Title

Equity Premium and Dividend Yield regressions: A lot of noise, little information, confusing results

Author
Valkanov, Rossen

Publication Date
1999-09-08

Equity Premium and Dividend Yield regressions: A lot of noise, little information, confusing results[1]

Rossen Valkanov[2]

08/08/1999

[1]I thank Ivo Welch for helpful discussions, and especially Yacine Aït-Sahalia and Mark Watson for detailed comments. Amit Goyal kindly provided the CRSP dataset. All remaining errors are my own.

[2]Contact: rossen.valkanov@anderson.ucla.edu or Anderson Graduate School of Management at UCLA, 110 Westwood Plaza, Box 951481, Los Angeles, CA 90095-1481, tel: (310) 825-7246.

Abstract

Suppose that the equity premium is forecasted by dividend yields. Even if such a relationship does exist, there is so much noise in the equity premium that estimation, inference and forecasting cannot be carried out using the faint signal coming from the dividend yields. For analyzing equity/dividend data, it is useful to quantify the signal in a given sample. We define an index of signal strength or information accumulation, by renormalizing the signal to noise ratio. The novelty in our parameterization is that the index of information accumulation explicitly influences rates of convergence and can even lead to inconsistent estimation, inconsistent testing, unreliable R^2's, and no out of sample forecasting power. Indeed, we prove that if the signal to noise ratio is close to zero, forecasts from the existing model will not do better than the simple unconditional mean. Thus, it is not surprising that dividend yield forecasts of the equity premium cannot outperform its mean. The analytic framework is general enough to capture most previous econometric findings related to the equity/dividend relationship.

1 Introduction

Empirical finance is full of puzzles. Take a plausible theory, bring it to the data, and you have yourself a puzzle. A case in point are the empirical tests of expected returns. The possibility that returns of risky assets could be forecasted, be it only partially, by other observable variables seemed to have been unthinkable until the early 1980's. It was commonly maintained that the efficient markets hypothesis implied unforecastable returns. In a series of articles, the most prominent of which are Fama and French (1988) and Campbell and Shiller (1988), it is argued that dividend yields should and do forecast expected returns with some success. It is fair to say that the pendulum has swung in the opposite direction, and it is now taken almost as a feature of the data that dividend yields have some power to predict stock returns. For a review of the literature, see Campbell, Lo and MacKinlay (1997). However, recent work by Nelson and Kim (1993) and Goetzmann and Jorion (1993) has put the empirical findings of these latest studies into question.

In the present paper, we argue that even if there exists a forecasting relationship between the equity premium and dividend yield, in cannot be exploited using simple regression techniques. Intuitively, even if the theoretical model leads the dividend yields to predict the equity premium, forecasting a very noisy variable (the equity premium) with a variable that, although persistent, has a very small variance (the dividend yield) will produce insignificant results. For concreteness, let's assume that we have the relationship

$$Y_{t+1} = \mu_y + \beta X_t + \varepsilon_{t+1} \tag{1}$$

where Y_{t+1} is the equity premium (risky return-Treasury bill return) and X_t is the dividend yield. Previous papers have tested this relationship under the null hypothesis of no relationship between the two variables. We ask the following question: If there is a relationship between the equity premium and dividend yields ($\beta \neq 0$), what estimates would we obtain, given the noisy data? The main goal of this paper is to explore the small sample properties of the OLS estimator of β, the resulting in-sample fit and out-of-sample forecasts, when the signal coming for X_t is very weak compared to the noise from ε_t, and X_t is a persistent process. The question above was partially motivated by figure 1 and table 1 below.

It is well-known that the equity premium has the time series properties of a very volatile, but stationary process. It is also known that the dividend yield is a persistent process, in the sense that its highest autoregressive root is close to, or at unity. However, as we can see from figure 1 above, the ratio $\frac{V(X_t)}{V(\varepsilon_t)}$, known as the "signal to noise ratio", is close to zero throughout the sample period. This empirical fact can be found in virtually any paper on the topic (for example, Fama and French (1988), Goyal and Welch (1999), Stambaugh (1999)) and is discussed further in the following sections.

We argue that if the signal to noise ratio (S/N) is very close to zero, asymptotic approximations can help up understand the small sample properties of $\hat{\beta}$. In asymptotic exercises, a small S/N is not considered to be a problem, because, given a sufficiently large dataset, we will eventually gather enough signal to get the OLS estimator $\hat{\beta}$ to be arbitrarily close to the true value of the parameter. In other words, the OLS estimator is consistent.

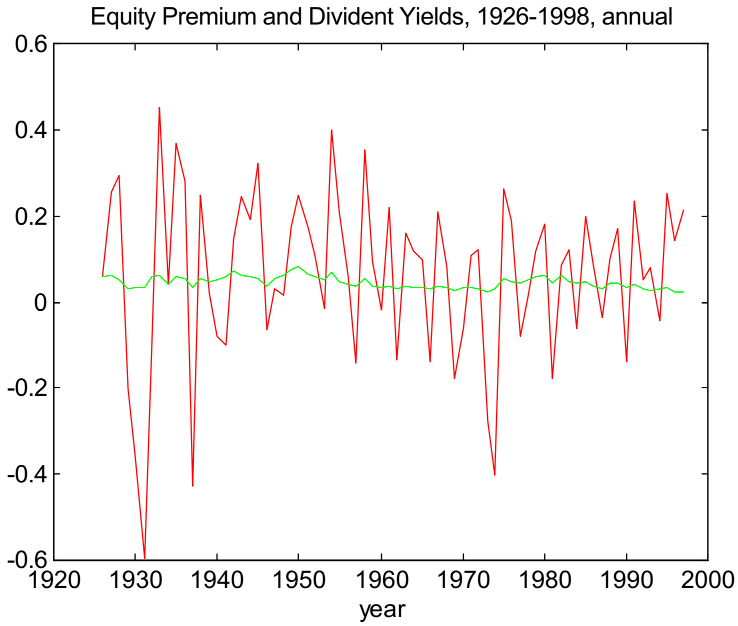

Figure 1: Equity Premium and Divident Yield, 1926-1998, annual data. The Equity Premium is many times more volatile than the Dividend Yield.

Also, no matter how dispersed the distribution is, it should be centered around the true value of the parameter, and after appropriate rescaling, inference will not be affected, asymptotically. However, we show, first through Monte Carlo simulations and then analytically, that in small samples, a S/N close to zero will affect estimation, inference, and forecasting. To understand the results, recall that for a precise estimate, we need not only a big number of observations, T, but also a large S/N. We can easily think of cases when T increases without a corresponding increase in signal. For example, going from yearly to monthly or from monthly to daily frequency, for a fixed time span, can conceivably add more noise than signal, and will not contribute to a more precise estimate. Indeed, going from yearly to monthly CRSP data does not seem to add much signal in the forecasting equation (1). A similar argument was made by Perron (1989), although in a different context and with a univariate model.

In practice, we have a fixed S/N and a fixed T. Therefore, we can think of a sequence of models defined by their S/N ratios: for each magnitude of S/N, we have a different model, given T. The higher is the S/N, the more information there is in the sample. We want to have a similar sequence of models, as T increases, in order to use asymptotic approximations to analyze the small sample properties of $\hat{\beta}$. Therefore, we create a rescaled sequence of models, where S/N is defined for different powers (not necessarily integers) of T. The focus will then shift from the magnitude of S/N to the scaling power of T, call it α.

Intuitively, α can be viewed as an index of the signal strength coming from X_t in the sample. For $\alpha = 0$, the signal is very powerful. This is the case when X_t is I(1). For $\alpha \in (0, 1/2)$, the signal is still stronger than the noise and $\hat{\beta}$ converges at rates higher than $T^{1=2}$. In the borderline case $\alpha = 1/2$, the signal to noise ratio is a constant, and this corresponds to the "usual", stationary case, where $\hat{\beta}$ converges at rate $T^{1=2}$. For $\alpha > 1/2$, $\hat{\beta}$ will converge at a rate slower than 1/2. For $\alpha > 1$, the estimator and the customary t-statistic will not be consistent. Interestingly, in the case $\alpha > 1/2$, the coefficient of determination R^2 will converge to zero in probability, even when there is a relationship between the equity premium and the dividend yields. Moreover, forecasts produced with the correctly specified model will not do better (in a mean square error sense) than, say, the unconditional mean. This last implication is borne out by the data, as demonstrated by Goyal and Welch (1999), who find that forecasts using the dividend yield model cannot outperform those from the unconditional mean.

Here is a brief summary of the main contributions of this article. For the case of equity/dividend equations, where a persistent variable with very small variance is thought to forecast a very noisy variable, it is useful to quantify the S/N in a given sample. We renormalize the S/N ratio by a power of the sample size, allowing us to find an index of the signal to noise ratio, that will not change asymptotically, and that will provide a measure of the signal in the sample. The novelty in our parameterization is that the rate of information accumulation explicitly influences the rate of convergence and can even lead to inconsistent estimation, inconsistent testing, unreliable R^2's, and no out of sample forecasting power.

The paper is structured as follows. Section 2 presents the model and provides a series of Monte-Carlo simulations, demonstrating that if the signal to noise ratio is close to zero, as is the case in the equity premium/dividend yield equation, estimation, inference and forecasting will be problematic, even for samples of fairly large size. The analytic results introducing the index α and deriving the asymptotic distributions of $\hat{\beta}$ and its t-statistic are presented in section 3. In section 4, we provide a brief overview of the econometric literature on this topic to help clarify the contribution of the paper. Section 5 contains results on the in-sample fit and out-of-sample forecasting with small signal to noise ratio. Section 6 concludes.

2 Model

Here is the essence of my point. Suppose we have

$$Y_{t+1} = \mu_y + \beta X_t + \varepsilon_{t+1} \tag{2}$$

$$X_{t+1} = \mu_x + \phi X_t + u_{t+1} \tag{3}$$

where $\phi = 1$, $\mu_x = 0$, and ε_t and u_t are random variables with mean zero, and variances σ_ε and σ_u, respectively. We are interested in the case when the ratio $\sigma_u/\sigma_\varepsilon$ is close to zero. If we rescale u_t as

$$\begin{bmatrix} \varepsilon_t \\ u_t \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & \tau \end{bmatrix} \begin{bmatrix} \varepsilon_t \\ v_t \end{bmatrix} = \Upsilon w_t \tag{4}$$

where w_t is a martingale difference sequence with $E(w_t w_t' | w_{t-1}, w_{t-2}, ...) = [\sigma_\varepsilon \, \sigma_{\varepsilon v}; \, \sigma_{\varepsilon v} \, \sigma_v] = \Sigma$ and finite fourth moments, then a small $\sigma_u/\sigma_\varepsilon$ ratio implies a small τ^2. The above setup, where X_t is the dividend yield and Y_t the equity premium (or the expected risky return), has been analyzed by Mankiw and Shapiro (1986), Stambaugh (1986, 1999), Nelson and Kim (1993), Goetzmann and Jorion (1993), Cavanagh et al. (1994), and Goyal and Welch (1999), among others. The first four papers focus on the small sample biases in $\hat{\beta}$, coming from the predetermined and persistent regressor. Cavanagh et al. (1994) look at the same problem with different tools, to reach similar conclusions. However, the small signal to noise ratio problem has not been treated by any papers related to this literature.

Model (2-4) was chosen as most closely describing the characteristics of the US stock market data. Table 1 presents summary statistics from CRSP annual and monthly time-series for the period 1926-1998, and several sub-periods. Two points are worth making. First, the dividend yield is a persistent process with a highest autoregressive root very close to unity. An Augmented Dickey Fuller test mostly cannot reject the null of a unit root in yearly and monthly data. Second, the variance of the equity premium is orders of magnitude greater than the variance of the dividend yield. The entry $V(DP)/V(EP)$ shows the signal to noise ratio, and τ^2 is an estimate of the parameter defined in (4). Both measures convey the same message: the signal coming from the dividend yield is many orders of magnitude weaker than the noise in the equity premium regression, despite the fact that the dividend yield is a more persistent variable. More interestingly, increasing the number of observations (from yearly to monthly data) seems to decrease, and not increase, the signal coming from the data. These are precisely the features of the data we want to capture with the above system. Figure 1 provides a picture of the same facts.

We assume that $\beta = \beta_0 \neq 0$, or that the equity premium is forecastable by the dividend yield, with cointegrating parameter β_0. For simplicity, we also assume that $\phi = 1$ and known. If ϕ is unknown and close to, but not exactly at unity, we might model it as local to unity, or $\phi = 1 + \frac{c}{T}$ as in Cavanagh, et al (1994). Introducing this further generality into $(2-3)$ will not affect our main point. The only difference will be that the distributions of interest will depend on an additional nuisance parameter, c. We relegate this issue to future papers.

2.1 Monte Carlo Results

2.1.1 No correlation between ε_t and v_t

To motivate the new results, we start off by performing the following Monte Carlo experiment. System $(2-3)$ is simulated for various values of T and τ, namely $T = 75, 200, 500$ and $\tau = 1, 0.1, 0.01, ..., 1 * 10^{-5}$. The case $T = 75$ corresponds to a sample with annual data, $T = 200$ is for quarterly data, and $T = 500$ is for a typical sample of monthly data. We set $\beta = 1, \sigma^2_{\varepsilon} = \sigma^2_{v} = 1$, where ε_t and ν_t are iid normal variates and $\sigma_{\varepsilon,v} = 0$. The system for each specification of (τ, T) is simulated $10,000$ times. At every simulation, we regress Y_{t+1} on X_t, X_t on X_{t-1}; and Y_t on Y_{t-1}; producing estimates $\hat{\beta}$, $\hat{\phi}$, and $\hat{\phi}_2$, respectively. The estimated means and variances of the three OLS estimates are shown in table 2 below. The results from these simulations are quite interesting, and it is worth discussing them.

Looking at table 2a, the entries in the first row, $\tau = 1$, (i.e $\sigma_u^2 = \sigma_\varepsilon^2 = 1$) show that the distribution of the estimate is centered exactly on the true value of β. However, as τ decreases, the estimates worsen considerably, even for samples as large as $T = 500$. Note that we have restricted ε_t and u_t to be independent. Therefore, we cannot have a small sample bias from predetermined lagged endogenous variable, as discussed in Stambaugh (1986, 1999).

But if this is not small sample bias, how can we account for such a poor performance of the estimator even in reasonably large samples? Recall that the asymptotic theory predicts that the distribution will be centered around the true parameter $\beta = 1$, but this is not clear from the table. The simulations suggest that as τ decreases, the estimates diverge farther and farther from the true value. In fact, we observe a simulation error, magnified by the small signal to noise ratio, $1/\tau$. Hence, should we be satisfied with the asymptotic approximations above when the signal to noise ratio is small? More importantly, can we trust the estimates from the returns/dividend equations? Can we conduct inference in the usual way? What about out of sample forecasts? Those are all questions that will be addressed below.

The table 2b shows the estimates of the autoregressive root in X_t. As expected, the parameter ϕ is estimated with its usual downward bias, but as T increases, the bias disappears. The estimate of ϕ is unaffected by the small signal to noise ratio, since its distribution is invariant to τ^2. Since X_t has a unit root, and the relationship (2) holds, then Y_t must also have a unit root.

Table 2c shows the estimates of the autoregressive root of Y_t. Surprisingly, for small τ's, the estimates are not anywhere close to 1. To understand this table, notice that if τ is small, we can write $(1 - L)\, Y_{t+1} = (1 + \theta L)\, r_{t+1}$ with θ very close to -1. In other words, Y_{t+1} is very close to being white noise. It is known that the battery of unit root tests will have a size close to 1 under such circumstances (Perron (1988), Schwert (1989), and Pantula (1991)). Hence, it is not surprising that the null of unit root in Y_t is rejected when σ_u^2 is small.

As shown in table 2d, inference is also problematic when τ is small, given the sample sizes of interest. The entries in the table show the mean of the distribution of the t-statistic for $H_0 : \beta = 1$ versus $H_a : \beta = 0$. For τ small, the distributions under the null and under the alternative are both centered at zero, implying that the test has almost no power even for samples of reasonable size (power equal to size). When the signal to noise ratio is small in a given sample, the R^2 must also be small, by definition, but as T increases, R^2 must converge to 1 since we assume a relationship between Y_t and X_t. The $R^{2\prime}s$ obtained in the simulations are shown in table 2e. Lastly, the out-of-sample forecast using the correct model is compared to a forecast, from the unconditional mean in table 2f. For τ big, the forecast from the model outperforms the mean in the MSE sense. As τ decreases, the two forecasts produce similar results, even for reasonably large sample sizes.

2.1.2 Correlation between ε_t and v_t

If ε_t and v_t are contemporaneously correlated, the OLS estimator of the slope coefficient will be to biased, as discussed in Stambaugh (1986, 1999) and Cavanagh et al (1994). The bias will depend on the magnitude of the fraction $\frac{cov(\varepsilon_t, u_t)}{cov(u_t)} = \frac{\sigma_{\varepsilon,v}}{\sigma_v}\frac{1}{\tau}$. Recall that $\sigma_v = 1$ and now we let $\sigma_{\varepsilon,v} = 0.5$. A decrease in τ leads to an increase of the bias. To illustrate this case, we have run the same simulations as before, except for $\sigma_{\varepsilon,v} = 0.5$. The results can be seen in table 3. The bias magnified by the small signal to noise ratio produces highly inaccurate estimates, in table 3a. Moreover, the t-statistic seems to be centered at around -1.45 under the null, and its distribution under the null and under the alternative are very similar. The rest of the results are similar to the case $\sigma_{\varepsilon,v} = 0$.

3 Asymptotic approximations when τ is close to zero

3.1 No correlation in residuals

For clarity of exposition, we first let $\sigma_{\varepsilon,v} = 0$ as was done in the Monte Carlo simulations. In this instance, ε_t is uncorrelated with X_t, and $\Sigma^{1/2}$ is diagonal. This assumption is not realistic in the equity-dividend case, but we are imposing it for now to emphasize that the conclusions do not depend on it. It is later relaxed to show that our results are even more dramatic. Under the above specification, $1/\sqrt{T}\sum_{j=1}^{[sT]} w_j \Rightarrow \Upsilon\Sigma^{1/2}W(s)$, $1/\sqrt{T}\sum_{j=1}^{[sT]}(w_j - \overline{w}) \Rightarrow \Upsilon\Sigma^{1/2}\underline{W}(s)$, where $W(s) = [W_1(s)\ W_2(s)]'$ is bivariate standard Weiner process on $D[0,1] \times D[0,1]$, $\underline{W}(s) = W(s) - \int_0^1 W(s)ds$, $t = [sT]$, and $\Rightarrow$ denotes convergence in distribution. To capture the small variance of u_t (relative to ε_t) we can write

$$\tau = \frac{1}{T^{\alpha}} \tag{5}$$

where, $\alpha \geq 0$. We will use this parameterization to explain the results in the simulations above. For T fixed, and τ small, one can always find an α that satisfies the relationship. The interpretation of (5) is that, although X_t has a unit root, its variance is still much smaller than the variance of ε_t, or

$$\begin{aligned} \frac{V(X_t)}{V(\varepsilon_t)} &= \frac{\sigma_u^2 t}{\sigma_\varepsilon^2} = \frac{t}{T^{2\alpha}} = \frac{t}{T} T^{(1-2\alpha)} \\ &= sT^{(1-2\alpha)} \end{aligned}$$

where $s = t/T$.

The parameter α can be interpreted as an index of the signal strength coming from X_t relative to the noise ε_t. For $\alpha = 0$, $\tau = 1$, or $\sigma_\varepsilon^2 = \sigma_u^2$. For $\alpha < 1/2$, $\frac{V(X_t)}{V(\varepsilon_t)}$ diverges for $T \to \infty$, as expected. However, if $\alpha > 1/2$, the signal to noise ratio vanishes, $\frac{V(X_t)}{V(\varepsilon_t)} \to 0$, as $T \to \infty$. If $\alpha = 1/2$, then $\frac{V(X_t)}{V(\varepsilon_t)} = s$. This is the borderline case in which the variance of the increments of X_t is so small that it offsets the signal coming from the higher stochastic order.

Given the parameterization above, we can show the following result

Proposition 1 *Under the assumptions above, if* $\sigma_{\varepsilon\nu} = 0$ *and* $\tau = \frac{1}{T^\alpha}$, *the OLS estimator* $\hat{\beta}$ *converges at rate* $T^{-(1-\alpha)}$ *to a functional of diffusion processes, or as* $T \to \infty$

$$T^{(1-\alpha)}\left(\hat{\beta} - \beta\right) \Rightarrow \frac{\int_0^1 W_2^\mu(s) dW_1(s)}{\int_0^1 \left(W_2^\mu(s)\right)^2 ds} \qquad (6)$$

In a more informal way, we can say

$$\hat{\beta} \overset{a}{\sim} \beta + ZT^{(\alpha-1)}$$

where Z is a mean zero random variable with a mixed-normal distribution. If $\alpha < 1$, the variance of $\hat{\beta}$ will decrease as $T \to \infty$, and we have consistency.

Otherwise, the estimator is inconsistent. However, note that $E\left(\hat{\beta}\right) = \beta$, since we assumed $\sigma_{\varepsilon\nu} = 0$. The fact that we have observed a big simulation error in table 2a when computing $E\left(\hat{\beta}\right)$ underscores the importance of proposition 1.

The above result is unusual, but not surprising. Unusual, because $\hat{\beta}$ does not converge to β at rates $T^{1/2}$ or T. In fact, the rate changes with α. However, this is not surprising, because we parameterized the model so that α controls the rate at which the signal, coming from X_t, accumulates. As α increases, the signal from X_t is decreasing compared to the noise ε_t, in the given sample. Therefore, the parameter β cannot be estimated precisely. The novelty in our parameterization is that the rate of information accumulation explicitly influences the rate of convergence and can even lead to inconsistent estimates.

The usual t test converges in distribution to the standard normal distribution, as before:

Proposition 2 *Under the assumptions above, if* $\sigma_{\varepsilon\nu} = 0$ *and* $\tau = \frac{1}{T^{\alpha}}$*, the t-statistic has the following distribution under the null, as* $T \to \infty$

$$t_{\hat{\beta}} = \frac{\hat{\beta} - \beta}{se(\hat{\beta})} \Rightarrow \frac{\int_0^1 W_2^{\mu}(s) dW_1(s)}{\left(\int_0^1 \left(W_2^{\mu}(s)\right)^2 ds\right)^{1/2}} \tag{7}$$

However, the t-test is not consistent for $\alpha \geq 1$. To see that, let $Ha : \beta = 0$. Then

$$\begin{aligned} t_{\hat{\beta}} &= \frac{\hat{\beta} - 0}{se(\hat{\beta})} = \frac{\hat{\beta} - \beta}{se(\hat{\beta})} + \frac{\beta - 0}{se(\hat{\beta})} = \\ &= \frac{T^{1-\alpha}\left(\hat{\beta} - \beta\right)\frac{1}{T^{1-\alpha}}\left(\sum_{t=1}^{T} X_t^2\right)^{1/2}}{\left(\hat{\sigma}_e^2\right)^{1/2}} + \frac{\left(T^{1-\alpha}\beta\right)\frac{1}{T^{1-\alpha}}\left(\sum_{t=1}^{T} X_t^2\right)^{1/2}}{\left(\hat{\sigma}_e^2\right)^{1/2}} \end{aligned} \tag{8}$$

The first term converges in distribution to $\left(\int_0^1 \left(W_2''(s)\right)^2 ds\right)^{-1/2} \left(\int_0^1 W_2''(s) dW_1(s)\right)$ as in (7). The second term explodes for $\alpha < 1$ (as needed for a consistent test), but converges in probability to 0 for $\alpha \geq 1$. Hence, for $\alpha \geq 1$, the distribution of the t-test is the same under the null and under the alternative, yielding power equal to size.

To see why parameterization (5) is useful, notice that in practice we have a fixed T, and a fixed τ. Given those values, $\alpha = -\frac{\log \tau}{\log T}$. When τ is small, the above expressions will provide a better explanation for the small sample behavior of $\hat{\beta}$ and its t-statistic than the usual approximations. Using the above parameterization, we can understand the puzzling results from the first set of simulations. Given a sample size T and τ, we can solve for α. This can be interpreted as how much information there is in the signal, coming from X_t. As shown, the new parameterization is useful to account for these results. Note that $\hat{\beta}$ is unbiased for all α, since $\sigma_{\varepsilon,u} = 0$. However, it is inconsistent, for $a \geq 1$.

In figure 2, we translate the signal to noise ratio from the CRSP monthly and yearly series into the index α. The first figure displays the results from the monthly data, with 864 observations. The value of α is computed for periods of 3, 5 and 10 years, as well as for the entire sample. The second figure displays the results from yearly data, with 72 observations. Similarly, the value of α is computed for the entire sample and for periods of 10 and 20 years. In both figures, increasing the span of the data results in a decrease of α (addition of more signal). However, it is important to notice that increasing the sample size twelve-fold by sampling data more frequently produces a relatively small decrease in α. This finding is consistent with the results from table 1 and implies that even with 864 observations in the sample size, the signal to noise ratio is very small. In other words, using monthly instead of yearly data will provide some useful information, but also a lot more noise.

One might be tempted to rescale the regressors in order to increase the signal in X_t. Suppose we rescale X_t by T^{α} in an attempt to make it "behave" as an I(1) process, or

$$Y_{t+1} = \mu_y + \tilde{\beta}\tilde{X}_t + \varepsilon_{t+1}$$

where $\tilde{\beta} = \frac{\beta}{T^{\alpha}}$ and $\tilde{X}_t = X_t T^{\alpha}$. Note that in order to preserve the relationship, we also have to rescale the coefficient. The rescaled coefficient $\tilde{\beta}$ is in a T^{α} neighborhood of 0. This scheme will lead to exactly the same problems. Namely, the OLS estimator of $\tilde{\beta}$ will converge at different rates depending on α, and will be inconsistent for $\alpha > 1$. Similarly, the corresponding t-statistic will be inconsistent for $\alpha > 1$. Rescaling does not work, because we are not adding any additional information.

3.2 Correlation in residuals

If the assumption $\sigma_{\varepsilon v} = 0$ is relaxed, the conclusions from above are only reinforced. Using results in Cavanagh et al. (1994), we can show that

Proposition 3 *Under the assumptions above, if $\tau = \frac{1}{T^{\alpha}}$, as $T \to \infty$*

$$T^{(1-\alpha)}\left(\hat{\beta} - \beta\right) \Rightarrow \left(1-\omega^2\right)^{1/2} \frac{\int_0^1 W_2^{\mu}(s)dW_{\perp}(s)}{\int_0^1 \left(W_2^{\mu}(s)\right)^2 ds} + \omega \frac{\int_0^1 W_2^{\mu}(s)dW_2(s)}{\int_0^1 \left(W_2^{\mu}(s)\right)^2 ds} \tag{9}$$

$$t_{\beta} \Rightarrow \left(1-\omega^2\right)^{1/2} \frac{\int_0^1 W_2(s)dW_{\perp}(s)}{\left(\int_0^1 \left(W_2^{\mu}(s)\right)^2 ds\right)^{1/2}} + \omega \frac{\int_0^1 W_2(s)dW_2(s)}{\left(\int_0^1 \left(W_2^{\mu}(s)\right)^2 ds\right)^{1/2}} \tag{10}$$

where $corr(\varepsilon_t, v_t) = \omega$ and $W_{\perp}(s)$ is a Wiener process obtained by projecting $W_1(s)$ on $W_2(s)$, with $E\left(W_{\perp}^2(s)\right) = \left(1-\omega^2\right)$.

The above proposition can be generalized to accommodate error terms with more general autocorrelation and heteroskedasticity (Hansen (1992), Stock and Watson (1993)). By construction, $W_{?}(s)$ and $W_2(s)$ are statistically independent. To help the intuition a bit further, we can write heuristically

$$\hat{\beta} \overset{a}{\sim} \beta + \left(1 - \omega^2\right)^{1/2} ZT^{(\alpha - 1)} + \omega RT^{(\alpha - 1)}$$

where Z is a mean zero random variable with a mixed-normal distribution, and R is a stochastic process with a defined density and a negative mean. Note that, if $\omega = 0$, the last term disappears in all of the above expressions. The higher the correlation between the errors is, the more dominant is the last term. Since we know (from simulation) that the density of R has most of its mass on negative values, this generates a negative bias in finite samples, which is a result also obtained by Stambaugh(1986, 1999) with different methods. Moreover, the smaller the signal to noise ratio is, the bigger the bias. For $\alpha > 1$, the bias does not disappear asymptotically and the variance of $\hat{\beta}$ increases, as discussed above.

Returning to our equity premium/ dividend yield equations, we can understand why we might have a situation when the coefficient seems to be significant (correct inference), but β is so poorly estimated that it produces forecasts no better than, say the mean of Y_t. The issues of in-sample fit and out of sample forecasting will be addressed in section 5.

3.3 Monte Carlo Again

The results in tables 2 and 3 were obtained by arbitrarily decreasing the value of τ. Tables 4 and 5 present the outcomes of the same set of simulations, but for $\tau = \frac{1}{T^{\alpha}}$, and $\alpha = 0, 0.20, 0.5, 0.67, 1$, and 2. Table 4 is for $\sigma_{\varepsilon,\nu} = 0$, whereas $\sigma_{\varepsilon,\nu} = 0.5$ is in table 5. As expected from the previous propositions, $\hat{\beta}$ is consistent for $\alpha < 1$, but is inconsistent for $\alpha > 1$ (tables 4a and 4b). The t-statistic for the null $\beta = 1$ is shown to be inconsistent against the alternative $\beta = 0$ (tables 4d and 5d). The inconsistency of the t-statistic does not depend on the choice of alternative, as can be seen from (8). The other results in the tables will be discussed below.

4 Relation to existent literature

4.1 Results with fixed signal to noise ratio

If $\sigma_{\varepsilon,\mu} = 0$ and τ is fixed, we can show that

$$T\left(\hat{\beta} - \beta\right) \Rightarrow \frac{1}{\tau} \frac{\int_0^1 W_2''(s) dW_1(s)}{\int_0^1 \left(W_2''(s)\right)^2 ds} \tag{11}$$

and

$$t_{\hat{\beta}} = \frac{\hat{\beta} - \beta_0}{se(\hat{\beta})} \Rightarrow \frac{\int_0^1 W_2''(s) dW_1(s)}{\left(\int_0^1 \left(W_2''(s)\right)^2 ds\right)^{1/2}}$$

As discussed above, we are interested in the case when the variance of u_t is orders of magnitude smaller than the variance of ε_t, which is captured by the constant τ, whose value is potentially very close to zero. The distribution of $\hat{\beta}$ will have a very big variance around its center, β, even for large sample sizes, but the estimator will be consistent. It is well known (Stock (1994)) that the asymptotic representations above approximate the finite sample distributions very closely. This is one of the reasons for their fast acceptance in time series econometrics. If $\sigma_{\varepsilon,u} \neq 0$ and $\phi = 1 + c/T$, similar results can be obtained, but the limiting distributions will depend on the correlation $\sigma_{\varepsilon,u}$ and on the parameter c. (Cavanagh et al. (1994)).

The system (2) has been studied by Stambaugh (1986,1999), Mankiw and Shapiro (1986), Nelson and Kim (1993), and Goetzmann and Jorion (1993). Those authors have investigated the small sample properties of $\hat{\beta}$, when $\sigma_{\varepsilon,u} \neq 0$ and ϕ is close to one. Stambaugh (1986, 1999) shows that the downward bias in $\hat{\phi}$, magnified by $\sigma_{\varepsilon,u}/\sigma_u^2$ results in a biased estimate of β. The magnitude and the direction of the bias depend on ϕ and $\sigma_{\varepsilon,u}/\sigma_u^2$. More precisely, $E\left(\hat{\beta} - \beta\right) = \frac{\sigma_{u,\varepsilon}}{\sigma_u^2} E\left(\hat{\phi} - \phi\right)$. Using Kendall's (1954) result that $E\left(\hat{\phi} - \phi\right) = -\left(1 + 3\phi\right)/T + O(T^{-2})$, and keeping in mind that the ratio $\sigma_{\varepsilon,u}/\sigma_u^2$ can be estimated consistently, one can compute the magnitude of the bias (up to $O(T^{-2})$). Interestingly, Cavanagh et al. (1994) use asymptotic tools to derive essentially the same result.

Our point can directly be related to the simulations and bootstrap work of Nelson and Kim (1993) and Goetzmann and Jorion (1993). By conducting the simulations or the bootstrap, the authors implicitly take into account the small signal to noise ratio. However, if the reality is that β is not zero, we showed that the usual t test has very little power to reject the null in their case. Note that in our set-up, the null is of an existing relationship. However, since the power is equal to size, the distribution under null and alternative are the same. The authors fail to reject their null (our alternative) of no prediction and control perfectly for the size of the test. However, since the signal to noise ratio is small, they also have virtually zero power to test against close alternatives, namely the alternatives of interest.

4.2 Small signal to noise or decreasing interval length with a fixed data span.

Perron (1987) shows that increasing the number of observations (frequency), while keeping the sample span fixed, will not necessarily lead to an increase of power in unit-root tests. The author achieves this result by drawing the innovations of the process from a distribution with an increasing variance. The higher the frequency, the higher the variance of the innovations. The increase in the variance and the increase in observations is done at the same rate T. In effect, increasing the length of the sample increases only the noise, without adding any signal.

Our setup has a different motivation, since we are considering a bivariate, cointegrated system. However, modeling the signal to noise ratio as a decreasing function of T is similar in spirit to Perron's work. Indeed, for $\alpha = 1$, noise will accumulate at the same rate as information does, and we will retrieve similar results in the present setup. In both studies, increasing the number of observations will not lead to more precise estimates of the parameter of interest. However, we are not interested in fixing the rate of noise accumulation. Our goal is precisely to investigate the impact of different signal to noise ratios on the statistics of interest. Therefore, we place the emphasis on α.

4.3 Alternative local to zero parameterization

Instead of the parameterization (5), we could have chosen $\tau = \frac{\lambda}{T}$, as in Stock and Watson (1998). Here is a quick digression to their (simplified) model: $y_t = \beta_t + u_t$, $\beta_t = \beta_{t-1} + v_t$ and $v_t = \tau \nu_t$, where $\tau = \lambda/T$. If we take $\lambda = 1$ then $y_t = \beta_t + u_t$, which is our set-up, assuming our $\beta = 1$ (except for timing). In that case, similar calculations yield $\left(\hat{\beta} - \beta\right) \Rightarrow$ $\frac{1}{,} \int_0^1 W_2''(s)dW_1(s) \left(\int_0^1 (W_2''(s))^2 ds\right)^{-1}$ and $t_{\hat{\beta}} \Rightarrow \int_0^1 W_2''(s)dW_1(s) \left(\int_0^1 (W_2''(s))^2 ds\right)^{-1/2}$.

However, the time-varying parameter (TVP) model in Stock and Watson (1998) is different in interpretation from the equity premium/dividend yield equations. The TVP β_t is not observable, and hence τ cannot be estimated with a simple regression. A Kalman filter approach can yield an estimator of τ, but it is known that this ML estimator has a large pile-up probability at zero. In our model, the series X_t are readily observable and we are not looking to model the fact that a parameter is so close to the boundary of its parameter space that it cannot be estimated reliably. We have a fixed parameter model, where the regressor has a small variance, but this variance can be estimated from a simple regression. The main interest here is to capture the effect of small τ on the small sample properties of $\hat{\beta}$, given a sample size. We are building a sequence of models; one for each α, given T. In that sense, our modelling is more akin to the one used by Pantula (1991).

5 In-Sample Fit and Long-Term Forecasting

5.1 In Sample Fit

When estimating and testing with equation (2), the t-statistic is usually only marginally significant and the coefficient of determination R^2 is very low. This is not surprising, given figure 1. However, using normal asymptotics, one would expect the R^2 to increase to 1 with the sample size, given that we postulated the existence of a relationship between Y_t and X_t. In the simple regression above, we can show that

Proposition 4 *Under the assumptions above, if* $\tau = \frac{1}{T^{\alpha}}$*, as* $T \to \infty$

$$R^2 \to_p \begin{cases} 1 & \alpha < 1/2 \\ 0 & \alpha > 1/2 \end{cases}$$

For the borderline case $\alpha = 1/2$, $R^2 = O_p(1)$.

For a small signal to noise ratio, an increase in the number of observations will not result in an increase in R^2, even if there is a relationship between the two variables. Table 6 illustrates that this result is consistent with the data. Increasing the number of observations (going from yearly to monthly frequency) might not result in higher R^2. In fact, we see that the R^2 in the monthly regressions is considerably lower than the ones in yearly data. However, since the yearly samples are very small and we don't know whether there truly is a relationship between the equity premium and dividend yields in the data[1], we can only argue that our setup is consistent with the evidence.

To show explicitly the impact of the previous proposition, we turn to the simulations in tables 2e, 3e, 4e, and 5e. The first two tables present the results of the simulations of R^2 for the case of τ fixed, $\sigma_{\varepsilon,\nu} = 0$ and $\sigma_{\varepsilon,\nu} = 0.5$, respectively. No matter what the correlation of the errors is, for a big τ, the R^2 increases to 1 as the sample size rises. However, for very small values of τ, the R^2 seems to converge to zero. This fact cannot be explained by usual, fixed τ results, but is in exact accord with the above proposition. Tables 4e and 5e present results for various α's, $\sigma_{\varepsilon,\nu} = 0$ and $\sigma_{\varepsilon,\nu} = 0.5$, respectively. We can clearly see that, for $a < 1/2$, the R^2 increases with the sample size. For $\alpha > 1/2$, R^2 converges to 0, as $T \to \infty$.

[1]Note that if X is stationary, a similar result can be obtained for the system. Then, $R^2 \to_p \frac{\beta^2 \frac{\sigma_x}{\sigma_\varepsilon}}{\beta^2 \frac{\sigma_x}{\sigma_\varepsilon} + 1}$, where $E(X_t^2) = \sigma_x$. If σ_x is much smaller than σ_ε, the ratio is not close to 1.

5.2 Long-Term Forecasting

Given that β is not estimated precisely, one has to wonder how good are the out of sample forecasts produced by (2-4). In the equity premium literature, the outcome of an out of sample forecasting exercise is often seen as the most relevant measure for a successful model. In light of the discussion above, we do not expect to be able to forecast with great accuracy, despite the postulated relationship between equity return and dividend yields. Goyal and Welch (1999) point out that the equity forecast produced from historical annual data does not perform any better than the unconditional mean. However, the authors attribute this lack of forecasting power to parameter instability rather than to small signal to noise ratio.

We will prove that, if the signal to noise ratio is small, forecasting using the estimated model will not do better than the simple unconditional mean. First, we show that asymptotics with fixed τ cannot give us insights into the problem. Then, we use the local to zero parameterization to derive the analytic results that would help explain the outcomes from the simulations.

We will compare two competing long-run forecasts: $\overline{Y} = \frac{1}{T}\sum_{t=1}^{T} Y_t$ and $\hat{Y}_{T+k|T} = \hat{\mu}_y + \hat{\beta} X_T$, where $k = [\kappa T]$, and $\kappa \in (0,1)$. Given τ small, we can show that both forecasts are asymptotically unbiased, or

$$E\left(T^{-1/2}\left(Y_{T+k} - \overline{Y}\right)\right) \to 0$$
$$E\left(T^{-1/2}\left(Y_{T+k} - \hat{Y}_{T+k|T}\right)\right) \to 0$$

More importantly, the asymptotic variances are:

$$E\left(T^{-1}\left(Y_{T+k}-\overline{Y}\right)^2\right) \to \tau^2\beta^2(\kappa+1/3)$$
$$E\left(T^{-1}\left(Y_{T+k}-\hat{Y}_{T+k|T}\right)^2\right) \to \tau^2\beta^2\kappa$$

Therefore, we can conclude that, asymptotically,

$$MSE(\overline{Y}) > MSE(\hat{Y}_{T+k|T})$$

This result might be expected. However, referring back to our battery of simulations, tables 2g, 3g, 4g, 5g, we come to the realization that for small τ, the MSE's from both forecasts are almost identical, even for relatively large sample sizes. This fact can well be captured using (5). Under that parameterization, $\hat{Y}_{T+k|T}$ does not always produce better forecasts than $\overline{Y}$.

Proposition 5 *Under the assumptions above, suppose* $\tau = \frac{1}{T^{\alpha}}$ *and* $k{=}[\kappa T]$ *where* κ *is a fixed number. Let* $\overline{Y}$ *be the sample mean of* Y_t *and let* $\hat{Y}_{T+k|T} = \widehat{\mu}_y + \hat{\beta}X_T$. *Then, as* $T \to \infty$

$$E\left(T^{-(1/2-\alpha)}\left(Y_{T+k}-\overline{Y}\right)\right) \to 0 \text{ and } E\left(T^{-(1/2-\alpha)}\left(Y_{T+k}-\hat{Y}_{T+k|T}\right)\right) \to 0 \quad , \quad \alpha < 1/2$$
$$E\left(\left(Y_{T+k}-\overline{Y}\right)\right) \to 0 \text{ and } E\left(\left(Y_{T+k}-\hat{Y}_{T+k|T}\right)\right) \to 0 \quad , \quad \alpha \geq 1/2$$

or both forecasts are asymptotically unbiased for all α's. *However,*

$$E\left(T^{-(1-2\alpha)}\left(Y_{T+k}-\overline{Y}\right)^2\right) \to \beta^2(\kappa+1/3) \quad , \quad \alpha < 1/2$$
$$E\left(\left(Y_{T+k}-\overline{Y}\right)^2\right) \to \beta^2(\kappa+1/3)+1 \quad , \quad \alpha = 1/2$$
$$E\left(\left(Y_{T+k}-\overline{Y}\right)^2\right) \to 1 \quad , \quad \alpha > 1/2$$

and

$$
\begin{aligned}
E\left(T^{-(1-2\alpha)}\left(Y_{T+k}-\hat{Y}_{T+k|T}\right)^2\right) &\to \beta^2\kappa \quad , \quad \alpha < 1/2 \\
E\left(\left(Y_{T+k}-\hat{Y}_{T+k|T}\right)^2\right) &\to \beta^2\kappa + 1 \quad , \quad \alpha = 1/2 \\
E\left(\left(Y_{T+k}-\hat{Y}_{T+k|T}\right)^2\right) &\to 1 \quad , \quad \alpha > 1/2
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
MSE(\overline{Y}) &> MSE(\hat{Y}_{T+k|T}) \quad , \quad \alpha < 1/2 \\
MSE(\overline{Y}) &= MSE(\hat{Y}_{T+k|T}) \quad , \quad \alpha \geq 1/2
\end{aligned}
$$

In the case $\alpha > 1/2$, we have $MSE(\overline{Y}) = MSE(\hat{Y}_{T+k|T})$. In other words, if the signal to noise ratio is not strong enough in a finite sample, the MSE's from both forecasts will be equal. The simulations in tables 4f and 5f illustrate the analytical results. The entries show the simulated MSE in the Monte Carlo experiments discussed above. For $\alpha < 1/2$, all values are rescaled by a power of the sample size, following the proposition. Whenever $\alpha < 1/2$, the $\hat{Y}_{T+k|T}$ is a better predictor for Y_{T+k} than $\overline{Y}$, but for $\alpha \geq 1/2$, the MSE's from the two forecasts are equal (except for simulation error). In other words, forecasts from the true model won't necessarily outperform the unconditional mean, if the signal to noise ratio is extremely low. Therefore, the findings in Goyal and Welch (1999) could very well be explained by the small signal to noise ratio.

6 Conclusion

Previous papers have tested the existence of relationship (2) under the null of $\beta = 0$. Here we conduct the opposite exercise. Suppose that the equity premium is forecasted by dividend yields. Even if such a relationship does exist, there is so much noise in the equity premium that estimation, inference and forecasting cannot be carried out using the faint signal coming from the dividend yields. It might well be the case that the equity premium is excessively volatile because of its dependence on too many economic variables. There is certainly evidence that other (macro and micro) factors, apart from the dividend yield, might be related to the risky return as argued in Fama (1981), Lamont (1998) and Lee, Myers, and Swaminathan (1999), among others.

The signal to noise ratio is so small in the equity premium/dividend yield data that it deserves particular attention. Simulations and analytical results, showed that even in fairly large samples, the signal to noise ratio is not large enough to conduct reliable estimation, inference, and forecasting. We have proposed an index of the signal contained in a sample. This index has an explicit impact on estimation, inference, in-sample fit and out of sample forecasting. We argue that adding more (frequent) observations does not guarantee an increase in this index, and the presented evidence (simulations and data) supports such a conclusion. Lastly, we demonstrate that a small signal to noise ratio can, in and of itself, lead to negatively correlated returns.

In this paper, we have focused on regression (1). However, since most of the recent empirical work is looking at forecasting long run returns (sum of short returns), in a companion paper we analyze the econometric issues that arise from regressing long returns on the dividend yield. Another direction for future work might be the explicit modelling of the equity/dividend equations with time-varying parameters, an area that has largely been left unexplored. Since it is hard to identify one parameter with noisy data, attempting to fit more complicated models might seem a futile exercise. However, a successful step in that direction is the paper by Goyal and Welch (1999) who let the slope and the intercept follow a deterministic function of time. A more general TVP framework, that explicitly takes into account the small signal to noise ratio, might in fact improve the performance of both estimates and forecasts. Such a setup might also cast light onto the reasons of predictability, if predictability is in fact in the data.

Appendix

Calculations for equation 11, with τ fixed. We have

$$\begin{aligned} Y_{t+1} &= \mu_y + \beta X_t + \varepsilon_{t+1} \\ X_{t+1} &= X_t + u_{t+1} \\ u_t &= \tau v_t \end{aligned}$$

Then $\hat{\beta} = \left(\sum_{t=1}^{T} Y_{t+1}\left(X_t - \overline{X}\right)\right)\left(\sum_{t=1}^{T}\left(X_t - \overline{X}\right)^2\right)^{-1}$

$= \beta + \left(\sum_{t=1}^{T}\left(X_t - \overline{X}\right)\varepsilon_{t+1}\right)\left(\sum_{t=1}^{T}\left(X_t - \overline{X}\right)^2\right)^{-1}$

$= \beta + \left(\sum_{t=1}^{T}\left(\sum_{j=1}^{t} u_j - \left(\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{t} u_j\right)\right)\varepsilon_{t+1}\right) \times$

$\left(\sum_{t=1}^{T}\left(\sum_{j=1}^{t} u_j - \left(\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{t} u_j\right)\right)^2\right)^{-1}$

$= \beta + \left(\tau\sum_{t=1}^{T}\left(\sum_{j=1}^{t} v_j - \left(\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{t} v_j\right)\right)\varepsilon_{t+1}\right) \times$

$\left(\tau^2\sum_{t=1}^{T}\left(\sum_{j=1}^{t} v_j - \left(\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{t} v_j\right)\right)^2\right)^{-1}.$

Then

$$\begin{aligned} T\left(\hat{\beta} - \beta\right) &= \frac{\frac{\tau}{T}\sum_{t=1}^{T}\left(\sum_{j=1}^{t} v_j - \left(\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{t} v_j\right)\right)\varepsilon_{t+1}}{\frac{\tau^2}{T^2}\sum_{t=1}^{T}\left(\sum_{j=1}^{t} v_j - \left(\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{t} v_j\right)\right)^2} \\ &\Rightarrow \frac{1}{\tau}\frac{\int_0^1 W_2''(s)dW_1(s)}{\int_0^1 \left(W_2''(s)\right)^2 ds} \end{aligned}$$

Define the usual t-statistic:

$$t_{\beta} = \frac{\hat{\beta} - 0}{se(\hat{\beta})} = \frac{\left(\hat{\beta} - 0\right)\left(\sum_{t=1}^{T}\left(X_t - \overline{X}\right)^2\right)^{1/2}}{\left(\frac{1}{T-1}\sum_{t=1}^{T}\left(Y_{t+1} - \hat{\beta}X_t\right)^2\right)^{1/2}} = \frac{A * B}{C}$$

First, under the null that $\beta = \beta_0$, from above, $T\left(\hat{\beta} - \beta_0\right) \Rightarrow \frac{1}{¿}\frac{\int_0^1 W_2(s)dW_1(s)}{\int_0^1 W_2^2(s)ds}$. Second $\frac{1}{T}\left(\sum_{t=1}^{T} X_t^2\right)^{1=2} = \left(\frac{1}{T^2}\sum_{t=1}^{T} X_t^2\right)^{1=2} \Rightarrow \left(\tau^2 \int_0^1 W_2^2(s)ds\right)^{1=2}$. Third, $C \rightarrow_p \left(\sigma_{"}^2\right)^{1=2}$. Putting things together, we have the usual result:

$$t_{fl} \Rightarrow \frac{\int_0^1 W_2''(s)dW_1(s)}{\left(\int_0^1 \left(W_2''(s)\right)^2 ds\right)^{1=2}}$$

Proposition 1 Proof: Suppose $\tau = 1/T^{fi}$. Then, similarly to the previous calculations, $\hat{\beta} = \frac{1}{T^{\alpha}}\left(\sum_{t=1}^{T}\left(\sum_{j=1}^{t} v_j\right)\varepsilon_{t+1}\right)\left(\frac{1}{T^{2\alpha}}\sum_{t=1}^{T}\left(\sum_{j=1}^{t} v_j\right)^2\right)^{i\,1}$

$$\begin{aligned}
T^{(1\,i\,fi)}\left(\hat{\beta}-\beta\right) &= T^{(1\,i\,fi)}\frac{\frac{1}{T^{\alpha}}\sum_{t=1}^{T}\left(\sum_{j=1}^{t} v_j - \left(\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{t} v_j\right)\right)\varepsilon_{t+1}}{\frac{1}{T^{2\alpha}}\sum_{t=1}^{T}\left(\sum_{j=1}^{t} v_j - \left(\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{t} v_j\right)\right)^2} \\
&= \frac{\frac{1}{T^{\alpha}}\sum_{t=1}^{T}\left(\sum_{j=1}^{t} v_j - \left(\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{t} v_j\right)\right)\varepsilon_{t+1}}{\frac{1}{T^{2\alpha}}\sum_{t=1}^{T}\left(\sum_{j=1}^{t} v_j - \left(\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{t} v_j\right)\right)^2}\frac{\frac{1}{T^{(1-\alpha)}}}{\frac{1}{T^{2(1-\alpha)}}} \\
&= \frac{\frac{1}{T}\sum_{t=1}^{T}\left(\sum_{j=1}^{t} v_j - \left(\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{t} v_j\right)\right)\varepsilon_{t+1}}{\frac{1}{T^2}\sum_{t=1}^{T}\left(\sum_{j=1}^{t} v_j - \left(\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{t} v_j\right)\right)^2} \\
T^{(1\,i\,fi)}\left(\hat{\beta}-\beta\right) &\Rightarrow \frac{\int_0^1 W_2''(s)dW_1(s)}{\int_0^1 \left(W_2''(s)\right)^2 ds}
\end{aligned}$$

■

Proposition 2 Proof: Under the null, the t-statistic is:

$$t_{fl} = \frac{\left(\hat{\beta}-0\right)\left(\sum_{t=1}^{T} X_t^2\right)^{1=2}}{\left(\frac{1}{T\,i\,1}\sum_{t=1}^{T}\left(Y_{t+1}-\hat{\beta}X_t\right)^2\right)^{1=2}} = \frac{A * B}{C}$$

First, from above, $T^{(1\,i\,fi)}\left(\hat{\beta}-\beta\right) \Rightarrow \frac{\int_0^1 W_2^{\mu}(s)dW_1(s)}{\int_0^1 \left(W_2^{\mu}(s)\right)^2 ds}$. Second, $\frac{1}{T^{1-\alpha}}\left(\sum_{t=1}^{T}\left(X_t - \overline{X}\right)^2\right)^{1=2} =$ $\frac{1}{T^{1-\alpha}}\left(\frac{1}{T^{2\alpha}}\sum_{t=1}^{T}\left(\sum_{j=1}^{t} v_j - \left(\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{t} v_j\right)\right)^2\right)^{1=2} =$

$$\left(\frac{1}{T^2}\sum_{t=1}^T\left(\sum_{j=1}^t v_j-\left(\frac{1}{T}\sum_{t=1}^T\sum_{j=1}^t v_j\right)\right)^2\right)^{1/2}\Rightarrow\left(\int_0^1\left(W_2^{\mu}(s)\right)^2ds\right)^{1/2},$$

and $C\rightarrow_p\left(\sigma_v^2\right)^{1/2}$. Finally, as above

$$t_{\hat{\alpha}}\Rightarrow\frac{\int_0^1 W_2^{\mu}(s)dW_1(s)}{\left(\int_0^1\left(W_2^{\mu}(s)\right)^2ds\right)^{1/2}}$$

■

Proposition 3 Proof: Let $v_t=\varepsilon_t-\text{Proj}(\varepsilon_t|u_t)=\varepsilon_t-\delta u_t$, where Proj() is the linear projection of ε_t on u_t and $\delta=\frac{\sigma_{\varepsilon u}}{\sqrt{\sigma_u\sigma_\varepsilon}}$. Then, $\frac{1}{(1-\delta^2)^{1/2}}\frac{1}{T^{1/2}}\sum_{i=0}^t v_t\Rightarrow W_\perp(s)$, where $W_\perp(s)$ is a standard Wiener process, dstributed independently of $W_2(s)$ by construction. We can also write it as $W_\perp(s)=W_1(s)-\delta W_2(s)$. Using exactly the same steps as above, it is straight forward to show that $T^{(1-\alpha)}\left(\hat{\beta}-\beta\right)\Rightarrow\frac{\int_0^1 W_2^{\mu}(s)dW_1(s)}{\int_0^1\left(W_2^{\mu}(s)\right)^2ds}=\left(1-\delta^2\right)^{1/2}\frac{\int_0^1 W_2^{\mu}(s)dW_\perp(s)}{\int_0^1\left(W_2^{\mu}(s)\right)^2ds}+\delta\frac{\int_0^1 W_2^{\mu}(s)dW_2(s)}{\int_0^1\left(W_2^{\mu}(s)\right)^2ds}$. Similarly $t_{\hat{\alpha}}\Rightarrow\frac{\int_0^1 W_2^{\mu}(s)dW_1(s)}{\left(\int_0^1\left(W_2^{\mu}(s)\right)^2ds\right)^{1/2}}=\left(1-\delta^2\right)^{1/2}\frac{\int_0^1 W_2^{\mu}(s)dW_\perp(s)}{\left(\int_0^1\left(W_2^{\mu}(s)\right)^2ds\right)^{1/2}}+\delta\frac{\int_0^1 W_2^{\mu}(s)dW_2(s)}{\left(\int_0^1\left(W_2^{\mu}(s)\right)^2ds\right)^{1/2}}$, and that completes the proof.■

Proposition 4 Proof: Recall that $R^2=\frac{\hat{\alpha}^2\sum_{t=1}^T\left(X_t-\overline{X}\right)^2}{\sum_{t=1}^T\left(Y_t-\overline{Y}\right)^2}$. The denominator is $\sum_{t=1}^T\left(Y_t-\overline{Y}\right)^2=\hat{\beta}^2\sum_{t=1}^T\left(X_t-\overline{X}\right)^2+\sum_{t=1}^T\varepsilon_t^2+LOT$, where LOT denotes terms of lower stochastic order for any α. For $\alpha<1/2$, the first term dominates, for $\alpha>1/2$ the second term dominates, and for $\alpha=1/2$ the two terms are of the same O_p order. Then, for $\alpha<1/2$, $R^2=\frac{\hat{\alpha}^2\frac{1}{T^{2-2\alpha}}\sum_{t=1}^T\left(X_t-\overline{X}\right)^2}{\hat{\alpha}^2\frac{1}{T^{2-2\alpha}}\sum_{t=1}^T\left(X_t-\overline{X}\right)^2+o_p(1)}\rightarrow^p 1$. In the case $\alpha>1/2$, $R^2=\frac{o_p(1)}{op(1)+\frac{1}{T}\sum_{t=1}^T\varepsilon_t^2}\rightarrow^p 0$. For the borderline case $\alpha=1/2$, we have $R^2=\frac{\hat{\alpha}^2\frac{1}{T}\sum_{t=1}^T\left(X_t-\overline{X}\right)^2}{\hat{\alpha}^2\frac{1}{T}\sum_{t=1}^T\left(X_t-\overline{X}\right)^2+\frac{1}{T}\sum_{t=1}^T\varepsilon_t^2}\Rightarrow\frac{\alpha^2\int_0^1\left(W_2^{\mu}(s)\right)^2ds}{\alpha^2\int_0^1\left(W_2^{\mu}(s)\right)^2ds+\sigma_\varepsilon^2}$, which completes the proof.

■

Proposition 5 Proof: Before proceeding, note that $Y_{T+k} = \mu_y + \beta X_{T+k-1} + \varepsilon_{T+k} = \mu_y + \beta X_T + \beta \sum_{i=1}^{k-1} u_{T+i} + \varepsilon_{T+k}$, where $u_t = \frac{1}{T^{\alpha}} v_t$ and $k = [\kappa T]$. Then

$$T^{-(1/2-\alpha)} Y_{T+k} = T^{-(1/2-\alpha)} \left(\mu_y + \beta X_T + \beta \sum_{i=1}^{k-1} u_{T+i} + \varepsilon_{T+k} \right) \Rightarrow \beta W(1) + \beta R \quad , \alpha < 1/2$$

$$Y_{T+k} \Rightarrow \mu_y + \beta W(1) + \beta R + \varepsilon_{T+k} \quad , \alpha = 1/2$$

$$Y_{T+k} \Rightarrow \mu_y + \varepsilon_{T+k} \quad , \alpha > 1/2$$

where $R \sim N\left(0, \kappa \sigma_v^2\right)$ and is independently distributed from ε_{T+k}. Similarly, the mean of Y_t

$$T^{-(1/2-\alpha)} \overline{Y} \Rightarrow \beta \int_0^1 W(s)\, ds \quad , \alpha < 1/2$$

$$\overline{Y} \Rightarrow \mu_y + \beta \int_0^1 W(s)\, ds \quad , \alpha = 1/2$$

$$\overline{Y} \to^p \mu_y \quad , \alpha > 1/2$$

If we use $\hat{Y}_{T+k|T}$,

$$T^{-(1/2-\alpha)} \hat{Y}_{T+k} = T^{-(1/2-\alpha)} \left(\hat{\alpha} + \hat{\beta} X_T \right) \Rightarrow \beta W(1) \quad , \alpha < 1/2$$

$$\hat{Y}_{T+k} \Rightarrow \mu_y + \beta W(1) \quad , \alpha = 1/2$$

$$\hat{Y}_{T+k} \to^p \mu_y \quad , \alpha > 1/2$$

Then, the asymptotic bias from using $\hat{Y}_{T+k|T}$ is $E\left\{T^{-(1/2-\alpha)}\left(Y_{T+k} - \hat{Y}_{T+k|T}\right)\right\} \to 0$ for $\alpha < 1/2$ and $E\left\{T\left(Y_{T+k} - \hat{Y}_{T+k|T}\right)\right\} \to 0$ for $\alpha \geq 1/2$. Similarly, the asymptotic bias from using $\overline{Y}$ is $E\left\{T^{-(1/2-\alpha)}\left(Y_{T+k} - \overline{Y}\right)\right\} \to 0$ for $\alpha < 1/2$ and $E\left\{T\left(Y_{T+k} - \overline{Y}\right)\right\} \to 0$ for $\alpha \geq 1/2$. Therefore, both forecasts are asymptotically unbiased, for all α's.

The expressions for the asymptotic variances can be computed in the same fashion, using the fact that $E\left(\frac{X_T}{T^{1/2}}\frac{(X_1+...+X_T)}{T^{3/2}}\right)=\frac{1}{T^2}E\left(X_1X_T+...+X_T^2\right)=$ $\frac{1}{T^2}E\left(\varepsilon_1^2+\left(\varepsilon_1^2+\varepsilon_2^2\right)+...+\left(\varepsilon_1^2+\varepsilon_2^2+...+\varepsilon_T^2\right)+cross\ terms\right)=\frac{1}{T^2}\left(\sigma_\varepsilon^2+2\sigma_\varepsilon^2+...+T\sigma_\varepsilon^2\right)=$ $\frac{\sigma_\varepsilon^2}{T^2}\frac{1}{2}T\left(T+1\right)\to\frac{1}{2}\sigma_\varepsilon^2$. Similarly, $E\left(\left(\frac{X_1+...+X_T}{T^{3/2}}\right)^2\right)=\frac{1}{T^3}E\left(X_1^2+...+X_T^2+2\sum_{i>j}X_iX_j\right)=$ $\frac{1}{T^3}\left\{\sigma_\varepsilon^2+2\sigma_\varepsilon^2+...+T\sigma_\varepsilon^2\right\}+2\sigma_\varepsilon^2\left(T+2\left(T-1\right)+3\left(T-2\right)+...+\left(T-1\right)\right)=$ $\frac{\sigma_\varepsilon^2}{T^3}\left\{\sum_{i=1}^T i+2\sum_{i=1}^T\left(T-i+1\right)i\right\} = \frac{\sigma_\varepsilon^2}{T^3}\left(\frac{T(T+1)}{2}+\frac{1}{6}T^3+\frac{1}{2}T^2+\frac{1}{3}T\right)\to$ $\frac{1}{3}\sigma_\varepsilon^2$. ■

References

[1] Campbell, J and R. Shiller, 1988, The Dividend-Price ratio and expecations of future dividends and discount factors, Review of Financial Studies, 1:3, 195-228.

[2] Campbell, J and A. Lo and A.C. MacKinlay, 1997, The econometrics of financial markets, Princeton University Press.

[3] Cavanagh, C., G. Elliott, and J. Stock, 1995, Inference in models with nearly integrated regressors, Econometric Theory, (11), 1131-1147.

[4] Fama, E., 1981, Stock Returns, Real Activity, Inflation, and Money, American Economic Review, 71:4, 545-565.

[5] Fama, E. and K. French, 1988, Dividend yields and expected stock returns, Journal of Financial Economics 22, 3-25.

[6] Elliott, G and J. Stock, 1994, Inference in time series regression when the order of integration of a regressor is unknown, Econometric Theory (10), 672-700.

[7] Goetzmann, W. and P. Jorion, 1993, Testing the Predictive Power of Dividend Yields, Journal of Finance 48:2, 663-679.

[8] Goyal, A. and I. Welch, 1999, Predicting the Equity Premium, working paper #4-99, Anderson School of Management.

[9] Hansen, B., 1992, Convergence to stochastic integrals for dependent heterogeneous processes, Econometric Theory, 8, 485-500.

[10] Hodrick, R., 1992, Dividend yields and expected stock returns: alternative procedures for inference and measurement, Review of Financial Studies (5:3), 357-386.

[11] Kim, M-J, C. Nelson and R. Startz, 1991, Mean reversion in stock prices: A reappraisal of the empirical evidence, Review of Economic Studies 58, 515-528.

[12] Kendall, M., 1954, Note on bias in the estimation of autocorrelation, Biometrika (41), 403-404.

[13] Lamont, O., 1998, Earnings and expected returns, Journal of Finance (53:5), 1563-1587.

[14] Lee, C., J. Myers, and B. Swaminathan, 1999, What is the intrinsic value of the Dow?, Journal of Finance, forthcoming.

[15] Mankiw, N, and M. Shapiro, 1986, Do we reject too often? Small sample properties of tests of rational expectations models, Economics Letters (20), 139-145.

[16] Nelson, C. and M. Kim, 1993, Predictable stock returns: the role of small sample bias, Journal of Finance (48:2), 641-661.

[17] Pantula, S., 1991, Asymptotic distributions of unit-root tests when the process in nearly stationary, Journal of Business and Economic Statistics, 9:1, 63-71.

[18] Perron, Pierre, 1989, Testing for a random walk: A simulation experiment of power when the sampling interval is varied, in B. Raj, ed: Advances in Econometrics and Modeling (KluwerAcademic Publishers, Inc., Norwell, Mass.).

[19] Perron, Pierre, 1988, Trends and random walks in macroeconomic time series, Journal of Economic Dynamics and Control, 12, 297-332.

[20] Poterba, J. and L. Summers, 1988, Mean Reversion in Stock Prices, Evidence and Implications, Journal of Financial Economics 22, 27-59.

[21] Richardson, Matthew and James H. Stock, 1989, Drawing inferences from statistics based on multi-year asset returns, Journal of Financial Economics 25, 323-348.

[22] Schwert, G.W., 1989, Tests for unit roots: a monte carlo investigation, Journal of Business and Economic Statistics, 7, 147-160.

[23] Stambaugh, R., 1986, Bias in Regressons with Lagged Stochastic Regressors, Chicago working paper.

[24] Stambaugh, R., 1999, Predictive Regressons, Journal of Financial Economics, forthcoming.

[25] Summers, L., 1986, Does the Stock Market Rationally Reflect Fundamental Values, Journal of Finance, (41:3), 591-603.

[26] Stock, J., 1994, Unit roots, structural breaks and trends, in Engle, R. and D. McFadden (eds.), Handbook of Econometrics, Vol IV, Amsterdam: Elsevier, Chapter 46.

[27] Stock, J. and M. Watson, 1993, A simple estimator of cointegrating vectors in higher order integrated systems, Econometrica (61:4), 783-820.

[28] Stock, J. and M. Watson, 1998, Median Unbiased Estimation of Coefficient Variance in a Time-Varying Parameter Model, Journal of the American Statistical Association, 93 (441), 349-358.

[29] Watson, M., 1994, Vector Autoregressions and Cointegration, Handbook of Econometrics, Volume IV, R. Engle and D. McFadden eds, Elsevier Science, Chapter 47.

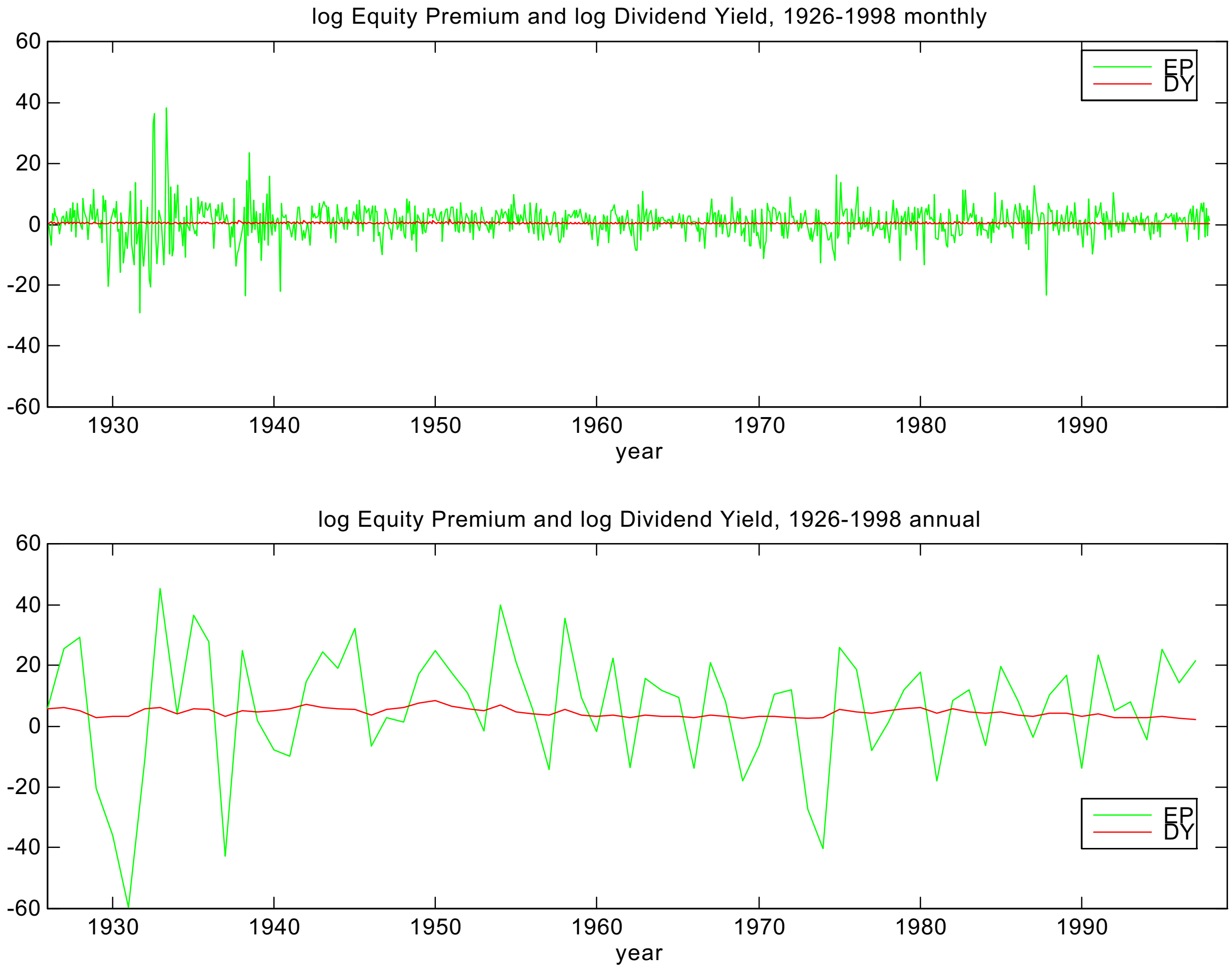

Figure 1

The monthly and annual data is from the CRSP database. All series are in logs and multiplied by 100. EP is the log of Equity Premium and DY is the log of Dividend Yield. The Dividend Yield is much less volatile than the Equity Premium.

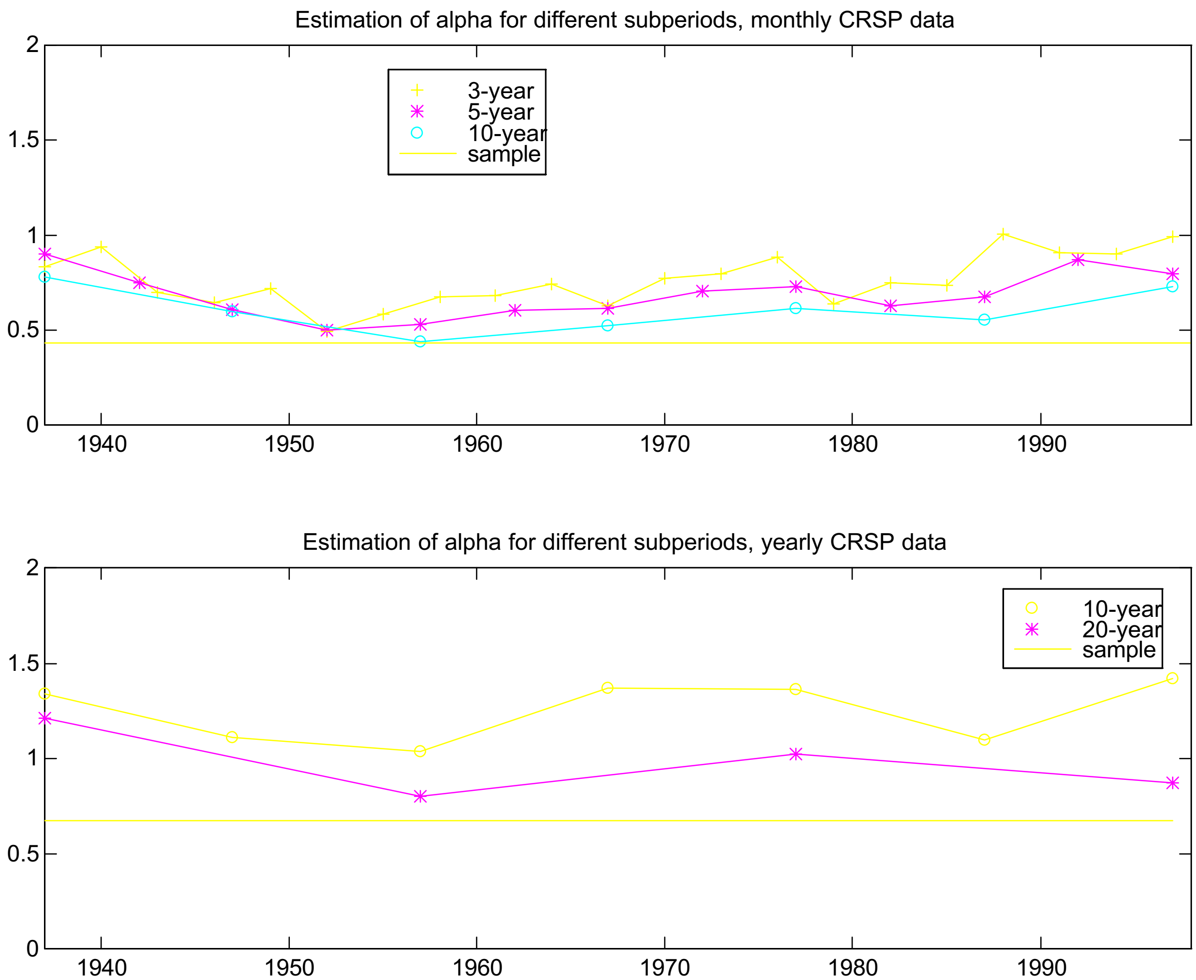

Figure 2

The value of α is obtained from: $\alpha = -\log \tau / \log T$, where τ^2 is obtained in the following way: Regress the Equity Premium on lagged Dividend Yield and estimate the variance of the residuals, call it σ_ε. Regress the Dividend Yield on itself (or several lags), and estimates the variance of the residuals, call it σ_u. Then $\tau^2 = \sigma_u / \sigma_\varepsilon$. The value of α is computed for different intervals, as indicated. Remarkably, it is fairly constant and sometimes higher than 1. Values higher than 0.5 indicate that the signal to noise ratio is extremely small. The OLS estimator of β in the forecasting regression will be very inaccurate.

Table 1

Period: 1926-1998

	Yearly Summary Statistics					Monthly Summary Statistics				
	Mean	Std.Dev	V(DY)/V(EP)	τ^2	AR	Mean	Std.Dev	V(DY)/V(EP)	τ^2	AR
EP	6.32	19.897	0.0049	0.0032	0.161	0.68	5.504	0.0016	0.0029	0.123
DY	4.49	1.387			0.964*	0.35	0.219			0.997

Period: 1926:1-1945:12

	Yearly Summary Statistics					Monthly Summary Statistics				
	Mean	Std.Dev	V(DY)/V(EP)	τ^2	AR	Mean	Std.Dev	V(DY)/V(EP)	τ^2	AR
EP	5.24	28.475	0.0017	0.0026	0.256	0.76	8.059	0.0007	0.0014	0.150
DY	5.15	1.178			0.966	0.42	0.220			0.997

Period: 1946:1-1978:12

	Yearly Summary Statistics					Monthly Summary Statistics				
	Mean	Std.Dev	V(DY)/V(EP)	τ^2	AR	Mean	Std.Dev	V(DY)/V(EP)	τ^2	AR
EP	6.22	17.683	0.0072	0.0037	0.116	0.57	4.038	0.0035	0.0067	0.100
DY	4.41	1.504			0.973	0.35	0.239			0.997

Period: 1979:1-1985:12

	Yearly Summary Statistics					Monthly Summary Statistics				
	Mean	Std.Dev	V(DY)/V(EP)	τ^2	AR	Mean	Std.Dev	V(DY)/V(EP)	τ^2	AR
EP	5.77	14.787	0.0031	0.0088	-0.632	0.66	4.484	0.0019	0.0047	0.069
DY	5.02	0.824			0.917*	0.39	0.198			0.999

Period: 1986:1-1998:12

	Yearly Summary Statistics					Monthly Summary Statistics				
	Mean	Std.Dev	V(DY)/V(EP)	τ^2	AR	Mean	Std.Dev	V(DY)/V(EP)	τ^2	AR
EP	9.35	12.595	0.0034	0.0037	0.123	0.87	4.178	0.0005	0.0009	0.050
DY	3.23	0.738			0.954	0.25	0.092			0.999

Notes: The monthly and annual data is from the CRSP database. 'EP' is the equity premium and 'DY' is the dividend yield. All series are in logs and multiplied by 100. The column 'V(DY)/V(EP)' is variance of the Dividend Yield over the variance of the Equity Premium. τ^2 is obtained in the following way: Regress the Equity Premium on lagged Dividend Yield and estimate the variance of the residuals, call it σ_ε. Regress the Dividend Yield on itself (or several lags), and estimates the variance of the residuals, call it σ_u. Then $\tau^2 = \sigma_u/\sigma_\varepsilon$. AR is the largest root of the corresponding variable. All equations were run with one lag, except for the monthly Divident Yield, which was run with six lags (chosen by sequential t-tests). For the Dividend Yield, * signifies that the Augmented Dickey Fuller test cannot accept the null of the coefficient being equal to 1 at the 0.05 level.

Table 2a

The OLS estimator of β in forecasting regression

	T= 75		T= 200		T= 500	
τ	$E(\hat{\beta})$	$Var(\hat{\beta})$	$E(\hat{\beta})$	$Var(\hat{\beta})$	$E(\hat{\beta})$	$Var(\hat{\beta})$
1	1.00	2.02e-003	1.00	2.75e-004	1.00	4.19e-005
1.0e-001	1.01	1.92e-001	1.00	2.63e-002	1.00	4.35e-003
1.0e-002	0.98	1.96e+001	0.98	2.68e+000	1.00	4.37e-001
1.0e-003	1.14	1.96e+003	0.91	2.69e+002	1.01	4.34e+001
1.0e-004	4.20	1.93e+005	2.39	2.72e+004	0.78	4.35e+003
1.0e-005	-2.43	2.00e+007	0.63	2.69e+006	6.68	4.51e+005

Table 2b

The OLS estimator of the root in X_t

	T= 75		T= 200		T= 500	
τ	$E(\hat{\phi})$	$Var(\hat{\phi})$	$E(\hat{\phi})$	$Var(\hat{\phi})$	$E(\hat{\phi})$	$Var(\hat{\phi})$
1	0.98	1.69e-003	0.99	2.51e-004	1.00	4.05e-005
1.0e-005	0.98	1.74e-003	0.99	2.51e-004	1.00	4.05e-005

Table 2c

The OLS estimator of the root in Y_t

	T= 75		T= 200		T= 500	
τ	$E(\hat{\phi}_2)$	$Var(\hat{\phi}_2)$	$E(\hat{\phi}_2)$	$Var(\hat{\phi}_2)$	$E(\hat{\phi}_2)$	$Var(\hat{\phi}_2)$
1	0.82	1.43e-002	0.93	2.99e-003	0.97	5.91e-004
1.0e-001	0.08	1.98e-002	0.21	2.19e-002	0.39	3.00e-002
1.0e-002	-0.01	1.35e-002	-0.00	4.98e-003	0.01	2.07e-003
1.0e-003	-0.01	1.33e-002	-0.00	5.03e-003	-0.00	2.01e-003
1.0e-004	-0.01	1.33e-002	-0.01	5.07e-003	-0.00	1.99e-003
1.0e-005	-0.01	1.36e-002	-0.01	4.91e-003	-0.00	2.03e-003

Notes: The system above is simulated, where τ is a fixed small number. The correlation between u_t and ϵ_t is zero, so there is no bias. The system is simulated 10000 times, for each specification of (τ, T). The true value of β is 1.

Table 2d

Mean of t-stat under Null and Alternative

τ	T= 75 null	T= 75 alt	T= 200 null	T= 200 alt	T= 500 null	T= 500 alt
1	0.02	28.41	-0.03	75.50	0.00	188.19
1.0e-001	0.01	2.82	0.00	7.56	-0.01	18.84
1.0e-002	0.01	0.28	-0.03	0.74	0.01	1.89
1.0e-003	-0.01	0.02	-0.00	0.07	0.01	0.19
1.0e-004	0.02	0.01	0.01	0.02	-0.01	0.01
1.0e-005	-0.01	-0.00	0.01	0.01	0.03	0.02

Table 2e

Mean of R^2

τ	T= 75	T= 200	T= 500
1	1.01	1.06	1.09
1.0e-001	0.13	0.25	0.44
1.0e-002	0.02	0.01	0.01
1.0e-003	0.02	0.01	0.00
1.0e-004	0.02	0.01	0.00
1.0e-005	0.02	0.01	0.00

Table 2f

Comparison of MSE's from $\bar{Y}$ and $\hat{Y}_{T+k|T}$

	T= 75			T= 200			T= 500		
τ	$MSE(\bar{Y})$	$MSE(\hat{Y}_{T+k\|T})$	Ratio	$MSE(\bar{Y})$	$MSE(\hat{Y}_{T+k\|T})$	Ratio	$MSE(\bar{Y})$	$MSE(\hat{Y}_{T+k\|T})$	Ratio
1	0.393	0.055	7.156	0.377	0.049	7.622	0.377	0.050	7.591
1.0e-001	0.018	0.015	1.205	0.009	0.006	1.546	0.006	0.002	2.358
1.0e-002	0.013	0.014	0.971	0.005	0.005	0.993	0.002	0.002	1.011
1.0e-003	0.013	0.014	0.972	0.005	0.005	0.992	0.002	0.002	0.997
1.0e-004	0.013	0.014	0.972	0.005	0.005	0.988	0.002	0.002	0.993
1.0e-005	0.014	0.014	0.970	0.005	0.005	0.988	0.002	0.002	0.995

Notes: The system above is simulated, where τ is a fixed small number. The correlation between u_t and ϵ_t is zero, so there is no bias. The system is simulated 10000 times, for each specification of (τ, T). The true value of β is 1.

Table 3a

The OLS estimator of β in forecasting regression

	T= 75		T= 200		T= 500	
τ	$E(\hat{\beta})$	$Var(\hat{\beta})$	$E(\hat{\beta})$	$Var(\hat{\beta})$	$E(\hat{\beta})$	$Var(\hat{\beta})$
1	0.97	2.22e-003	0.99	3.17e-004	0.99	5.36e-005
1.0e-001	0.65	2.28e-001	0.87	3.35e-002	0.95	5.12e-003
1.0e-002	-2.58	2.40e+001	-0.34	3.28e+000	0.46	5.37e-001
1.0e-003	-34.00	2.37e+003	-12.33	3.32e+002	-4.41	5.33e+001
1.0e-004	-346.55	2.28e+005	-132.20	3.18e+004	-53.32	5.44e+003
1.0e-005	-3415.52	2.25e+007	-1299.23	3.16e+006	-531.64	5.23e+005

Table 3b

The OLS estimator of the root in X_t

	T= 75		T= 200		T= 500	
τ	$E(\hat{\phi})$	$Var(\hat{\phi})$	$E(\hat{\phi})$	$Var(\hat{\phi})$	$E(\hat{\phi})$	$Var(\hat{\phi})$
1	0.98	1.57e-003	0.99	2.38e-004	1.00	4.16e-005
1.0e-005	0.98	1.54e-003	0.99	2.39e-004	1.00	3.96e-005

Table 3c

The OLS estimator of the root in Y_t

	T= 75		T= 200		T= 500	
τ	$E(\hat{\phi}_2)$	$Var(\hat{\phi}_2)$	$E(\hat{\phi}_2)$	$Var(\hat{\phi}_2)$	$E(\hat{\phi}_2)$	$Var(\hat{\phi}_2)$
1	0.87	8.60e-003	0.95	1.50e-003	0.98	2.81e-004
1.0e-001	0.09	2.03e-002	0.22	2.30e-002	0.40	2.99e-002
1.0e-002	-0.01	1.33e-002	-0.00	5.04e-003	0.01	2.11e-003
1.0e-003	-0.01	1.33e-002	-0.00	4.92e-003	-0.00	2.07e-003
1.0e-004	-0.01	1.33e-002	-0.01	4.93e-003	-0.00	1.99e-003
1.0e-005	-0.01	1.31e-002	-0.01	5.08e-003	-0.00	1.97e-003

Notes: The system above is simulated, where τ is a fixed small number. The correlation between v_t and ϵ_t is 0.50, so there is bias. The system is simulated 10000 times, for each specification of (τ, T). The true value of β is 1.

Table 3d

Mean of t-stat under Null and Alternative

τ	T= 75 null	T= 75 alt	T= 200 null	T= 200 alt	T= 500 null	T= 500 alt
1	-1.45	27.59	-1.46	75.39	-1.44	187.85
1.0e-001	-1.47	2.07	-1.40	6.80	-1.44	18.12
1.0e-002	-1.48	-0.49	-1.45	-0.01	-1.41	1.13
1.0e-003	-1.46	-0.73	-1.43	-0.68	-1.44	-0.58
1.0e-004	-1.48	-0.76	-1.42	-0.76	-1.44	-0.75
1.0e-005	-1.46	-0.75	-1.40	-0.74	-1.47	-0.77

Table 3e

Mean of R^2

τ	T= 75	T= 200	T= 500
1	1.00	1.06	1.09
1.0e-001	0.09	0.22	0.42
1.0e-002	0.02	0.01	0.01
1.0e-003	0.02	0.01	0.00
1.0e-004	0.02	0.01	0.00
1.0e-005	0.02	0.01	0.00

Table 3f

Comparison of MSE's from $\bar{Y}$ and $\hat{Y}_{T+k|T}$

	T= 75			T= 200			T= 500		
τ	$MSE(\bar{Y})$	$MSE(\hat{Y}_{T+k\|T})$	Ratio	$MSE(\bar{Y})$	$MSE(\hat{Y}_{T+k\|T})$	Ratio	$MSE(\bar{Y})$	$MSE(\hat{Y}_{T+k\|T})$	Ratio
1	0.379	0.054	7.052	0.387	0.049	7.853	0.380	0.050	7.574
1.0e-001	0.016	0.014	1.170	0.009	0.006	1.529	0.006	0.003	2.277
1.0e-002	0.013	0.014	0.976	0.005	0.005	0.995	0.002	0.002	1.010
1.0e-003	0.013	0.014	0.966	0.005	0.005	0.989	0.002	0.002	0.996
1.0e-004	0.013	0.014	0.969	0.005	0.005	0.986	0.002	0.002	0.996
1.0e-005	0.013	0.014	0.975	0.005	0.005	0.993	0.002	0.002	0.996

Notes: The system above is simulated, where τ is a fixed small number. The correlation between v_t and ϵ_t is 0.50, so there is bias. The system is simulated 10000 times, for each specification of (τ, T). The true value of β is 1.

Table 4a

The OLS estimator of β in forecasting regression

	T= 75		T= 200		T= 500	
α	$E(\hat{\beta})$	$Var(\hat{\beta})$	$E(\hat{\beta})$	$Var(\hat{\beta})$	$E(\hat{\beta})$	$Var(\hat{\beta})$
0.00	1.00	2.02e-003	1.00	2.69e-004	1.00	4.39e-005
0.20	1.00	1.11e-002	1.00	2.28e-003	1.00	5.10e-004
0.50	1.00	1.52e-001	1.00	5.40e-002	1.00	2.20e-002
0.67	1.00	6.50e-001	0.99	3.19e-001	0.99	1.74e-001
1.00	0.97	1.15e+001	1.01	1.08e+001	0.94	1.05e+001
2.00	2.79	6.61e+004	-2.18	4.51e+005	18.79	2.67e+006

Table 4b

The OLS estimator of the root in X_t

	T= 75		T= 200		T= 500	
α	$E(\hat{\phi})$	$Var(\hat{\phi})$	$E(\hat{\phi})$	$Var(\hat{\phi})$	$E(\hat{\phi})$	$Var(\hat{\phi})$
0.00	0.98	1.61e-003	0.99	2.45e-004	1.00	4.18e-005
2.00	0.98	1.64e-003	0.99	2.62e-004	1.00	4.03e-005

Table 4c

The OLS estimator of the root in Y_t

	T= 75		T= 200		T= 500	
α	$E(\hat{\phi}_2)$	$Var(\hat{\phi}_2)$	$E(\hat{\phi}_2)$	$Var(\hat{\phi}_2)$	$E(\hat{\phi}_2)$	$Var(\hat{\phi}_2)$
0.00	0.82	1.44e-002	0.93	2.92e-003	0.97	5.80e-004
0.20	0.56	3.92e-002	0.70	2.35e-002	0.80	1.30e-002
0.50	0.11	2.21e-002	0.12	1.33e-002	0.13	1.02e-002
0.67	0.02	1.49e-002	0.02	5.64e-003	0.02	2.28e-003
1.00	-0.01	1.32e-002	-0.00	5.00e-003	-0.00	2.00e-003
2.00	-0.01	1.32e-002	-0.01	5.08e-003	-0.00	1.95e-003

Notes: The system above is simulated, where α is fixed. The correlation between u_t and ϵ_t is zero, so there is no bias. The system is simulated 10000 times, for each specification of (α, T). The true value of β is 1.

Table 4d

Mean of t-stat under Null and Alternative

	T= 75		T= 200		T= 500	
α	null	alt	null	alt	null	alt
0.00	-0.01	27.98	-0.01	75.81	-0.01	188.97
0.20	0.01	12.01	0.01	26.03	-0.03	54.61
0.50	-0.02	3.28	0.05	5.35	0.02	8.43
0.67	0.02	1.58	-0.01	2.16	-0.01	2.94
1.00	-0.02	0.37	0.01	0.38	-0.02	0.36
2.00	0.01	0.01	-0.02	-0.01	0.03	0.01

Table 4e

Mean of R^2

α	T= 75	T= 200	T= 500
0.00	1.01	1.07	1.09
0.20	0.70	0.81	0.90
0.50	0.16	0.15	0.15
0.67	0.06	0.03	0.02
1.00	0.02	0.01	0.00
2.00	0.02	0.01	0.00

Table 4f

Comparison of MSE's from $\bar{Y}$ and $\hat{Y}_{T+k|T}$

	T= 75			T= 200			T= 500		
α	$MSE(\bar{Y})$	$MSE(\hat{Y}_{T+k\|T})$	Ratio	$MSE(\bar{Y})$	$MSE(\hat{Y}_{T+k\|T})$	Ratio	$MSE(\bar{Y})$	$MSE(\hat{Y}_{T+k\|T})$	Ratio
0.00	0.382	0.054	7.055	0.389	0.050	7.743	0.381	0.050	7.604
0.20	0.460	0.122	3.768	0.416	0.087	4.775	0.401	0.070	5.719
0.50	1.405	1.089	1.290	1.362	1.056	1.289	1.410	1.063	1.327
0.67	1.108	1.042	1.063	1.081	1.033	1.047	1.033	0.994	1.039
1.00	1.029	1.048	0.982	1.030	1.044	0.987	1.022	1.026	0.996
2.00	1.008	1.040	0.969	1.016	1.025	0.991	0.980	0.981	0.998

Notes: The system above is simulated, where α is fixed. The correlation between u_t and ϵ_t is zero, so there is no bias. The system is simulated 10000 times, for each specification of (α, T). The true value of β is 1.

Table 5a

The OLS estimator of β in forecasting regression

	T= 75		T= 200		T= 500	
α	$E(\hat{\beta})$	$Var(\hat{\beta})$	$E(\hat{\beta})$	$Var(\hat{\beta})$	$E(\hat{\beta})$	$Var(\hat{\beta})$
0.00	0.97	2.22e-003	0.99	3.17e-004	0.99	5.36e-005
0.20	0.92	1.28e-002	0.96	2.79e-003	0.98	6.14e-004
0.50	0.69	1.80e-001	0.81	6.56e-002	0.88	2.68e-002
0.67	0.37	7.71e-001	0.54	4.03e-001	0.65	2.21e-001
1.00	-1.61	1.28e+001	-1.66	1.27e+001	-1.72	1.36e+001
2.00	-191.18	7.13e+004	-519.09	5.06e+005	-1330.59	3.27e+006

Table 5b

The OLS estimator of the root in X_t

	T= 75		T= 200		T= 500	
α	$E(\hat{\phi})$	$Var(\hat{\phi})$	$E(\hat{\phi})$	$Var(\hat{\phi})$	$E(\hat{\phi})$	$Var(\hat{\phi})$
0.00	0.98	1.57e-003	0.99	2.38e-004	1.00	4.16e-005
2.00	0.98	1.54e-003	0.99	2.39e-004	1.00	3.96e-005

Table 5c

The OLS estimator of the root in Y_t

	T= 75		T= 200		T= 500	
α	$E(\hat{\phi}_2)$	$Var(\hat{\phi}_2)$	$E(\hat{\phi}_2)$	$Var(\hat{\phi}_2)$	$E(\hat{\phi}_2)$	$Var(\hat{\phi}_2)$
0.00	0.87	8.60e-003	0.95	1.50e-003	0.98	2.81e-004
0.20	0.61	3.64e-002	0.74	2.07e-002	0.82	1.10e-002
0.50	0.11	2.30e-002	0.13	1.38e-002	0.13	1.05e-002
0.67	0.02	1.46e-002	0.02	5.56e-003	0.02	2.41e-003
1.00	-0.01	1.34e-002	-0.00	4.94e-003	-0.00	1.99e-003
2.00	-0.01	1.31e-002	-0.01	5.08e-003	-0.00	1.97e-003

Notes: The system above is simulated, where α is fixed. The correlation between v_t and ϵ_t is 0.50, so there is bias. The system is simulated 10000 times, for each specification of (α, T). The true value of β is 1.

Table 5d

Mean of t-stat under Null and Alternative

	T= 75		T= 200		T= 500	
α	null	alt	null	alt	null	alt
0.00	-1.45	27.59	-1.46	75.39	-1.44	187.85
0.20	-1.47	11.16	-1.40	25.43	-1.44	53.74
0.50	-1.48	2.47	-1.45	4.59	-1.41	7.68
0.67	-1.46	0.81	-1.43	1.41	-1.44	2.16
1.00	-1.48	-0.38	-1.42	-0.39	-1.44	-0.40
2.00	-1.46	-0.74	-1.40	-0.74	-1.47	-0.77

Table 5e

Mean of R^2

α	T= 75	T= 200	T= 500
0.00	1.00	1.06	1.09
0.20	0.66	0.79	0.90
0.50	0.12	0.12	0.13
0.67	0.03	0.02	0.02
1.00	0.02	0.01	0.00
2.00	0.02	0.01	0.00

Table 5f

Comparison of MSE's from $\bar{Y}$ and $\hat{Y}_{T+k|T}$

	T= 75			T= 200			T= 500		
α	$MSE(\bar{Y})$	$MSE(\hat{Y}_{T+k\|T})$	Ratio	$MSE(\bar{Y})$	$MSE(\hat{Y}_{T+k\|T})$	Ratio	$MSE(\bar{Y})$	$MSE(\hat{Y}_{T+k\|T})$	Ratio
0.00	0.379	0.054	7.052	0.387	0.049	7.853	0.380	0.050	7.574
0.20	0.439	0.119	3.692	0.409	0.087	4.690	0.403	0.072	5.590
0.50	1.297	1.064	1.220	1.357	1.060	1.280	1.348	1.043	1.293
0.67	1.071	1.056	1.014	1.049	1.020	1.029	1.047	1.020	1.027
1.00	1.007	1.040	0.968	1.002	1.015	0.986	0.995	0.999	0.996
2.00	1.005	1.031	0.975	1.008	1.015	0.993	1.003	1.007	0.996

Notes: The system above is simulated, where α is fixed. The correlation between v_t and ϵ_t is 0.50, so there is bias. The system is simulated 10000 times, for each specification of (α, T). The true value of β is 1.