LONG-RUN PPP MAY NOT HOLD
AFTER ALL

Charles Engel

LONG-RUN PPP MAY NOT HOLD
AFTER ALL

ABSTRACT

Recent tests using long data series find evidence in favor of long-run PPP (by rejecting either the null hypothesis of unit roots in real exchange rates or the null of no cointegration between nominal exchange rates and relative prices.) These tests may have reached the wrong conclusion. Monte Carlo experiments using artificial data calibrated to nominal exchange rates and disaggregated data on prices show that tests of long-run PPP have serious size biases. They may fail to detect a sizable and economically significant unit root component. For example, in the baseline case which is calibrated to actual price data, unit roots and cointegration tests with a nominal size of five percent have true sizes that range from .90 to .98 in artificial 100-year long data series, even though the unit root component accounts for 42% of the variance of the real exchange rate in sample. On the other hand, tests of stationarity are shown to have very low power in the same circumstances, so it is quite likely that a researcher would reject a unit root and fail to reject stationarity even when the series embodied a large unit root component.

Charles Engel
Department of Economics
University of Washington
Seattle, WA 98195
and NBER

Recent work on purchasing power parity (PPP) among high-income countries has found evidence in favor of the hypothesis that real exchange rates converge to their PPP level in the long run. This work (for example, Frankel (1986), Edison (1987)) reaches a conclusion opposite from earlier work that found real exchange rates to have unit roots. The more recent work uses longer sample periods (100 years or more), which imbue the tests with greater power to reject the null hypothesis of a unit root.

However, the more recent literature may have reached the wrong conclusion. While the power of these tests over long horizons is great, there may be a serious size bias in the tests. We argue that when a random variable evolves according to the sum of two processes -- a stationary but persistent component and a non-stationary component -- tests for unit roots are incorrectly sized. Here, we contend that there is a specific source of non-stationary behavior in real exchange rates which the unit roots tests generally fail to detect. The size bias is large even when the unit root component accounts for a significant share of the movement in the real exchange rate. This size bias is shown to occur not only in unit roots tests, but also in tests for cointegration of nominal exchange rates and prices.

Whether or not PPP holds in the long run is a prominent question in international macroeconomics. A wide variety of hypotheses about exchange rate movements hold that in the short run the exchange rate deviates from its long-run equilibrium value, but converges eventually to that equilibrium value. But, there is disagreement over what constitutes the equilibrium value. Some argue it is the PPP level, while others argue that there is a complex set of factors determining the long-run value, including such things as the relative labor productivities at home and abroad. In the latter set of models, the equilibrium value for the real exchange rate need not be a stationary variable; so, the real exchange rate itself may have a unit root. Policy decisions and

1

forecasts of the path of exchange rates may depend on getting the long-run value right.

The problem that is addressed here is closely related to earlier work in the time-series literature on testing for unit roots. Suppose the (log of the) real exchange rate, q_t, is composed of two components, x_t and y_t:

(1) $\quad q_t = x_t + y_t.$

Assume y_t is a non-stationary random variable -- for example, a simple random walk. Let x_t be a stationary but persistent random variable -- for example, an AR(1) with large first-order serial correlation. Then, q(t) is non-stationary. Its first difference is stationary with a moving average component. Schwert (1989) uses Monte Carlo results to show that standard unit roots tests can be grossly incorrectly sized when the moving average component is important. The problems that arise here are also examined in the literature on processes that are nearly stationary (for example Cochrane (1991) and Blough (1992)).

This paper makes three contributions which may be of interest not only to international economists, but to users of unit roots tests in general:

(1) We argue that the problem of "near observational equivalence" is not merely a theoretical curiosity, but in fact arises in the case of the real exchange rate. We are able to obtain data for disaggregated components of prices over a relatively short horizon (25 years). From that data, we can indeed identify one component (x_t), the relative price of traded goods across countries, that should be stationary on theoretical grounds. Theory does not preclude that the other component (y_t), which involves the relative price of traded to non-traded goods, is non-stationary. In our 25-year time series, y_t appears more persistent than x_t, and we cannot rule out that it has a unit root. This component has a much smaller innovation variance than x_t. We use parameters estimated from the 25-year time series of the disaggregated prices to simulate the behavior of the real exchange rate in 100-year samples.

2

(2) We find that the size bias in unit roots (and cointegration) tests is large, even when the unit root component, y_t, accounts for a large proportion of the variance in real exchange rate movements within the 100-year sample period. For some issues, the failure to detect a small unit root component does not matter. But, the results here are striking -- large unit roots components will go undetected with a very high probability. We argue that this size bias even has implications even for how we consider the short-run behavior of the real exchange rate.

(3) Recently, tests which have a null hypothesis of stationarity have been developed (for example, Kwiatkowski, Phillips, Schmidt and Shin (1992).) Researchers frequently have taken the position that if one simultaneously rejects a unit root with a Dickey-Fuller type test, and fails to reject stationarity with a KPSS test, that there is strong and mutually reinforcing evidence that the series being tested is stationary.[1] Here we show that under the same circumstances in which the unit roots tests have large size biases, the stationarity tests have very low power. One is quite likely to reject a unit root and fail to reject stationarity, even though there is a large unit root component embodied in the series.

Section 1 decomposes the real exchange rate into its x_t and y_t components using disaggregated data. We cannot reject the null that the real exchange rate or its two components have unit roots with our twenty-five year time series. Unfortunately, long time series on the disaggregated series do not exist, so we are unable to say whether this failure to reject a unit root arises because of the low power of the tests for non-stationarity. However, we discuss why it is plausible that the x_t component is stationary and the y_t component is non-stationary. We take disaggregated U.S./U.K. data for the 1970-1995 time period and estimate a simple time-series model.

---

[1] See, for example, Chen and Tran (1994) and Cheung and Chinn (1996); or Fischer and Park (1991) for the converse.

In section 2, we construct artificial 100-year time series using the model estimated in section 1. Using Monte Carlo simulations, we find the true size of several tests for long-run PPP that have nominal sizes of 5%. We examine the simple Dickey-Fuller test and the Augmented Dickey-Fuller test for unit roots in real exchange rates. We also simulate the behavior of the Error Correction Model test, and the Horvath-Watson test for cointegration between the nominal exchange rate and relative price levels.

We acknowledge that the behavior of exchange rates and prices in our 25-year sample may be different than the behavior over the past 100 years as a whole. The variance of nominal exchange rates may be different, and price-setting behavior may have changed, hence altering the persistence of the stationary component of the real exchange rate. There may also be measurement error in x_t and y_t which causes their relative variability to be mismeasured. Hence, we use the time series constructed from the model estimated in section 1 only as a benchmark. We perform Monte Carlo exercises for a wide variety of parameters. Indeed, one exercise calibrates parameters to actual 100-year data on nominal exchange rates and (non-disaggregated) price indexes.

Section 3 performs parallel Monte Carlo exercises on the KPSS test of stationarity. Here, since our artificial data contain unit roots, we are interested in assessing the power of the KPSS test.

Section 4 concludes with a discussion of the implications of the size bias in tests for long-run PPP.

1. A Decomposition of the Real Exchange Rate

The real exchange rate is the relative price of foreign goods to domestic goods corrected for the nominal exchange rate. In log terms,

4

(2) $\quad q_t = s_t + p_t^* - p_t,$

where p_t^* is the log of the foreign price level and p_t is the log of the domestic price

level.

If price indexes are geometric means of traded goods prices and non-traded goods

prices, we can write

$$p_t = (1-\alpha)p_t^T + \alpha p_t^N,$$

and,

$$p_t^* = (1-\beta)p_t^{T*} + \beta p_t^{N*}.$$

Then, the real exchange rate can be written as

(3) $\quad q_t = (s_t + p_t^{T*} - p_t^T) + \beta(p_t^{N*} - p_t^{T*}) - \alpha(p_t^N - p_t^T).$

In reference to the decomposition in equation (1), we have

$$x_t = s_t + p_t^{T*} - p_t^T,$$

and

$$y_t = \beta(p_t^{N*} - p_t^{T*}) - \alpha(p_t^N - p_t^T).$$

The relative price of traded goods, x_t, is likely to be a stationary random

variable. If all goods in the traded goods price indexes have the same weights at home

and abroad, then changes in x_t occur only because of deviations from the law of one

price. Although there is considerable evidence that deviations from the law of one price

can be large and persistent (see, for example, Engel (1993), Rogers and Jenkins (1995)

and Wei and Parsley (1995)), they are almost certainly stationary. Some type of goods

arbitrage would rule out the possibility that these deviations could become unbounded,

and would thus preclude a unit root in x_t.

Studies which have argued that there may be unit roots in the real exchange rate

have concentrated on the y_t term. Permanent shocks to productivity could impart a non-

stationary component to the relative price of non-traded to traded goods. This approach

builds on the early influential work of Balassa (1964) and Samuelson (1964). Cross-

sectional studies of prices show that there can be very large differences in non-traded goods prices across countries. Rogoff (1996) illustrates the point using data constructed from the Penn World Tables. Prices tend to me much lower in very low income countries as compared to industrialized countries.

Engel (1995) uses the decomposition in equation (1) to separate real exchange rate changes for the U.S. into their x_t and y_t components. That study uses a variety of price indexes for which data is available on sub-components that can be identified as traded and non-traded goods. Here we will pay special attention to one measure -- the GDP deflator for personal consumption expenditures for the U.S. and the U.K. We choose to examine these series, because there are 100-year-long annual series for both countries (used by Rogers (1995)) for the personal consumption deflators and the nominal exchange rate and shorter time series on the disaggregated data.

In this section, we make use of quarterly data on sub-categories of the personal consumption deflator for the years 1970 to 1995.[2] The sub-index for the deflator for personal consumption of commodities in each country is used as the price index for traded goods, and the deflator for personal consumption of services is used as the price index for non-traded goods. This is the assignment used by Engel (1995) and Stockman and Tesar (1995), although clearly these categories are not precise classifications of traded and non-traded goods.

We shall attempt to deal with some of the measurement error problems later. One problem that arises is that many services are traded, so that it is probably inaccurate to classify all spending on services as non-traded spending. Conversely, when consumers purchase commodities, at least part of what they are paying for is a non-traded marketing service. That is, the consumer price of a product includes payment for the advertising, retailing and distribution services that bring the good to market. Some of those

---

[2] All of the 25-year data was obtained from Datastream.

6

services are better classified as non-traded. We will only use our parameter estimates from this data as a benchmark, and will consider how measurement error is likely to affect the parameters.

We begin by performing the usual battery of tests for unit roots and cointegration on these variables. All of the tests we perform in this section on actual data are also performed on the artificial data we construct in section 2. The tests are described in detail in the Appendix.

The Table 1 summarizes the results. We cannot reject (at the 5% level) the null of a unit root in q_t using either a simple Dickey-Fuller (DF) or the Augmented Dickey-Fuller (ADF) test.

Next, we proceed to test for cointegration of s_t and $p_t-p_t^*$. A large number of authors have tested for long-run PPP using cointegration techniques, including Baillie and Selover (1987), Edison (1987), Taylor (1988), Mark (1990), Patel (1990), Kim (1990), Cheung and Lai (1993) and Edison, Gagnon and Melick (1994).

We consider two models of cointegration. The first is the Error-Correction Model (ECM). This is a single-equation test for no cointegration. The single-equation methodology incorporates in the null hypothesis the assumption that $p_t-p_t^*$ is weakly exogenous. We shall shortly report some evidence that sustains this assumption. The second test is the two-equation Horvath-Watson (HW) test for no cointegration. Note that in both instances, we impose the null that the cointegrating vector for s_t and $p_t-p_t^*$ is (1,-1).

Table 1 shows that we fail to reject the null hypothesis of no cointegration at the 5 per cent level using the test from the single-equation Error Correction Model (ECM). We reach the same conclusion using the Horvath-Watson two-equation model of cointegration.

All of our tests using the twenty-five year data sample fail to reject the null

7

that long run PPP does not hold. This is not too surprising. The motivation for economists such as Frankel (1986), Edison (1987) and Glen (1992) to use very long series on prices and exchange rates was that the tests for unit roots (and cointegration) have little power in time series as short as 25 years. Indeed, we shall see later when we look at a 100-year long series that is comparable to this data that we can reject a unit root in the real exchange rate using the ADF test.

We also perform unit roots tests on x_t. Table 1 indicates that we fail to reject a unit root at the 5 per cent level for x_t using either the DF or the ADF test. We also test for cointegration of s_t with $p_t^T - p_t^{T*}$. Both the ECM and HW test fail to reject the null of no cointegration at the 5 per cent level.

Finally we perform a DF and an ADF test for unit roots in y_t. Both fail to reject a unit root at the 5 per cent level..

We have argued that it is likely that x_t is stationary while y_t might have a unit root. Some support for this view comes from the estimated degree of persistence of the two variables. For example, the estimates of β from the DF test (equation (A1)) and the ADF test (equation (A2)) for x_t are, respectively, .9184 and .8721. For y_t, the corresponding estimates are .9677 and .9668. While the estimates are not statistically significantly different (for example, we could not reject a unit root in any case), they are consistent with the view that y_t is more persistent and is more likely to have a unit root.

Further support for that view comes from Figure 1. This Figure plots the variance ratio statistics, as in Cochrane (1988), for q_t, x_t and y_t.[3] That is, for x_t (and likewise for q_t and y_t), the Figure plots $Var(x_{t+k}-x_t)/Var(x_{t+1}-x_t)$ for horizons of $k = 1$ to $k = 75$. If a series follows a random walk, the (population) variance of the k-

---

[3] These variance ratios were calculated using Cochrane's formula, which corrects for small-sample bias.

8

difference in that series will be k times the variance of the first difference. For a stationary series, the variance ratio approaches a limit. From Figure 1, we can see that the variance ratio statistics for y_t rise much more steeply than for x_t or q_t, indicating more persistence.[4] It is also noteworthy that the variance ratios for x_t and q_t are nearly identical, particularly at the shorter horizons, which is a reflection of the small contribution that y_t makes to movements in q_t over the short and medium run.

Our model for the components of the real exchange rate is straightforward. We assume that the y_t component follows a simple random walk:[5]

(4) $y_{t+1} - y_t = au_{t+1}$,

where u_t is an i.i.d., $N(0,1)$ random variable. This equation determines the movements of the relative prices of non-traded to traded goods. We can think of the shock u_{t+1} as incorporating shocks to tastes and technology which cause this relative price to change permanently.

Define z_t to be the relative price levels, unadjusted for the exchange rate:

(5) $z_t \equiv p_t^T - p_t^{T*}$.

So, $x_t = s_t - z_t$.

We posit a simple error-correction representation for z_t and s_t:

(6) $s_{t+1} - s_t = - \delta(s_t - z_t) + bu_{t+1} + cv_{t+1}$,

(7) $z_{t+1} - z_t = \gamma(s_t - z_t) + d\varepsilon_{t+1} + fv_{t+1} + gu_{t+1}$.

The nominal exchange rate in our model can be affected by the u_t shock from above, as well as a monetary shock, v_t. The monetary shock does not affect the relative price, y_t, but is incorporated in the nominal exchange rate.

---

[4] At the very long horizons ($k > 75$) all of the statistics drop off significantly, but these statistics should probably be ignored since they were calculated using very few data points. So, they are not included in Figure 1.

[5] We suppress the intercept terms in the presentation of the model for expositional clarity. Intercept terms were included when the model was estimated, although all of them were insignificantly different from zero. The artificial data created for the Monte Carlo exercises does not include intercept terms.

z_t is a nominal variable whose dimensions are the same as the nominal exchange rate. We allow the u_t and v_t shocks to affect z_t. In addition, there is a shock, ε_t, which is a source of shocks to the PPP relationship in traded goods prices. It might represent shocks to the degree of market segmentation.

v_t and ε_t are also i.i.d., N(0,1) random variables.

Equations (6) and (7) imply that z_t and s_t are cointegrated, with cointegrating vector (-1,1). Together they imply that the relative price of traded goods, x_t, is stationary and follows a simple AR(1) process:

(8) $\quad x_{t+1} = (1-\rho)x_t - d\varepsilon_{t+1} + (c-f)v_{t+1} + (b-g)u_{t+1},$

where $\rho \equiv \delta + \gamma$.

The system (4), (6) and (7) was estimated by an iterative GLS procedure. The coefficients estimates are reported in Table 2. Standard errors are constructed from the inverse of the estimated information matrix.

There are several things of note about these estimates. First, among the coefficients on the random errors, $\hat{c}$ is nearly five times larger than the next largest coefficient. Nominal exchange rates are much more variable than nominal prices.

Then, $\hat{d}$, from the equation for z_t is next largest, while $\hat{a}$ and $\hat{g}$ are not much smaller. The other coefficients on the random errors -- $\hat{b}$ and $\hat{f}$ -- are nearly zero. The implication of these estimates is, first, that nominal prices are much less variable than the nominal exchange rate ($\hat{d}$ and $\hat{a}$ are much smaller than $\hat{c}$.) Second, there is some correlation between shocks to the two terms involving only nominal prices, y_t and z_t ($\hat{g}$ is non-zero.) Third, there is almost no correlation between innovations in nominal exchange rate and nominal prices ($\hat{b}$ and $\hat{f}$ are near zero.)

Also note that the estimate of γ is actually negative, but not significantly different than zero. The fact that it is near zero helps justify our assumption above when performing the ECM test that prices are weakly exogenous. The persistence of x_t can

10

be ascertained by the sum of $\hat{\delta}$ and $\hat{\gamma}$, with $\hat{\rho}$ = .07698. This implies a high degree of persistence, but lower than the usual measure for q_t. The estimated half-life of x_t is about nine quarters.

None of these results are surprising in the light of Engel (1995). That study does not estimate a formal model of exchange rates and prices, but does decompose the mean-squared error of q_{t+j}-q_t for j = 1, 2, ... , 100 (using quarterly data), into x_t's share and y_t's share. First, we note that such a decomposition is nearly unambiguous because x_t and y_t are nearly uncorrelated, which is consistent with the finding that $\hat{b}$ and $\hat{f}$ are close to zero.

The major finding of Engel (1995) is that the mean-squared error of q_{t+j}-q_t is almost completely attributable to movements in x_{t+j}-x_t, even when j is very large. This is compatible with the model presented in equations (4), (6) and (7), given the coefficient estimates. The model does imply that eventually, for j large enough, the variance of q_{t+j}-q_t must be dominated by the variance of y_{t+j}-y_t, since y_t is the unit root component of q_t. But, 25 years is too short a time span for the y_t component to dominate. That is, first, because the innovation variance of x_t is much larger than that of y_t. The innovation variance of x_t is $(c-f)^2$ + $(b-g)^2$ + d^2, which is estimated to be .002667, as compared to .0000328, which is the estimate of a^2, the innovation variance of y_t. Second, as we have noted, x_t is fairly persistent (though it is modeled to be stationary.)

The next section uses the model estimated above to simulate 100 years of data for s_t and p_t-p_t^*, and then conducts Monte Carlo tests to assess the true size of the DF, ADF, ECM and HW tests of long-run PPP.

11

2. Monte-Carlo Measurements of Size of Tests for Long-Run PPP

In this section, we report the results of Monte Carlo exercises to measure the true

size of the DF, ADF, ECM and HW tests for long-run PPP. The details of the Monte Carlo

exercises are in the Appendix. The baseline case we consider is based on the parameter

estimates from the 25 year data sample that we used to estimate the model of section 1.

Each artificial series we create has 400 data points. We choose 400 observations

because that corresponds to a 100-year data sample (given that our parameters are

estimated on quarterly data.) Several recent studies of real exchange rate behavior have

used comparable samples to conclude that long-run PPP holds. See, for example, Frankel

(1986), Edison (1987), Edison and Klovland (1987), Kim (1990), Abuaf and Jorion (1990),

Ardeni and Lubian (1991), Glen (1992) and Cheung and Lai (1994).[6]

For each series, we perform all four tests. We record whether we would reject the

unit root with a five per cent test.

For each set of parameters, we also calculate the fraction of the variance of

$q_{t+400} - q_t$ that should theoretically be attributed to the unit root component. While

there is some ambiguity about this decomposition in general, because y_t and x_t are

correlated, in practice the correlation is small enough that the decomposition is not

very dependent on how the correlation is treated. We will report the calculation for

---

[6] Two recent studies use even longer time series. Lothian and Taylor (1996) use 200 years of wholesale price index data, while Froot, Kim and Rogoff (1995) use 700 years of commodity price data. However, the data used in both of these studies are constructed from a limited number of goods. The price data do not incorporate the non-traded goods prices that, according to our model, account for the non-stationary component of the real exchange rate. So, while there rejection of a unit root is more convincing than with the 100-year data series, there tests should more properly be considered tests for unit roots in the x_t component.

$$\frac{\mathrm{Var}(y_{t+400}\text{-}y_t)}{\mathrm{Var}(x_{t+400}\text{-}x_t) + \mathrm{Var}(y_{t+400}\text{-}y_t)} = \frac{400a^2}{\dfrac{1\text{-}\rho^{800}}{1\text{-}\rho^2}\left[(b\text{-}g)^2+(c\text{-}f)^2+d^2\right]+400a^2} \ .$$

We will refer to the variance of the 400-quarter change as the "long-run variance". Of course, this phrase does not have its usual meaning in this context -- the limit of the variance in the k-period change as k goes to infinity. Since q_t has a unit root, that limit does not exist.

Our findings are summarized in Figures 2A-D, 3A-D and 4A-D. Before turning to those, it is useful to discuss the baseline case. For the parameters estimated in section 1, the unit root component, y_t, accounts for 42.12 per cent of the long-run variance. So, it is not insignificant in terms of the long-run movements of the real exchange rate. Yet, using a nominal size of five per cent, all four tests almost always reject a unit root (or, equivalently, no cointegration of s_t and $p_t\text{-}p_t^*$). The true size for the DF test is .9808; for the ADF test, .8978; for the ECM test, .9434; and for the HW test, .9066.

Thus, if the data generating process estimated in section 1 produced 100 years of quarterly data, we would almost always conclude that long-run PPP holds, even though it would not. What is happening here that makes the tests so wrong? Informally, of course, the issue is that there is a lot of persistence to the x_t component, and it has a large innovation variance. Movements in x_t are so large that they mask movements in y_t. At longer horizons, y_t contributes more to the movement in q_t, but the tests allow for only finite dynamics. If enough lags were allowed to capture the importance of y_t, the size of the tests would improve but the power would be significantly diminished. So, we can think of the problem as one of near observational equivalence (as in Cochrane (1991) and Blough (1992).) We are taking a stationary series, x_t, and adding to it a non-stationary series, y_t, that is relatively unimportant in terms of its contribution to changes in q_t.

13

Somewhat more formally, rewrite equation (4) as

$$y_{t+1} - y_t = w_{t+1},$$

where $w_{t+1} = au_{t+1}$, and write equation (8) as

$$x_{t+1} = \phi x_t + m_{t+1},$$

where $\phi = 1-\rho$ and $m_{t+1} = -d\varepsilon_{t+1} + (c-f)v_{t+1} + (b-g)u_{t+1}$. Then, the univariate ARMA representation for the real exchange rate is

$$\Delta q_t = \phi \Delta q_{t-1} + \zeta_t + \mu \cdot \zeta_{t-1},$$

where

$$\mu = -\frac{2\sigma_m^2 + (1+\phi^2)\sigma_w^2 + 2(1+\phi)\sigma_{mw} - \sqrt{(1-\phi^2)^2\sigma_w^4 + 4(1-\phi)^2\sigma_m^2\sigma_w^2 + 4(1-\phi^2)(1-\phi)\sigma_w^2\sigma_{mw}}}{2\sigma_m^2 + 2\phi\sigma_w^2 + 2(1+\phi)\sigma_{mw}}.$$

Here, σ_m^2 is the variance of m_t, σ_w^2 is the variance of w_t, and σ_{mw} is their covariance.

Inspection of this expression for μ shows that when the unit root component is small, μ approaches -1. Specifically, as σ_w^2 and σ_{mw} go to zero, μ goes to -1. μ also goes to -1 when the stationary component is very persistent -- that is, as ϕ goes to unity, irrespective of the values of σ_w^2 and σ_{mw}.

In our case, both of these circumstances are nearly true, so that Δq_t follows an ARMA(1,1), and the coefficient on the moving average component is nearly -1. That is exactly the situation investigated in detail by Schwert (1989). He found that there was a large size distortion in the Phillips-Perron test for a unit root. His findings about the size are very similar to ours -- even in very long samples, the tests reject a unit root over 90 per cent of the time when there is one present.

The baseline stochastic process that we simulate may not produce a representative 100-year series for a number of reasons. In the remainder of this section, we consider alternate parameterizations to get an idea of the scope of the size problems with tests for long-run PPP.

First, we consider the issue of whether the degree of nominal exchange rate variability that we estimate from our 1970-1995 sample is representative of the exchange rate variance over the 100-year sample. The dollar/pound rate over the latter sample has been quite volatile, but over the past 100 years there have been periods in which it was very quiet, and in which the real exchange rate also was quite stable. One approach to dealing with this issue is to model switches of regime from low volatility to high volatility states, and examine the consequences of this heteroskedasticity for the size of real exchange rate tests.[7]

Here, we undertake the simpler exercise of investigating the consequences of different values of the parameter c from equation (6). So, we fix all of the other parameters at their values reported in Table 2, but then conduct Monte Carlo exercises for various values of c. The results of those exercises are summarized in Figures 2A-2D.

For example, Figure 2A reports the results on the simple Dickey-Fuller test. We graph the true size of the test against the fraction of the long-run variance accounted for by the unit root component. The baseline case is $c = 0.051$. Most of the values of c that we investigate are smaller, allowing for the effect of more quiescent nominal exchange rates. Figure 2A is quite striking -- the probability of rejecting a unit root remains very high even when the unit root component accounts for a very large fraction of long-run real exchange rate movements. For example, when y_t accounts for 84.1% of the long-run variance of q_t, the probability of rejecting a unit root using a five per cent test is still 67.9 per cent.

Figures 2B, 2C and 2D show the comparable results for the ADF, ECM and HW tests, respectively. The ECM test has the worst size bias -- here, in the case where y_t accounts for 84.1% of long-run movements in q_t, the probability of rejection is 72.9%.

---

[7] Lothian and Taylor (1996) address this issue with long time series on wholesale prices.

15

The ADF test is the least size-biased, but the bias is still considerable. Again, when 84.1% of the movements of q_t in the long-run are attributable to y_t, the probability of rejection is 55.1%.

Next, we consider whether we have understated the variance of the y_t process. Engel (1995) discusses why this measure of the relative price of non-traded goods may understate the importance of y_t. In particular, if there is a large non-traded component in x_t (due perhaps to marketing and distribution costs, as discussed above), then the true variance of changes in y_t may be understated by the measure used here. So, we perform Monte Carlo exercises for various values of the parameter a from equation (8). These results are reported in figures 3A-3D.

As in Figures 2, there is a trade-off between the fraction of q_t's long-run variance attributed to y_t and the true size of the test. In all cases, the tests appear to be very badly sized. The worst size bias is in the simple Dickey-Fuller test in this case. The other three tests are fairly comparably sized. For all of the tests, the true size is above 50 per cent even when the unit root component accounts for over 85% of the long-run movements in the real exchange rate.

When a is set to zero, there is no unit root component in the real exchange rate. Then, the probability of rejection measures the power of the test -- the probability of rejecting a unit root when there is none. The simple Dickey-Fuller test has the greatest power for this data generating process, but all four of the tests have impressively high power. The worst of them, the ADF, still has a 95% chance of rejecting the null when the null is false.

Next, we allow for different values of δ from equation (6). In our baseline simulations, we set δ equal to .077. Figures 4A-4D trace the outcome from Monte Carlo simulations for values of δ ranging from .01 to .09. For values of δ that are quite small, the tests appear to have less size bias. So, when δ is equal to .01, the size of

the tests is around .10 for all of the tests. However, the reason the tests have smaller size bias in this instance undoubtedly is not because they detect the unit root component y_t. Certainly what is occurring is that x_t is very persistent in the case where δ is small. Even if the y_t component were not present, the tests would fail to reject the null of a unit root in x_t (or the null of no cointegration between s_t and z_t).

Given that we have set γ from equation (7) equal to zero (see the Appendix), when δ is also set to zero, both x_t and y_t are unit root processes. The real exchange rate follows a simple random walk. In this case, we find (not surprisingly) that the size of the tests is correct: .047 for the DF; .050 for the ADF; .049 for the ECM; and, .048 for the HW.

We also consider various values of the parameter d in equation (7). However, varying this parameter had little effect on our conclusions about the size of the tests for long-run PPP. Monte-Carlo experiments were performed for values of d ranging from zero to 15 times the baseline value, and the true size of all tests for long-run PPP were at least .89 when the nominal size is .05.

Finally, we consider some simulations in which the parameters of the stochastic process in (4), (6) and (7) are calibrated to a 100-year data sample. This data, from Rogers (1995), consists of the nominal dollar/pound exchange rate, and the personal consumption deflator in the U.S. and U.K. from 1892-1992. The description of the data in Rogers, and in the original source, Mitchell (1988), is sparse. The exchange rate is described as a period average, and the prices appear to be end-of-period data. All of the data is annual. In this one-hundred year data, we reject the null of a unit root in real exchange rates at the 5 per cent level using the ADF test.

We choose parameter values to match three moments in the data -- the first-order autocorrelation coefficient of the real exchange rate, and the variances of the annual difference in the relative price levels and the three-year difference in the nominal

17

exchange rate.[8] The Appendix describes how we create artificial series whose moments match those from the true data.

Table 3 shows the sets of parameter values we use to perform our analysis of the size of the long-run PPP tests. As it turns out, the value of the parameter c is about the same across all of the sets, but there is quite a bit of variation in the other parameter values. As the table shows, the unit root component, y_t, accounts for as little as 0.9% of the long-run variance of q_t to as much as 89.4%, depending on the particular parameter values.

The true size of the tests is, nonetheless, fairly consistent across all the sets of parameter values and all of the tests. As shown in Table 3, the size ranges only from 0.290 to 0.423. So, for all parameter values, there is considerable size bias in all four tests.

3. Tests of Stationarity

We have argued that standard unit roots tests for real exchange rates assume under the null hypothesis that the first difference of the real exchange rate has a stationary autoregressive representation. They are not designed to handle the case we are interested in, where the real exchange rate is composed of a sum of a unit root process and a persistent stationary process. In this case (repeating equation (1) for convenience:)

$$q_t = x_t + y_t,$$

the change in the real exchange rate has a moving average component.

---

[8] During the periods of "fixed" nominal exchange rates, the occasional devaluations yield distant outliers in the 1-year changes in exchange rates. The variance of the 1-year changes in these periods is quite large -- generally larger than for floating rate periods. Taking the variance of 3-year changes gives a more reasonable picture of the amount of exchange rate volatility. Of course, as we mention above, it would be best to model explicitly these jumps in the nominal exchange rate.

Intuitively, it is easy to understand why we cannot design a test whose null is that q_t has a unit root in this case. To reject the null, one would need to reject the hypothesis that y_t (the unit-root component) has a non-zero variance. But, as Blough (1992) and Cochrane (1991) have pointed out, it would be essentially impossible to rule out a y_t process with an arbitrarily small variance. In Blough's formulation, the maximum power of any such test against the alternative that y_t has zero variance is just the size of the test. Since a process for q_t in which y_t has zero variance looks just like one in which y_t has an arbitrarily small variance; if we are willing to tolerate, say, a 5 per cent chance of rejecting the null of a unit root when the null is true (the size is 5 per cent), then we can only have a 5 per cent chance of rejecting the null when the null is not true but is arbitrarily close to being true (the power is 5 per cent).

On the other hand, it seems quite reasonable to formulate a test whose null is that the variance of the y_t component is zero. While it seems like we could really never prove the y_t component does not exist, it does seem possible under some circumstances to prove the claim that it does exist. That is the motivation of the test for stationarity of Kwiatkowski, Phillips, Schmidt and Shin (1992). That test, in essence, is a test of the null hypothesis that y_t has a zero variance.

We consider, however, whether the KPSS test has much power to reject the null of stationarity in our case. As in section 2, we perform Monte Carlo simulations of the real exchange rate based on the model of section 1. However, since the null is stationarity and the artificial series really have a unit root, when we tabulate the fraction of times the null is rejected we are calculating the power of the KPSS test. The details of the Monte Carlo simulation are in the Appendix.

Unfortunately, the KPSS test has very low power to detect the unit root component. For the baseline model, we reject the null only 8.6 per cent of the time. Figures 5A-C show the power of the test for various values of the parameters c, a and δ (from the

model of equations (4), (6) and (7)). We also tried varying the parameter d, although these results are not reported graphically. In addition, we have constructed artificial series whose moments match those of our 100-year data (as we did at the end of Section 2 above). These results are reported in the last line of Table 3.

In the best case among all of the different parameterizations, the KPSS test rejected the null a mere 13.3 per cent of the time. In most cases the power was much lower, even though in some of the parameterizations the unit root component accounts for a very large share of the "long-run" variance.

It is interesting that the KPSS test has low power here. One might be tempted to conclude that since the DF, ADF, ECM and HW tests reject a unit root, and the KPSS test fails to reject stationarity, that there is mutually reinforcing evidence that there is no unit root. But, in fact, all of the tests lead to the wrong inference -- the size of the DF, ADF, ECM and HW tests are bad, and the KPSS test has low power.

4. Conclusions

We have found that there can be large size biases in tests for long-run PPP. There may be a significant unit root component that is not detected by these tests. We associate that component with the y_t term above, which represents the element corresponding to the relative price of non-traded goods in the real exchange rate.

There is additional evidence that the relative relative prices of non-traded goods are not stationary. If these prices were non-stationary, we would expect to see some large variation in a cross section of aggregate price levels (corrected for the nominal exchange rate), particularly because the non-stationary component accounts for such a large fraction of the real exchange rate movement in the long run in our model. That is exactly the finding of Rogoff (1996), who discovers enormous variation in price levels across countries using data from Summers and Heston (1991). In some instances, richer

countries have price levels that are an order of magnitude higher than small countries.

Indeed, consider the following experiment. The cross-sectional variance in the log of the relative prices of consumption deflators in the Summers-Heston data is 0.2686. If we fit an AR(1) model to our 25 years of U.S./U.K. real exchange rates, the variance of the quarterly innovations is 0.00263. Suppose each real exchange rate relative to the U.S. from the Summers-Heston panel follows an identical AR(1) with innovation variance of 0.00263. What value of the quarterly autocorrelation coefficient would produce a cross-sectional variance of 0.2686? The answer is 0.995 (that is, 0.00263/(1-0.995) = 0.2686.) So, while it is, of course, possible that the Summers-Heston data was generated by stationary real exchange rate series, it seems very possible that there is a unit root component in some of the real exchange rates.

But, if this unit root component is present, is it important? In our baseline case, we find that the unit root component accounts for 42.1 per cent of the variance of the change in the real exchange rate over a 100-year period. But over a 20-year period, this component would only account for 12.7 per cent of the real exchange rate variance. This is consistent with the Engel (1995), who finds that the y_t term does not account for very much of the mean-squared error of q_t, even for time intervals as long as 20 years.

However, the question of whether or not there is a unit root component in q_t is important, even for short-run forecasting. Consider rewriting equations (4)-(7) as:

(9) $\quad q_{t+1} - y_{t+1} = \phi(q_t - y_t) + \chi_{t+1},$

where $\chi_t = -d\varepsilon_{t+1} + (c-f)v_{t+1} + (b-g)u_{t+1}$, and $\phi = 1-\rho$. From equation (9) we can interpret the real exchange rate as following a stochastic process where q_t temporarily deviates from an equilibrium value, y_t, but is "mean-reverting". Under the theory of long-run PPP, the equilibrium value is a constant, but here we have argued that we cannot reject the hypothesis that y_t is a slow-moving process with a unit root.

Why might it make a difference whether y_t is a constant or a very slow-moving unit

root? For the moment, consider the case where we know y is constant. When we estimate equation (9), our estimate of y is the mean value of q_t in our data. If ϕ is small, then the mean value of q_t is a precise estimate of y. But, when ϕ is close to one, the sample mean of q_t is an imprecise measure of y. This is the case we are concerned with, since we believe ϕ is nearly unity.

We could improve our forecasts in this stationary case, however, if we brought in some outside information. Here, we have a good idea what the mean of q_t would be if it were stationary. Theory suggests that if the real exchange rate settles down to any long-run value, the most plausible one is the purchasing power parity value. So, for short-run forecasting purposes, we could take the PPP value of q_t as our measure of y.

However, if we do not know whether y_t is a constant or a slow-moving random walk, we cannot bring in this outside information. For if y_t is a random walk -- even one whose innovation variance is very small -- it could be very far from the PPP value of the real exchange rate. In concrete terms, suppose we try to forecast the yen/dollar exchange rate using the fact that the real exchange rate has a stationary component. Should we forecast that the exchange rate will regress to its PPP value, or to some other value?

Consider these additional bits of evidence. Both Mark (1995) and Chinn and Meese (1995) find that they can forecast the nominal exchange rate at long horizons by predicting it returns to a target level -- but the target level is not the PPP value.

Mark and Choi (1996) explicitly allow for the target component of the real exchange rate to move over time according to various models of long-run real exchange rate determination. They find that models in which long-run PPP holds are significantly outperformed, in terms of out-of-sample forecasting power, by models that allow the long-run real exchange rate to vary over time. In fact, their preferred models take exactly the form of equation (9), with the target real exchange rate having a unit root.

In this study, we do not at all address the recent tests for PPP, such as Frankel and Rose (1996) and Wei and Parsley (1995), that look at a cross-section of countries. It seems plausible that these tests would suffer from similar size problems as the ones addressed here. On the one hand, the size problems might be worse, since these studies use shorter sample periods which are less likely to pick up the movements of the y_t component. On the other hand, if the y_t component were correlated across real exchange rates, it might be more detectable in the cross-sectional studies.

So, what the tests for long-run PPP probably can tell us is that there is some sort of "mean-reversion" in real exchange rates. What is not yet clear is whether there is convergence to PPP in the long run.

Appendix

The Monte-Carlo tests of section 2 (and the tests on the actual data in section 1) examine long-run PPP using four different tests.

First is the simple Dickey-Fuller (DF) test for unit roots in the real exchange rate, q_t. We estimate the equation:

$$(A1) \quad q_t = \alpha + \beta q_{t-1} + \nu_t$$

The null hypothesis is $\beta = 1$.

Next is the Augmented Dickey-Fuller test. Here, we estimate the equation

$$(A2) \quad q_t = \alpha + \beta q_{t-1} + \gamma_1 \Delta q_{t-1} + \gamma_2 \Delta q_{t-2} + \dots + \gamma_j \Delta q_{t-j} + \nu_t.$$

The lag length j, was chosen by an iterative data-based procedure, as recommended by Ng and Perron (1995). We start with a maximum number of lags (12) and test for the significance of γ_{12}. If it is significantly different from zero, then $j = 12$. Otherwise, we drop the 12th lag, reestimate the regression, and proceed until γ_j is significantly different from zero. The null hypothesis is again $\beta = 1$.

In the test performed in section 1, we end up choosing a lag length of six. For each iteration of the Monte Carlo procedure of section 2, we do the iterative procedure to choose the lag length. The number of lags actually chosen varied from zero to twelve. Approximately 35-40 per cent of the time (depending on the parameters used to generate the artificial data), a lag length of zero was chosen.

Our first cointegration test derives from the Error Correction Model (ECM) test proposed by Kremers, Ericsson and Dolado (1992). We estimate the equation

$$(A3) \quad \Delta s_t = \alpha - \beta q_{t-1} + \delta_0 \Delta(p_t\text{-}p_t^*) + \delta_1 \Delta(p_{t-1}\text{-}p_{t-1}^*) + \delta_2 \Delta(p_{t-2}\text{-}p_{t-2}^*) +$$
$$\dots + \delta_j \Delta(p_{t-j}\text{-}p_{t-j}^*) + \gamma_1 \Delta s_{t-1} + \gamma_2 \Delta s_{t-2} + \dots + \gamma_j \Delta s_{t-j} + \nu_t.$$

Note that we constrain the number of lags of $\Delta(p_t\text{-}p_t^*)$ and Δs_t to be equal. The lag length is chosen by the same type of data-based iterative procedure as in the ADF test.

24

Here, in each iteration we test the joint null $\delta_j=0$, $\gamma_j=0$. Again, when we do the Monte Carlo procedure, we use this iterative procedure to choose the lag length for each set of artificial data.

The ECM test is valid in this case if $p_t-p_t^*$ is weakly exogenous. In general, the ECM method allows for estimation of the cointegrating vector, although here we have imposed that it is (1,-1).

The null hypothesis in this case is $\beta = 0$. Following Zivot (1995) and Hansen (1995), the test statistic for the ECM test depends on the long-run covariance matrix of v_t and $v_t + \sum_{i=0}^{j} \delta_i \Delta(p_{t+i}-p_{t+i}^*)$. This matrix is calculated in the standard way, using a Bartlett kernel, with the selection rule for the order of the kernel weight function chosen as in Andrews (1991). The critical values are presented in Hansen (1995).

In each iteration of the Monte Carlo, then, we compute the long-run covariance matrix and use it to compare the test statistic to the critical value from the Hansen table.

The second cointegration test is based on the procedure suggested by Horvath and Watson (1994). Here we estimate the system of equations given by:

(A4) $\quad \Delta s_t = \alpha_1 - \beta_1 q_{t-1} + \delta_{11}\Delta(p_{t-1}-p_{t-1}^*) + \delta_{12}\Delta(p_{t-2}-p_{t-2}^*) + \ldots +$

$\quad\quad\quad\quad \delta_{1j}\Delta(p_{t-j}-p_{t-j}^*) + \gamma_{11}\Delta s_{t-1} + \gamma_{12}\Delta s_{t-2} + \ldots + \gamma_{1j}\Delta s_{t-j} + v_{1t},$

$\quad\quad \Delta(p_t-p_t^*) = \alpha_2 - \beta_2 q_{t-1} + \delta_{21}\Delta(p_{t-1}-p_{t-1}^*) + \delta_{22}\Delta(p_{t-2}-p_{t-2}^*) + \ldots +$

$\quad\quad\quad\quad \delta_{2j}\Delta(p_{t-j}-p_{t-j}^*) + \gamma_{21}\Delta s_{t-1} + \gamma_{22}\Delta s_{t-2} + \ldots + \gamma_{2j}\Delta s_{t-j} + v_{2t}.$

Note that we impose that the cointegrating vector is (1, -1). We also impose that the lag length on $\Delta(p_t-p_t^*)$ and Δs_t are the same in for both variables in both regressions. We choose the lag length again by an iterative data based criterion. Here, in each iteration we test the joint null that $\delta_{1j} = 0$, $\gamma_{1j} = 0$, $\delta_{2j} = 0$, and $\gamma_{2j} = 0$.

The null hypothesis here is that $\beta_1 = 0$ and $\beta_2 = 0$. We compare the test statistic

25

to the critical values reported in Horvath and Watson.

We construct 5000 replications of data series with 400 observations each. In each case, we actually constructed 450 data points, and took the last 400 points to avoid any bias from start-up values. The start-up values for all the variables is zero, which is equal to their unconditional mean given that there are no intercept terms included in the simulations. The error terms u_t, v_t and ε_t are assumed to be $N(0,1)$, and were created using the "rndn" command in Gauss version 3.01. For each of these 5000 artificial series, we perform the four tests.

One issue arises in our Monte Carlo simulations for cointegration using the Horvath-Watson test. The estimated value of γ from equation (7) is actually negative, but not significantly different from zero. If we construct our artificial data using this negative value, there would be some problems of interpretation with the HW test. Suppose for example that $-\gamma = \delta$. Then z_t, and therefore q_t, has a unit root even if y_t were zero. But the HW test would conclude that the variables were cointegrated, because it tests the joint hypothesis that γ and δ are non-zero. So, in our simulations, we set γ equal to zero, and set our measure of δ equal to the estimated value of $\delta + \gamma$. This leaves the persistence of the x_t component unchanged.

At the end of section 2, we perform simulations based on our 100-year data sample. We search for combinations of the four parameters -- a, c, d and δ -- that produce 100-year artificial series whose moments match those of the data. To construct the 100-year series for the nominal exchange rate, we produce 400 data points and averaged every four together. This is meant to replicate the construction of the actual data, which is an annual average. For the prices, we also produce 400 data points, and use every fourth number, since the actual price data is end-of-year data. The Monte Carlo exercises used to choose these values of a, c, d and δ employ 5000 replications of each series.

We use a hill-climbing technique to find these parameters. We begin with an

initial guess at a set of parameter values, construct 5000 artificial time series, and calculate the statistics of interest. We adjust the parameters, and construct new artificial time series, until we find a set of parameters that produce the desired statistics.

To construct the KPSS test, we calculate $u_t \equiv q_t - \bar{q}$, where $\bar{q}$ is the sample average of q_t. Then, define $S_t = \sum_{i=1}^{t} u_t$. The KPSS statistic is calculated as:

$$\sum_{t=1}^{T} S_t^2 \Big/ \left(T^2 \hat{\sigma}^2 \right),$$

where $\hat{\sigma}^2$ is an estimate of the long-run variance of q_t. We calculate this variance as we did in calculating the long-run variance of v_t for the HW statistic described above. Again, each Monte Carlo experiment involved construction of 5000 artificial data series of length 400.

References

Abuaf, Niso and Phillipe Jorion, 1990, Purchasing power parity in the long run, Journal of Finance 45, 157-174.

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

Ardeni, Pier Giorgio and Diego Lubian, 1991, Is there trend reversion in purchasing power parity?, European Economic Review 35, 1035-1055.

Baillie, Richard T. and David D. Selover, 1987, Cointegration and models of exchange rate determination, International Journal of Forecasting 3, 43-51.

Balassa, Bela, 1964, The purchasing power parity doctrine: A reappraisal, Journal of Political Economy 72, 584-596.

Blough, Steven R., 1992, The relationship between power and level for generic unit root tests in finite samples, Journal of Applied Econometrics 7, 295-308.

Chen, Baizhu, and Kien C. Tran, 1994, Are we sure that the real exchange rate follows a random walk? A reeximation, International Economic Journal 8, 33-44.

Cheung, Yin-Wong and Menzie D. Chinn, 1996, Further investigation of the uncertain unit root in GNP, Journal of Business and Economic Statistics, forthcoming.

Cheung, Yin-Wong and Kon Lai, 1993, Long-run purchasing power parity during the recent float, Journal of International Economics 34, 181-192.

Cheung, Yin-Wong and Kon Lai, 1994, Mean reversion in real exchange rates, Economics Letters 46, 251-256.

Chinn, Menzie D. and Richard A. Meese, 1995, Banking on currency forecasts: How predictable is change in money?, Journal of International Economics 38, 161-178.

Cochrane, John H., 1988, How big is the random walk in GNP?, Journal of Political Economy 96, 893-920.

Cochrane, John H., 1991, A critique of the application of unit root tests, Journal of Economic Dynamics and Control 15, 275-284.

Edison, Hali J., 1987, Purchasing power parity in the long run: A test of the dollar/pound exchange rate (1890-1978), Journal of Money, Credit and Banking 19, 376-387.

Edison, Hali J. and Jan T. Klovland, 1987, A quantitative reassessment of the purchasing power parity hypothesis: Some evidence on Norway and the United States, Journal of Applied Econometrics 2, 309-334.

28

Edison, Hali J.; Joseph E. Gagnon; and, William R. Melick, 1994, Understanding the empirical literature on purchasing power parity: The post-Bretton Woods era, Board of Governors of the Federal Reserve System, International Finance Discussion Papers, no. 465.

Engel, Charles, 1993, Real exchange rates and relative prices: An empirical investigation, Journal of Monetary Economics 32, 35-50.

Engel, Charles, 1995, Accounting for U.S. real exchange rate changes, National Bureau of Economic Research, working paper no. 5394.

Fischer, Eric O'N., and Joon Y. Park, 1991, Testing purchasing power parity under the null hypothesis of cointegration, Economic Journal 101, 1476-1484.

Frankel, Jeffrey A., 1986, International capital mobility and crowding out in the U.S. economy: Imperfect integration of financial markets or goods markets?, in R.W. Hafer, ed., How open is the U.S. economy?, (Lexington: Lexington Books).

Frankel, Jeffrey A., and Andrew Rose, 1996, A panel project on purchasing power parity: Mean reversion within countries and between countries, Journal of International Economics 40, 209-224.

Froot, Kenneth A.; Michael C. Kim; and Kenneth Rogoff, 1995, The law of one price over 700 years, National Bureau of Economic Research, working paper no. 5132.

Glen, Jack D. 1992, Real exchange rates in the short, medium, and long run, Journal of International Economics 33, 147-166.

Hansen, Bruce E., 1995, Rethinking the univariate approach to unit root testing: Using covariates to increase power, Department of Economics, Boston College.

Horvath, Michael T. and Mark Watson, 1994, Testing for cointegration when some of the cointegrating vectors are known, National Bureau of Economic Research, technical working paper no. 171.

Kim, Yoonbai, 1990, Purchasing power parity in the long run: A cointegration approach, Journal of Money, Credit and Banking 22, 491-503.

Kremers, Jeroen J.M.; Neil R. Ericsson; and Juan J. Dolado, 1992, The power of cointegration tests, Oxford Bulletin of Economics and Statistics 54, 325-348.

Kwiatkowski, Denis; Peter C.B. Phillips; Peter Schmidt; and, Yongcheol Shin, 1992, Testing the null hypothesis of stationariy against the alternative of a unit root: How sure are we that economic time series have a unit root?, Journal of Econometrics 54, 159-178.

Lothian, James R., and Mark P. Taylor, 1996, Real exchange rate behavior: The recent float from the perspective of the past two centuries, Journal of Political Economy 104, 488-509.

Mark, Nelson C., 1990, Real and nominal exchange rates in the long run, Journal of International Economics 28, 115-136.

Mark, Nelson C., 1995, Exchange rates and fundamentals: Evidence on long-horizon predictability, American Economic Review 85, 201-218.

Mark, Nelson C. and Doo-Yull Choi, 1996, Real exchange-rate prediction over long horizons, Department of Economics, The Ohio State University.

Mitchell, Brian R., 1988, British Historical Statistics (Cambridge: Cambridge University Press).

Ng, Serena and Pierre Perron, 1995, Unit root tests in ARMA models with data dependent methods for the truncation lag, Journal of the American Statistical Association 90, 268-281.

Patel, Jayendu, 1990, Purchasing power parity as a long-run relation, Journal of Applied Econometrics 5, 367-379.

Rogers, John H., 1995, Real shocks and real exchange rates in really long-term data, Board of Governors of the Federal Reserve.

Rogers, John H. and Michael A. Jenkins, 1995, Haircuts or hysteresis: Sources of movements in real exchange rates, Journal of International Economics 38, 339-360.

Rogoff, Kenneth, 1996, The purchasing power parity puzzle, Journal of Economic Literature, forthcoming.

Samuelson, Paul A., 1964, Theoretical notes on trade problems, Review of Economics and Statistics 46, 145-154.

Schwert, G. William, 1989, Tests for unit roots: A Monte Carlo investigation, Journal of Business and Economics Statistics 7, 147-159.

Stockman, Alan C. and Linda L. Tesar, 1995, Tastes and technology in a two-country model of the business cycle: Explaining international comovements, American Economic Review 85, 168-185.

Summers, Robert and Alan Heston, 1991, The Penn World Table (Mark 5): An expanded set of international comparisons, 1950-88, Quarterly Journal of Economics 106, 327-368.

Taylor, Mark P., 1988, An empirical examination of long run purchasing power parity using cointegration techniques, Applied Economics 20, 1369-1381.

Wei, Shang-Jin and David C. Parsley, 1995, Purchasing power disparity during the floating rate period: Exchange rate volatility, trade barriers and other culprits, National Bureau of Economic Research, working paper no. 5032.

Zivot, Eric, 1995, The power of single equation tests for cointegration when the cointegrating vector is known, Department of Economics, University of Washington.

Table 1

Unit-Roots Tests and Cointegration Tests on Disaggregated Data

Tests on s_t and p_t-p_t^*:

	Test Statistic	5% Critical Value
Dickey-Fuller	2.120	2.89
Augmented Dickey-Fuller	2.802	2.89
Error-Correction Model	2.631	2.81
Horvath-Watson	7.921	10.18

Tests on s_t and p_t^T-p_t^{T*}:

	Test Statistic	5% Critical Value
Dickey-Fuller	2.115	2.89
Augmented Dickey-Fuller	2.823	2.89
Error-Correction Model	2.653	2.81
Horvath-Watson	6.387	10.18

Tests on y_t:

	Test Statistic	5% Critical Value
Dickey-Fuller	1.710	2.89
Augmented Dickey-Fuller	1.732	2.89

Table 2

Estimates of the Coefficients from Equations (4), (6), and (7)

Coefficient	Estimate	standard error
δ	0.080382	0.03812
γ	-0.003415	0.00849
a	0.005725	0.00038
b	0.001088	0.00509
c	0.050770	0.00361
g	0.006109	0.00113
f	0.000632	0.00113
d	0.011286	0.00085

Table 3

Size of Long-Run PPP Tests with Parameters Calibrated to Long-Run Data

	1	2	3	4	5	6	7	8
a	0.0010	0.0030	0.0060	0.0090	0.0120	0.0150	0.0180	0.0210
d	0.0189	0.0193	0.0195	0.0193	0.0186	0.0174	0.0155	0.0126
c	0.0415	0.0415	0.0417	0.0420	0.0427	0.0438	0.0456	0.0484
δ	0.0240	0.0244	0.0257	0.0281	0.0320	0.0382	0.0476	0.0609
var1	0.0092	0.0773	0.2587	0.4591	0.6289	0.7550	0.8405	0.8941
DF	0.3074	0.2900	0.3048	0.3130	0.3436	0.3560	0.3938	0.4226
ADF	0.3078	0.2986	0.3120	0.3152	0.3358	0.3264	0.3478	0.3566
ECM	0.3886	0.3810	0.3954	0.3726	0.3738	0.3626	0.3432	0.3200
HW	0.3116	0.3002	0.3164	0.2984	0.3096	0.3040	0.3040	0.3020
KPSS	0.0062	0.0046	0.0054	0.0128	0.0138	0.0204	0.0350	0.0502

Notes:

The letters a, d, c and δ refer to parameter values from equations (4), (6) and (7).

"var1" refers to the fraction of the 100-year change in the real exchange rate accounted for by the unit root component for each set of parameter values.

The numbers reported in the rows DF, ADF, ECM, and HW are the true size for a test of nominal size of 5 per cent for the Dickey-Fuller, Augmented Dickey-Fuller, Error Correction Model and Horvath-Watson tests, respectively.

The numbers reported in the KPSS row are the power of a KPSS test for stationarity with a size of 5 per cent.

33
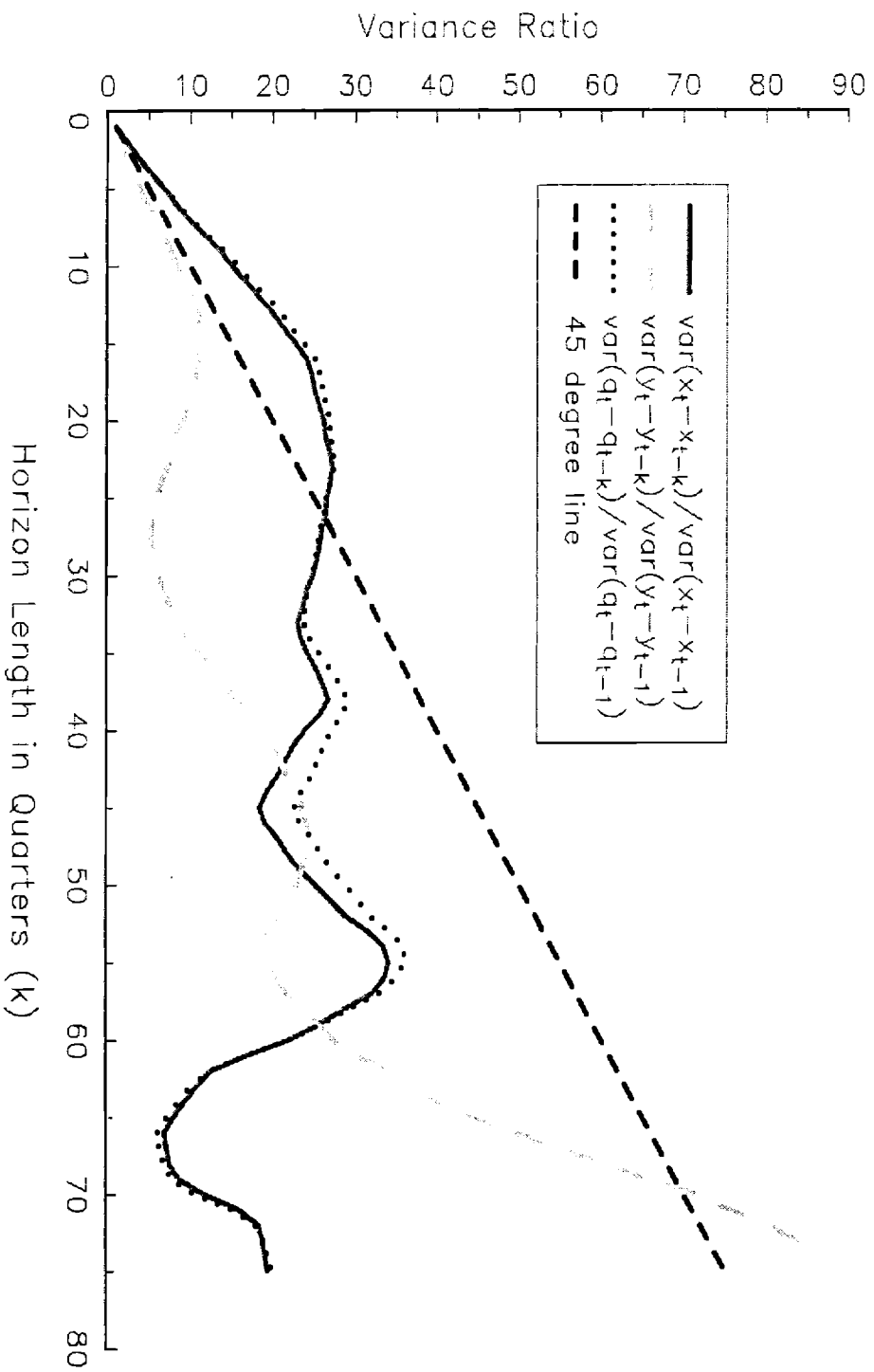
Figure 1 US/UK Real Exchange Rate Variance Component Ratios
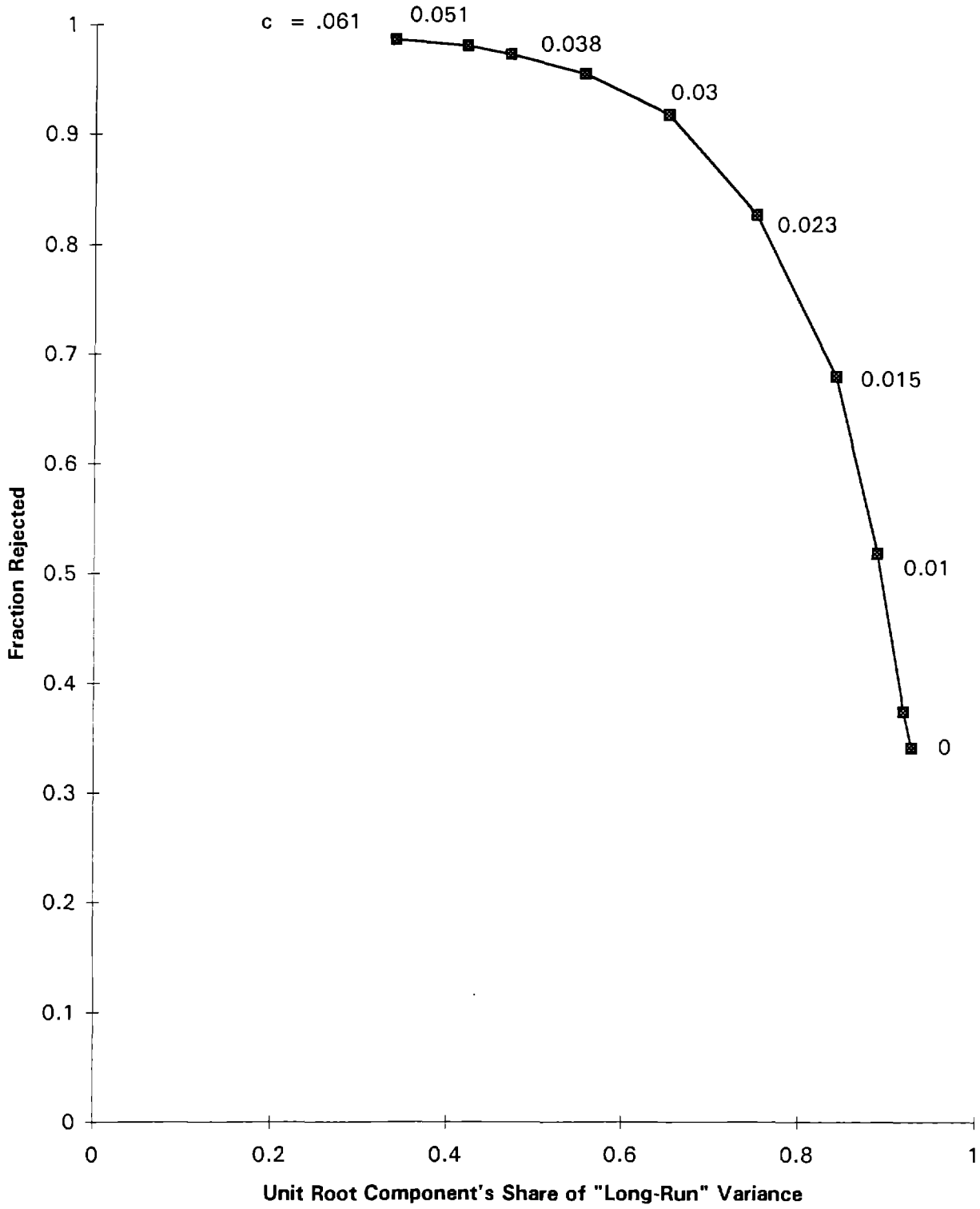Quarterly Deflator Data from 1970:1 to 1994:4

Variance Ratio

Horizon Length in Quarters (k)

var($x_t - x_{t-k}$)/var($x_t - x_{t-1}$)
var($y_t - y_{t-k}$)/var($y_t - y_{t-1}$)
var($q_t - q_{t-k}$)/var($q_t - q_{t-1}$)
45 degree line

Figure 2A: Dickey-Fuller Test as c Varies

Figure 2B: Augmented Dickey-Fuller Test as c Varies

Fraction Rejected (y-axis)

Unit Root Component's Share of "Long-Run" Variance (x-axis)

c = .061
0.051
0.038
0.03
0.023
0.015
0.01
0
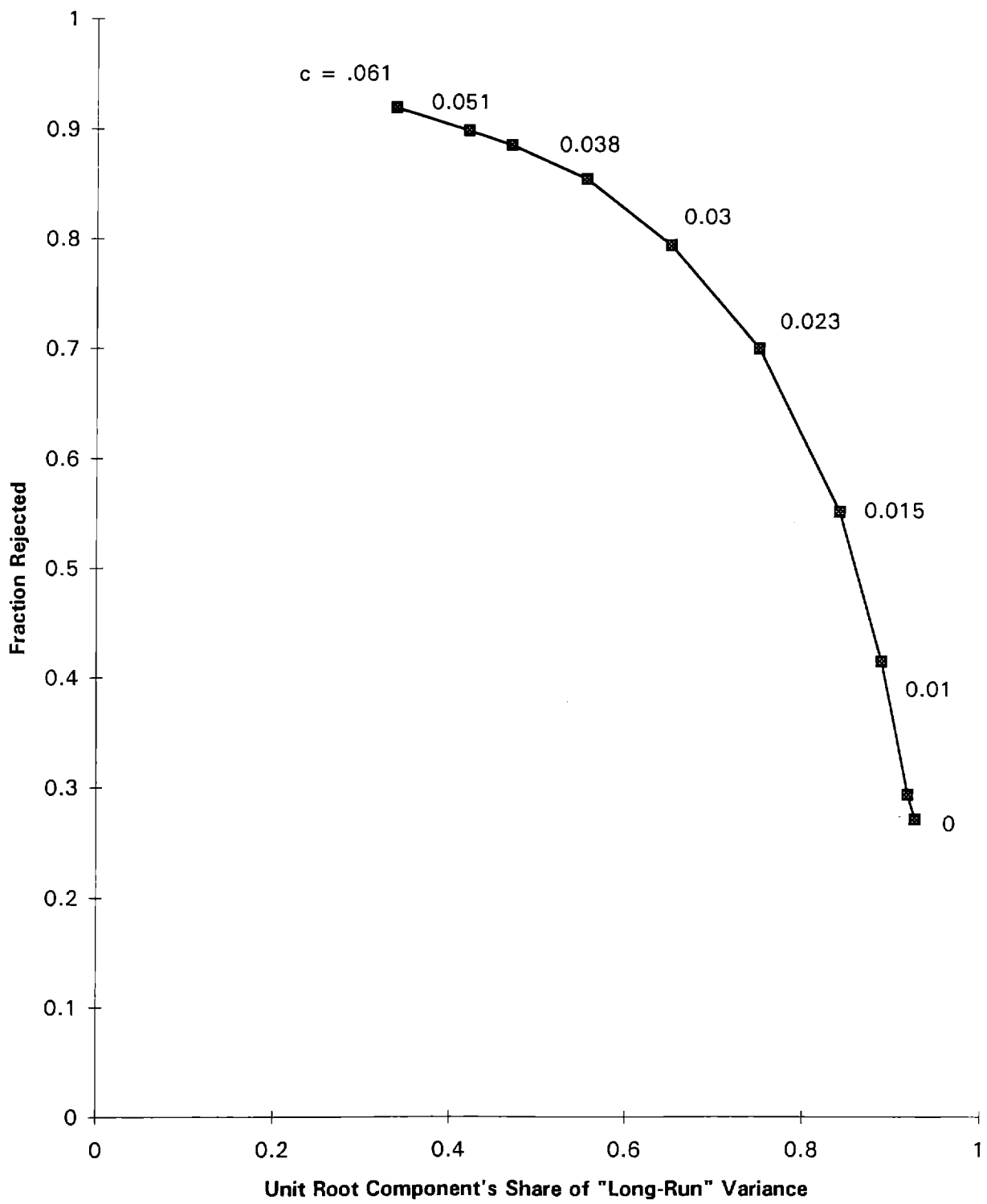
Figure 2C: Error-Correction Model Test as c Varies
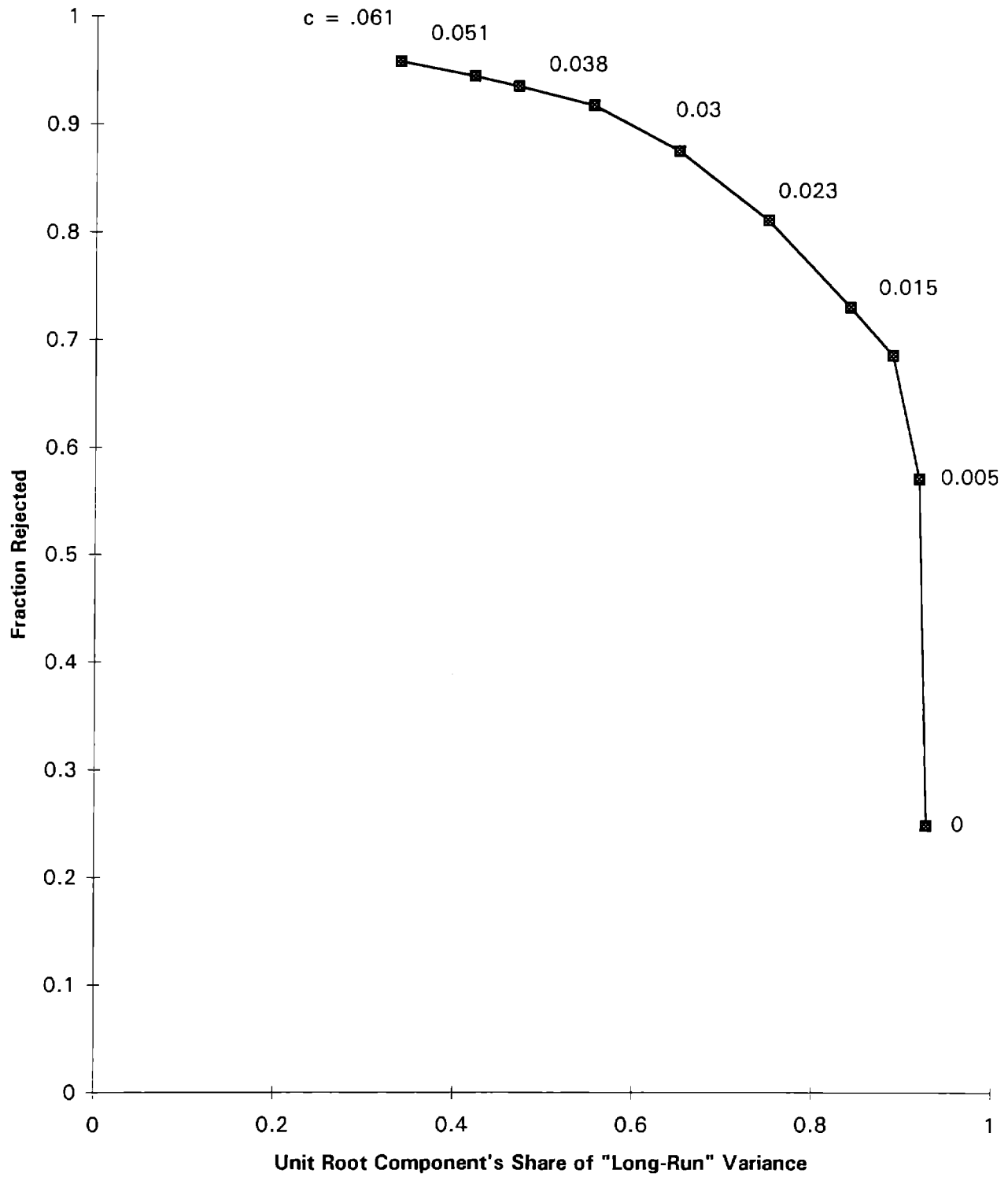
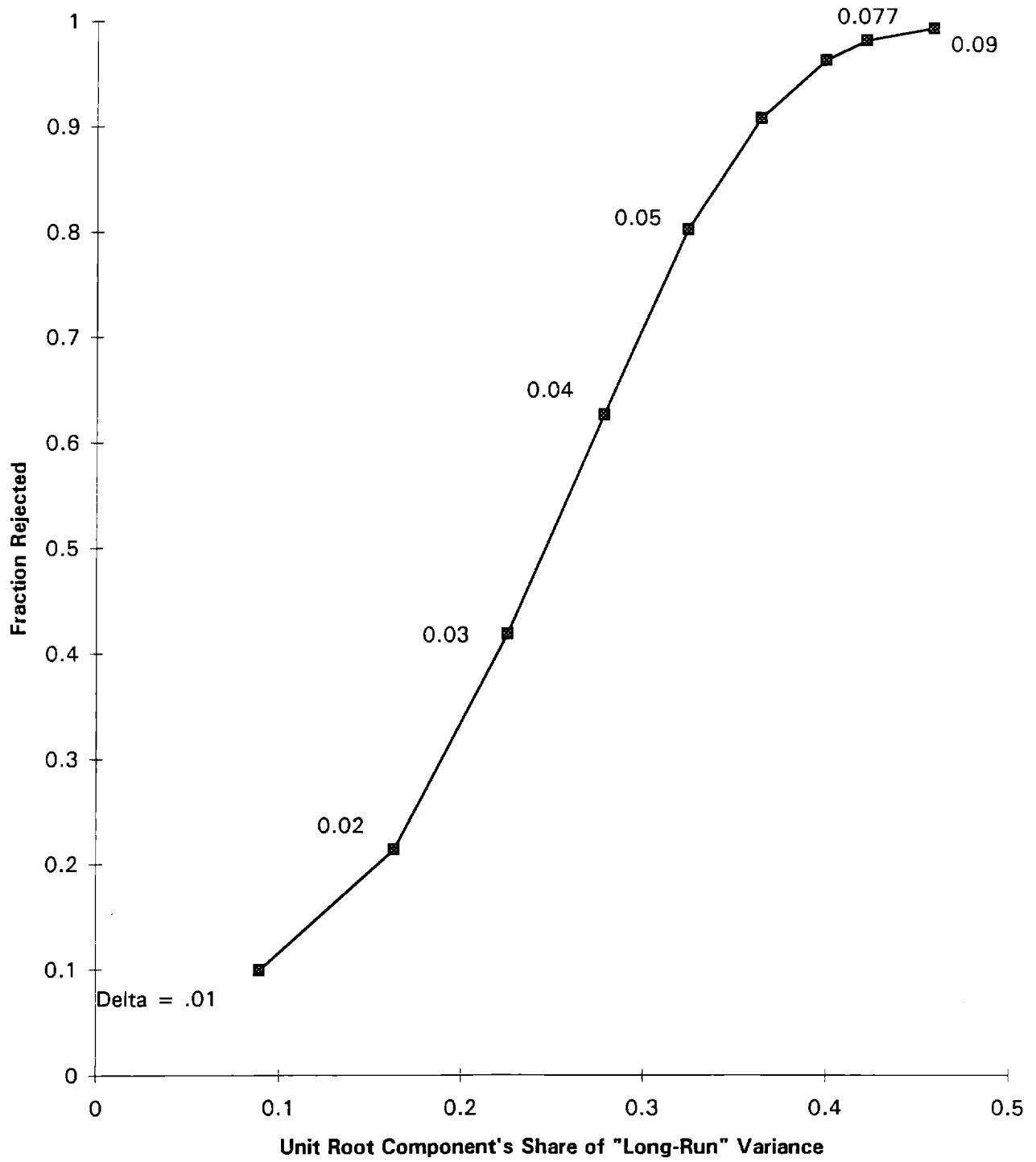Figure 2D: Horvath-Watson Test as c Varies
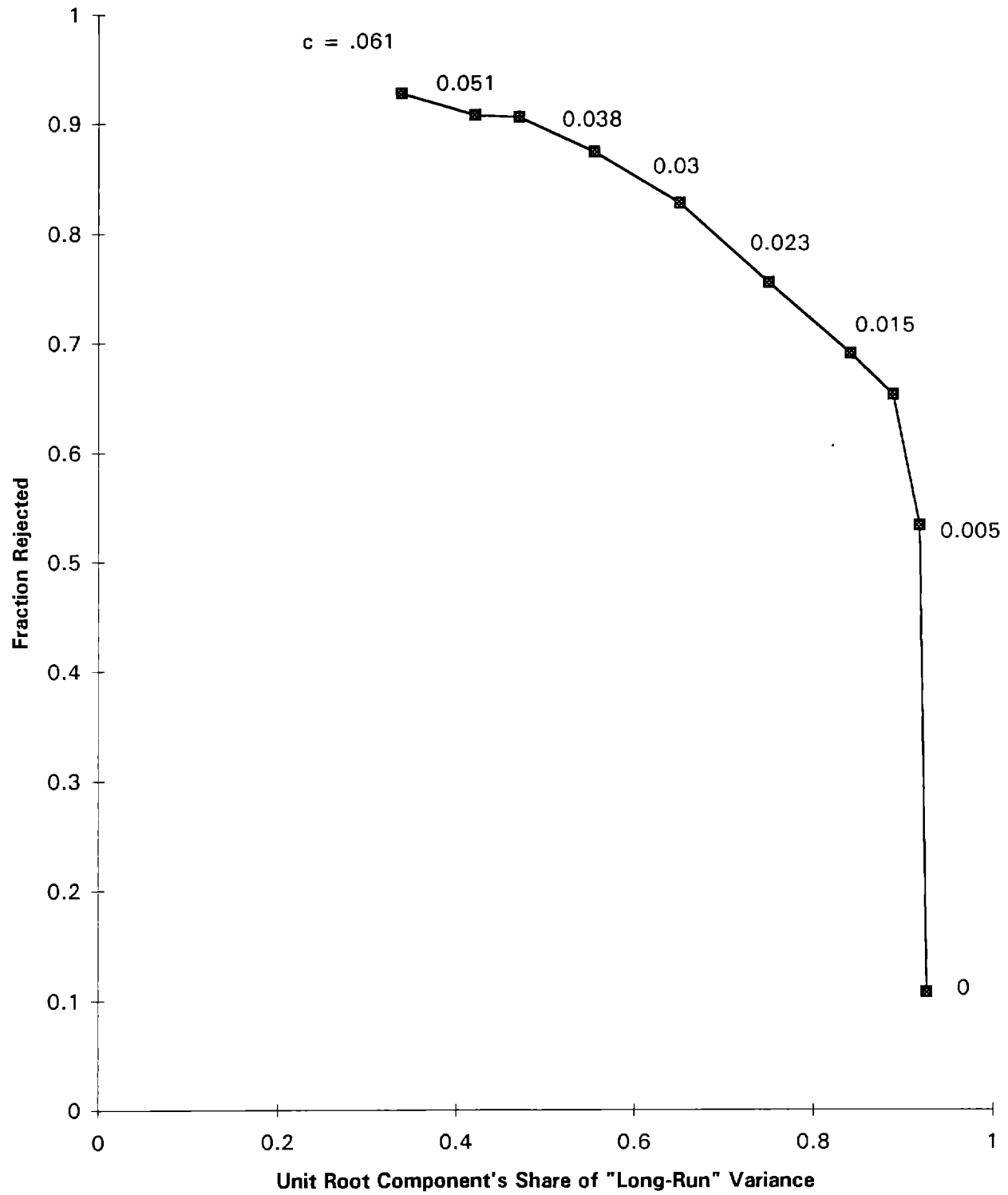
Figure 3A: Dickey-Fuller Test as a Varies
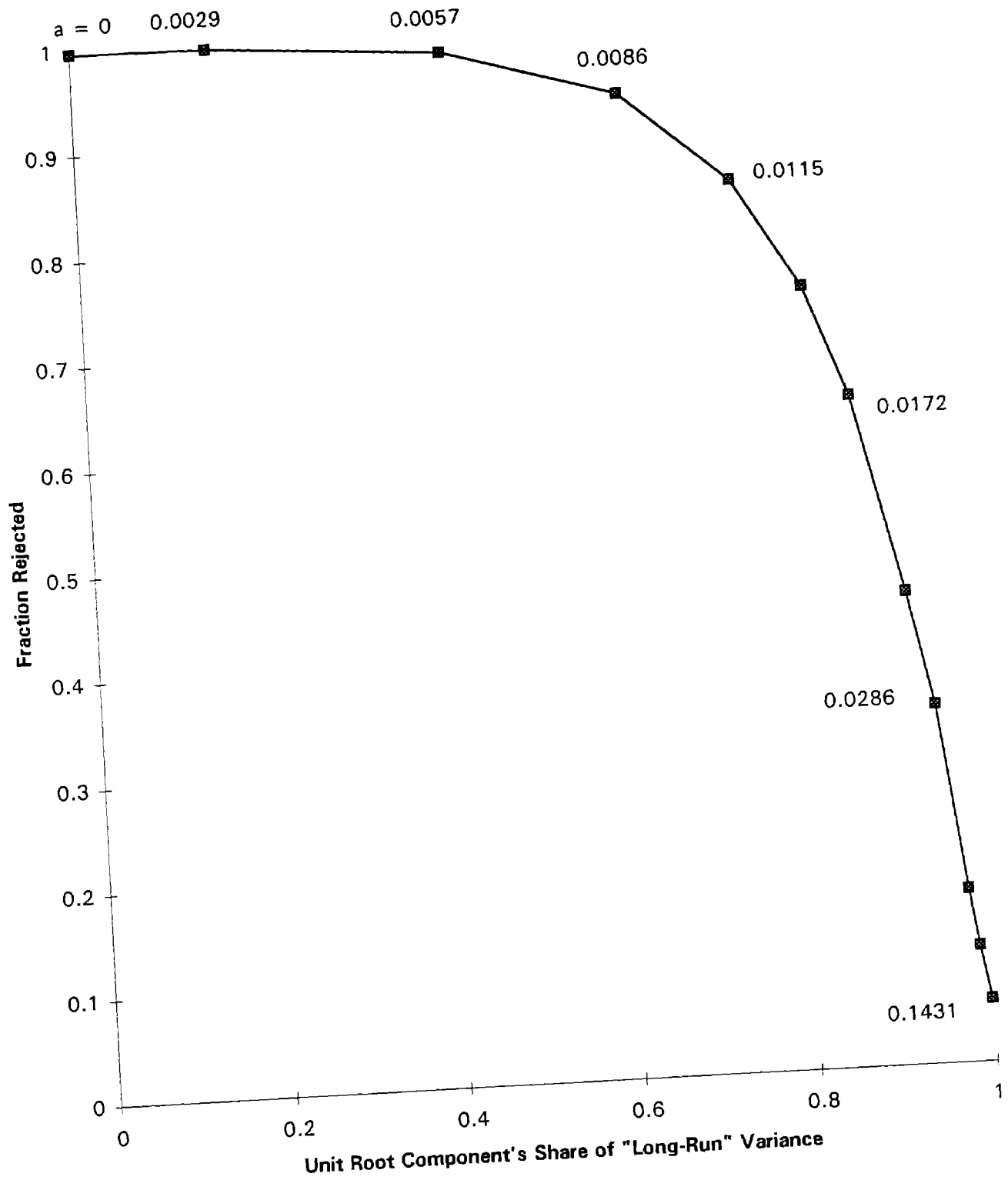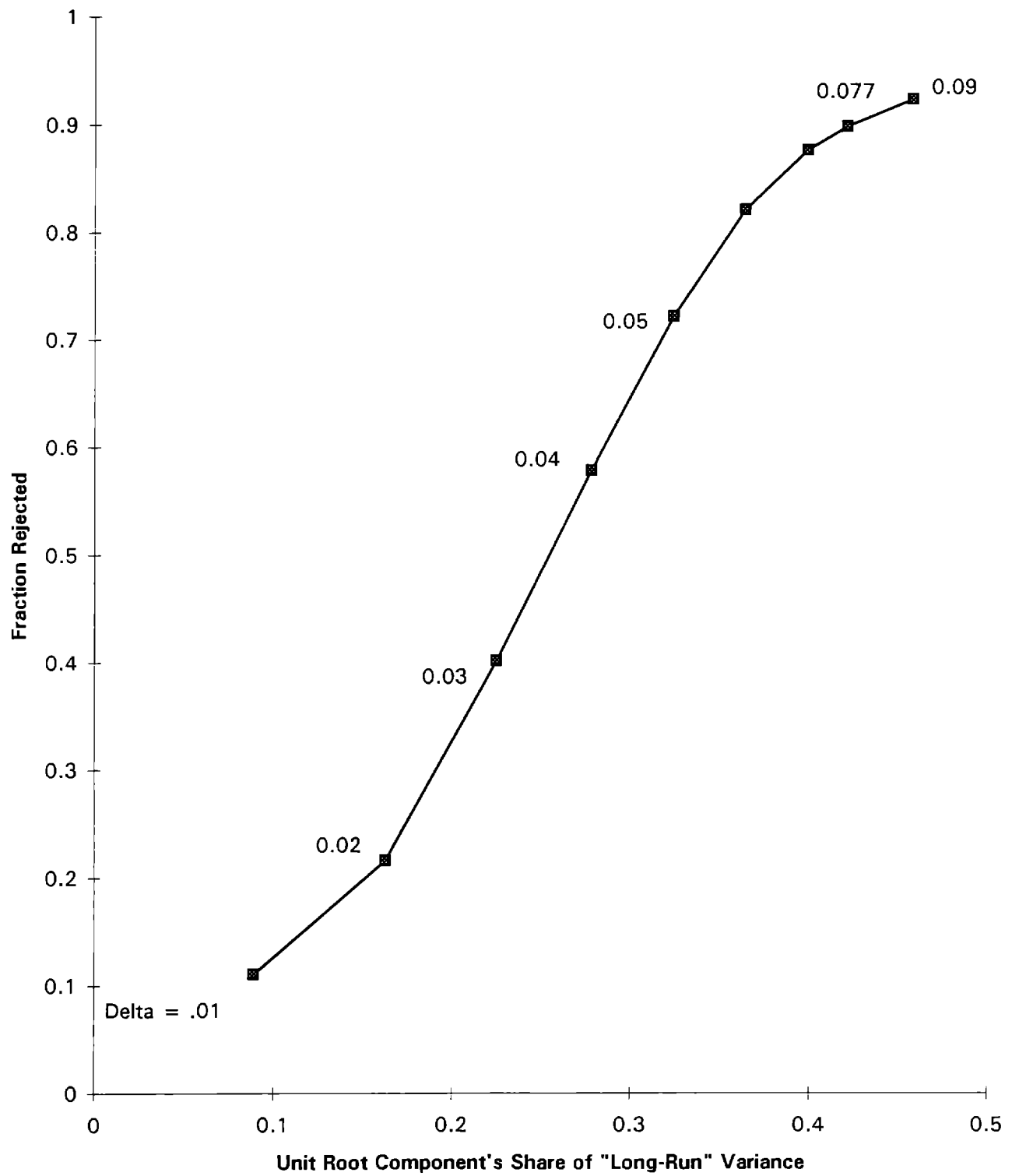
Figure 3B: Augmented Dickey-Fuller Test as a Varies
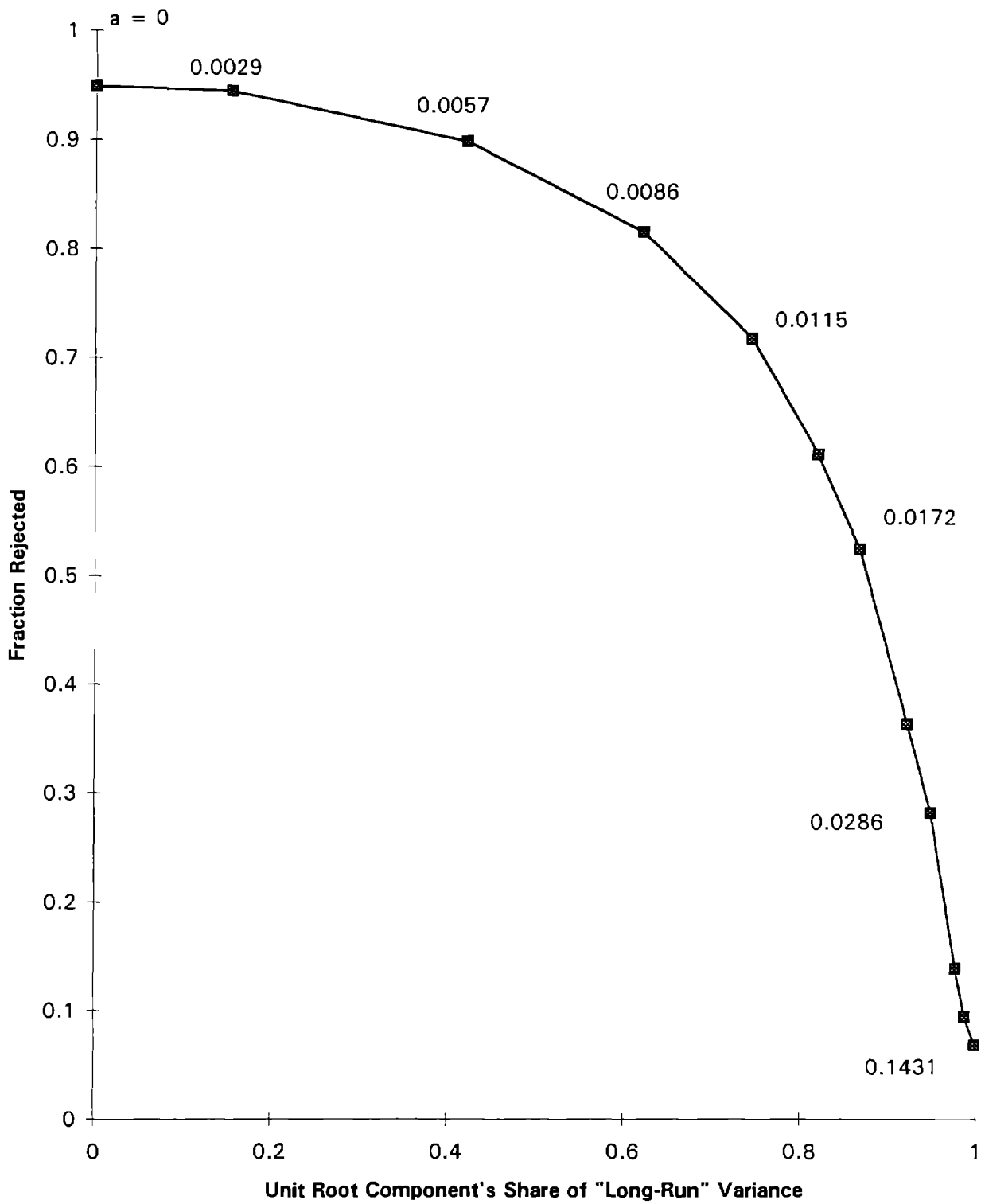
Figure 3C: Error-Correction Model Test as a Varies
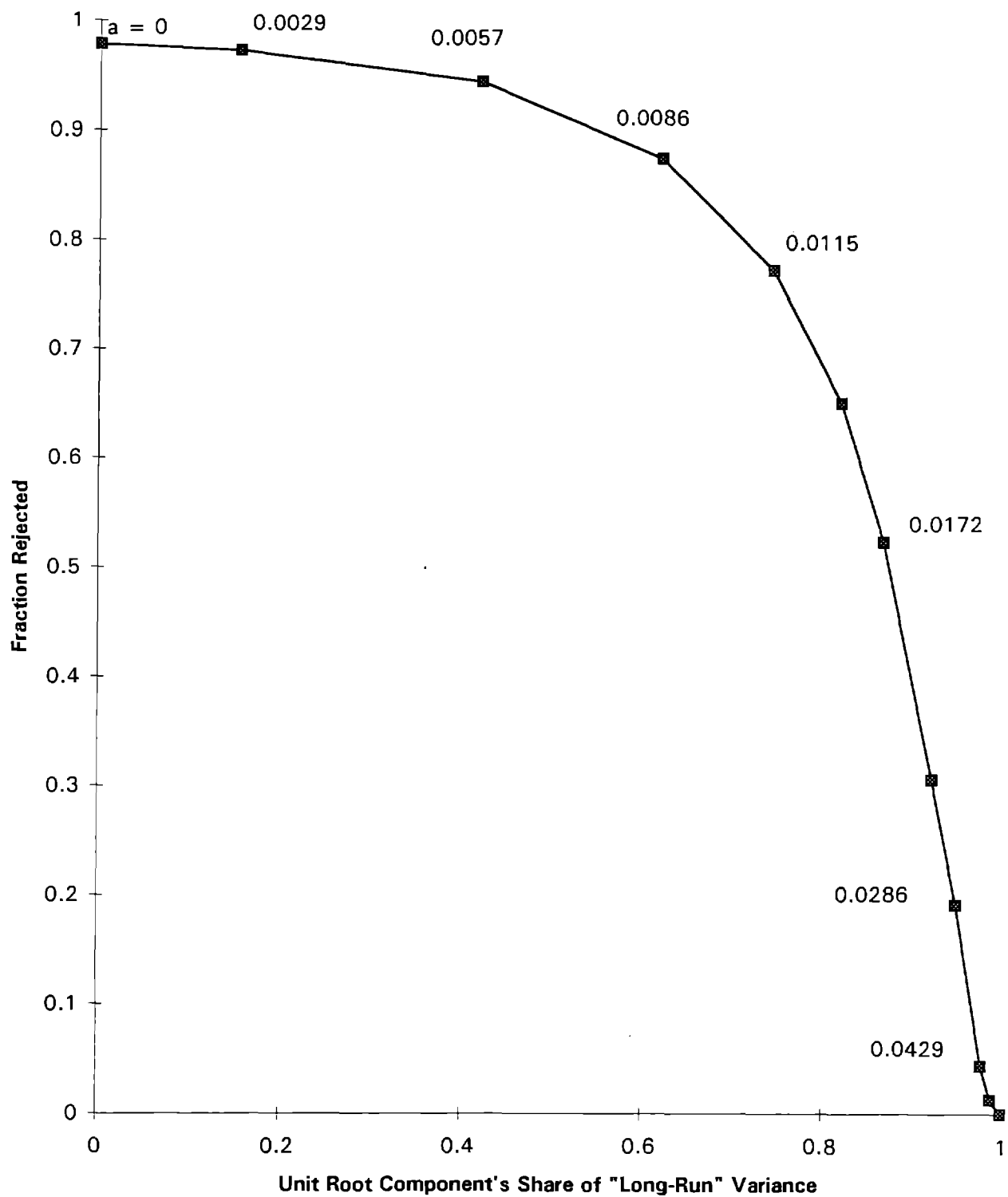
Figure 3D: Horvath-Watson Test as a Varies

a = 0   0.0029

0.0057

0.0086

0.0115

0.0172

0.0229

0.0429

Fraction Rejected

Unit Root Component's Share of "Long-Run" Variance
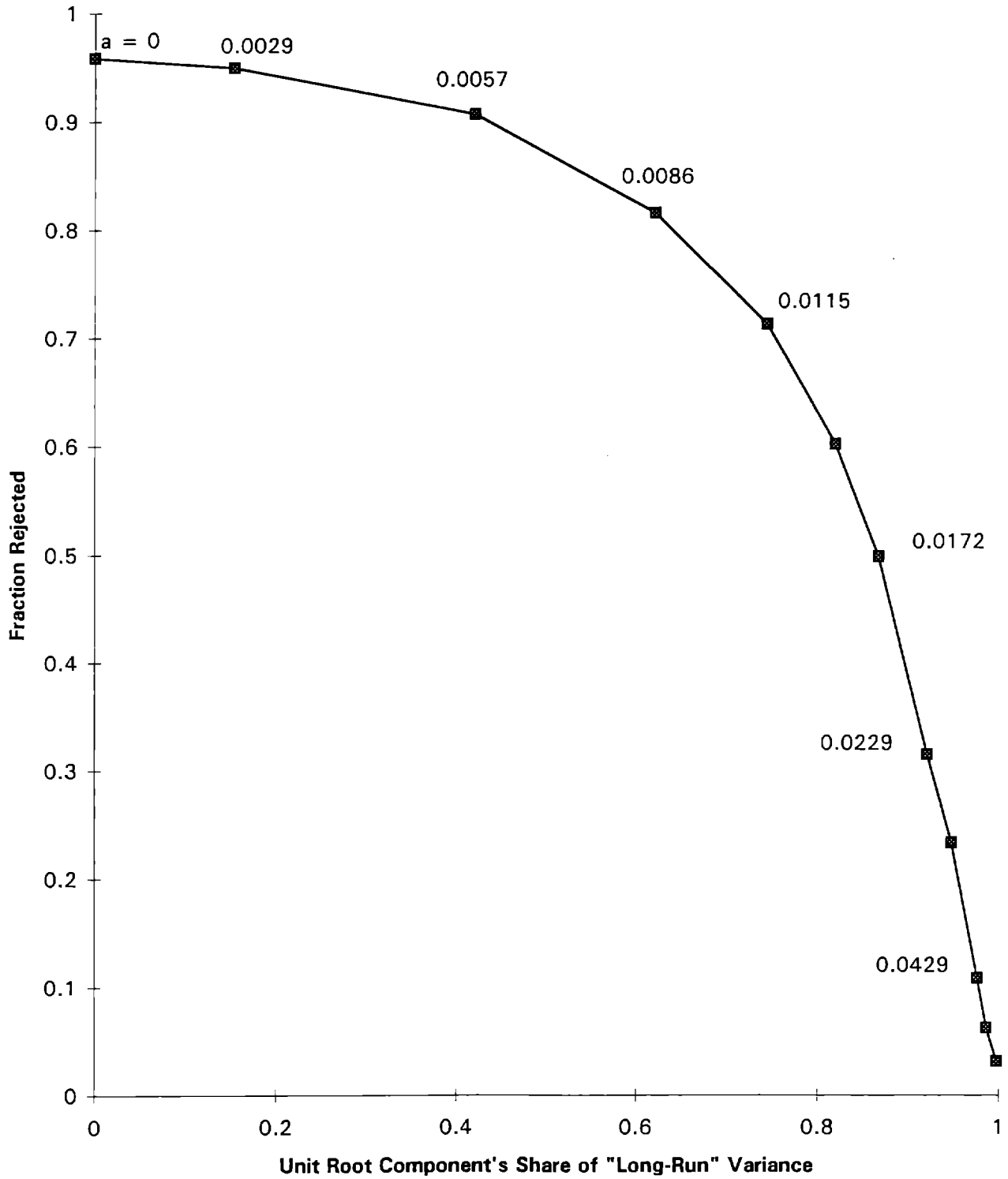
Figure 4A: Dickey-Fuller Test as Delta Varies

Fraction Rejected (y-axis)

Unit Root Component's Share of "Long-Run" Variance (x-axis)

0.077

0.09

0.05

0.04

0.03

0.02

Delta = .01

Figure 4B: Augmented Dickey-Fuller Test as Delta Varies

Fraction Rejected

Unit Root Component's Share of "Long-Run" Variance

Figure 4C: Error-Correction Model Test as Delta Varies
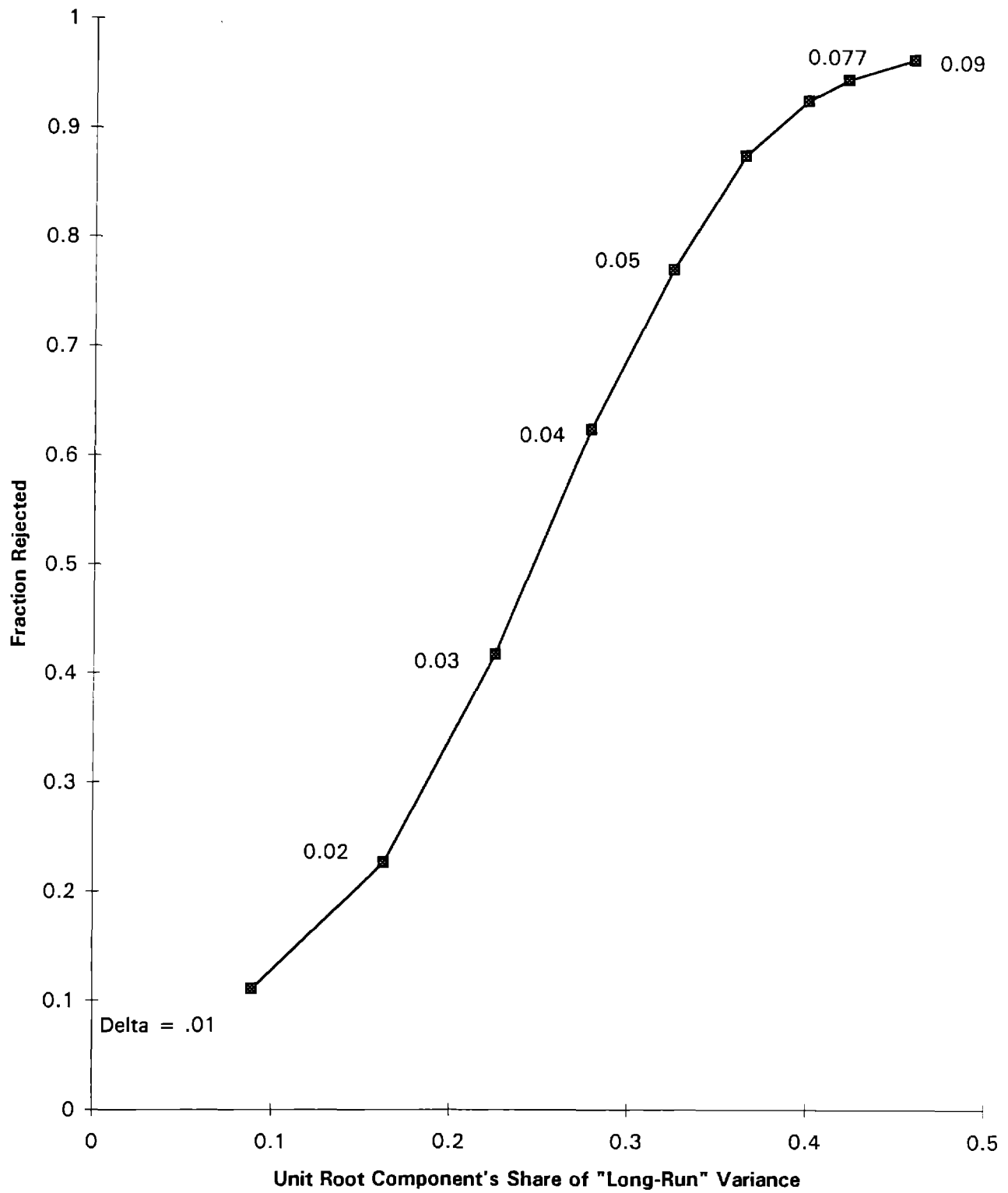
Figure 4D: Horvath-Watson Test as Delta Varies
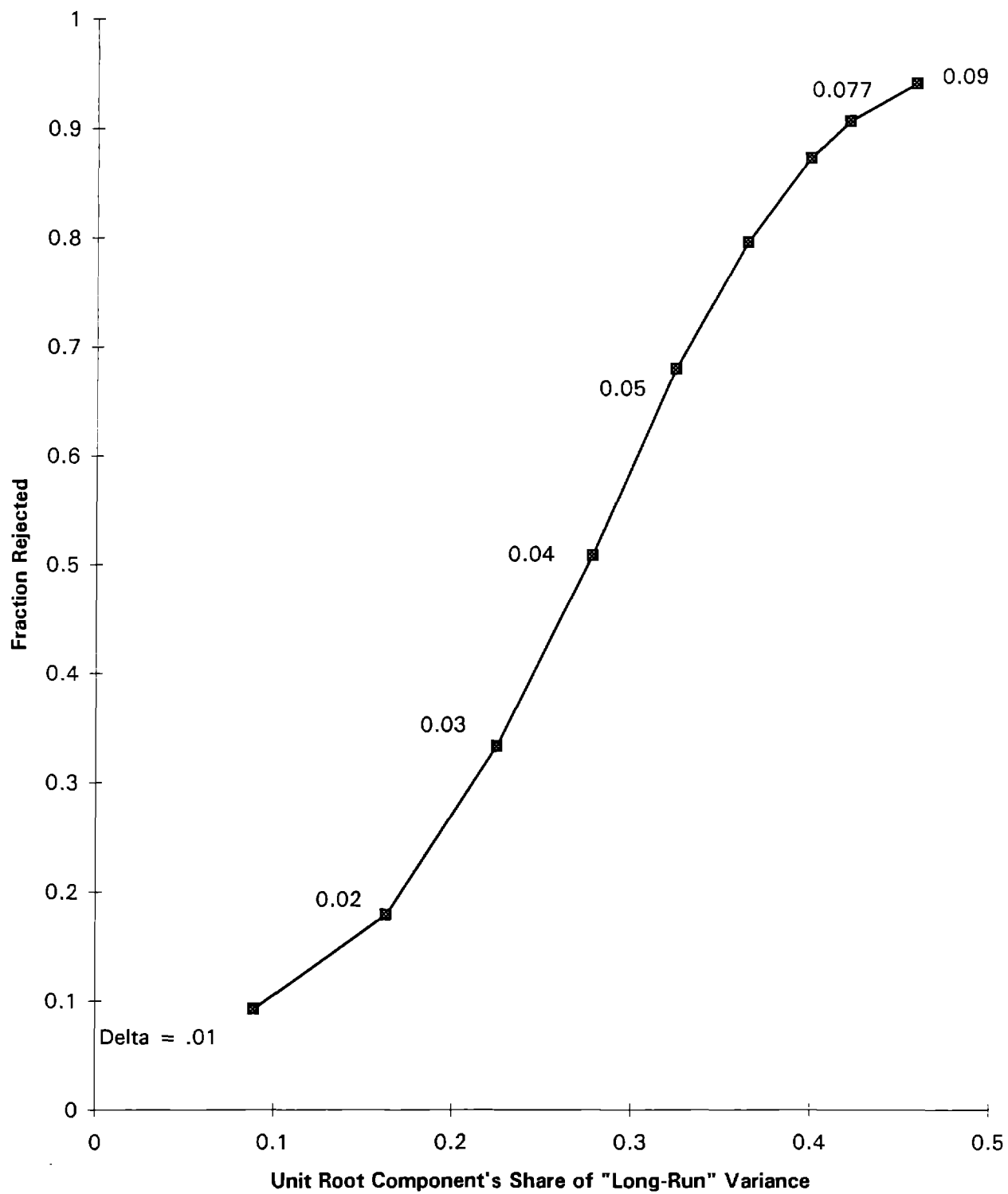
Figure 5A: KPSS Test as c Varies

Fraction Rejected
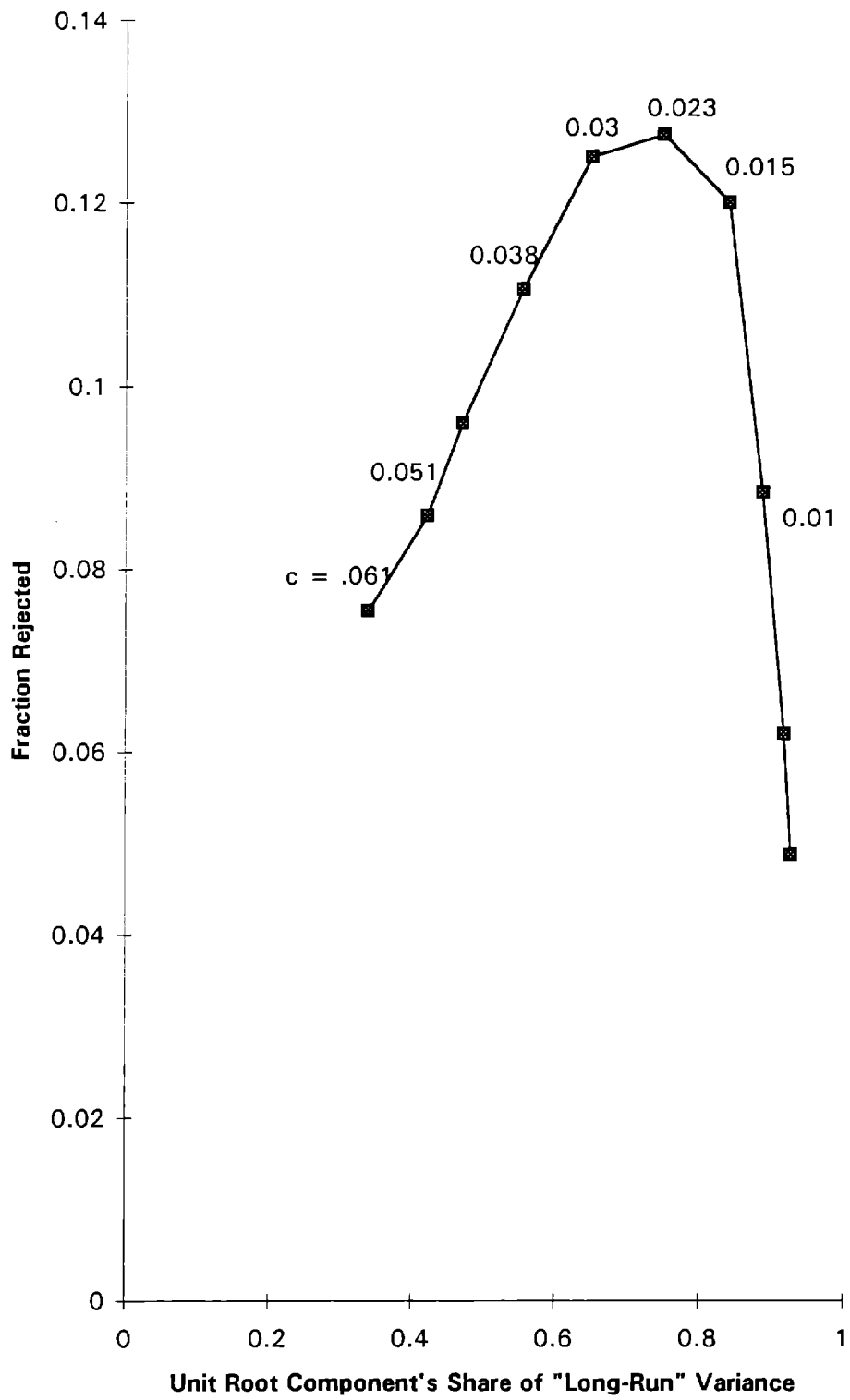
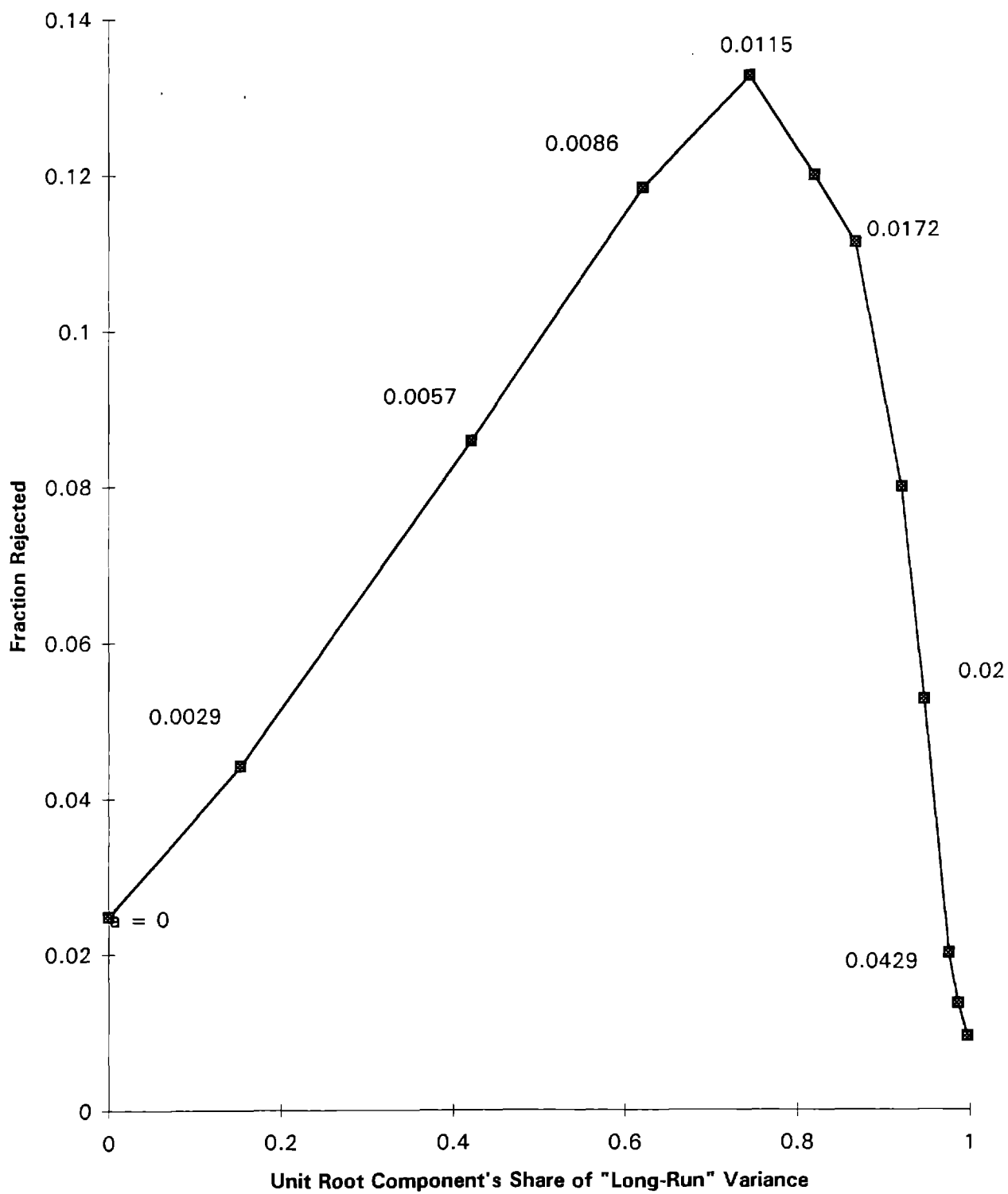Unit Root Component's Share of "Long-Run" Variance

Figure 5B: KPSS Test as a Varies

Figure 5C: KPSS Test as Delta Varies