}^2 & 0 \\ 0 & \sigma_{2.1}^2 \end{pmatrix}$, where $\sigma_{2.1}^2$ is a long-run variance of $u_{2.1t}$. Thus, $u_{2.1t} = u_{2t} - \delta\frac{\sigma_{22}}{\sigma_{11}}u_{1t}$, and $E(u_{1t}u_{2.1t}) = 0$,where $u_{1t} = (1-\rho L)x_t$. And x_t will be orthogonal to $u_{2.1t}$. This method is similar to that of Saikkonen (1991) and Stock and Watson (1993), which adds first difference of x_t to the cointegration equation. We cannot add leads and/or lags of the first fist difference here because ρ is allowed to deviate from exact unity.

Now we allow serial correlation in u_t. Denoting $E(u_t u_{t+k}') = \Gamma(k)$, we add another assumption (See Saikkonen (1991)).

A4: $\sum_{j=-\infty}^{\infty} ||\Gamma(k)|| < \infty$

Then, in general, we have $u_{2t} = \sum_{j=-\infty}^{\infty} \pi_j u_{1t-j} + u_{2.1t}$, where $\sum_{j=-\infty}^{\infty} ||\pi_j|| < \infty$, $u_{1t-j} = (1-\rho L)x_{t-j}$ and $u_{2.1t}$ is a stationary process such that $E(u_{1t+k}u_{2.1t}) = 0,\ \ k = 0, \pm 1, \pm 2, ...$ The spectral density of $u_{2.1t}$ is $f_{2.1}(\lambda) = f_{u_2u_2}(\lambda) - f_{u_2u_1}(\lambda)f_{u_1u_1}^{-1}(\lambda)f_{u_1u_2}(\lambda)$ and $2\pi f_{2.1}(0) = \Omega_{2.1} = \Omega_{22} - \Omega_{21}\Omega_{11}^{-1}\Omega_{12}$.

Now we have

$$y_t = \beta x_t + \sum_{j=-\infty}^{\infty} \pi_j u_{1t-j} + u_{2.1t} \tag{10}$$

[2] Because we assume no serial correlation in error terms, we add one quasi-difference term. If we allow serial correlation, we need to add more lags and leads of the quasi-difference term.

Because $\{\pi_j\}$ is absolute summable, $\pi_j \approx 0$ for $|j| > K$ and K sufficiently large. Thus, we can truncate the sum at $|j| = K$ in practice. Now

$$y_t = \beta x_t + \sum_{j=-K}^{K} \pi_j u_{1t-j} + \widetilde{v}_t \tag{11}$$

We also need to assume;

A5: $T^{-1/3}K \to 0$ and $K \to \infty$ as $T \to \infty$.

Thus, we need knowledge of ρ to apply fixed regressor bootstrap. In practice, however, we do not know the true value of ρ. ρ is estimable by the usual OLS, and the estimate $\widehat{\rho}$ converges at a faster rate if true value is unity. But in finite sample, $\widehat{\rho}$ tends to be biased. Furthermore, there is discontinuity between $\rho = 1$ and$\rho < 1$ under the usual asymptotic theory. Thus, we propose constructing a confidence interval for ρ with the local-to-unity asymptotic theory instead of using $\widehat{\rho}$. A valid confidence interval for c can be constructed by inverting the ADF t-statistic on ρ and by collecting the set of values for c that cannot be rejected in a hypothesis test that $c = c_o$. In the next section, we explain how an equally tailed confidence interval for ρ (and c) is constructed by inverting the augmented Dickey-Fuller (ADF) test or DF-GLS test (Stock (1991), Elliott, Rothenberg, and Stock (1996)). We should keep in mind that because true value of ρ is unknown, accuracy of test might be lost as the number of leads and lags increases.

5 Finite Sample Performance

We evaluate both size and power to investigate finite sample performance. We study different combinations of nuisance parameters c and δ. $c = 0, -1, -5$,or -10. $\delta = 0.3$, or 0.7. We replicate the simulated (unconditional) distribution of $\sup F$ statistics. For each experiment, simulation replication is 1000. Under the null hypotheses, $\beta_o = 1$. Equations

(1) and (2) include constant terms. We set $\alpha_1 = \alpha_2 = 0.2$. For no serial correlation case, the sample size is $T = 100$. For serial correlation case, we have $T = 100$ and $T = 300$.

We generate homoskedastic fixed regressor bootstrap distribution $\sup F(b)$ conditional on x_t, and each bootstrap replication is 1000. For each bootstrap distribution, we obtain the critical value and compute how many percent of the Monte Carlo replication exceeds it. For testing a structural change, we truncate first and last 15% of data.

We have the following cases;

i) the original cointegration equation $y_t = \alpha_2 + \beta x_t + u_{2t}$.

ii) $y_t = \alpha_2 + \beta x_t + (1 - \rho L)x_t + u_{2t}$, assuming that true value of ρ in $(1 - \rho L)x_t$ is known.

iii) $y_t = \alpha_2 + \beta x_t + (1 - \widehat{\rho} L)x_t + u_{2t}$, where $\widehat{\rho}$ is the OLS estimate.

iv) $y_t = \alpha_2 + \beta x_t + (1 - \underline{\rho} L)x_t + u_{2t}$, where $\underline{\rho}$ can be any value within a confidence interval constructed by inversion of ADF $t-$statistics.

v) $y_t = \alpha_2 + \beta x_t + (1 - \underline{\rho} L)x_t + u_{2t}$, where $\underline{\rho}$ can be any value within a confidence interval constructed by inversion of DF_GLS $t-$statistics.

Cases iii) and iv) are application of Bonferroni's inequality. We will discuss the procedure when we report size assuming no serial correlation. We evaluate all the cases for no serial correlation. When serial correlation is taken into account, we consider only cases iv) and v), which are empirically desirable.

5.1 No Serial Correlation in u_t

First we assume no serial correlation in u_t. Then, $E\left(u_t u_t'\right) = E\left(\varepsilon_t \varepsilon_t'\right) = \Sigma = \begin{pmatrix} 1 & \delta \\ \delta & 1 \end{pmatrix}$.

5.1.1 Size

We compute the actual size against the nominal size 10% in cases i) ii) iv) and v). In all experiments, we maintain the null hypothesis of $\beta_o = 1$ for the entire sample. Thus, if the performance is good, the actual size should be 10%. In each experiment, the sample size T is 100.

Results are shown in Table 1. First of all, the size does not seem to depend on the magnitude of local-to-unity parameter, while it does depend on δ. This is because the procedure conditions on x_t.

In case i), we simply apply the fixed regressor bootstrap to the original cointegration equation $y_t = \alpha_2 + \beta x_t + u_{2t}$. If $\delta = 0.7$, which is likely to be observed in a cointegration relationship, the actual size exceeds 20%. This is because the fixed regressor bootstrap generates the distribution of $\sup F(b)$ assuming exogeneity of the regressor, while the regressor is actually highly correlated with the regression error. As we have higher value of δ, the size distortion will be larger. The results tell us that the regressor needs to be orthogonalized to the error term to reduce the size distortion.

Case ii): we add an extra term $(1 - \rho L)x_t$ to the cointegration equation, where ρ is the true value. Now the cointegration equation is $y_t = \alpha_2 + \beta x_t + \delta(1 - \rho L)x_t + u_{2.1t}$, and the error $u_{2.1t}$ is orthogonal to the regressor. This procedure is empirically infeasible since true value of ρ is unknown in practice. Nevertheless we present the results to show the size distortion is eliminated by orthogonalization. If true value of ρ is taken into the equation, the actual size stays around 10%.

Cases iv) and v) (Bonferroni's test): A feasible procedure when true value of ρ is unknown and true distribution of $\sup F$ is also unknown will be application of Bonferroni's inequality. Let $C_c(\eta_1)$ denote a $100(1 - \eta_1)\%$ confidence region for c, and $C_{\sup F|c}(\eta_2)$ denote $100(1 - \eta_2)\%$ confidence region for $\sup F$ that depends on c. Then, if we obtain a confidence region for any possible c, a $100\,(1-\eta)\,\%$ valid confidence region for β which is independent of c can be constructed as;

$$C^{B}_{\sup F}\,(\eta) = \cup_{c\in C_c(\eta_1)} C_{\sup F|c}\,(\eta_2)$$

By Bonferroni's inequality, the confidence region $C^{B}_{\sup F}\,(\eta)$ has confidence level of at least $100(1-\eta)\%$, where $\eta = \eta_1 + \eta_2$. Thus, by applying Bonferroni's inequality we can obtain a valid confidence region for $\sup F$ and can effectively control the size.

We construct a confidence interval for c (or ρ equivalently) in two different methods; inverting the usual augmenting Dickey-Fuller (ADF) t-statistic and the DF-GLS t-statistic. We construct a confidence interval for ρ in either way. Then, we generate the fixed regressor bootstrap distribution $\sup F(b)$ for any value of ρ within $\left(\underline{\rho}, \overline{\rho}\right)$, where $\underline{\rho}$ and $\overline{\rho}$ are lower and upper bounds of a confidence interval respectively. We consider the following equation;

$$y_t = \alpha_2 + \beta x_t + \gamma(1-\underline{\rho}L)x_t + \widetilde{u}_{2t} \tag{12}$$

where $\underline{\rho}$ will be replaced with $\overline{\rho}$ or any value of ρ within the confidence interval.

For each generated distribution, the 90th percentile of $\sup F(b)$ is collected. The maximum one will be the valid critical value that control the size equal to or below 10%.

The sample test statistic of $\sup F$ is a draw from unconditional distribution that also depends on ρ. The distributions of $\sup F$ are generated by Monte Carlo for $\underline{\rho}$ and $\overline{\rho}$, which we denote as $\sup F(\underline{\rho})$ and $\sup F(\overline{\rho})$ respectively and percentage that exceeds above 10% level of critical value.

Case iv): we first construct a confidence interval for ρ by inverting the usual ADF t-statistic as suggested by Stock (1991), using a monotonic relationship between c and the distribution of ADF t-statistics. We construct an equally tailed confidence interval of c

for a given value of ADF $t-$statistic. Then, by local-to-unity asymptotic theory, we can obtain a confidence interval for ρ.

Overall the size is effectively controlled. The actually size stays around 10% though it exceeds 12% when $c = -10$ and $\delta = 0.7$.

Case v): a more accurate confidence interval for ρ (and more accurate Bonferroni test) can be constructed by inverting the DF-GLS t-statistic proposed by Elliott, Rothenberg, and Stock (1996). While there is no uniformly most powerful (UMP) test for testing a unit root, it is possible to construct a power envelope by obtaining the upper bound of power against any fixed alternative $\rho = \widetilde{\rho}$. Each unit root test testing the null of $\rho = 1$ against a given fixed alternative $\widetilde{\rho}$ yields the power curve which is tangent to the envelope at $\rho = \widetilde{\rho}$ (equivalently $c = \widetilde{c}$ in finite sample). Although the power of the usual ADF $t-$statistic is substantially below the power envelope when a deterministic term is included, it is possible to improve the test so that its power is very close to the bound by efficiently estimating the deterministic term. Elliott and Stock (2001) showed it is possible to construct a confidence interval for ρ by DF-GLS test since there is a one-to-one relationship between c and the distribution of DF-GLS statistics. By inverting this efficient test (DF-GLS test), we should be able to have a more accurate confidence interval for ρ. In practice we pick up a fixed alternative $c = \widetilde{c}$ where power is one-half since such a value of $c = \widetilde{c}$ yields the power curve that stays close to the power envelope over a long range of c (King, 1987). ERS recommend $\widetilde{c} = -7$ when the deterministic term is constant.

The size is around 10%. In terms of size, there is no large difference between ADF t-test and DF-GLS test except for $c = -10$ and $\delta = 0.7$. It is because our deterministic term is only constant, and α_1 is relatively small. If we include both constant and trend terms, DF-GLS test will outperform the ADF t-test.

Local Alternative Power We compare the local alternative power for the cases i) iii),iv), and v). The local alternative hypothesis is;

$$\beta = \begin{array}{cc} \beta_o & \text{if } t < \frac{T}{2} \\ \beta_o + \frac{g}{T} & \text{if } t \geq \frac{T}{2} \end{array} \tag{13}$$

where $g = 0, 1, 2, ...20$ and $T = 100$. True timing of a structural change is at $\frac{T}{2}$.

We replicate the distribution of $\sup F_n$ under the local alternative hypothesis. We also replicate the null distribution $\sup F(b)$ by fixed regressor bootstrap to obtain the critical value, and compute how many percent of $\sup F_n$ exceeds the critical value. The nominal size is set as 10%. For case i), we adjust critical values so that the size becomes 10%, because the size distortion is large.

Figures 1a and 1b present the results. First of all, in contrast to size, power tends to be lost as the local-to-unity parameter c goes further away from 0. As a result, test loses power for testing the local alternative $\beta = \beta_o + \frac{g}{T}$, where $g > 0$. However, by taking into account information about ρ, power is improved.

For any combination of c and δ, there is power gain from including the quasi-difference term. There is a large power gain by using a confidence interval for ρ rather than using the point estimate $\widehat{\rho}$. By taking into account any plausible value of ρ within a confidence interval, accuracy of test largely improves. There is not much power gain by using the point estimate of ρ if $\delta = 0.3$ and $c = -5$ and -10.

If we use confidence interval of ρ, power gain is large. Power gain is especially large when endogeneity is large ($\delta = 0.7$). For any value of c, power is substantially improved when $\delta = 0.7$. Even though $c = -10$, power achieves around 90% at $g = 20$ for $\delta = 0.7$.

By adding the quasi-difference term, the cointegration equation is

$$y_t = \alpha_2 + \beta x_t + \delta(1 - \rho L)x_t + u_{2.1t} \tag{14}$$

The coefficient on the quasi-difference term is δ. If δ is large, weight on the quasi-difference term is large. Thus including the quasi-difference term substantially improves power. There is some power gain by including point estate $\widehat{\rho}$ when $\delta = 0.7$. But it is not as big as using a confidence interval for ρ . Test becomes more accurate by taking into account a certain range of feasible ρ rather than including a point estimate.

There is no large difference between using ADF $t-$statistics and DF_GLS t-statistics because the deterministic term is negligibly small. Since they produce similar confidence intervals, power is also similar.

From the experiments without serial correlation, we conclude that cases iv) and v) are empirically desirable.

5.2 Serial Correlation in u_t

Now we allow a serial correlation in u_t and investigate finite sample performance for the empirically feasible cases iv) and v). As in no serial correlation case, we evaluate size and local alternative power with the nominal size 10% for $T = 100$ and $T = 300$.

u_t is modeled as VAR (1) process $\Phi(L)u_t = \varepsilon_t$, and the scaled long run variance-covariance matrix of u_t is $\Omega = \Phi(1)^{-1}\Sigma\Phi(1)^{-1\prime} = \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix}$, where $\Phi(1) = \sum_i \Phi_i$ and $\Sigma = E\left[\varepsilon_t \varepsilon_t'\right] = \begin{bmatrix} 1 & \delta \\ \delta & 1 \end{bmatrix}$. We impose a mild serial correlation $\Phi(L) = \begin{bmatrix} 1 - a_{11}L & 0 \\ 0 & 1 - a_{22}L \end{bmatrix} =$

$\begin{bmatrix} 1-0.2L & 0 \\ 0 & 1-0.4L \end{bmatrix}$. We do not assume a highly persistent u_{2t} because if a_{22} is near one, then variables may not be cointegrated

In order to replicate the null distribution $\sup F(b)$ by the fixed regressor bootstrap, we modify the cointegration equation as follows;

$$y_t = \alpha_2 + \beta x_t + \sum_{j=-K}^{K} \pi_j (1-\rho L) x_{t-j} + v_t \tag{15}$$

for $t = 1, .., T$. v_t is 'approximately' orthogonal to $(1-\rho L)x_{t-j}$ for $j = -K, ..., K$.

In order to generate the fixed regressor bootstrap $\sup F(b)$, we set K,and regress $y_t(b)\tilde{}i.i.d.N(0,1)$ on the right hand variables. We also replicate the actual (unconditional) distribution of $\sup F$. To compute $\sup F$, we estimate $\sigma_{1.2} = \Omega_{22} - \Omega_{21}\Omega_{11}^{-1}\Omega_{12}$ and use the equation (4), because v_t is an approximation of $u_{2.1t}$. We estimate Ω as VAR (1) representation following Berk (1974).

5.2.1 Size

Size is reported in table 1d for $T = 100$ and 1e for $T = 300$. We set $K = 3$. With $K = 3$, actual sizes are around the nominal size 10% for any combination of c and δ. To control the size, the choice of K is important. When we set $K = 1$, actual sizes go up to about 20%. There is no big difference in size whether using confidence intervals by ADF or $DF - GLS$ t-statistics in our simulation. Size is not affected by nonzero c even in the presence of serial correlation.

5.2.2 Local Alternative Power

The local alternative power is reported in figures 2a and 2b for $T = 100$ and 3a and 3b for $T = 300$. By allowing serial correlation, power is lost. Loss of power is especially large for $\delta = 0.3$. Why power is lost? There is a trade-off between power and size. It is large when cointegration error is serially correlated. We do not have true value of ρ, but use a confidence interval from the sample. As the number leads and lags increases, v_t goes close to the orthogonal part of error $u_{2.1t}$. As a result, the size distortion is effectively eliminated. At the same time, however, accuracy of test decreases because ρ is unknown. If we reduce leads and lags, power will improve while size distortion is large. Thus, we have to be cautious when cointegration error is serially correlated especially for small δ.

6 Empirical Study

In the previous section, we have proposed modification of fixed regressor bootstrap taking into account the sample information about the regressor in the cointegration equation. We have found that adding the quasi-difference term substantially improves power of the test. In this section, we apply our proposed test to empirical data. The present value model in the stock market and the expectations hypothesis are our empirical examples. Both hypothesis are often tested in a cointegration framework, and they are usually found to be rejected in empirical study. Under our procedure, we should be able to investigate whether there is the long run stable relationship among data even though data may or may not have an exact unit root. Rejection of test will correctly indicate the relationship is unstable, that is, the hypothesis is rejected. We truncate first and last 15% of data.

We examined the test with different lead- and lag-lengths of a quasi-difference term and with different truncations. As we have learned in simulation, the fixed regressor null distribution changes and thus critical value changes as lag- and lead- length changes. But a timing of a structural change is not affected.

6.1 The Present Value Model in the Stock Market

First we study the present value model.

6.1.1 Theoretical Framework

The present value model in the stock market says that a stock price today is the expectation of the present values of all future dividends. Following Campbell and Shiller (1988) and Bansal and Lundblad (2002), the log stock price - dividend ratio can be expressed as a function of the expected dividend growth rate and the ex-ante stock return.

$$p_t - d_t = \frac{\kappa_0}{1-\kappa_1} + E\left[\sum_{i=0}^{\infty} \kappa_1^j \left(g_{t+1+j} - r_{t+1+j}\right)\right]$$

where p_t and d_t are the logs of stock price and dividends respectively, g_t is the continuous dividend growth rate, and r_t is the log of total return.[3] The stock price and the dividend are often assumed to be $I(1)$ (for example, Mankiw, Romer, and Shapiro (1985, 1991) or Cochrane (1992, 1994), and unit root tests fail to reject the null hypothesis of unit root in empirical studies (Campbell and Shiller (1988), Timmermann (1995)). Thus, the logs of stock price and dividend are also assumed to have a unit root. Then, the continuous growth rate of dividend and stock return, i.e. the first difference of the log dividend and stock price should be stationary. Consequently, if the present value model holds, the logs of stock price and dividend are cointegrated with the cointegrating vector, $(1, -\beta_o) = (1, -1)$, and the cointegration equation is

$$p_t - \beta_o d_t = u_{2t}$$

[3] $\kappa_1 = \frac{1}{1+\exp\left(\overline{d_i - p_i}\right)}$, $\kappa_0 = -\log(\kappa_1) - (1-\kappa_1)\left(\overline{d_i - p_i}\right)$, where $\overline{d_i - p_i}$ is the mean of logs of dividend yield.

or

$$p_t = \beta_o d_t + u_{2t}$$

where $\beta_o = 1$ under the null hypothesis.[4] Let us denote the estimated β as $\widehat{\beta}$. If $\widehat{\beta}$ is significantly larger than one, that is, the log of stock price is far more volatile than dividend.

6.1.2 Previous empirical findings

Testing the present value model with logs of stock price and dividend was first proposed by Campbell and Shiller (1988). A typical empirical finding is against the present value model. The (log of) stock price is far too volatile relative to dividend for the present value model to hold, which is the so called "excess volatility puzzle". Barsky and Delong (1993) estimated β, $\widehat{\beta} = 1.61$, and the associated $t-$statistic was positive and significant. Gonzalo, Lee, and Yang (2007) found that, by the Johansen test, the estimated reciprocal of β is around 0.6, and the estimated cointegrating vector is significantly different from $(1, -1)$. Barksy and Delong (1993) and Bansal and Lundblad (2002) attributed this high volatility of stock prices to nonstationarity or near nonstationarity of dividend growth rate, g. On the other hand, Timmermann (1995) showed by Monte Carlo experiment that the logs of stock price and dividend fail to be cointegrated because the rate of return, r, is highly persistent and not because r is highly volatile.

[4] Levels of stock price and dividend are also tested in a cointegration framework (Campbell and Shiller (1987), for example). In this case, the cointegrating vector is a function of discount factor, the value of which is not specified by the model. Since our concern is the hypothesis testing on the cointegrating vector, cointegration between stock price and dividend in levels is beyond scope of this paper.

6.1.3 Our empirical analysis

For our empirical analysis we use annual data. The original data are annual stock returns with and without dividends from CRSP. From the original data set, I constructed the log real stock price and the log real dividends. A detailed description about data is found in appendix. The entire sample is from 1926 to 2007. Figure 4 plots log real stock price and log real dividend. Over all they are moving in a similar manner. If we take a closer look, however, we find some difference. For example, in earlier 1970's, there is a sharp drop in stock price, while dividend also drops but not as sharply as stock price does. The difference is more obvious during 1990's. While there is a sharp increase in stock price, dividend increases only slightly. There is a relatively sharp drop in stock price around year 2001 recession.

Table 2 presents our empirical results. The first column presents 95% confidence intervals for the largest root in stock price and dividend. Both are tight around unity, which implies that the largest root is near unity. The long run correlation δ is estimated from the parametric estimation of the long run variance covariance proposed by Berk (1974). Because δ exceeds 0.9 and the confidence interval for ρ is relatively tight, we expect a large power gain from adding quasi-difference terms.

Last two columns present $\sup F$ statistic and critical values obtained from the fixed regressor bootstrap. The $\sup F$ statistic is computed as we have discussed in the previous section. The fixed regressor bootstrap critical value is computed with 3 leads and 3 lags. The stable relationship between stock price and dividend is rejected. The test indicates structural change in the relationship in 1972 around the first oil crisis.

6.2 Expectations Hypothesis

Now we turn to the expectations hypothesis of term structure.

6.2.1 Theoretical Framework

We denote p_t^n as log price of n-period zero-coupon bond and r_t^n as yield to maturity of n-period of zero-coupon bonds. Then, $r_t^n = -\frac{1}{n}p_t^n$. We denote f_t^n as the return from contract at time t to buy one-period zero-coupon bond that matures at time $t+n$. Then, $f_t^n = p_t^{n-1} - p_t^n$, and $f_t^1 = r_t^1$. A version of the expectations hypothesis says that forward rate is expected future spot rate plus risk premium;

$$f_t^j = E_t\left(r_{t+j-1}^1\right) + \Lambda(j,t) \tag{16}$$

where $\Lambda(j,t)$ is risk premia which depends on risk considerations or preferences about liquidity.

Then, we have the following relationship;

$$r_t^n = \frac{1}{n}\sum_{j=1}^{n} E_t\left[r_{t+j-1}^1\right] + L(n,t) \tag{17}$$

where $L(n,t) = \frac{1}{n}\sum_{j=1}^{n}\Lambda(j,t)$.

Under the pure version of the expectations hypothesis, risk premia are zero, while under the other versions of the hypothesis, the risk premia are constant over time.

By rearranging the equation above, we have;

$$r_t^n - r_t^1 = \frac{1}{n}\sum_{j=1}^{n}(n-j)E_t\left[\Delta r_{t+j-1}^1\right] + L(n,t) \tag{18}$$

Assuming yields to maturities are unit root processes, the first difference of the one-period yields Δr_{t+j-1}^1 is stationary, which implies that $r_t^n - r_t^1$ is stationary. If r_t^n and r_t^1 are cointegrated with the cointegrating vector $(1, -\beta_o) = (1, -1)$, the expectations hypothesis holds. The hypothesis was first examined in a cointegration framework by Campbell and

Shiller (1987).

6.2.2 Previous Empirical Findings

The expectations hypothesis is rejected in most empirical study (see Engle and Granger 1987 and Hall, Anderson, and Granger, 1992 for cointegration test). And the rejection is often attributed to the period between 1979:9 and 1982:9 when the Fed ceased targeting. Hall et al.(1992) and Shea (1992) concluded that test is rejected because risk premia were highly volatile during this period.

A structural change in the term structure has been studied as well. For example, Hamliton (1988) detected the period 1979:9 to 1982:9 as a separate regime by a Markov switching model.

6.2.3 Our Empirical Analysis

We use U.S. zero-coupon bond yield provided by McCulloch and Kwon (1993) for empirical study (see appendix). The entire sample is 1970:1 - 1991:2, which includes change in the Federal Reserve's operating procedure at the end of 1970's. The Federal Reserve ceased targeting a short-term interest rates in the fall of 1979 and resumed targeting in the fall of 1982. We test the relationship between yield to maturity of one-month r_t^1 and relatively short-period yield r_t^n, where n is 2 months, 3 months, 4 months and 6 months.

Figure 5 plots yields. They seem to be moving in a similar manner. They are all highly volatile while the Fed ceased targeting. Another thing to note is that r_t^1 especially experienced a sharp drop around October 1987.

Table 3 presents results. 95% confidence intervals are relatively wide compared to those for the present value model. Long-run correlations δ are relatively small. Thus, we should keep in mind that power of the test might be low. We added to the cointegration equation 3 leads and 3 lags of a quasi-difference term. For relatively short yields, we fail to identify a structural change around the late 1970's or early 1981.s when the Federal Reserve changed the operating procedure. For the relationship between 1-month and 2-, 3-, or 4-month yields, we found a structural change in 1986:11, which is close to so called 'Black Monday' when r_t^1 experienced a sharp drop. For the one-month yield and six-month yield, we found a structural change in 1981:1, which we might be able to relate to the change in the operating procedure by the Fed.

7 Conclusion

We have proposed a modified fixed regressor bootstrap so that test is applicable to a cointegration framework. The test is robust to the presence of a near unit root. We have found that by modification, size distortion is corrected and power gain is large especially when an endogeneity problem is large. But trade-off between size and power is large when there is a serial correlation in the cointegration equation error. As the number of leads and lags increases, power is lost especially when the magnitude of endogeneity is small. Power possibly improves if we use more powerful unit root test to construct confidence intervals for ρ. The possible test would be the test that includes stationary covariates proposed by Hansen (1995) and extended by Elliott and Jansson (2003).

We examined the present value model and the expectations hypothesis in empirical study. Our test is designed for a single structural break. We detected a structural change in the early 1970s for the present value model and the late 1980's for the expectations hypothesis. We have found that a stable relationship is not maintained for both examples. However, in the long run, it is possible that the relationship among economic variables experience multiple structural changes in the long run. For further research, we might need to consider testing multiple breaks.

A Proof

Proof. Proof for Lemma 1.

i) and ii) follow from Lemma 1 in Phillips (1987).

Proof for iii) is the following. Let $W_1(r) = \frac{1}{\sigma_{11}\sqrt{T}}\sum_1^{Tr} u_{1t}$, $W_{2.1}(r) = \frac{1}{\sigma_{2.1}\sqrt{T}}\sum_1^{Tr} u_{2.1t}$ and $W_2(r) = \frac{1}{\sigma_{22}\sqrt{T}}\sum_1^{Tr} u_{2t}$.

We have $W_2 = W_{2.1} + \delta\frac{\sigma_{22}}{\sigma_{11}}W_1$ and $\sigma_{2.1}^2 = (1-\delta^2)\sigma_{22}^2$, where $\sigma_{2.1}^2$ is the long-run variance of $u_{2.1t}$.

$u_{2.1t}$ is orthogonal to u_{1t-j} for all j.

Then, we have

$$\frac{1}{\sigma_{22}T}\sum_{t=1}^{[Tr]} x_t u_{2t} = \frac{1}{\sigma_{22}T}\sum_{t=1}^{[Tr]}\left[x_{t-1}u_{2t} + u_{1t}u_{2t} + \frac{c}{T}x_{t-1}u_{2t}\right] \tag{19}$$

$\frac{1}{\sigma_{22}T}\sum_{t=1}^{[Tr]} x_{t-1}u_{2t} \Rightarrow \sigma_{11}\int_0^r J_c(\lambda)dW_{2.1} + \frac{\delta\sigma_{2.1}}{\sqrt{(1-\delta^2)}}\int_0^r J_c(\lambda)dW_1$

$\frac{1}{\sigma_{22}T}\sum_{t=1}^{[Tr]} u_{1t}u_{2t} + \frac{c}{T}\frac{1}{\sigma_{22}T}\sum_{t=1}^{[Tr]} x_{t-1}u_{2t} \rightarrow r\delta\frac{\sigma_{2.1}}{\sqrt{(1-\delta^2)}}\frac{S_{11}^2}{\sigma_{11}^2} + O_p(T^{-1})$ almost surely by law of large number,

where $S_{11}^2 = \lim_{T\to\infty}\frac{1}{[Tr]}\sum_1^{[Tr]} u_{1t}^2$.

$\frac{1}{\sigma_{22}T}\sum_{t=1}^{[Tr]} u_{1t}u_{2t} = \frac{[Tr]}{\sigma_{22}T}\frac{1}{[Tr]}\sum_{t=1}^{[Tr]}\left[\sigma_{22}\sqrt{T}\frac{1}{\sigma_{2.1}\sqrt{T}}u_{2.1t}u_{1t} + \sigma_{22}\sqrt{T}\delta\frac{\sigma_{22}}{\sigma_{11}}\frac{1}{\sigma_{11}\sqrt{T}}u_{1t}^2\right]$

Proof. Proof for Thorem 2 follows Theorem 2 in Hansen (2000). ■ ■

B The Data Set

B.1 The Present Value Model

The original data is from the CRSP value-weighted NYSE portfolio from 1926 to 2007, in which annual total stock return, r_t^{total}, and return without dividend, r_t^o, are provided . We construct the series of real stock price and real dividend as follows. The nominal stock price is computed as investing 1 in this portfolio at the end of 1925 and multiplying by $(1 + r_t^o)$ stock price at $t-1$ to have a stock price at the end year t. Annual dividend at t is $\left(r_t^{total} + r_t^o\right) P_{t-1}$. We divide them by the Consumer Price Index with 1982-1984 base year to obtain real stock price, P_t and real dividend D_t. We take the natural logs of P_t and real D_t.

B.2 The Expectations Hypothesis

The data for this study is the zero-coupon yield curve. The data is from McCullock and Kwon (1993).

References

[1] Ahn, S. K. and G. C. Reinsel, "Estimation for Partially Nonstationary Autoregressive Models," Journal of the American Statistical Association 85 (1990), pp. 813-823.

[2] Andrews, D. W. K., "Tests for Parameter Instability and Structrual Change with Unknown Change Point," Econometrica 61 (1993), pp. 821-856.

[3] Andrews, D. W. K. and W. Ploberger, "Optimal Tests When a Nuisance Parameter Is Present Only Under the Alternative," Econometrica 62 (1994), pp. 1383-1414.

[4] Andrews, D. W. K. and J. C. Monahan, "An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator," Econometrica 60 (1991), pp. 953-966.

[5] Bai, J. and P. Perron, "Estimating and Testing Linear Models with Multiple Structural Changes," Econometrica 66 (1998), pp. 47-78.

[6] Basawa, I. V., A. K. Mallik, W. P. McCormick, J. H. Reeves, and R. L Taylor, "Bootstrapping Unstable First-Order Autoregressive Processes," Annals of Statistics 19 (1991), pp. 1098-1101.

[7] Berk, K. N., "Consistent Autoregressive Spectral Estimates," the Annals of Statistics 2 (1974), pp. 489-502.

[8] Bobkoski, M. J., "Hypothesis Testing in Nonstationary Time Series," unpublished PhD. thesis (1983) (Department of Statistics, University of Wisconsin, Madison, WI).

[9] Campbell, J. Y. and P. Perron, "Pitfalls and oportunities: What Macroeconomists Should Know about Unit Roots," NBER Macroeconomics Annual (1991), Cambridge and London: MIT Press.

[10] Campbell, J. Y., and R. J. Shiller, "Cointegration and Tests of the Present Value Models," the Journal of Political Economy 95 (1987), pp. 1062-1088.

[11] Cavanagh, C., "Roots Local to Unity," manuscript (1985), (Department of Economics, Harvard University, Cambridge, MA).

[12] Cavanagh, C., G. Elliott, and J. H. Stock, "Inference in Models with Nearly Integrated Regressors," Econometric Theory 11 (1995), pp. 1131-1147.

[13] Elliott, G., "On the Robustness of Cointegration Methods When Regressors Almonst Have Unit Roots," Econometrica 66 (1998), pp. 149-158.

[14] Elliott, G. and Jansson, M., "Testing for Unit Roots With Stationary Covariates," Journal of Econometrics 115 (2003), pp. 75-89.

[15] Elliott, G., T.J. Rothenberg, and J.H. Stock, "Efficient Tests for an Autoregressive Unit Root," Econometrica 64 (1996), pp.813-836.

[16] Elliott, G. and J. H. Stock, "Confidence Intervals for Autoregressive Coefficients near One," Journal of Econometrics 103 (2001), pp. 155-181.

[17] Engle, R. F. and C. W. Granger, "Cointegration and Error Correction: Representation, Estimation, and Testing," Econometrica 55 (1987), pp. 251-276.

[18] Gregory, A. W. and B. E. Hansen, "Residual -based tests for cointegraion in models with regime shifts," Journal of Econometrics 70 (1996), pp. 99-126.

[19] Hall, A. D., H. M. Anderson, and C. W. J. Granger, "A Cointegration Analysis of Treasury Bill Yields," Review of Economics and Statistics 74 (1992), pp. 116-126.

[20] Hall, P., "Methodology and Theory for the Bootstrap," (1994), In Engle and McFadden, D. L. (Eds), Handbook of Econometrics. Vol. IV., Elsevier, New York.

[21] Hamilton, J.D., "Rational-Expectaions Econometric Analysis of Changes in Regime: An Investigation of the Term Structure of Interest Rates," Journal of Economic Dynamics and Control 12 (1988), pp.385-423.

[22] Hansen, B. E., "Tests for Parameter Instability in Regressions with I(1) processes," Journal of Business and Economic Statistics 10 (1992), pp.321-335.

[23] Hansen, B. E., "Rethinking the Univariate Approach to Unit Root Testing: Using Covariates to Increase Power," Econometric Theory 11 (1995), pp. 1148-1171.

[24] Hansen, B. E., "Testing for Structural Change in Conditional Models," Journal of Econometrics 97 (2000), pp. 93-115.

[25] Hansen, P. R., "Structural Changes in the Cointegrated Vector Autoregressive Model," Journal of Econometrics 114 (2003), pp. 261-295.

[26] Inoue, A., "Tests of cointegrating rank with a trend break," Journal of Econometrics 90 (1999), pp. 215-237.

[27] Kejriwal, M. and P. Perron, "Testing for Multiple Structural Changes in Cointegrated Regression Models," (2008a) working paper.

[28] Kejriwal, M. and P. Perron, "The Limit Distribution of the Estimates in Cointegrated Regression Models with Multiple Structural Changes," (2008b) forthcoming in Journal of Econometrics.

[29] King, M.L., "Towards a Theory of Point Optimal Testing," Econometric Reviews 6 (1988), pp.169-218.

[30] Newey, W. K. and K. D. West, "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," Econometrica 55 (1987) pp. 703-708.

[31] Phillips, P. C. B., "Towards a Unified Asymptotic Theory for Autoregression," Biometrica 65 (1987), pp. 533-547.

[32] Phillips, P. C. B. and B. E. Hansen, "Statistical Inference in Instrumental Variables Regression with I(1) Processes," Review of Economic Studies 59 (1990), pp. 99-125.

[33] Quandt, R., "Tests of the Hypothesis That a Linear Regression Obeys Two Separate Regimes," Journal of the American Statistical Association 55 (1960), pp. 324-330.

[34] Saikkonen, P. "Asymptotically Efficient Estimation of Cointegration Regressions," Econometric Theory 7 (1991), pp. 1-21.

[35] Stock, J. H., "Confidence Intervals for the Largest Autoregressive Root in Macroeconomic Time Series," Journal of Monetary Economics 28 (1991), pp. 435-459.

[36] Stock, J. H. and M. W. Watson, "A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems," Econometrica 61 (1993), pp. 783-820.

[37] Seo, B., "Tests for Structural Change in Cointegrated Systems," Econometric Theory 14 (1998), pp. 222-259.

$T=100$	δ							
	$i)$no quasi difference term		$ii)\rho$ is known		$iv)ADF$		$v)DF\_GLS$	
c	**0.3**	**0.7**	**0.3**	**0.7**	**0.3**	**0.7**	**0.3**	**0.7**
0	0.135	0.224	0.099	0.081	0.108	0.091	0.089	0.104
-1	0.132	0.227	0.114	0.115	0.095	0.075	0.089	0.067
-5	0.132	0.227	0.125	0.091	0.160	0.090	0.106	0.090
-10	0.115	0.200	0.096	0.092	0.099	0.124	0.103	0.101

Table 1: Size. No serial correlation Nominal size is 10 percent..

	$\boldsymbol{\delta}$							
	$T=100$				$T=300$			
c	$iv)ADF$		$v)DF\_GLS$		$iv)ADF$		$v)DF\_GLS$	
	0.3	**0.7**	**0.3**	**0.7**	**0.3**	**0.7**	**0.3**	**0.7**
0	0.125	0.099	0.129	0.094	0.096	0.094	0.125	0.099
-1	0.094	0.097	0.125	0.114	0.107	0.098	0.107	0.104
-5	0.101	0.098	0.106	0.105	0.107	0.100	0.119	0.108
-10	0.107	0.092	0.085	0.091	0.111	0.120	0.104	0.106

Table 2: Size. Errors are represented as VAR(1). Nominal size is 10 percent. 3 leads and 3 lags

T 1926-1927 82	95% CI for ρ		δ	$supF$	Upper 5% c.v.	Date
	Price		0.995	234	31.25	1972
	(0.982	1.059)				
	Dividend					
	(0.973	1.059)				

Table 3: Empirical Study: the Present Value Model. 3 leads and 3 lags

T 1970:1-1991:2 254	95% CI for ρ		δ	$supF$	Upper 5% c.v.	Date
1-2 month	1 month					
	(0.920	1.009)	0.174	54.703	28.473	1986 : 11
	2 months					
	(0.931	1.011)				
1-3 month	3 months					
	(0.934	1.012)	0.164	49.402	28.735	1986 : 11
1-4 month	4 months					
	(0.937	1.012)	0.151	38.078	28.545	1986 : 11
1-6 month	6 months					
	(0.936	1.012)	0.096	42.641	28.033	1981 : 3

Table 4: Empirical Study: the Expectations Hypothesis. 3 leads and 3 lags

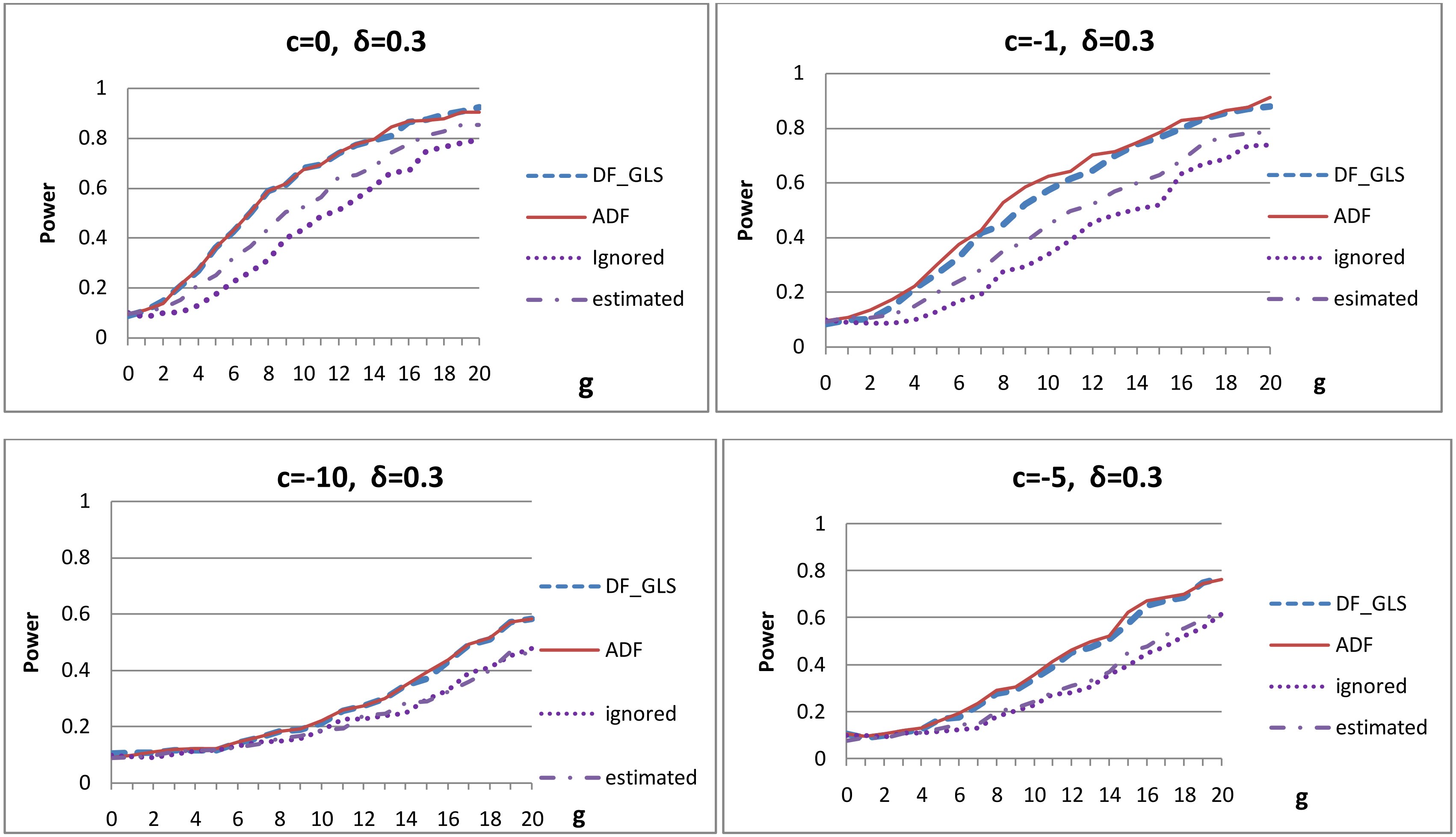

Figure 1a: δ=0.3 (no serial correlation in errors) Local alternative power $\beta_o + \frac{g}{T}$, where T=100. ("DF_GLS" is confidence interval for ρ with DF_GLS , "ADF" is confidence interval for ρ with ADF, "ignored" is the one that ignores endogeneity, and "estimated" is ρ estimated.)

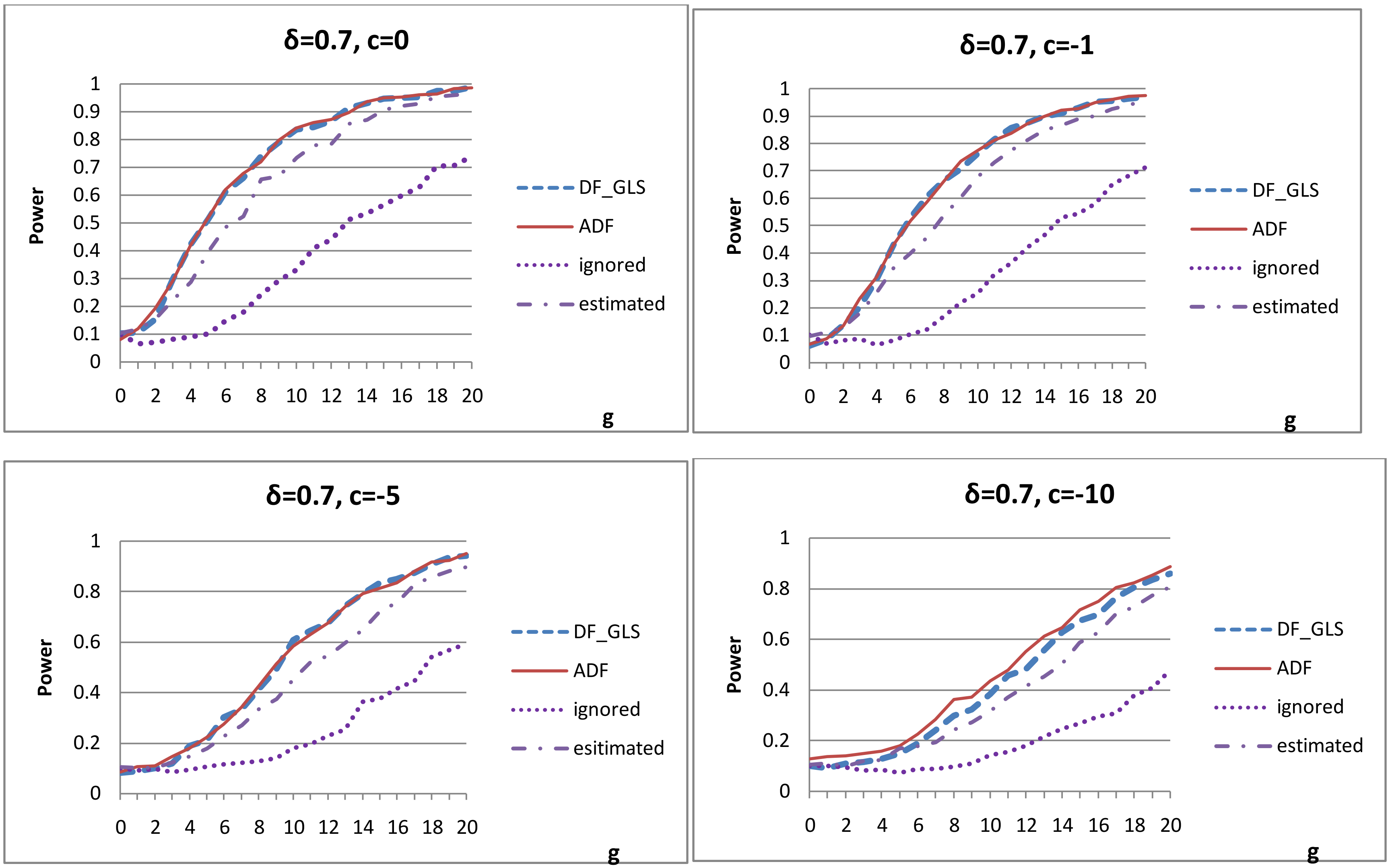

Figure 1b:δ=0.7 (no serial correlation in residuals)

Local alternative power $\beta_o+\frac{g}{T}$, where T=100. (“DF_GLS” is confidence interval for ρ with DF_GLS , “ADF” is confidence interval for ρ with ADF, “ignored” is the one that ignores endogeneity, and “estimated” is ρ estimated.)

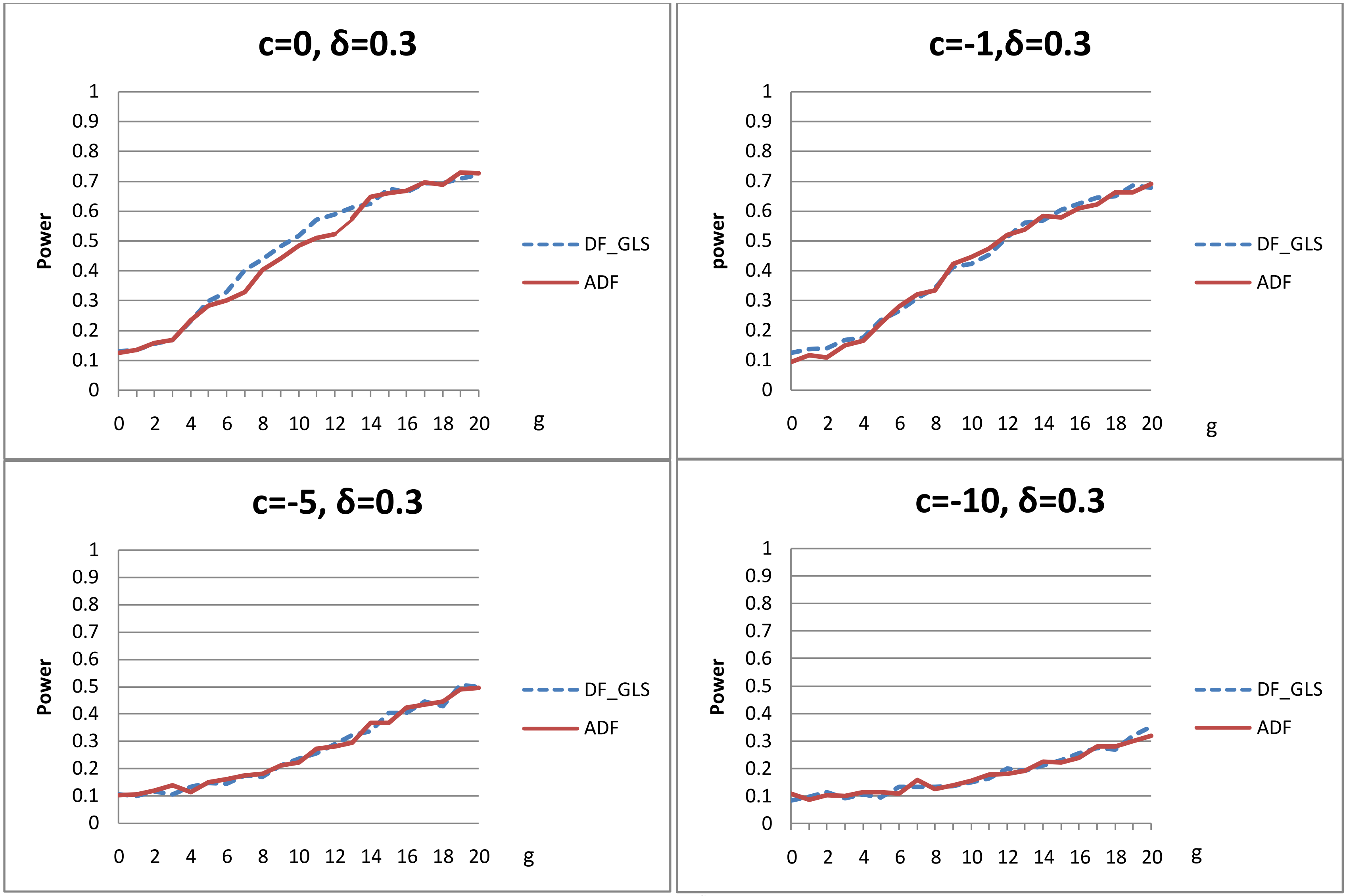

Figure 2a: δ=0.3 (serial correlation in errors) Local alternative power $\beta_o+\frac{g}{T}$, where T=100. ("DF_GLS" is confidence interval for ρ with DF_GLS , "ADF" is confidence interval for ρ with ADF.)

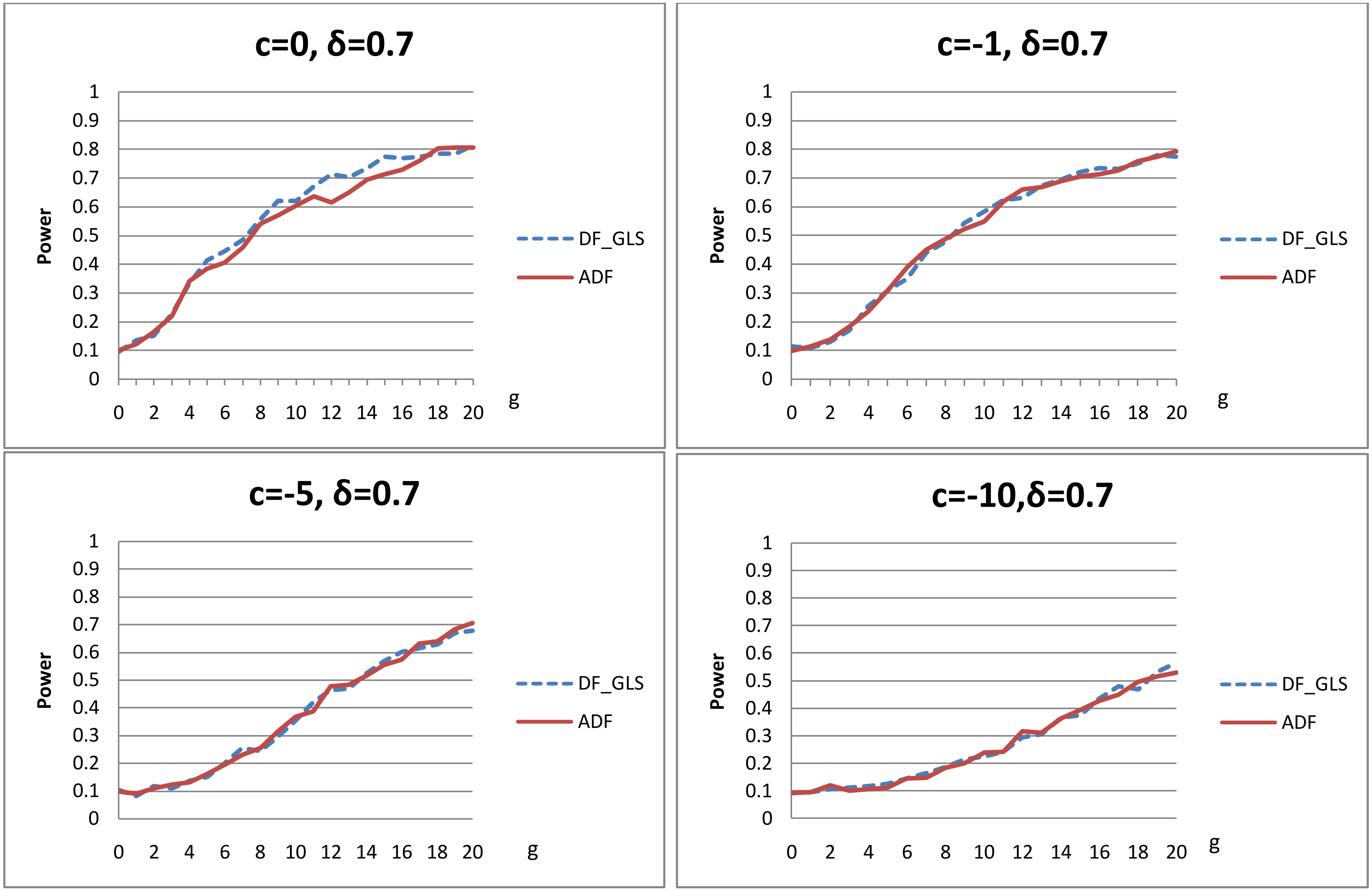

Figure 2b: δ=0.7 (serial correlation in errors) Local alternative power $\beta_o+\frac{g}{T}$, where T=100. ("DF_GLS" is confidence interval for ρ with DF_GLS , "ADF" is confidence interval for ρ with ADF.)

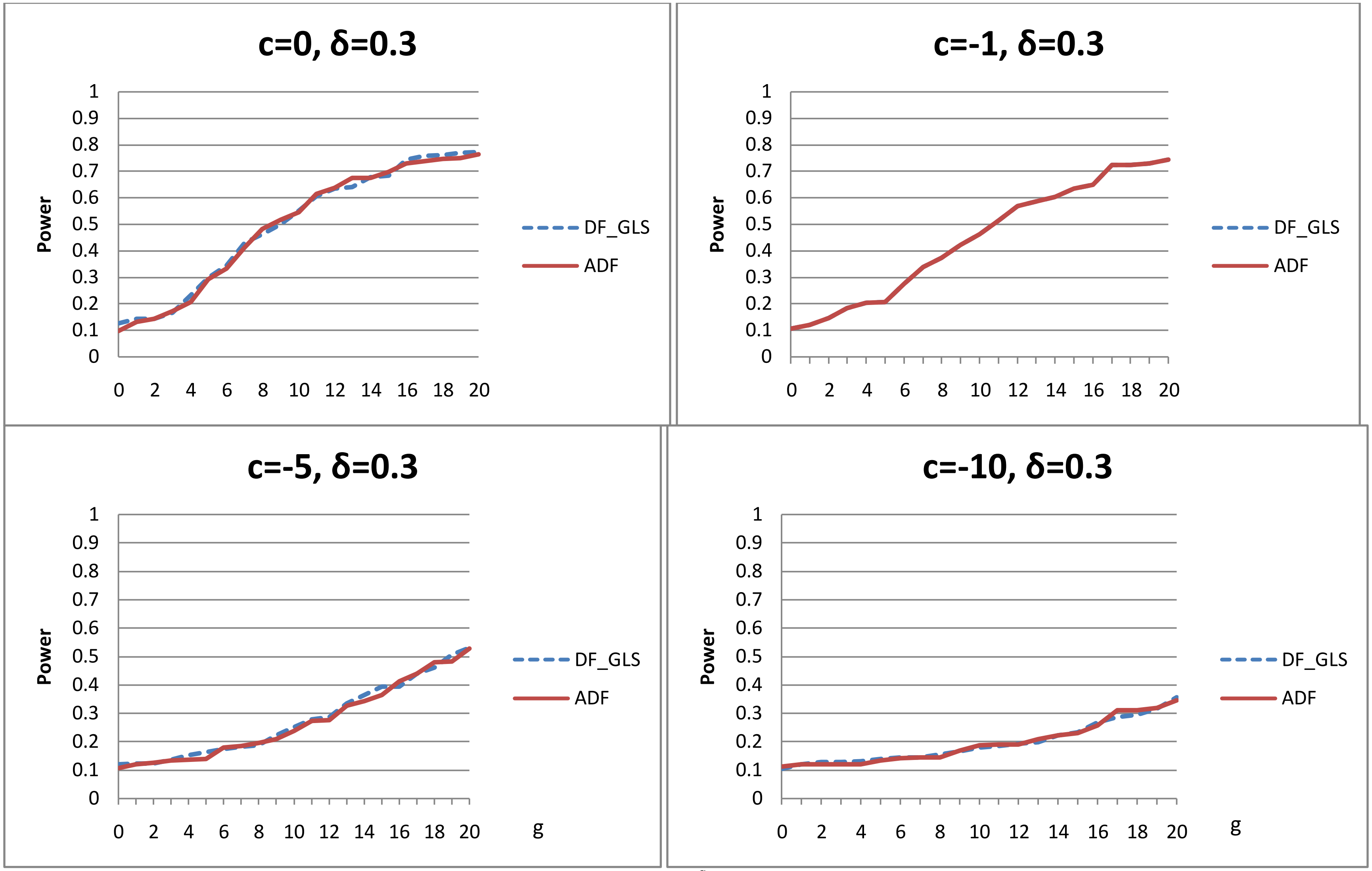

Figure 3a: δ=0.3 (serial correlation in errors) Local alternative power $\beta_o + \frac{g}{T}$, where T=300. ("DF_GLS" is confidence interval for ρ with DF_GLS , "ADF" is confidence interval for ρ with ADF.)

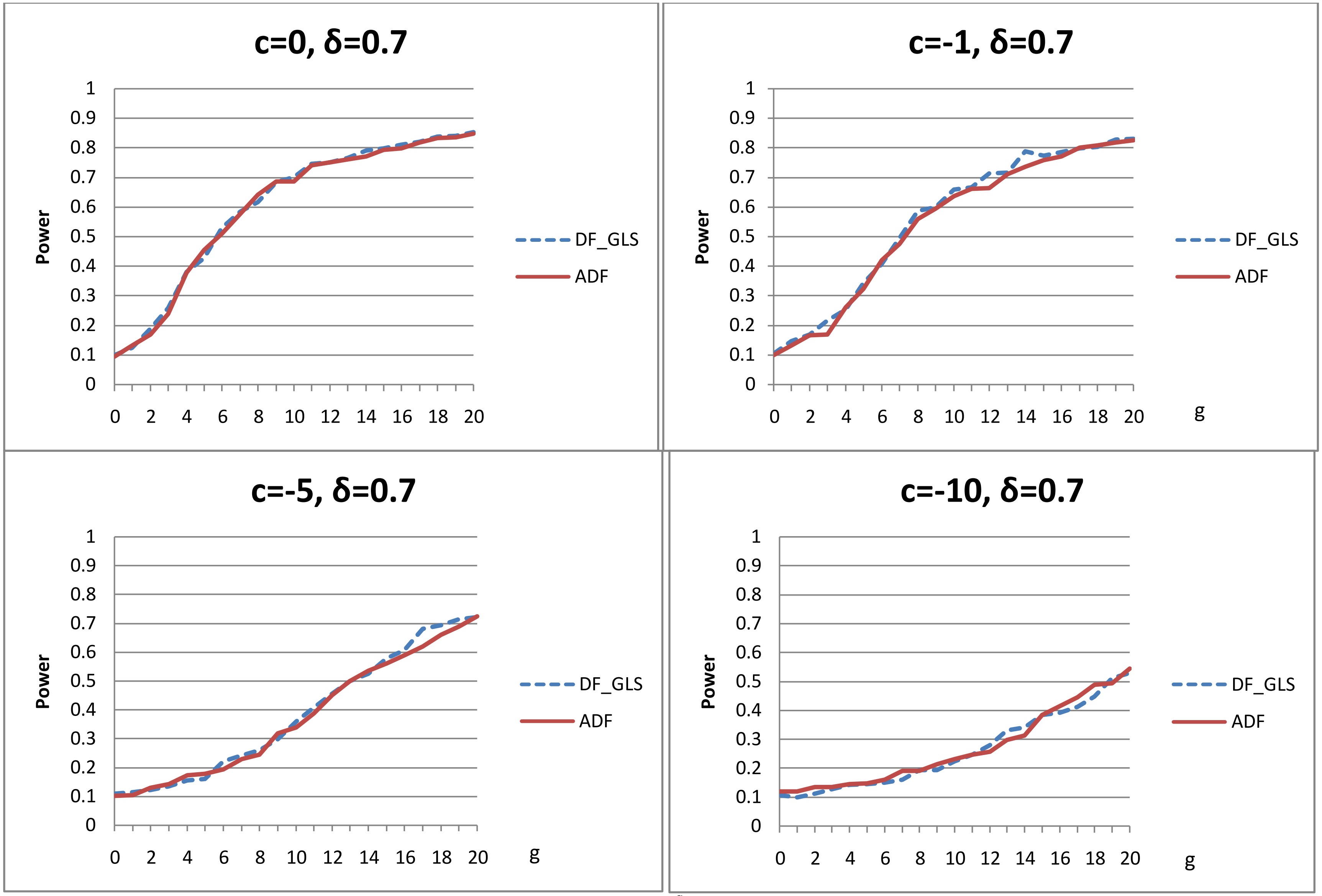

Figure 3b: δ=0.7 (serial correlation in errors) Local alternative power $\beta_o+\frac{g}{T}$, where T=300. ("DF_GLS" is confidence interval for ρ with DF_GLS , "ADF" is confidence interval for ρ with ADF.)

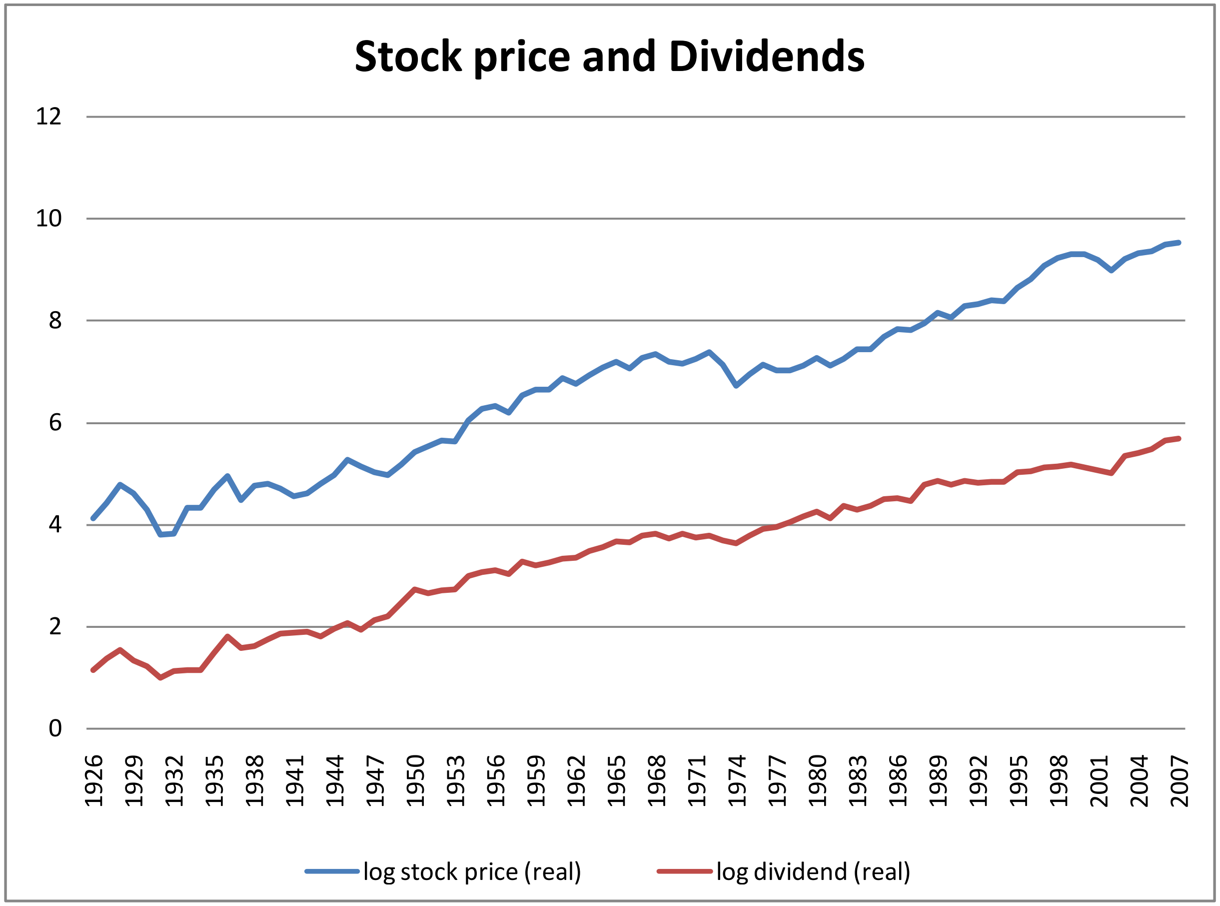

Figure 4: Annual data of log real stock price and log real dividend. 1926-2007.

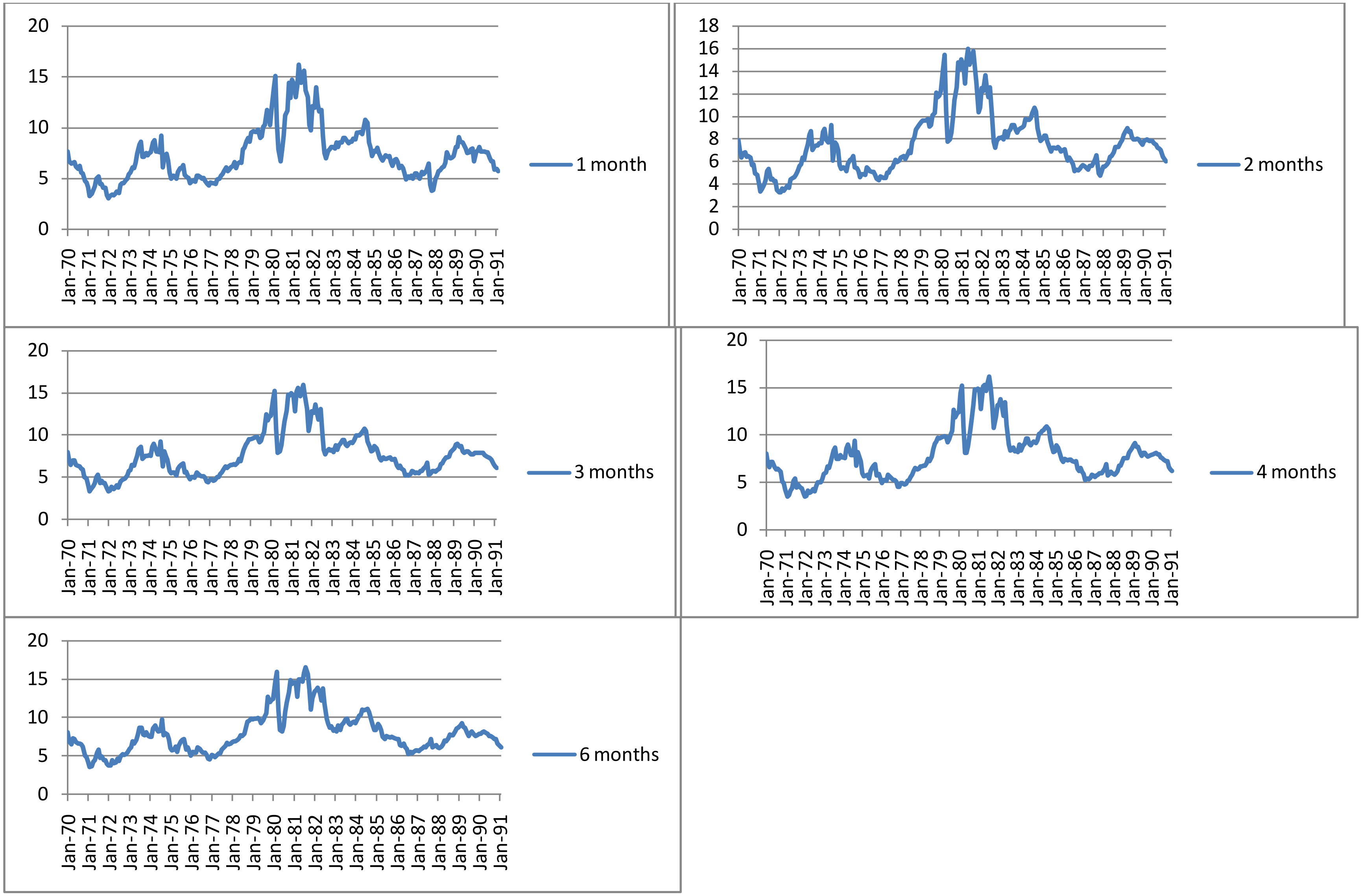

Figure 5: U.S. zero-coupon yields with maturities 1, 2, and 3 months.