

# Unit Root Tests using GLS Detrended Data and Covariates in Structural Change Models\*

Hui Liu                      Gabriel Rodríguez<sup>†</sup>  
University of Ottawa      University of Ottawa

This Version: February 14, 2006

## Abstract

This paper extends the results of Elliott and Jansson (2003) to the context of structural change models. We show that when testing for a unit root against stationarity with an unknown structural break, substantial power gains can be achieved by incorporating extra information contained in an arbitrary number of covariates. The power gains are dependent on the long-run correlation between the shocks of covariates and quasi-differences of potentially correlated time series to be tested. The higher correlation between covariates and quasi-differences of time series, the higher power gains can be realized. A derived feasible statistic has asymptotic power function tangent to the power envelope. In finite sample experiments, Monte-Carlo simulation results confirm large power improvements without size deteriorations.

**Keywords:** Unit Root, GLS detrended data, Feasible Point Optimal Test, Covariates, Power envelope.

**JEL:** C2, C5.

---

\*This paper is drawn from the third chapter of the PhD Dissertation of H. Liu at the Department of Economics of the University of Ottawa. Rodríguez thanks Financial support from the Faculty of Social Sciences of the University of Ottawa.

<sup>†</sup>Address for Correspondence: Gabriel Rodríguez, Department of Economics, University of Ottawa, P. O. Box 450, Station A, Ottawa, Ontario, Canada, K1N 6N5. E-mail address: gabrielr@uottawa.ca, Telephone: +613-562-5800 (1750), Fax: +613-562-5999.

# 1 Introduction

There has been large literature on how to improve the power of unit root tests. Hansen (1995) showed that it can be costly by ignoring the information in related time series by using univariate unit root tests, as power gains can be achieved by including correlated stationary covariates in the regression equation. He reconstructed an ADF statistic (Dickey and Fuller, 1979; Said and Dickey, 1984) using multivariate framework. Comparing the power performance of the CADF (covariate ADF) and the ADF tests, he found substantial power gains in both large and small samples.

Recently, Elliott and Jansson (EJ 2003, hereafter) generalized the approach of Hansen (1995) by including different sets of deterministic components and an arbitrary number of covariates. They derived a family of multivariate point optimal tests and a feasible test when nuisance parameters are unknown. Their results have confirmed the importance of covariates in improving the power of unit root tests. In fact, the limiting distribution of the feasible point optimal tests depends on the long-run correlation between the covariates and the quasi-differences of the potentially integrated time series. The higher the correlation coefficient between the covariates and the quasi-differences, the larger improvements in the power of unit root tests.

This paper contributes to the existing literature in two ways. First it borrows the approach of EJ (2003) and applies it to the unknown structural change models. Two types of models are considered following Perron and Rodríguez (PR 2003, hereafter). Type I Model only includes a break in the slope of the trend function, while type II model considers a break in both intercept and time trend. The date of the break point is treated as unknown in both models<sup>1</sup>. Introducing an arbitrary number of covariates to a family of P-tests derived in PR (2003), this paper will show that the tests of two types of models will converge to the same limiting distribution. In other words, the rule that a break in intercept is a slowly evolving deterministic component still applies even when covariates are present. The limiting distributions of the P test in PR (2003) are retrieved when the correlation coefficient between

---

<sup>1</sup>An ADF in the spirit of Hansen (1995) has been proposed by Shin (2001) but dealing with known break dates.

covariates and the quasi-differences of the potentially integrated time series is zero. When the correlation is between zero and one, asymptotic and finite sample simulations show that large power gains are achieved when using covariates.

Second, this experiment allows us to derive a feasible test when nuisance parameters are unknown. It takes into account one additional nuisance parameter compared to that of EJ (2003) since the structural break is unknown. It will show that, although different from that of EJ (2003), the feasible test proposed here still achieves the power envelope asymptotically.

The rest of the paper is organized as follows. Section 2 defines the model and derives the limiting distribution of the test. Section 3 presents the feasible test when all the nuisance parameters are unknown. Section 4 presents results obtained from simulations and evaluates the test performance in both large and small samples. Section 5 concludes. Proofs are contained in the appendix.

## 2 The Model and Asymptotic Theory

The data generating process (DGP) considered is of the following form:

$$y_t = d_{yt} + u_{yt}, \quad (1)$$

$$x_t = d_{xt} + u_{xt}, \quad (2)$$

$$A(L) \begin{pmatrix} [1 - \rho L]u_{y,t} \\ u_{x,t} \end{pmatrix} \equiv A(L) u_t(\rho) = e_t \quad (3)$$

$$\rho = 1 + cT^{-1} \quad (4)$$

where (4) is the local to unity framework established in Phillips (1987), Chan and Wei (1987).  $A(L)$  is a matrix polynomial of finite order  $k$  and  $L$  the lag operator,  $t = 1, 2, \dots, T$ ;  $x_t$ , an  $m \times 1$  vector, is an arbitrary number of stationary covariates containing extra information of  $y_t$  (a scalar), the variable to be tested. The deterministic component of  $y_t$  is denoted as  $d_{yt}$ . Two cases are considered following PR (2003). The first case includes an

intercept, a time trend and a break in slope of the trend function. Hence  $d_{yt} = \beta_1 + \beta_3 t + \beta_5 (t - T_B) \mathbf{1}(t > T_B)$ , where  $\mathbf{1}(\cdot)$  is the indicator function and  $T_B$  the unknown break date. The second case considers an intercept, a time trend, and a break in both intercept and the slope. Therefore  $d_{yt} = \beta_1 + \beta_3 t + \beta_5 (t - T_B) \mathbf{1}(t > T_B) + \beta_6 \mathbf{1}(t > T_B)$ . The deterministic component of covariates include an intercept and a time trend in both cases, that is,  $d_{xt} = \beta_2 + \beta_4 t$ . Notice that the first case can be obtained by setting  $\beta_6 = 0$  in the second case. Therefore Case 1 and 2 can be simplified using the restriction  $[I_{2(m+2)} - S_i]\beta^i = 0$ , where  $i = 1, 2$ ,  $S_1 = \begin{bmatrix} I_{2m+3} & 0 \\ 0 & 0 \end{bmatrix}$ , and  $S_2 = I_{2(m+2)}$ .

Following EJ (2003), we have the following assumptions for the model: i)  $|A(z)| = 0$  has roots outside of the unit circle; ii)  $E_{t-1}(e_t) = 0$ ,  $E_{t-1}(e_t e_t') = \Sigma$  and  $\sup_t E\|e_t\|^{2+\eta} < \infty$  (a.s.) for some  $\eta > 0$ , where  $\Sigma$  is positive definite and  $E_{t-1}(\cdot)$  denotes conditional expectation with respect to  $\{e_{t-1}, e_{t-2}, \dots\}$ ; iii)  $u_0, u_1, \dots, u_{t-k}$ , are  $O_p(1)$ .

Furthermore, using the same notation as in EJ (2003), the expression

$$\Omega = A(1)^{-1} \Sigma A(1)^{-1'} = \begin{bmatrix} \omega_{yy} & \omega_{yx} \\ \omega'_{yx} & \Omega_{xx} \end{bmatrix} \quad (5)$$

is defined as the spectral density at the frequency zero (scaled by  $2\pi$ ) of  $u_t(\rho)$ . Therefore  $R^2 = \omega_{yy}^{-1} \omega_{yx} \Omega_{xx}^{-1} \omega'_{yx}$  is a measure of the long-run correlation between shocks to  $x_t$  and quasi-differences of  $y_t$  at the frequency zero. The value of  $R^2$  represents the contribution of covariates to the explanation of  $y_t$ , and the value of  $R^2$  ranges from zero to unity. When  $R^2 = 0$ , covariates are irrelevant to the potentially integrated series and we have the traditional case where no covariates are present in testing for unit roots. When  $R^2 = 1$ , there is perfect correlation between covariates and quasi-differences of  $y_t$ , and the partial sums of  $x_t$  cointegrates with  $y_t^2$ . In most cases,  $R^2$  ranges from zero to one, and we are concerned with the performance of unit root tests as  $R^2$  increases.

---

<sup>2</sup>If there exists cointegration, the model should be set up using another framework, unless the coefficients of the cointegrating vector are assumed to be known, which is the case analyzed by Elliott, Jansson and Pesavento (2005). Therefore, the range of correlation is  $0 \leq R^2 < 1$  in this paper.

To summarize, we study a VAR model including covariates  $x_t$  and quasi-differences of  $y_t$ . Our interest is to test whether the parameter  $\rho$  is equal to unity, or in other words, whether  $y_t$  has a unit root under the null against the alternative that  $y_t$  has a root less than unity. Using an univariate framework, King (1980, 1988) and Dufour and King (1991) showed that the unit root hypothesis can be examined using Neyman-Pearson tests in a univariate framework. ERS (1996) further developed a family of point optimal tests using quasi-differenced data without considering structural break. PR (2003) extended this type of statistics to models with an unknown structural change. In this paper, we extend them using a multivariate framework originally proposed by Hansen (1995).

We first derive an optimal statistic (denoted by P test) assuming all the nuisance parameters are known except the date of structural break. Accordingly we have the following assumptions: i)  $A(L) = I$ , therefore  $\Omega = \Sigma$ ; ii)  $e_t$  is normally distributed with  $u_{y_0} = 0$ ; iii)  $T_B = T\lambda$  for some  $\lambda \in (0, 1)$ . Based on these assumptions, we test the null hypothesis of  $c = 0$  against the local alternative that  $c = \bar{c} < 0$  where  $c = T(\rho - 1)$  and  $\bar{c} = T(\bar{\rho} - 1)$ . PR (2003) showed that when the break point is unknown, the limiting distribution of the P test can be defined by:

$$P^i(1, \bar{\rho}) = \inf_{\lambda \in [0, 1]} \sum_{t=1}^T \hat{u}_t^i(\bar{\rho}, \lambda)' \Sigma^{-1} \hat{u}_t^i(\bar{\rho}, \lambda) - \inf_{\lambda \in [0, 1]} \sum_{t=1}^T \hat{u}_t^i(1, \lambda)' \Sigma^{-1} \hat{u}_t^i(1, \lambda) - \bar{c}. \quad (6)$$

where in the present case,  $\hat{u}_t^i(r) = z_t(r) - d_t(r)' \hat{\beta}^i$  and  $r = \bar{\rho}, 1$ . When  $t > 1$ , we have:

$$z_t(r) = [(1 - rL)y_t, x_t']'$$

$$d_t(r)' = \begin{bmatrix} 1 - r & 0 & (1 - rL)t & 0 & (1 - rL)(t - T_B) \mathbf{1}(\cdot) & (1 - r) \mathbf{1}(\cdot) \\ 0 & I_m & 0 & I_m t & 0 & 0 \end{bmatrix}$$

while, when  $t = 1$ , we have:

$$z_1(r) = [y_1, x_1']'$$

$$d_1(r)' = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & I_m & 0 & I_m & 0 & 0 \end{bmatrix}.$$

Furthermore,

$$\hat{\beta}^i = [S_i(\sum_{t=1}^T d_t(r) \Sigma^{-1} d_t(r)') S_i]^{-1} [S_i \sum_{t=1}^T d_t(r) \Sigma^{-1} z_t(r)],$$

where  $D^-$  is the Moore Penrose inverse of  $D$ . The next Theorem establishes the limiting distribution of the statistic.

**Theorem 1** *Let  $\{y_t\}$  and  $\{x_t\}$  be generated by (1) to (4), with  $A(L) = I$ ,  $e_t \sim i.i.d. N(0, \Sigma)$ , assumptions (i) to (iii) hold. To test the null hypothesis of  $\rho = 1$  ( $c = 0$ ) against the alternative hypothesis of  $\rho = \bar{\rho} = 1 + \bar{c}/T$  with  $c = \bar{c}$  fixed as  $T \rightarrow \infty$ , the P test have the following asymptotic distribution for Case  $i = 1$  and 2 :*

$$P^i(1, \bar{\rho}) \Rightarrow \Lambda_1(c, \bar{c}, R^2) + \Lambda_2^i(c, \bar{c}, \lambda, R^2) - \bar{c} \quad (7)$$

where  $\Lambda_1(c, \bar{c}, R^2) = (\bar{c}^2 - 2c\bar{c})(1 + \frac{R^2}{1-R^2}) \int W_{1c}^2 - 2\bar{c} \int W_{1c} dW_1 + 2\bar{c} \frac{R}{\sqrt{1-R^2}} \int W_{1c} dW_2$ ,  $W_1$  and  $W_2$  are independent univariate standard Brownian motions.  $W_{1c}$  is an Ornstein-Uhlenbeck process defined by  $W_{1c}(s) = c \int_0^s e^{c(s-h)} W_1(h) dh + W_1(s)$ . The expression  $\Lambda_2(c, \bar{c}, \lambda, R^2)$  is defined in the appendix.

The above limiting distribution has several indications. First, the asymptotic power depends on  $\bar{c}$ , which corresponds to one particular point under the alternative hypothesis. This means that the asymptotic power functions are tangent to the highest power envelope at point  $c = \bar{c}$ . Second, the limiting distribution is nonstandard. Third, the distribution of the P-test also depends on the parameter  $R^2$ . When  $R^2 = 0$ , there is no covariate correlated with the quasi-differences of  $y_t$  and consequently we retrieve the same asymptotic distribution as that derived in PR (2003). When  $R^2$  is greater than zero, the limiting distribution is a function of  $R^2$ , indicating that extra information contained in the covariates may make a difference on the performance of the test.

Using Theorem 1,  $T = 1,000$ , and 10,000 replications, the asymptotic power functions and power envelopes at different  $R^2$  can be calculated. The simulation results for  $R^2 = 0.0, 0.3, 0.5, 0.7, 0.9$  are graphed in Figure 1. We observe that the power envelopes reach their lower bound when

$R^2 = 0$ . When  $R^2$  is greater than zero, the power attainable increases considerably. For example, when  $c = -10$ , the maximum power of  $R^2 = 0.0, 0.3, 0.5, 0.7, 0.9$  are 10.61%, 21.27%, 35.14%, 55.33%, and 85.95% respectively. That is, the power increases by 100% from  $R^2 = 0.0$  to  $R^2 = 0.3$ ; by 65.2% from  $R^2 = 0.3$  to 0.5; by 57.5% from  $R^2 = 0.5$  to 0.7; and by 55.3% from  $R^2 = 0.7$  to 0.9. These evidence suggests that incorporating the information in covariates can substantially increase the power of unit root test in large samples. This result is consistent with what has been found in Hansen (1995) and EJ (2003).

### 3 A Feasible Test

In the previous section, we assumed all the nuisance parameters except the break point are known and the errors are normally distributed. In practice,  $R^2$ , the correlation between  $x_t$  and the variable to be tested is not known, and it is likely that there is serial correlation in the time series. Therefore it is necessary to develop a feasible test to estimate all the unknown nuisance parameters. In this section, we derive this test by adopting the approach of EJ (2003) in the first three steps when constructing the test, and we modify the last step in order to estimate the break point. The construction of the test is as follows:

1. Estimate  $R^2$ . It is performed by running a VAR model:  $A(L)z_t(1, \lambda) = d_t + e_t$ , where  $\lambda$  is the break point,  $d_t$  is the deterministic components specified according to the model under analysis. Then we use the residuals generated from the VAR model to construct  $\hat{\Sigma} = T^{-1} \sum_{t=k+1}^T \hat{e}_t(\lambda) \hat{e}_t(\lambda)'$ ,  $\hat{\Omega} = \hat{A}(1, \lambda)^{-1} \hat{\Sigma} \hat{A}(1, \lambda)^{-1'}$ , and obtain  $\hat{R}^2 = \hat{\omega}_{yx} \hat{\Omega}_{xx}^{-1} \hat{\omega}'_{yx} / \hat{\omega}_{yy}$  where  $\hat{A}(1, \lambda) = I + \sum_{i=1}^k A_i$  is the  $(i+1)th$  matrix element of  $A(L)$ .
2. Estimate the nuisance parameters for quasi-differencing and detrend the data under the null and alternative hypothesis, respectively. That is, construct  $\tilde{u}_t^i(r, \lambda) = z_t(r, \lambda) - d_t(r, \lambda)' \tilde{\beta}^i(r, \lambda)$  when  $r = 1, \rho$ , and

$$\tilde{\beta}^i(r, \lambda) = \left[ \sum_{t=1}^T d_t(r, \lambda) \hat{\Omega}^{-1} d_t(r, \lambda)' \right]^{-1} \left[ \sum_{t=1}^T d_t(r, \lambda) \hat{\Omega}^{-1} z_t(r, \lambda) \right].$$

3. Run the VAR model  $\tilde{A}(L)\tilde{u}_t^i(r, \lambda) = \tilde{e}_t^i(r, \lambda)$ , using  $\tilde{u}_t^i(r, \lambda)$  obtained from the previous step, and construct the variance-covariance matrices

$$\tilde{\Sigma}^i(r, \lambda) = T^{-1} \sum_{t=k+1}^T \tilde{e}_t^i(r, \lambda) \tilde{e}_t^i(r, \lambda)'$$

4. Select the break point and construct the test statistic. The break point is selected by choosing the variance-covariance matrices  $\tilde{\Sigma}(1)$  and  $\tilde{\Sigma}(\bar{\rho})$  minimizing  $\sum_{t=k+1}^T \tilde{e}_{y,t}^i(r, \lambda) \tilde{e}_{y,t}^i(r, \lambda)'$  for  $r = 1, \bar{\rho}$ , respectively (see appendix for proof). Then we can construct the test statistic

$$\tilde{P}^i(1, \bar{\rho}) = T\{tr[\tilde{\Sigma}(1)^{-1}\tilde{\Sigma}(\bar{\rho})] - [m + \bar{\rho}]\}. \quad (8a)$$

This test has the asymptotic power that achieves the power envelope bound at  $\bar{c}$  under assumptions (i)-(iii). The following Theorem establishes its limiting distribution.

**Theorem 2** *Let  $\{y_t\}$  and  $\{x_t\}$  be generated by (1) to (4), assumptions (i) to (iii) hold, also assuming deterministic components are correctly specified for  $i = 1, 2$ . To test the null hypothesis of  $\rho = 1$  ( $c = 0$ ) against the alternative hypothesis of  $\rho = \bar{\rho} = 1 + \bar{c}/T$  with  $c = \bar{c}$  fixed as  $T \rightarrow \infty$ , the limiting distribution of the test defined in (8a) has asymptotic distribution:*

$$\tilde{P}^i(1, \bar{\rho}) \Rightarrow \Lambda_1(c, \bar{c}, R^2) + \Lambda_2^i(c, \bar{c}, \lambda, R^2) - \bar{c}. \quad (9)$$

As what we have observed before, the limiting distributions of the feasible test and hence the asymptotic results depend on the selection of  $\bar{c}$  for detrending and  $R^2$ . We have shown that the feasible test asymptotically achieves the highest possible power at  $\bar{c}$ . Ideally we should calculate a power envelope corresponding to each  $R^2$  and our choice of  $\bar{c}$  for quasi-differencing should be dependent on  $R^2$ . However, in practice we may use  $\bar{c} = -22.5$ , which is the same value selected by PR (2003) when  $R^2 = 0$ , for *GLS* detrending. Because as  $R^2$  increases, the power of the test increases dramatically and the choice of  $\bar{c}$  becomes less important; see EJ, 2003.

We present the asymptotic critical values in Table 1. The simulations of the limiting distribution are done by using 1,000 steps to approximate the Wiener process on  $[0, 1]$  as the partial sums of *i.i.d.*  $N(0, 1)$  random variables. As a comparison, we also present finite sample critical values for both cases analyzed in the paper. We use  $T = 100$  and  $200$  with data generated by a random walk with zero initial value and *i.i.d.*  $N(0, 1)$  errors. The lag length is set to zero in the estimation of the VAR model in step one.

## 4 Evaluation of the Test

Figure 2 graphs the asymptotic power of the feasible test for both cases and  $R^2 = 0.0, 0.3, 0.5, 0.7, 0.9$ . We first observe that the feasible test has power tangent to the power envelope at one point and never falls far below it. This suggests that very little asymptotic power is lost at points away from the tangent point, especially for lower values of  $R^2$ . Overall, substantial power gains are realized due to covariates. Consider the asymptotic power gains from using covariates when  $R^2 = 0.5$  and the local alternative  $c = -10$ . The asymptotic power rises by 217% (power is 34.52% when  $R^2 = 0.5$ , and 10.86% when  $R^2 = 0.0$ ). Using the local alternative  $c = -4$ , power gains are smaller, but still an increase of approximately 87% (power is 8.98% when  $R^2 = 0.5$ , while it is 4.8% when  $R^2 = 0.0$ ).

The test is also evaluated using small sample Monte Carlo analysis. To simulate the critical values, power and size,  $\{y_t\}$  and  $\{x_t\}$  are generated using (1) to (4) with normal errors and known variance-covariance matrix, whose variances equal to one and covariances equal to 0.3, 0.5, 0.7, 0.9 respectively. Using 10,000 replications,  $T = 100, 200$ , finite sample critical values are calculated and tabulated in Table 1 along with the asymptotic critical values. Power and size are presented in Table 2 (Case 1) and Table 3 (Case 2). There are two factors that influence the finite sample power performance. One is the number of observations, power increases as  $T$  increases. For example, when  $\rho = 0.90$ ,  $R^2 = 0.81$ , the power increases 15% when  $T$  increases from 100 to 200 for Case 1 and 26% for Case 2. Another factor is the increase of correlation between the covariate and the variable to be tested. As  $R$  (or  $R^2$ ) is closer to 1, power increases dramatically. For example, power

increases 19% as  $R$  increases from 0.3 to 0.5 in Table 2 when  $\rho = 0.90$ .

In terms of the overall power gains in finite sample, we have the following observations. For Case 1, when  $T = 100$  and  $\rho = 0.96$ , the test has a power of 7.36% without covariates, while power increases to 11.57% when  $R^2 = 0.49$ , approximately a 57.2% power gain; while for Case 2, approximately a 52.4% power gain. For Case 1, when  $T = 200$ ,  $\rho = 0.96$ , the power is 11.87% without covariates, while it increases to 29.19% when  $R^2 = 0.49$ , approximately a 146% power gain; while for Case 2, approximately a 120% power gain.

## 5 Conclusions

Following previous research (Hansen 1995; EJ 2003; Shin 2001; PR 2003), we introduce an arbitrary number of covariates to unknown structural change models. We have derived the test statistic that has asymptotic power function tangent to the power envelope at one point under the alternative hypothesis. We have found that the proposed test is a function of both  $\bar{c}$  and  $R^2$  (the correlation of covariates and quasi-differences of  $y_t$ ). Large sample simulations have shown substantial power gains when  $R^2$  is greater than zero, especially when  $R^2$  is close to 1. Monte Carlo simulations have also shown dramatic power improvements in finite samples without size deterioration. To be of more empirical relevance, we have developed a feasible test when there are serial correlations, and when  $R^2$  and break point are unknown. Our study has shown that this test achieves the power envelope when  $T \rightarrow \infty$ .

## References

- [1] Chan, N. H. and C. Z. Wei (1987), "Asymptotic Inference for Nearly Nonstationary AR(1) Processes," *Annals of Statistics* **15**, 1050-63.
- [2] Dickey, D. A. and W. A. Fuller (1979), "Distribution of the Estimator for Autoregressive Time Series with a Unit Root," *Journal of the American Statistical Association* **74**, 427-431.

- [3] Dufour, J.-M. and M. King (1991), "Optimal Invariant Tests for the Autocorrelation Coefficient in Linear Regressions with Stationary or Nonstationary Errors," *Journal of Econometrics* **47**, 115-143.
- [4] Elliott, G. and M. Jansson (2003), "Testing for Unit Roots with Stationary Covariate," *Journal of Econometrics* **115**, 75-89.
- [5] Elliott, G., M. Jansson, and E. Pesavento (2005), "Optimal Power for Testing Potential Cointegration Vectors with Known Parameters for Nonstationarity," *Journal of Business & Economic Statistics* **23**, 34-48.
- [6] Hansen, B. E. (1995), "Rethinking the Univariate Approach to Unit Root Testing," *Econometric Theory* **11**, 1148-1171.
- [7] King, M. L. (1980), "Robust Test for Spherical Symmetry and Their Application to Least Squares Regression," *Annals of Statistics* **8**, 1265-1271.
- [8] King, M. L. (1988), "Towards a Theory of Point Optimal Testing," *Econometrics Review* **6**, 169-218.
- [9] Perron, P. and G. Rodríguez (2003), "GLS Detrending, Efficient Unit Root Tests and Structural Change," *Journal of Econometrics* **115**, 1-27.
- [10] Phillips, P. C. B. (1987), "Time Series Regression with Unit Roots," *Econometrica* **55**, 277-302.
- [11] Said, S. E. and D. A. Dickey (1984), "Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order," *Biometrika* **71**, 599-608.
- [12] Shin, S. (2001), "Covariate Unit Root Test Allowing for a Structural Change," Department of Economics Working Paper #01-14, The Ohio State University.

## Appendix

**Proof of Theorem 1.** We derive the limiting distribution of the P-test for Case  $i = 1, 2$ . The P-test statistic is given by

$$P^i(1, \bar{\rho}) = \inf \sum_{t=1}^T \hat{u}_t^i(\bar{\rho})' \Sigma^{-1} \hat{u}_t^i(\bar{\rho}) - \inf \sum_{t=1}^T \hat{u}_t^i(1)' \Sigma^{-1} \hat{u}_t^i(1) - \bar{c},$$

where  $\hat{u}_t^i(r) = z_t(r) - d_t(r)' \hat{\beta}^i$  for  $r = \bar{\rho}, 1$ .

When  $t > 1$ ,

$$\begin{aligned} z_t(r) &= \begin{bmatrix} (1-rL) y_t \\ x_t \end{bmatrix}, \\ d_t(r)' &= \begin{pmatrix} 1-r & 0 & (1-rL)t & 0 & (1-rL)(t-T_B) \mathbf{1}(\cdot) & (1-r) \mathbf{1}(\cdot) \\ 0 & I_m & 0 & I_m t & 0 & 0 \end{pmatrix}; \end{aligned}$$

while when  $t = 1$ ,

$$\begin{aligned} z_1(r) &= \begin{bmatrix} y_1 \\ x_1 \end{bmatrix}, \\ d_1(r)' &= \begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & I_m & 0 & I_m & 0 & 0 \end{pmatrix}, \end{aligned}$$

where  $\mathbf{1}(\cdot)$  is the indicator function. Using OLS,

$$\hat{\beta}^i = [S_i (\sum_{t=1}^T d_t(r) \Sigma^{-1} d_t(r)') S_i]^{-1} [S_i \sum_{t=1}^T d_t(r) \Sigma^{-1} z_t(r)].$$

Therefore the P-test has the following form:

$$\begin{aligned} P^i(1, \bar{\rho}) &= \sum_{t=1}^T e_t(\bar{\rho})' \Sigma^{-1} e_t(\bar{\rho}) - \sum_{t=1}^T e_t(1)' \Sigma^{-1} e_t(1) - \bar{c} \\ &\quad \inf [S_i N_T(1)]' [S_i D_T(1) S_i]^{-1} [S_i N_T(1)] - \\ &\quad \inf [S_i N_T(\bar{\rho})]' [S_i D_T(\bar{\rho}) S_i]^{-1} [S_i N_T(\bar{\rho})] \end{aligned}$$

where  $D_T(r) = \sum_{t=1}^T [\Psi_T^{-1} d_t(r) \Sigma^{-1} d_t(r)' \Psi_T^{-1}]$ ,  $N_T(r) = \sum_{t=1}^T [\Psi_T^{-1} d_t(r) \Sigma^{-1} e_t(r)]$ ,  $\Psi_T$  is a scaling matrix. Because  $s_1 \Sigma^{-1} s_1' = (1 + \bar{\delta}' \bar{\delta}) \omega_{yy}^{-1}$  and  $s_1 \Sigma^{-1/2'} =$



$$\lim_{T \rightarrow \infty} \left[ \sup_{\frac{2}{T} \leq s \leq 1} T^{1/2} \Psi_T^{-1} d_{[Ts]}(r) \Sigma^{-1/2'} \right] \Rightarrow \begin{pmatrix} 0 & 0 \\ 0 & I_m \\ 1 - c_r s & -(1 - c_r s) \bar{\delta}' \\ 0 & s I_m \\ 1 - c_r (s - \lambda) & -\bar{\delta}' [1 - c_r (s - \lambda)] \\ 0 & 0 \end{pmatrix}, \quad (\text{A.11})$$

where  $\lambda = T_B/T$  is the break point. Combining the above results (A.10) and (A.11) we have

$$\begin{aligned} D_T(r) &= \sum_{t=1}^T [\Psi_T^{-1} d_t(r) \Sigma^{-1} d_t(r)' \Psi_T^{-1}] \\ &= \Psi_T^{-1} d_1(r) \Sigma^{-1} d_1(r)' \Psi_T^{-1} + \sum_{t=2}^T [\Psi_T^{-1} d_t(r) \Sigma^{-1} d_t(r)' \Psi_T^{-1}] \\ &\Rightarrow \begin{bmatrix} 1 + \bar{\delta}' \bar{\delta} & 0 & 0 \\ 0 & I_m & -(1 - c_r/2) \bar{\delta} \\ 0 & -(1 - c_r/2) \bar{\delta}' & (1 + c_r^2/3 - c_r)(1 + \bar{\delta}' \bar{\delta}) \\ 0 & I_m/2 & -(1/2 - c_r/3) \bar{\delta} \\ 0 & \bar{\delta}' a & (1 + \bar{\delta}' \bar{\delta}) v \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ I_m/2 & \bar{\delta} a & 0 \\ -(1/2 - c_r/3) \bar{\delta} & (1 + \bar{\delta}' \bar{\delta}) v & 0 \\ I_m/3 & \bar{\delta} b & 0 \\ \bar{\delta}' b & (1 + \bar{\delta}' \bar{\delta}) d & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad (\text{A.12}) \end{aligned}$$

where

$$\begin{aligned} a &= 1 - \lambda - c_r \lambda + c_r \lambda^2/2, \\ b &= -\frac{1 - \lambda^2}{2} + \frac{c_r}{3} + \frac{c_r \lambda^3}{6} - \frac{c_r \lambda}{2}, \\ v &= 1 - \lambda - c_r + c_r \lambda - c_r^2 \lambda/2 + c_r^2 \lambda^3/2 + c_r^2 (1 - \lambda^3)/3, \\ d &= 1 - \lambda - c_r - c_r \lambda^2 + 2c_r \lambda + c_r^2/3 - c_r^2 \lambda^3/3 - c_r^2 \lambda + c_r^2 \lambda^2. \end{aligned}$$

Similarly, we can calculate the limiting distribution for  $N_T(r) = \sum_{t=1}^T [\Psi_T^{-1} d_t(r) \Sigma^{-1} e_t(r)]$ ,

$$\begin{aligned}
N_T(r) &= \sum_{t=1}^T [\Psi_T^{-1} d_t(r) \Sigma^{-1} e_t(r)] \\
&= \sum_{t=1}^T \Psi_T^{-1} [d_t(r) \Sigma^{-1/2'}] [\Sigma^{-1/2} e_t(r)] \\
&= \Psi_T^{-1} [d_1(r) \Sigma^{-1/2'}] [\Sigma^{-1/2} e_t(r)] + \sum_{t=2}^T \Psi_T^{-1} [d_t(r) \Sigma^{-1/2'}] [\Sigma^{-1/2} e_t(r)] \\
&= \Psi_T^{-1} [d_1(r) \Sigma^{-1/2'}] [\Sigma^{-1/2} e_t(r)] + \\
&\quad T^{-1/2} \sum_{t=2}^T [T^{1/2} \Psi_T^{-1} d_t(r) \Sigma^{-1/2'}] [\Sigma^{-1/2} e_t(r)] \tag{A.13}
\end{aligned}$$

$$\Rightarrow \begin{pmatrix} \epsilon_{y,1} - \bar{\delta}' \epsilon_{x,1} \\ \epsilon_{x,t} + (\rho - r) \Sigma^{-1/2} s_1' u_{y,t-1} \\ \int_0^1 (1 - c_r s) [dW_1(s) - \bar{\delta}' V(s)] + \\ (c - c_r)(1 + \bar{\delta}' \bar{\delta}) \int_0^1 (1 - c_r s) W_{1c}(s) ds \\ \int_0^1 s dV(s) - (c - c_r) \bar{\delta} \int s W_{1c}(s) ds \\ G \\ 0 \end{pmatrix}, \tag{A.14}$$

where

$$\begin{aligned}
G &= \int_{\lambda}^1 (1 - c_r s) [dW_1(s) - \bar{\delta}' V(s)] + (c - c_r)(1 + \bar{\delta}' \bar{\delta}) \int_{\lambda}^1 (1 - c_r s) W_{1c} ds \\
&\quad + c_r \lambda \int_{\lambda}^1 d[W_1(s) - \bar{\delta}' V(s)] + c_r \lambda (c - c_r)(1 + \bar{\delta}' \bar{\delta}) \int_{\lambda}^1 W_{1c}(s) ds.
\end{aligned}$$

Thus the limiting distribution for Case 2 is:

$$P^i[1, \bar{\rho}] \Rightarrow \Lambda_1(c, \bar{c}, R^2) + \Lambda_2^i(c, \bar{c}, R^2) - \bar{c}$$

where expression  $\Lambda_2^i(c, \bar{c}, R^2)$  is the limiting distribution of  $\inf [S_i N_T(1)]' [S_i D_T(1) S_i]^{-1} [S_i N_T(1)] - \inf [S_i N_T(\bar{\rho})]' [S_i D_T(\bar{\rho}) S_i]^{-1} [S_i N_T(\bar{\rho})]$ , which can be calculated using (A.12) and (A.14). Proof for Case 2 completes.

Next we want to show that Case 1 has the same limiting distribution as Case 2. For Case 1,

$$\begin{aligned} D_T(r) &= \sum_{t=1}^T [\Psi_T^{-1} d_t(r) \Sigma^{-1} d_t(r)' \Psi_T^{-1}] \\ &= \Psi_T^{-1} d_1(r) \Sigma^{-1} d_1(r)' \Psi_T^{-1} + \sum_{t=2}^T [\Psi_T^{-1} d_t(r) \Sigma^{-1} d_t(r)' \Psi_T^{-1}] \end{aligned}$$

When  $t > 1$ ,

$$d_t(r)' = \begin{pmatrix} 1-r & 0 & (1-rL)t & 0 & (1-rL)(t-T_B)\mathbf{1}(\cdot) & 0 \\ 0 & I_m & 0 & I_m t & 0 & 0 \end{pmatrix},$$

and

$$\lim_{T \rightarrow \infty} [ \sup_{\frac{2}{T} \leq s \leq 1} T^{1/2} \Psi_T^{-1} d_{[Ts]}(r) \Sigma^{-1/2'} ] \Rightarrow \begin{pmatrix} 0 & 0 \\ 0 & I_m \\ 1 - c_r s & -(1 - c_r s) \bar{\delta}' \\ 0 & s I_m \\ 1 - c_r (s - \lambda) & -\bar{\delta}' [1 - c_r (s - \lambda)] \\ 0 & 0 \end{pmatrix}.$$

Therefore,  $D_T(r)$  and  $N_T(r)$  have the same asymptotic distributions for Case 1 and 2, and so is the limiting distribution for P test. Proof completes ■

**Proof of Theorem 2.** Following EJ (2003),

1.  $\sum_{t=k+1}^T \tilde{e}_t^i(\bar{\rho}) \tilde{e}_t^i(\bar{\rho})' - \sum_{t=k+1}^T \tilde{e}_t^i(1) \tilde{e}_t^i(1)' = \sum_{t=k+1}^T \tilde{e}_t^i(\bar{\rho}) \tilde{e}_t^i(\bar{\rho})' - \sum_{t=k+1}^T \tilde{e}_t^i(1) \tilde{e}_t^i(1)' + o_p(1)$ , where  $\tilde{e}_t^i(r) = A(L) \tilde{u}_t^i(r)$ .
2.  $\sum_{t=k+1}^T \tilde{e}_t^i(r)' \Sigma^{-1} \tilde{e}_t^i(r) - \sum_{t=k+1}^T e_t(r)' \Sigma^{-1} e_t(r) \Rightarrow -[S_i N(c, \bar{c}, R^2)]' [S_i D(\bar{c}, R^2) S_i]^{-1} [S_i N(c, \bar{c}, R^2)]$ ,
3.  $\sum_{t=k+1}^T e_t(\bar{\rho})' \Sigma^{-1} e_t(\bar{\rho}) - \sum_{t=k+1}^T e_t(1)' \Sigma^{-1} e_t(1) \Rightarrow \Lambda_1^i(c, \bar{c}, R^2)$ .

Using the identity  $tr[S \times e \times e'] = e' \times S \times e$ ,  $S = \Sigma^{-1} = \Sigma(1)^{-1}$ , we can transform the equation 1:

$$\begin{aligned}
& \inf_{\lambda \in [0,1]} [tr[\Sigma^{-1} \sum_{t=k+1}^T \tilde{e}_t^i(\bar{\rho}, \lambda) \tilde{e}_t^i(\bar{\rho}, \lambda)']] - \inf_{\lambda \in [0,1]} [tr[\Sigma^{-1} \sum_{t=k+1}^T \tilde{e}_t^i(1, \lambda) \tilde{e}_t^i(1, \lambda)']] \\
&= \inf_{\lambda \in [0,1]} [tr[\Sigma^{-1} \sum_{t=k+1}^T \hat{e}_t^i(\bar{\rho}, \lambda) \hat{e}_t^i(\bar{\rho}, \lambda)']] - \inf_{\lambda \in [0,1]} [tr[\Sigma^{-1} \sum_{t=k+1}^T \hat{e}_t^i(1, \lambda) \hat{e}_t^i(1, \lambda)']] + o_p(1) \\
&= \inf_{\lambda \in [0,1]} \sum_{t=k+1}^T \hat{e}_t^i(\bar{\rho}, \lambda)' \Sigma^{-1} \hat{e}_t^i(\bar{\rho}, \lambda) - \inf_{\lambda \in [0,1]} \sum_{t=k+1}^T \hat{e}_t^i(1, \lambda)' \Sigma^{-1} \hat{e}_t^i(1, \lambda) + o_p(1). \quad (\text{A.15})
\end{aligned}$$

From equation 2 and 3, we have

$$\begin{aligned}
& \sum_{t=k+1}^T \hat{e}_t^i(\bar{\rho})' \Sigma^{-1} \hat{e}_t^i(\bar{\rho}) - \sum_{t=k+1}^T \hat{e}_t^i(1)' \Sigma^{-1} \hat{e}_t^i(1) \\
&= \sum_{t=1}^T e_t(\bar{\rho})' \Sigma^{-1} e_t(\bar{\rho}) - \sum_{t=1}^T e_t(1)' \Sigma^{-1} e_t(1) \\
&\quad - [S_i N(c, \bar{c}, R^2)]' [S_i D(\bar{c}, R^2) S_i]^{-1} [S_i N(c, \bar{c}, R^2)] \\
&\quad + [S_i N(c, 0, R^2)]' [S_i D(\bar{c}, R^2) S_i]^{-1} [S_i N(c, 0, R^2)] \quad (\text{A.16})
\end{aligned}$$

Now, substitute (A.16) into (A.15), we get

$$\begin{aligned}
& \inf_{\lambda \in [0,1]} [tr[\Sigma^{-1} \sum_{t=k+1}^T \tilde{e}_t^i(\bar{\rho}, \lambda) \tilde{e}_t^i(\bar{\rho}, \lambda)']] - \inf_{\lambda \in [0,1]} [tr[\Sigma^{-1} \sum_{t=k+1}^T \tilde{e}_t^i(1, \lambda) \tilde{e}_t^i(1, \lambda)']] \\
&\Rightarrow \sum_{t=1}^T e_t(\bar{\rho}, \lambda)' \Sigma^{-1} e_t(\bar{\rho}, \lambda) - \sum_{t=1}^T e_t(1, \lambda)' \Sigma^{-1} e_t(1, \lambda) \\
&\quad - \sup_{\lambda \in [0,1]} [S_i N(c, \bar{c}, R^2, \lambda)]' [S_i D(\bar{c}, R^2) S_i]^{-1} [S_i N(c, \bar{c}, R^2, \lambda)] \\
&\quad + \sup_{\lambda \in [0,1]} [S_i N(c, 0, R^2, \lambda)]' [S_i D(\bar{c}, R^2, \lambda) S_i]^{-1} [S_i N(c, 0, R^2, \lambda)] \\
&\equiv P^i(1, \bar{\rho}).
\end{aligned}$$

In our model, only  $y_t$  contains a structural break. Therefore we use  $\tilde{\Sigma}(r) = \inf_{\lambda \in [0,1]} \sum_{t=k+1}^T \tilde{e}_{y,t}^i(r, \lambda)' \Sigma^{-1} \tilde{e}_{y,t}^i(r, \lambda)$  to select the break point, and the feasible test statistic is,

$$\tilde{P}^i(1, \bar{\rho}) = T \{tr[\tilde{\Sigma}(1)^{-1} \tilde{\Sigma}(\bar{\rho})] - [m + \bar{\rho}]\} \blacksquare$$

Table 1. Asymptotic and Finite Critical Values for  $P$  test choosing  $T_B$  minimizing the statistic;  
 $(\bar{c} = -22.5$  when constructing the test)

	Size	$T = \infty$	$T = 100$		$T = 200$	
			Case 1	Case 2	Case 1	Case 2
$R^2 = 0.0$	.01	6.6206	7.0831	8.2017	6.8656	7.9770
	.025	7.7184	8.3496	9.6151	8.2074	9.3505
	.05	8.7784	9.6171	10.7726	9.4273	10.4317
	.10	10.2240	11.2630	12.2715	11.0206	12.0643
	.20	12.4409	13.2986	14.1866	13.0273	14.0473
$R^2 = 0.1$	.01	6.3306	7.1161	7.7447	6.8647	7.4366
	.025	7.5704	8.3208	8.9815	8.1843	8.8018
	.05	8.8655	9.5971	10.3549	9.3553	10.0709
	.10	10.7418	11.1881	12.2124	11.0294	11.8583
	.20	13.3284	13.6536	14.7468	13.5258	14.5056
$R^2 = 0.2$	.01	6.3366	7.0755	7.6763	6.7214	7.5043
	.025	7.7627	8.5707	9.2372	8.3291	9.0627
	.05	9.3670	10.0312	10.9802	9.8132	10.6892
	.10	11.5830	11.9907	13.1029	11.8267	12.7795
	.20	14.6210	14.8797	16.2305	14.7595	15.9071
$R^2 = 0.3$	.01	6.3418	7.2735	8.1307	6.7738	7.4329
	.025	8.1433	9.1303	9.8774	8.7816	9.6521
	.05	10.1212	10.9046	11.8302	10.6501	11.5858
	.10	12.7136	13.2524	14.4177	12.9126	14.0307
	.20	16.2374	16.6372	18.0740	16.4752	17.6493
$R^2 = 0.4$	.01	6.4881	7.8499	8.7390	7.2775	8.2742
	.025	8.8891	10.0639	10.9070	9.7519	10.5740
	.05	11.3466	12.2461	13.3519	11.8917	12.9659
	.10	14.3757	15.0278	16.2870	14.6182	15.8075
	.20	18.4936	19.0608	20.5460	18.7793	20.4266

Table 1 (continues). Asymptotic and Finite Critical Values for  $P$  test choosing  $T_B$   
 minimizing the statistic;  
 ( $\bar{c} = -22.5$  when constructing the test)

	Size	$T = \infty$	$T = 100$		$T = 200$	
			Case 1	Case 2	Case 1	Case 2
$R^2 = 0.5$	.01	7.1952	9.1102	9.6599	8.4968	9.2217
	.025	10.0659	11.7718	12.8048	11.3839	12.3373
	.05	13.0346	14.3242	15.6562	13.9767	14.9083
	.10	16.7505	17.6416	19.1295	17.1750	18.7120
	.20	21.5898	22.5719	24.3645	22.2750	23.9609
$R^2 = 0.6$	.01	8.3182	11.2399	12.0004	10.8230	12.0457
	.025	12.1112	14.5488	15.8906	14.0533	15.1350
	.05	15.7378	17.7247	19.2281	17.3959	18.5054
	.10	20.2065	21.9735	23.5982	21.3516	23.1323
	.20	26.1940	28.0226	30.1219	27.5030	29.5977
$R^2 = 0.7$	.01	10.0794	15.3051	16.5035	15.1582	16.4011
	.025	15.6598	19.7340	21.1399	19.2412	20.3855
	.05	20.0063	24.0217	25.8066	23.3866	24.6636
	.10	26.0101	29.2556	31.4516	28.7144	30.6317
	.20	33.8148	37.2448	40.0105	36.5297	39.0657
$R^2 = 0.8$	.01	13.9343	24.6019	26.0516	24.7058	26.6857
	.025	22.0148	30.7935	32.8000	30.3721	32.2811
	.05	28.8717	36.7978	39.1595	36.0741	38.1094
	.10	37.6966	44.5746	47.0921	43.7420	46.6231
	.20	48.6718	55.6362	59.6670	54.7744	58.7771
$R^2 = 0.9$	.01	25.6335	54.3674	56.9218	55.6351	58.4940
	.025	43.7594	64.9650	69.1188	65.2765	68.7000
	.05	55.7402	76.4949	80.8711	75.554	80.1414
	.10	72.5938	90.7439	95.3365	89.8564	95.8853
	.20	92.7436	111.4939	119.3781	110.4932	118.0968

Table 2. Finite Sample Size and Power; Case 1  
( $\bar{c} = -22.5$  when constructing the test)

	$T = 100$					$T = 200$				
$R =$	0.0	0.30	0.50	0.70	0.90	0.0	0.30	0.50	0.70	0.90
$R^2 =$	0.0	0.09	0.25	0.49	0.81	0.0	0.09	0.25	0.49	0.81
$\rho$										
1.00	0.0501	0.0501	0.0501	0.0501	0.0501	0.0501	0.0501	0.0500	0.0500	0.0500
0.98	0.0565	0.0567	0.0610	0.0755	0.1033	0.0733	0.0752	0.0873	0.1206	0.2121
0.96	0.0736	0.0750	0.0795	0.1157	0.2082	0.1187	0.1378	0.1679	0.2919	0.6508
0.94	0.0896	0.0967	0.1127	0.1838	0.3956	0.2029	0.252	0.3214	0.5796	0.9661
0.92	0.1344	0.1462	0.1634	0.2894	0.6482	0.3455	0.3555	0.5306	0.8325	0.9997
0.90	0.1839	0.1948	0.2312	0.4284	0.8633	0.5218	0.5409	0.7441	0.9525	1.0000
0.88	0.2390	0.2592	0.3252	0.5825	0.9685	0.6796	0.7223	0.8821	0.9902	1.0000
0.86	0.3156	0.3777	0.4226	0.7150	0.9942	0.8030	0.8511	0.9498	0.9986	1.0000

Number of lags in VAR is zero; 10,000 replications.

Table 3. Finite Sample Size and Power; Case 2  
( $\bar{c} = -22.5$  when constructing the test)

	$T = 100$					$T = 200$				
$R =$	0.0	0.30	0.50	0.70	0.90	0.0	0.30	0.50	0.70	0.90
$R^2 =$	0.0	0.09	0.25	0.49	0.81	0.0	0.09	0.25	0.49	0.81
$\rho$										
1.00	0.0500	0.0501	0.0501	0.0500	0.0501	0.0501	0.0501	0.0500	0.0501	0.0501
0.98	0.0585	0.0595	0.0615	0.0775	0.0952	0.0628	0.0674	0.0832	0.1078	0.1882
0.96	0.0765	0.0775	0.0798	0.1166	0.1886	0.1239	0.1323	0.1706	0.2726	0.5875
0.94	0.0965	0.0978	0.1136	0.1788	0.3513	0.2269	0.2588	0.3209	0.5329	0.8996
0.92	0.1427	0.1440	0.1613	0.2738	0.5768	0.3646	0.4100	0.5417	0.7947	0.9726
0.90	0.1921	0.1954	0.2262	0.4048	0.7867	0.5269	0.6557	0.7408	0.9251	0.9914
0.88	0.2529	0.2616	0.3216	0.5549	0.9114	0.6597	0.7198	0.8669	0.9735	0.9961
0.86	0.3277	0.3519	0.4168	0.6798	0.9550	0.7750	0.8580	0.9473	0.9892	0.9992

Number of lags in VAR is zero; 10,000 replications.

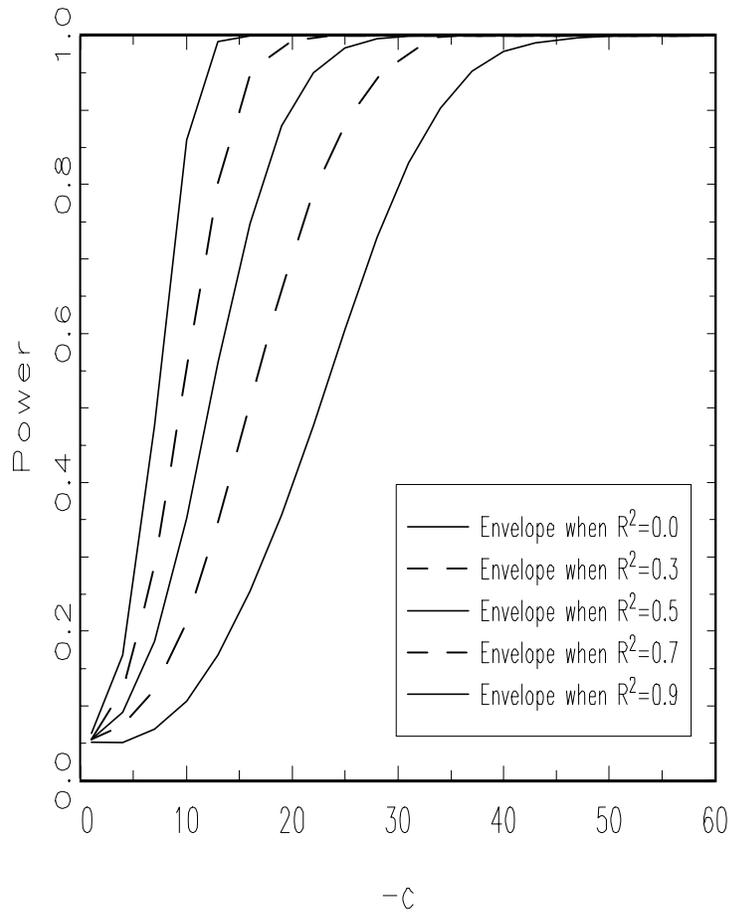


Figure 1. Power Envelopes for  $R^2 = 0.0, 0.3, 0.5, 0.7, 0.9$ .

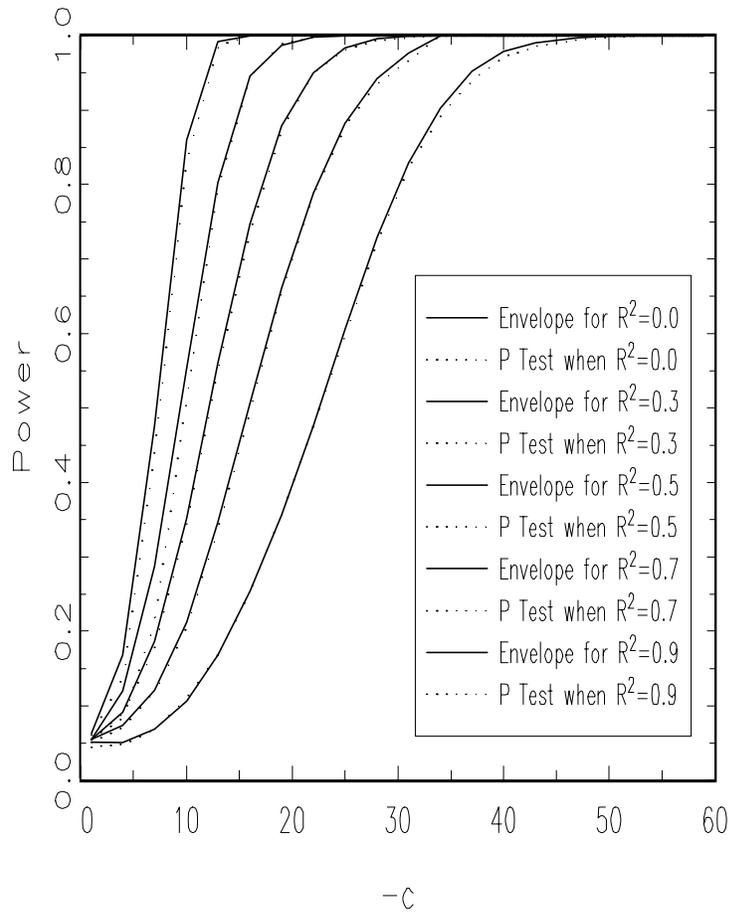


Figure 2. Power Envelopes and Asymptotic Power Functions of the Feasible Point Optimal Test