Testing the Present Value Model

Masako Miyanishi
Department of Economics, University of Georgia,
Athens, GA
October 19, 2008

Abstract

This paper considers the "excess volatility puzzle" of the present value model in the stock market. In an empirical study, stock price is found to be far too volatile for the present value model to hold. Dividend and stock price are often assumed to have a unit root, and the model is tested in a cointegration framework. Because the dividend data may have not an exact, but a near unit root, we suspect the test is biased. Even if an efficient estimator is used, the test will be severely biased when data does not have an exact unit root and the innovation of data is correlated with the cointegration equation. We show that dividend data could be a near unit root process. We propose a valid test in the presence of a near unit root. After we take into account the sample information of the largest root, the test is marginally rejected with the 5 percent significance level in the long run data. The presence of a near unit root may be one of factors that cause the test to be rejected under the usual normality assumption.
1 Introduction

This paper considers the present value model in the stock market, which states that a stock price today is the expectation of the present value of all future dividends. Empirical studies have tested the model in a cointegration framework, assuming that (logs of) stock price and dividend are unit root processes (for example, Mankiw, Romer, and Shapiro (1985, 1991) and Cochrane (1992, 1994). If data supports the present value model, logs of stock price and dividend are cointegrated with the cointegrating vector (1,-1). The typical empirical finding is against the present value model. The (log of) stock price is found to be far too volatile relative to dividend, which is the so called "excess volatility puzzle" (Barsky and Delong (1993)). In other words, the hypothetical cointegrating vector is rejected in empirical study. We re-examine this "excess volatility puzzle".

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

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

The "excess volatility" in the present value model is found even if an efficient class of estimator is used. We suspect that the "excess volatility" is due to the presence of a near unit root. If so, the standard normal assumption is not applicable to the test statistics of the cointegrating vector. We examine this possibility using the annual data of stock price and dividend from 1926 to 2007. Then, we will consider a valid test in the presence of a near unit root. We make four contributions in this paper. First, we show that the largest root of dividend (and stock price) may not have an exact unit root. We estimate confidence intervals for the largest autoregressive root as suggested by Stock (1991), instead of having point estimates. Confidence intervals contain unity but, there is the possibility that the largest root is below unity, that is, the data may be a near unit root process. If the largest root is smaller than one and if there is endogeneity in the cointegration equation, test statistics will be biased. By estimating confidence intervals and correlations, we claim that the presence of a near unit root might be at least one of causes of the "excess volatility puzzle"; the test statistics on the cointegrating vector are systematically biased because the data do not have an exact unit root, although it may not fully explain the rejection of test.

Second, we propose a valid test in the presence of a near unit root. We apply Bonferroni’s inequality to take into account the sample information on the largest root. Application of Bonferroni’s inequality was first proposed by Cavanagh, Elliott and Stock (1995) for a predictive regression. Torous et al. (2004) and Campbell and Yogo (2006) applied
the method to test stock return predictability. Since a predictor variable such as dividend-price ratio is highly persistent, the test of predictability is rejected too often. We apply Bonferroni’s inequality to Saikkonen’s (1992) efficient estimator. For a given value of correlation and possible values of the largest root, we show how to obtain valid critical values. The method is simple and applicable to any efficient estimator.

Third, we evaluate finite sample performance of the proposed test. Since the data is always finite in an empirical study, we examine how the test is improved in a finite sample rather than looking at asymptotics. By numerical work, we compare the size and power of our test to an efficient test without any treatment and another valid test proposed by Wright (2000). Compared to Saikkonen’s test without any treatment, our test greatly improves both size and power. It is obvious that an efficient class of estimator leads to invalid inference when the data does not have a unit root. And the performance of our test is comparable to that of the method of Wright (2000).

Fourth, we apply our test to empirical study. Based on an estimated correlation and the sample information about the largest root, we obtain valid critical values. We conduct a test for the entire sample from 1926 to 2007 and various subsamples. For the entire sample, test statistic is marginal, slightly exceeds critical values. For subsamples, the present value model is still rejected by data. For example, in a subsample from 1926 to 1951, during which U.S. economy experienced the Great Depression, the behavior of stock price seems different from that of dividend. While dividend is nearly a unit root process, the largest root of stock price seems well below unity. For a later period, 1974-2007, the linear combination of stock price and dividend (cointegration error) becomes highly persistent.

The rest of paper is organized as follows. The second section presents the model we examine and two parameters that cause the test statistics to be biased. We discuss
the estimation method as well. In the third section, we propose a valid test which we call Bonferroni-Saikkonen, and also discuss another testing method in the presence of a near unit root. The fourth section evaluates finite sample performance of our test by simulations. The fifth section is empirical application, where we apply our test to the present value model. The sixth section concludes.

2 Test Statistics in the Presence of Near Unit Root

2.1 Model

We consider the following simple triangle model.

\[
\begin{align*}
    x_t &= \alpha_1 + \rho x_{t-1} + u_{1t} \\
    y_t &= \alpha_2 + \beta x_t + u_{2t}
\end{align*}
\]

The second equation is the cointegration equation. We are interested in testing the cointegrating coefficient \( \beta \) under

\[
H_0: \beta = \beta_0 \quad \text{vs.} \quad H_1: \beta \neq \beta_0
\]

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

\( \alpha_1 \) and \( \alpha_2 \) are deterministic terms. \( u_t = [u_{1t}, u_{2t}]' \) can be serially correlated and \( \Phi(L) \)
\( u_t = \varepsilon_t \), where \( \varepsilon_t = [\varepsilon_{1t}, \varepsilon_{2t}]' \) and \( E[\varepsilon_t\varepsilon_t'] = \Sigma \). \( \Phi(L) \) is a lag polynomial of known order with all roots outside the unit circle, that is, \( u_t \) is stationary. \( 2\pi \) times spectral density at frequency zero of \( u_t \) is \( \Omega = \Phi(1)^{-1}\Sigma \Phi(1)^{-1}' = \begin{pmatrix} \sigma_1^2 & \delta\sigma_1\sigma_2 \\ \delta\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \), where \( \Phi(1) = \sum_i \Phi_i \) and \( \delta \) represents zero frequency correlation.

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

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

Here we are going to explain bias in \( t \)-statistics on \( \beta \) in a univariate case. By Theorem 1 in Elliott (1998),

\[
\hat{V}^{-1/2}\{T(\hat{\beta} - \beta) - B\}^{-A}N(0, 1)
\]
where ~a indicates ‘approximately distributed, and

\[ B = -\Omega_{21}\Omega_{11}^{-1}c \]  
\[ \hat{V} = V = \left[ \Omega_{11}^{1/2} \left( \int (J_c^d)^2 \right) \Omega_{11}^{1/2} \right]^{-1} \Omega_{21} \]  
\[ \Omega_{21} = \Omega_{22} - \Omega_{21}\Omega_{11}^{-1}\Omega_{12} \]

⇒ indicates weak convergence, and \( dJ_c^d(\lambda) \) is a detrended Ornstein Uhlenbeck process

where

\[ dJ_c(\lambda) = cJ_c(\lambda)d\lambda + dW_1(\lambda) \]  
\( J_c(0) = 0, W_1(\lambda) \) is a standard Brownian Motion associated with \( x_t \), and \( J_c^d = J_c(\lambda) - \int_0^1 J_c(s)ds \).

Thus, t-statistic has a bias term \( \hat{V}^{-1/2}B \), and,

\[ \hat{V}^{-1/2}B \Rightarrow -\frac{c\delta \int (J_c^d)^2}{\sqrt{(1 - \delta^2)}} \]  

If \( \delta = 0 \), there is no bias even if \( x_t \) does not have an exact unit root, though \( \delta \) is highly unlikely to be zero. If \( x_t \) does have an exact unit root (i.e., \( c = 0 \)), there is no bias, even if \( \delta \neq 0 \). As \( \delta \) gets close to one and \( c < 0 \), bias will go to infinity. If \( c < 0 \), the sign of \( \delta \) determines the direction of bias. If \( \delta > 0 \), t-statistic will be negatively biased.

2.3 Two nuisance parameters

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

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

An alternative approach would be to place a confidence interval on the value for \( \rho \) (or \( c \) interchangeably in a finite sample). Since the estimator of \( \rho \) does not have asymptotically a symmetric distribution, we cannot obtain equally tailed confidence intervals by adding and subtracting a proportion of the estimated standard error of the estimator (Stock (1991)). Instead, a valid confidence interval can be constructed by collecting the set of values for \( c \) that cannot be rejected in a hypothesis test that \( c = c_0 \). We explain how an equally tailed confidence interval for \( \rho \) (and \( c \)) is constructed. Stock (1991) first provided such a method through the inversion of the augmented Dickey-Fuller (ADF) test for a unit root of a first or higher order autoregressive (AR) process. Consider a higher-order AR model,

\[
x_t = \mu_1 + v_t
\]

where \( a(L)v_t = \varepsilon_t, \ a(L) = b(L)(1 - \rho L), \ b(L) = \sum_{i=0}^k b_i L^i, \ b_0 = 1 \) and \( L \) is the lag operator; \( b(1) \neq 0 \) is assumed, and \( \varepsilon_t \) is a martingale difference sequence with a finite

\(^1\)Note that if \( u_{1t} \) is independent and identically distributed, then, as \( T \to \infty \), the estimator of \( c \), \( \hat{c} = T(\hat{\rho} - 1) \Rightarrow c + \frac{I_n}{I_0^2(J_c(r))} dW(r) \), where \( J_c(r) \) is Ornstein-Uhlenbeck process with \( dJ_c(r) = cJ_c(r) dr + dW(r), 0 \leq r \leq 1 \) and \( W(r) \) is a standard Brownian motion. See Phillips (1987) for detailed discussion. Furthermore, \( T(\hat{\rho} - \rho) = O_p(T) \) if \( \hat{\rho} = 1 + \frac{c}{T} \), then, \( \hat{c} - c = O_p(1) \). Hence, \( c \) is not consistently estimable.
fourth moment. \( \rho \) is the largest root factored out from \( a(L) \) and may be equal to or near one. Local-to-unity asymptotic theory is applied. Then, the ADF regression model is

\[
\Delta x_t = \bar{\mu}_1 + (\alpha(1) - 1)x_{t-1} + \sum_{j=1}^{k} \alpha^*_j \Delta x_{t-1} + \varepsilon_t
\]

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

As shown in Stock (1991), the ADF \( t \)-statistic, which we denote as \( \hat{\tau} \), depends only on the local-to-unity parameter \( c \) and is continuous in \( c; \)

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

where \( \Rightarrow \) indicates weak convergence, and \( J(.) \) is a diffusion process which satisfies \( dJ(s) = cJ(s)ds + dW(s) \), where \( W(.) \) is a Brownian motion. If \( x_t \) is exactly a unit root process, \( J(.) \) is a Brownian motion.

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

\[
S (\hat{\tau}) = \left\{ c : f_{l_{1/2} \alpha} (c) \leq \hat{\tau} \leq f_{u_{1/2} \alpha} (c) \right\}
\]

where \( f_{l_{1/2} \alpha} (c) \) and \( f_{u_{1/2} \alpha} (c) \) are respectively the lower and upper \( \frac{1}{2} \alpha \) percentiles of \( \hat{\tau} \) as a function of \( c \). And because the function \( f(.) \) is strictly monotonically increasing in \( c \), by inverting the equally tailed confidence set of \( \hat{\tau} \), it is possible to obtain an equally tailed
confidence interval for $c$ such that

$$S(\hat{\tau}) = \left\{ c : f^{-1}_{ui/2\alpha}(c) \leq c \leq f^{-1}_{i/2\alpha}(c) \right\}$$

for each $\hat{\tau}$, there is a closed confidence interval of $c$, and thus an equally tailed confidence interval for $\rho$ can be constructed in a finite sample. Because the ADF test has low power, the confidence set of $c$ based on the ADF test statistics tends to have a wide confidence interval. Since the properties of the confidence interval are directly linked to the properties of the test used in the inversion, i.e., shorter confidence intervals can be related to the inversion of higher power tests, Elliott, and Stock (2001) provided methods that enable construction of confidence intervals based on inverting asymptotically efficient tests for a unit root, which is suggested by Elliott, Stock, and Rothenberg (1996). Elliott and Stock (2001) show that confidence intervals provided the initial condition (the value of $x_t$ after removing deterministics) are asymptotically narrow.

The empirically study below examines the confidence intervals based on these methods.

### 2.3.2 Zero frequency correlation

Unless long-run correlation $\delta$ is zero, the test statistic of $\beta$ will be biased in the presence of a near unit root. If economic variables are closely related and moving together, $\delta$ is likely to be non-zero. To determine whether the test statistics are biased, it is necessary to estimate $\delta$. A variety of consistent estimators is available. The estimators can be categorized into two approaches; the sum of autocovariance approach (nonparametric approach) and the autoregressive approximation approach (parametric approach). In both approaches, the method is to first estimate the simple regression models in equations (1) and (2) ignoring any serial correlation and then to construct the OLS residuals from these regressions. Denote these estimated residuals as $\hat{u}_t = [\hat{u}_1, \hat{u}_{2t}]'$ for $t = 1, ..., T$. 
The sums of autocovariances approach arises naturally out of the definition of the zero frequency and includes popular methods such as the Newey and West (1987) estimator. We define the \( j \)-the lag autocovariance matrix as \( \Gamma(j) = E[\tilde{u}_t \tilde{u}'_{t+j}] \). The asymptotic covariance between \( u_{1t} \) and \( u_{2t} \), \( \Omega \), is the infinite sum of autocovariance matrices, which is equivalent to \( 2\pi \) times spectral density at frequency zero (for a univariate case, Theorem 8.3.1 in Anderson (1971)).

\[
\Omega = \sum_{j=-\infty}^{\infty} \Gamma(j)
\]

The sum of covariances approach then would use \( \hat{\Gamma}(j) = T^{-1} \sum_{t=|j|+1}^{T} \tilde{u}_t \tilde{u}'_{t+j} \) to estimate \( \hat{\Omega} \). Two practical problems arise, to which various solutions account for the wide variety of estimators of this matrix.

First, since we have only \( T \) observations, we cannot compute the variance-covariance matrices outside of this range. Even then, when \( j \) is large compared to \( T \) we would have few observations for estimating the variance-covariance matrix and so the estimates will be very poor. This means that we have to choose a "cutoff" point and only compute the covariances up to some largest value for \( j \), denoted \( j_{\text{max}} \). Different methods for choosing \( j_{\text{max}} \), called the 'window,' result in different estimators. Estimators would then be

\[
\Omega = \sum_{j=-j_{\text{max}}}^{j_{\text{max}}} T^{-1} \sum_{t=|j|+1}^{T} \tilde{u}_t \tilde{u}'_{t+j}
\]

The second problem is that \( \Omega \) should be a positive definite matrix (as it is the long run version of a variance covariance matrix). However, estimates formed this way need not be positive definite estimate (\( \hat{\Gamma}(0) = T^{-1} \sum_{t=|j|+1}^{T} \tilde{u}_t \tilde{u}'_t \)) for the covariances (any other choice for \( j \)) we do not. If we had the true (instead of estimated) covariances they would add up to a positive definite matrix. But the estimated ones, especially poorly estimated
covariances, which become worse as \( j \) gets far from 0 as we have less observation can be so error ridden as to outweigh the variance covariance and result in a negative definite matrix. The response to this, borrowed from the frequency domain (less than one, so downweight) on the covariances so that they are smaller. The estimator is of the form

\[ \Omega = \sum_{j=-j_{\max}}^{j_{\max}} \omega_j T^{-1} \sum_{t=|j|+1}^{T} \hat{u}_t \hat{u}'_{t+j} \]

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

Autoregressive approximations take a different approach, based on the results of Berk (1974). In the case of stationary \( u_t \), the Wold representation theorem states that there is a moving average representation \( u_t = C(L)\varepsilon_t \), where \( E[\varepsilon_t \varepsilon'_t] = \Sigma \). The theoretical relationship of interest here is \( C(1)\Sigma C(1)' \), where \( C(1) = \sum_{i=0}^{\infty} C_i \). In practice it is much easier to note that for invertible moving averages this implies that we have a vector autoregression \( A(L)u_t = \varepsilon_t \), which can be inverted as \( u_t = A(L)^{-1} \varepsilon_t \) and \( \Omega = A(1)^{-1} \Sigma A(1)^{-1}' \). In this sense a two step approach to estimation can be considered. First run a vector autoregression on the estimated residuals, that is, run \( \hat{u}_t = A^*(L)\hat{u}_{t-1} + \varepsilon_t \) and save the estimates \( \hat{A}_i, \hat{\varepsilon}_t \). We then have the estimator

\[ \hat{\Omega} = \left( I - \sum_{i=1}^{k_{\max}} \hat{A}_i^* \right)^{-1} \left( T^{-1} \sum_{t=1}^{T} \hat{\varepsilon}_t \hat{\varepsilon}'_t \right) \left( I - \sum_{i=1}^{k_{\max}} \hat{A}_i^* \right)^{-1}' \]

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

We have to keep in mind that we are estimating the long-run variance for a cointegration relationship. The spectrum density at frequency zero cannot be defined if (error term of) a variable has a unit root. If two variables are not cointegrated with the specified cointegrating vector, $u_{2t}$ becomes highly persistent, which may make $\Omega$ unstable.

### 3 Alternative Procedures

We propose the application of Bonferroni’s inequality to Saikkonen’s (1992) efficient class estimator. We then compare numerical performance to other procedures.

#### 3.1 Bonferroni test

In the presence of a near unit root, t-statistics will not have the standard normal distribution. The conventional critical values under the normal assumption may lead to an invalid statistical inference. Suppose the nominal size is 10%, then, the conventional critical values for a two-sided test are ±1.645. Since the actual asymptotic distribution may be different from the standard normal distribution, the size will exceed 10%. As the value of $c$ becomes further away from zero while $\delta$ is nonzero, the size distortion will become larger. If $c \ll 0$ and $\delta < 0$, the actual t-statistics distribution will be skewed to the right, and the actual size of the right tail above 1.645 will exceed 5%.

To make a correct statistical inference, we need valid critical values. In this section, we obtain them by Bonferroni’s inequality. The famous Bonferroni’s inequality is;

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

$$C^B_{\beta}(\eta) = \bigcup_{c \in C_c(\eta_1)} C_{\beta|c}(\eta_2)$$

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

We propose applying the above idea to Saikkonen’s (1992) efficient estimator. Bonferroni’s inequality has been applied to a near unit root problem in a predictive regression setting. Cavanagh, Elliott, and Stock (1995) first suggested the application to the one-period-ahead predictive regression. Torous, Valkanov, and Yan (2004) and Campbell and Yogo (2006) applied the procedure to testing the predictability of stock returns. In testing predictability, the dividend-price ratio or the earning-price ratio is often used as a predictor variable and it is often found to have a near unit root. In a predictive regression, the regressor which is supposed to be stationary is found to have a near unit root. In a cointegration setting which we are now considering, the regressor which is supposed to have a unit root may have a near but not exact unit root.

We would like to have valid critical values so that we may test with level at most $10\%$ ($\eta = 0.1$). We first obtain a confidence interval of $\rho$. We discuss $\rho$ rather than $c$ because $\rho$ and $c$ are interchangeable in finite sample, which is always the case in empirical study. We construct $100(1 - \eta_1)$% confidence interval for $\rho$. From equation (1), we compute
the augmented Dickey-Fuller (ADF) $t$-statistic of $\rho$. Then, as suggested by Stock (1991), we obtain the corresponding confidence interval for $c$, $(\underline{c}, \overline{c})$. In a finite sample, we have an equally tailed confidence region for $\rho$, $(\underline{\rho}, \overline{\rho})$ as $(1 + \frac{\hat{\rho}}{T}, 1 + \frac{\overline{\rho}}{T})$. Now we obtain valid critical values for $t$-statistic of $\beta$ denoted as $t_\beta$. For each value of $\rho$ in the confidence interval, we obtain the distribution of $t_\beta$ based on Saikkonen (1992)'s estimator. For each distribution of $t_\beta$, we obtain $5th$ and $95th$ percentiles of $t_\beta$ denoted as $d_{t_\beta, 0.05}$ and $d_{t_\beta, 0.95}$ respectively. Then, we pick the minimum of $d_{t_\beta, 0.05}$ and the maximum of $d_{t_\beta, 0.95}$ among all the realized distributions of $t_\beta$, which are valid critical values for test with level at most $10\%$. That is,

$$(d_{l}^B (\eta_1, \eta_2), d_{u}^B (\eta_1, \eta_2)) = \left( \min_{\rho \leq \rho \leq \overline{\rho}} d_{t_\beta, 1/2 \eta_2}, \max_{\rho \leq \rho \leq \overline{\rho}} d_{t_\beta, 1-1/2 \eta_2} \right)$$

where $d_{l}^B (\eta_1, \eta_2)$ and $d_{u}^B (\eta_1, \eta_2)$ are the valid lower and upper critical values and $\eta_2 = 0.1$.

It is known that tests based on Bonferroni's inequality tend to be conservative, that is, below $10\%$ for the nominal size $10\%$, especially when $\delta$ is small. We may correct the Bonferroni test so that the size is close enough to $10\%$. Basically we refine the test by tightening the confidence interval for $\rho$ depending on $\delta$. For example, Cavanagh et al. (1995) recommend, to yield a test with size $10\%$, $\eta_1$ should be 30\% for $\delta = 0$ or 0.5 and $\eta_1$ should be 24\% for $\delta = 0.7$. Other discussion is found in Campbell and Yogo (2006).

For empirical application, we first use data to estimate $\delta$ and a confidence interval for $\rho$ as suggested by Stock (1991), since an empirical distribution of $t_\beta$ depends on $\delta$ and $\rho$. Given $\delta$, for each value of $\rho$ between $\underline{\rho}$ and $\overline{\rho}$, the null distribution $(\hat{\beta} = \beta_o)$ of $t_\beta$ is generated. Valid critical values are the minimum of the 5th percentile and the maximum of the 95th percentile of a set of distributions of $t_\beta$. 

15
3.2 Inference based on stationarity tests

Wright (2000) suggested an inference method based on the stationarity test proposed by Nyblom (1989), and Kwitkowski, Phillips, Schmidt, and Shin (1992). The stationarity test is a unit root test, testing the null hypothesis of stationarity against the alternative hypothesis of a unit root. Wright (2000) applied the test to testing for stationarity of the residuals of a cointegration equation such that

\[ \xi_t = y_t - \beta_o x_t \]

If \( \xi_t \) is stationary, the variables are cointegrated with the hypothesized cointegration coefficient, \( \beta_o \). This approach is actually testing the joint null hypothesis of \( \beta = \beta_o \) and cointegration. The stationarity tests are the locally most powerful invariant (LMPI) tests, which are free nuisance parameters under the null hypothesis. Thus, even if \( x_t \) may not have an exact unit root, the test can effectively avoid a size distortion problem.

The test statistic is a Lagrange multiplier (LM) test, denoted as \( L(\cdot) \). In a simple form without deterministic terms, the test statistic is

\[ L_1(\beta_o) = \frac{T^{-1} \left( \sum_{t=1}^{T} \xi_t \right)^2}{\hat{\sigma}_2^2} \]

where \( \xi_t = y_t - \beta_o x_t \) and \( \hat{\sigma}_2^2 \) is a consistent estimate of zero frequency variance of \( u_{2t} \). Under the null hypothesis, \( L_1(\beta_o) \) has the limiting chi-square distribution with degree of freedom one, and is free of nuisance parameters. Under the alternative of \( \beta \) such that \( \beta = \beta_o + g/T \), \( L_1(\beta) \) is a function of a diffusion process and depends on the nuisance parameters.

---

\(^2\)In practice, Wright (2000) recommends to choose the minimum one between \( \kappa \hat{\sigma}_2^2(\hat{\beta}) \) and \( \hat{\sigma}_2^2(\beta_o) \) as \( \hat{\sigma}_2^2 \), that is, \( \hat{\sigma}_2^2 = \min \left( \kappa \hat{\sigma}_2^2(\hat{\beta}), \hat{\sigma}_2^2(\beta_o) \right) \), where \( \kappa \) is some constant, and typically is 15, and \( \hat{\sigma}_2^2(\cdot) \) is the estimated zero frequency variance with \( \hat{\beta} \) or \( \beta_o \) imposed.
parameters, $c$ and $\delta$.

4 Finite Sample Performance

This section evaluates finite sample performance. Size and power for testing $\beta = \beta_o$ are reported. The sample sizes are $T=100, 500$. Finite sample performance of three procedures is presented. They are denoted as "Saikkonen" which estimate a cointegration coefficient by using Saikkonen’s (1992) method without any treatment for a near unit root, "Bonferroni-Saikkonen," which applies Bonferroni’s inequality to Saikkonen (1992), and "Wright," which applies the stationarity test based method suggested by Wright (2000).

Table 1 presents sizes for these procedures as a function of true values of $c$ and $\delta$ with nominal size 10%. The size is $\Pr[\text{reject } H_0: \beta = \beta_o | H_0 \text{ is true}]$. Saikkonen’s size is computed as $\Pr[W_\beta > \chi^2_{0.9} (1) | H_0 \text{ is true}]$, where $W_\beta$ is the Wald statistic and $\chi^2_{0.9} (1)$ is the 90th percentile of a chi-square distribution with degree of freedom one. Wright’s size is computed as $\Pr[L(\beta) > \chi^2_{0.9} (1) | H_0 \text{ is true}]$. Bonferroni-Saikkonen’s size is $\Pr[W_\beta > d_B^{0.9} (0.9) | H_0 \text{ is true}]$, where $d_B^{0.9} (0.9)$ is the 90th percentile of the actual distribution of the Wald statistic in the presence of a near unit root.

"Saikkonen" tells us how large the actual size can be when the possibility of a near unit root is ignored. When $T=100$ and $\delta = 0.7$, the actual size is as large as 44% against the nominal size of 10%. Typical estimation methods may totally lack robustness when the largest root deviates from unity. "Wright" handles the size well. Because this procedure is based on the LMPI test, it is stable for any value of $c$ and $\delta$ when the null hypothesis holds. "Bonferroni-Saikkonen" also controls size well, although the size tends to be conservative. As the sample size becomes large, the actual size becomes relatively stable around 10%.

Next we evaluate local alternative power. Figures 1a and 1b present local alternative power for 10% level test of $\beta = \beta_o$ against alternative of $\beta = \beta_o + g/T$. Power is evaluated
for the sample sizes of $T=100$, 500, local to unity parameter $c$ of 0, -5, and -10, and zero frequency correlation $\delta$ of 0.3 and 0.7. Figure 1a presents power for $T=100$. Once the largest root deviates from unity, both procedures lose power. Especially "Wright" loses its power substantially. This relatively large loss of power is attributable to the fact the null hypothesis of the LMPI test is existence of stationarity. Once $c$ becomes below zero and variables become stationary, the test tends to fail to reject alternatives of $\beta$.

Figure 1b presents local alternative power for $T=500$. As the sample size increases, power is improved for both "Wright" and "Bonferroni-Saikkonen".

5 Empirical Study

In this section we apply the Bonferroni test to the present value model in the stock market. If the model holds, the logs of stock price and dividend should move together. However, it is empirically known that stock price is far more volatile than dividend. We examine whether the presence of a near unit root leads the test statistics to be biased.

5.1 Present value model in the stock market

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

$$p_t - d_t = \frac{\kappa_0}{1 - \kappa_1} + E \left[ \sum_{i=0}^{\infty} \kappa_1^i (g_{t+1+i} - r_{t+1+i}) \right]$$
where \( p_t \) and \( d_t \) are the logs of stock price and dividends respectively, \( g_t \) is the continuous dividend growth rate, and \( r_t \) is the log of total return.\(^3\) The stock price and the dividend are often assumed to be \( I(1) \) (for example, Mankiw, Romer, and Shapiro (1985, 1991) or Cochrane (1992, 1994), and unit root tests fail to reject the null hypothesis of unit root in empirical studies (Campbell and Shiller (1988), Timmermann (1995)). Thus, the logs of stock price and dividend are also assumed to have a unit root. Then, the continuous growth rate of dividend and stock return, i.e. the first difference of the log dividend and stock price should be stationary. Consequently, if the present value model holds, the logs of stock price and dividend are cointegrated with the cointegrating vector, \((1, -\beta_o) = (1, -1)\), and the cointegration equation is

\[
p_t - \beta_o d_t = u_{2t}
\]

or

\[
p_t = \beta_o d_t + u_{2t}
\]

where \( \beta_o = 1 \) under the null hypothesis.\(^4\) Let us denote the estimated \( \beta \) by Saikkonen (1992) or other efficient methods as \( \hat{\beta} \). Then, in a typical empirical study, \( \hat{\beta} \) is significantly larger than one, that is, the log of stock price is far more volatile than dividend.

### 5.2 Previous empirical findings

Testing the present value model with logs of stock price and dividend was first proposed by Campbell and Shiller (1988). A typical empirical finding is against the present value

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

\[^4\] Levels of stock price and dividend are also tested in a cointegration framework (Campbell and Shiller (1987), for example). In this case, the cointegrating vector is a function of discount factor, the value of which is not specified by the model. Since our concern is the hypothesis testing on the cointegrating vector, cointegration between stock price and dividend in levels is beyond scope of this paper.
model. The (log of) stock price is far too volatile relative to dividend for the present value model to hold, which is the so called "excess volatility puzzle". Barsky and Delong (1993) estimated $\beta, \tilde{\beta} = 1.61$, and the associated $t$–statistic was positive and significant. Gonzalo, Lee, and Yang (2007) found that, by the Johansen test, the estimated reciprocal of $\beta$ is around 0.6, and the estimated cointegrating vector is significantly different from $(1, -1)$. Barsky and Delong (1993) and Bansal and Lundblad (2002) attributed this high volatility of stock prices to nonstationarity or near nonstationarity of dividend growth rate, $g$. On the other hand, Timmermann (1995) showed by Monte Carlo experiment that the logs of stock price and dividend fail to be cointegrated because the rate of return, $r$, is highly persistent and not because $r$ is highly volatile.

5.3 Our empirical study

Table 2 presents our empirical study. The original data are annual stock returns with and without dividends from CRSP. From the original data set, I constructed the log real stock price and the log real dividends. A detailed description about data is found in appendix. The entire sample is from 1926 to 2007. I also examined three subsamples; 1926-1951, 1952-2007, and 1974-2007.

We estimate $\beta$ by Saikkonen’s (1992) estimator. Pesavento (2004) proved that power of test is improved by VAR approach such as Saikkonen’s (1992) when data are highly correlated. For the present value model, stock price and dividend are supposed to be highly correlated. Saikkonen’s (1992) method takes the autoregressive error correction representation and estimates $\beta$ from the (approximately) spectral density estimator of the data at the zero frequency. Thus, it takes the following form:

$$\Delta Y_t = \Psi Y_{t-1} + K(L)\Delta Y_{t-1} + \epsilon_t$$  (8)
where $Y_t = [\Delta x_t, \Delta y_t]'$ and $\epsilon_t$ is an error term. Then define $\hat{S} = \hat{\Psi}'\hat{\Sigma}^{-1} \hat{\Psi}$, where $\hat{\Sigma}$ is the estimated long run variance covariance matrix for $\hat{\epsilon}_t = \Delta Y_t - \hat{\Psi}Y_{t-1} + \hat{K}(L)\Delta Y_{t-1}$. Then, $\hat{\beta} = -\hat{S}_{22}^{-1}\hat{S}_{21}$. And theoretically, $\hat{S}_{22}^{-1}$ is an orthogonal part of the variance in the cointegration equation error, i.e., $\hat{\Omega}_{2,1} = \hat{S}_{22}^{-1}$, where $\Omega_{2,1} = \Omega_{22} - \Omega_{21}\Omega_{11}\Omega_{12}$ for $\Omega$ defined in section 2. For our estimation, however, we computed $\Omega_{2,1}$ (and $\delta$) from $\hat{\Omega}$ imposing the null hypothesis;

$$u_{1t}^o = x_t - \rho_o x_{t-1}$$
$$u_{2t}^o = y_t - \beta_o x_t$$

where $\rho_o = 1$, $\beta_o = 1$. And we estimated $\Omega$ for $u_t^o = [u_{1t}^o, u_{2t}^o]'$ by Berk’s (1975) VAR approach and Newey and West (1987) sum of autocovariances approach and obtain $\hat{\Omega}_{2,1}$ and $\hat{\delta}$. We compute $t-$statistics with $\hat{\Omega}_{2,1}$. In Table 2, $t-$statistics and standard errors for $\beta$ are from $\hat{\Omega}$ by Berk’s (1974). We imposed the null hypothesis because power of test will be improved by imposing the hypothetical coefficients (Elliott, Jansson and Pesavento (2005)).

Table 2 presents 95 % confidence intervals for the largest autoregressive root in $p_t$ and $d_t$. All the confidence intervals contain unity. The intervals are very similar between $p_t$ and $d_t$ except for the subsample 1926-1951. For the 1926-1951 subsample, the confidence interval for $p_t$ is relatively wide. It is partially because the sample size is small. Since the interval depends on the sample size, it tends to be wise in a small sample. But for the same sample period, the interval for $d_t$ is quite tight around unity. Thus, the largest root in $p_t$ is likely to be further away from unity than any other sample, and $p_t$ and $d_t$ may behave differently during this period.

$\hat{\delta}$’s are all positive. Especially $\hat{\delta}$’s by VAR approximation are close to one. We expect the test statistics are biased in the positive direction if the largest root is not exactly
For the entire sample and subsample 1952-2007, tests are positive and significant under the standard normal assumption. On the other hand, for the subsamples 1926-1951 and 1974-2007, tests are negative contrary to our prediction. For 1926-1951, \( t \)-statistic is marginally significant with 10% significance level, and it is negative. There should be other factors than the presence of a near unit root. The economy experienced the Great Depression during this period, which caused a huge drop in stock price. Some authors specify some economic data have a broken trend around the Great Depression rather than a unit root (e.g. Perron (1989)). Though we do not test for a break here, the confidence interval tells us that the behavior of stock price is quite different from the latter period. For another subsample 1974-2007, \( t \)-statistic is significant with 2.5% significant level. \( t \)-statistics are unstable for this period. Though \( t \)-statistic presented here is negative, the one with Newey and West (1987) estimator is positive, while all the \( t \)-statistics with Berk’s or Newey and West have the same sign for any other sample. It is because the cointegration error \( u_{2t} \) becomes highly persistent in the latter period. This result supports the explanation in previous empirical studies that the dividend growth rate or stock return is highly persistent.

5.4 Making a valid inference in the presence of a near unit root

We obtain valid critical values for each sample using the Bonferroni method mentioned in the previous section. Distributions of \( t \)-statistics of \( \beta \) are produced by 1000 replications for \( T=100 \), confidence intervals of \( \rho \), and \( \delta \). While \( \delta' \)'s are estimated by the two different methods, I used \( \delta' \)'s estimated with VAR representation in Berk (1974), because autocorrelation indicates \( u_{2t} \) is well represented as an autoregressive process. In Table 2 we present 5% and 2.5% critical values on the right hand tail (i.e., they are the 95 percentile and 97.5 percentile of the replicated distributions). Since \( \delta > 0 \), the actual distributions have size
distortions in the positive direction. Hence, critical values on the right tail are presented. They are quite different from the conventional critical values of 1.645 and 1.964.

For the entire sample 1927-2007, our $t$-statistic marginally exceeds the 97.5 percentile of the replicated distribution, while it is well above 2.5% significance level under standard normal assumption. Given a confidence interval for $\rho$ and $\delta$, the actual null distribution shifts in the positive direction. Thus, the presence of a near unit root may be at least one of factors for the rejection under the standard normal assumption, though it may not fully explain the rejection. If we look at each subsample, however, test is rejected with a 5% significance level. In the short run, due to a large economic incident such as the Great Depression or change in the behavior of dividend or stock price, one for one relationship between stock price and dividend may not always hold.

6 Conclusion

In this paper, we have examined the "excess volatility puzzle" of the present value model in the stock market. If the present value model holds, (logs) of stock price and dividend move together over time. A typical empirical study rejects this hypothesis; stock price is far too volatile. We examined whether the "excess volatility" is caused by a near unit root in dividend data. It seems to be likely that dividend does not have an exact unit root, though it is highly persistent in finite sample, and the innovation of dividend is correlated with the error of the cointegration equation. Thus, the $t$-test has nonstandard distribution and conventional critical value leads us to make an invalid inference.

We have proposed a valid test in the presence of near unit root. We have applied the test to empirical data. When we use the valid critical values, the test is marginally rejected in the long run data. Hence, the presence of a near unit root is one of causes for the rejection of test under the asymptotic normality assumption. But other factors also
cause the rejection of the present value model. The behavior of data changes over time. For a subsample 1926-1951, the largest root in stock price seems well below unity. For a later subsample 1974-2007, the hypothetical linear combination of stock price and dividend (i.e., cointegration error) is highly persistent. For further research, we need to investigate whether any structural change in the relationship between data, or dividend growth rate becomes temporarily or permanently highly persistent.

A Appendix: The Data Set

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

References


<table>
<thead>
<tr>
<th>$\delta$</th>
<th>$c$</th>
<th>Saikkonen</th>
<th>Wright</th>
<th>Bonferroni-Saikkonen</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>0.120</td>
<td>0.164</td>
<td>0.088</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.052</td>
<td>0.158</td>
<td>0.062</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>0.036</td>
<td>0.152</td>
<td>0.104</td>
</tr>
<tr>
<td>T=100</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.126</td>
<td>0.164</td>
<td>0.056</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.164</td>
<td>0.156</td>
<td>0.078</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>0.200</td>
<td>0.168</td>
<td>0.116</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.128</td>
<td>0.148</td>
<td>0.074</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.274</td>
<td>0.136</td>
<td>0.068</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>0.440</td>
<td>0.152</td>
<td>0.124</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.138</td>
<td>0.168</td>
<td>0.102</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.492</td>
<td>0.144</td>
<td>0.090</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>0.752</td>
<td>0.132</td>
<td>0.146</td>
</tr>
<tr>
<td>T=150</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.086</td>
<td>0.158</td>
<td>0.100</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.106</td>
<td>0.146</td>
<td>0.074</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>0.040</td>
<td>0.144</td>
<td>0.082</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.102</td>
<td>0.146</td>
<td>0.106</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.106</td>
<td>0.144</td>
<td>0.108</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>0.028</td>
<td>0.184</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.116</td>
<td>0.132</td>
<td>0.100</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.066</td>
<td>0.148</td>
<td>0.114</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>0.038</td>
<td>0.154</td>
<td>0.112</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.118</td>
<td>0.132</td>
<td>0.096</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.222</td>
<td>0.138</td>
<td>0.084</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>0.404</td>
<td>0.146</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.110</td>
<td>0.164</td>
<td>0.082</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.420</td>
<td>0.144</td>
<td>0.082</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>0.704</td>
<td>0.110</td>
<td>0.102</td>
</tr>
<tr>
<td>T=300</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.116</td>
<td>0.146</td>
<td>0.112</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.070</td>
<td>0.146</td>
<td>0.072</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>0.028</td>
<td>0.184</td>
<td>0.124</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.122</td>
<td>0.160</td>
<td>0.086</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.248</td>
<td>0.144</td>
<td>0.120</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>0.430</td>
<td>0.154</td>
<td>0.076</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.112</td>
<td>0.160</td>
<td>0.096</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.176</td>
<td>0.142</td>
<td>0.104</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>0.144</td>
<td>0.142</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.078</td>
<td>0.142</td>
<td>0.092</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.446</td>
<td>0.162</td>
<td>0.094</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>0.724</td>
<td>0.142</td>
<td>0.104</td>
</tr>
<tr>
<td>T=500</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.116</td>
<td>0.146</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.070</td>
<td>0.146</td>
<td>0.124</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>0.028</td>
<td>0.184</td>
<td>0.124</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.122</td>
<td>0.160</td>
<td>0.096</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.248</td>
<td>0.144</td>
<td>0.100</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>0.430</td>
<td>0.154</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.112</td>
<td>0.160</td>
<td>0.092</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.176</td>
<td>0.142</td>
<td>0.094</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>0.144</td>
<td>0.142</td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>95% CI for ρ</td>
<td>Correlation</td>
<td>β</td>
<td>t-stat</td>
</tr>
<tr>
<td>--------</td>
<td>---------------</td>
<td>-------------</td>
<td>-----</td>
<td>--------</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AR</td>
<td>NW</td>
<td>(s.e.)</td>
</tr>
<tr>
<td>1926-2007</td>
<td>Price (0.982 1.059)</td>
<td>0.995</td>
<td>0.408</td>
<td>1.912 (0.211)</td>
</tr>
<tr>
<td>1926-1951</td>
<td>Price (0.738 1.167)</td>
<td>0.968</td>
<td>0.389</td>
<td>0.433 (0.343)</td>
</tr>
<tr>
<td>1952-2007</td>
<td>Price (0.935 1.084)</td>
<td>0.981</td>
<td>0.465</td>
<td>4.562 (0.382)</td>
</tr>
<tr>
<td>1974-2007</td>
<td>Price (0.888 1.137)</td>
<td>0.957</td>
<td>0.506</td>
<td>-1.254 (0.151)</td>
</tr>
</tbody>
</table>

Table 2: Present Value Model in the Stock Market (Empirical Study)

Data: CRSP (annual)

The 95% confidence intervals for the largest autoregressive root in the log of real stock price (Price) and real dividend (Dividend) are constructed following Stock (1991).

β is estimated by Saikkonen (1992)

Bonferroni critical values: upper 10% is 90 percentile and upper 5% is 95 percentile of simulated Wald-statistic distributions.
Figure 1a: Local Alternative Power of 10% level testing $\beta = \beta_0 + g/T$, where $\beta_0 = 1$, $T=100$, $c = 0$, -5, -10, and $\delta = 0.3, 0.7$. 
Figure 1b: Local Alternative Power of 10 % level testing $\beta = \beta_0 + g/T$, where $\beta_0=1$, $T=500$, $c=0, -5, -10$, and $\delta=0.3, 0.7$. 