Maximum likelihood estimation of the equity premium*

Efstathios Avdis
University of Alberta

Jessica A. Wachter
University of Pennsylvania
and NBER

July 2, 2013

Abstract

The sample average of the expected return on stocks over bonds is 5.2% as measured by log returns over the postwar period. We write down a simple specification for the expected excess return that allows for predictability. Estimating this specification by maximum likelihood leads to an economically significant reduction in expected excess return: to 3.9%. Using simulations, we show that our method is substantially less noisy in finite samples than the traditional sample mean.

*First draft: July 2, 2013. Avdis: avdis@ualberta.ca; Wachter: jwachter@wharton.upenn.edu.
We are grateful for seminar participants at the Wharton School for helpful comments.

1 Introduction

The equity premium, namely the expected return on equities less the riskfree rate, is an important economic quantity for many reasons. It is an input into the decision process of individual investors as they determine their asset allocation between stocks and bonds. It is also a part of cost-of-capital calculations and thus investment decisions by firms. Finally, financial economists use it to calibrate and to test, both formally and informally, models of asset pricing and of the macroeconomy.[1]

The equity premium is usually estimated by taking the sample mean of stock returns and subtracting a measure of the riskfree rate such as the average Treasury Bill return. As is well known (Merton, 1980), it is difficult to estimate the mean of a stochastic process. If one is computing the sample average, a tighter estimate can be obtained only by extending the data series in time which has the disadvantage that the data are potentially less relevant to the present day.

Given these challenges, it is not surprising that a number of studies investigate how to estimate the equity premium using techniques other than taking the sample average. These include making use of survey evidence (Claus and Thomas, 2001; Graham and Harvey, 2005; Welch, 2000), as well as data on the cross section (Polk, Thompson, and Vuolteenaho, 2006). The branch of the literature most closely related to our work uses the accounting identity that links prices, dividends, and returns. This work includes Blanchard (1993), Fama and French (2002) and Donaldson, Kamstra, and Kramer (2010). The idea is simple in principle, but the implementation is inherently complicated by the fact that the formula for returns is additive, while incorporating estimates of future dividend growth requires multi-

[1]See, for example, the classic paper of Mehra and Prescott (1985), and surveys such as Kocherlakota (1996), Campbell (2003), DeLong and Magin (2009), Mehra and Prescott (2003).

year discount rates which are multiplicative.[2] As DeLong and Magin (2009) discuss in a survey of the literature, it is not clear why such methods would necessarily improve the estimation of the equity premium.

In this paper, we propose a method of estimating the equity premium that incorporates additional information contained in the time series of prices and dividends in a simple and econometrically-motivated way. Like the papers above, our work relies on a long-run relation between prices, returns and dividends. Our implementation is different, and grows directly out of maximum likelihood estimation of autoregressive processes. First, we show that our method yields an economically significant difference in the estimation of the equity premium. Taking the sample average of monthly log returns and subtracting the monthly log return on the Treasury bill over the postwar period implies a monthly equity premium of 0.43%. Our maximum likelihood approach implies an equity premium of 0.32%. In annual terms, these translate to 5.2% and 3.9% respectively. Assuming that returns are approximately lognormally distributed, we can also derive implications for the equity premium computed in levels: in monthly terms the sample average implies an equity premium of 0.53%, or 6.37% per annum, while maximum likelihood implies an equity premium of 0.42% per month, or 5.06% per annum.

Besides showing that our method yields economically significant differences, we also perform a Monte Carlo experiment to demonstrate that, in finite samples and under a number of different assumptions on the data generating process, the maximum likelihood method is substantially less noisy than the sample average. For example, under our baseline simulation, the sample average has a standard

[2]Fama and French (2002) have a relatively simple implementation in that they replace price appreciation by dividend growth in the expected return equation. We will discuss their paper in more detail below.

error of 0.087%, while our estimator has a standard error of only 0.050%.

Further, we derive formulas that give the intuition for our results. Maximum likelihood allows additional information to be extracted from the level of the predictor series. In the postwar sample, this additional information implies that shocks to the dividend-price ratio have on average been negative. In contrast, ordinary least squares (OLS) implies that the shocks are zero on average by definition. Because shocks to the dividend-price ratio are negatively correlated with shocks to returns, our results imply that shocks to returns must have been positive over the time period. Thus maximum likelihood implies an equity premium that is below the sample average.

Given this intuition, we show by Monte Carlo simulations that the effect of our procedure is greater the more persistent is the predictor variable. Interestingly, we also find that while the mean of the predictor variable is harder to estimate for greater persistence, there is a parameter region for which the equity premium becomes easier to estimate for greater persistence. Finally, we also use our framework to demonstrate that when there is a persistent component to the equity premium, finite-sample measures of return variance are biased downward; as the persistence increases this bias becomes severe.

The remainder of our paper proceeds as follows. Section 2 describes our statistical model and estimation procedure. Section 3 describes our results. Section 4 describes the intuition for our results and how they vary with the persistence of the state variable. Section 4 also describes the bias in variance of returns that results from the persistent component. Section 5 concludes.

2 Statistical Model and Estimation

2.1 Statistical model

Let R_{t+1} denote net returns on an equity index between t and $t+1$, and $R_{f,t+1}$ denote net riskfree returns between t and $t+1$. We let $r_{t+1} = \log(1+R_{t+1}) - \log(1+R_{f,t+1})$. We assume

$$r_{t+1} - \mu_r = \beta(x_t - \mu_x) + u_{t+1} \tag{1a}$$

$$x_{t+1} - \mu_x = \theta(x_t - \mu_x) + v_{t+1} \tag{1b}$$

where conditionally on $(r_1, \ldots, r_t, x_0, \ldots, x_t)$, the vector of shocks $[u_{t+1}, v_{t+1}]^\top$ has a bivariate normal distribution with zero mean and covariance matrix

$$\Sigma = \begin{bmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{bmatrix}.$$

We assume throughout that the dividend-price ratio follows a stationary process, namely, that $\theta < 1$. Note that our assumptions on the shocks imply that μ_r is the equity premium and that μ_x is the mean of the predictor variable. While we focus on the case that the shocks are normally distributed, we also explore robustness to alternative distributional assumptions. In our empirical implementation, we will take x_t to be the log of the dividend-price ratio.

Our assumptions on the return and predictor process are standard in the literature. Indeed, the first equation is equivalent to the standard ordinary least squares regression that has been a focus of measuring predictability in stock returns for 30 years (Keim and Stambaugh, 1986; Fama and French, 1989). We have simply rearranged the parameters so that the mean excess return μ_r appears explicitly as a parameter. In a time-series setting in which the shocks are generally correlated with future values of the right-hand-side variable, the properties of the OLS estimate of β depend on assumptions on x_t even if the estimate itself does not. The

first-order autoregressive assumption in (1b) is usual in cases where modeling x_t is necessary.[3] Our assumption that the dividend-price ratio is stationary is subject to debate and we do not make it lightly. However, we follow much recent work in financial econometrics in cautiously making this assumption and seeing how far it will take us.[4] Moreover, current economic models of the dividend-price ratio imply that this ratio is stationary. Prior studies of this system of equations have focused on the bias in estimating β; we focus on the estimation of the mean μ_r. We do not claim that standard estimates of μ_r are biased, but rather that incorporating information from the above system of equations can lead to improvements in efficiency.[5]

2.2 Estimation procedure

We estimate the parameters μ_r, μ_x, β, θ, σ_u^2, σ_v^2 and σ_{uv} by maximum likelihood. The assumption on the shocks implies that, conditional on the first observation x_0, the likelihood function is given by

$$p\left(r_1,\ldots,r_T;x_1,\ldots,x_T|\mu_r,\mu_x,\beta,\theta,\Sigma,x_0\right) = \\ |2\pi\Sigma|^{-\frac{T}{2}}\exp\left\{-\frac{1}{2}\left(\frac{\sigma_v^2}{|\Sigma|}\sum_{t=1}^{T}u_t^2-2\frac{\sigma_{uv}}{|\Sigma|}\sum_{t=1}^{T}u_tv_t+\frac{\sigma_u^2}{|\Sigma|}\sum_{t=1}^{T}v_t^2\right)\right\}. \quad (2)$$

Maximizing this likelihood function is equivalent to running ordinary least squares regression. Not surprisingly, maximizing the above requires choosing means and predictive coefficients to minimize the sum of squares of the error terms u_t and v_t.

[3]See, for example, Kandel and Stambaugh (1996), Campbell and Viceira (1999), Stambaugh (1999) and Barberis (2000).

[4]See, for example, Fama and French (2002), Lewellen (2004), Cochrane (2008), Van Binsbergen and Koijen (2010).

[5]Estimation of the mean and of the predictive coefficient are related, in that the bias in β arises from the bias in θ, which in turn arises from the need to estimate μ_x (Andrews, 1993).

This likelihood function, however, ignores the information contained in the initial draw x_0. For this reason, studies have proposed a likelihood function that incorporates the first observation (Box and Tiao, 1973; Poirier, 1978), assuming that it is a draw from the stationary distribution. In our case, the stationary distribution of x_0 is normal with mean μ_x and variance

$$\sigma_x^2 = \frac{\sigma_v^2}{1-\theta^2},$$

see Hamilton (1994). The resulting likelihood function is

$$p\left(r_1,\ldots,r_T;x_0,\ldots,x_T|\mu_r,\mu_x,\beta,\theta,\Sigma\right) = \left(2\pi\sigma_x^2\right)^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}\left(\frac{x_0-\mu_x}{\sigma_x}\right)^2\right\}\times$$
$$|2\pi\Sigma|^{-\frac{T}{2}}\exp\left\{-\frac{1}{2}\left(\frac{\sigma_v^2}{|\Sigma|}\sum_{t=1}^{T}u_t^2 - 2\frac{\sigma_{uv}}{|\Sigma|}\sum_{t=1}^{T}u_t v_t + \frac{\sigma_u^2}{|\Sigma|}\sum_{t=1}^{T}v_t^2\right)\right\}. \quad (3)$$

We follow Box and Tiao in referring to (2) as the conditional likelihood and (3) as the exact likelihood. Other recent work that makes use of the exact likelihood in predictive regressions includes Stambaugh (1999) and Wachter and Warusawitharana (2009, 2012), who focus on estimation of the predictive coefficient β.[6]

We derive the values of μ_r, μ_x, β, θ, σ_u^2, σ_v^2 and σ_{uv} that maximize this likelihood by solving a set of first-order conditions. We give closed-form expressions for each maximum likelihood estimate in the Appendix. Our solution amounts to solving a polynomial for the autoregressive coefficient θ, after which the estimate of every other parameter unravels easily. That this method does not feature numerical optimization makes our estimates computationally expedient.

The main comparison we carry out in this paper is between estimating the equity premium using the sample mean versus maximum likleihood. At the same

[6]Wachter and Warusawitharana (2009, 2012) also use Bayesian methods rather than maximum likelihood.

time, we compare estimates of the full parameter vector as follows. For the mean return and mean predictor, we compare the maximum likelihood estimates $\hat{\mu}_r$ and $\hat{\mu}_x$ to the sample means $\bar{\mu}_r$ and $\bar{\mu}_x$. For the predictability coefficient and the predictor persistence, we compare the maximum likelihood estimates $\hat{\beta}$ and $\hat{\theta}$ to the OLS estimates β^{OLS} and θ^{OLS}. For the covariance matrix, we report estimates of the standard deviations σ_u, σ_v and the correlation ρ_{uv} between u_t and v_t by backing them out of the maximum likelihood and OLS estimates for σ_u^2, σ_v^2 and σ_{uv}.[7]

2.3 Data

We calculate maximum likelihood estimates of the parameters in our predictive system for the excess return of the value-weighted market portfolio from CRSP. Motivated by Campbell and Shiller (1988) we apply our predictability framework on the logarithm of the return and with the logarithm of the dividend-price ratio as a predictor. The analysis we conduct, however, applies in principle to any variable used as a predictor of returns. In particular, we define r_t to be the logarithm of the gross return in excess of the risk free asset, $r_t = \log(1 + R_t) - \log(1 + R_t^f)$. We take R_t to be the cum-dividend monthly net return of the value-weighted market portfolio and R_t^f to be the monthly net return of the 30-day Treasury Bill. We construct the monthly dividends series of the market portfolio from CRSP data by summing over dividend payouts over the concurrent month and the previous eleven months.

The summary statistics of returns and the dividend-price ratio appear in Ta-

[7]Our maximum likelihood estimates for the entries of the covariance matrix are $\hat{\sigma}_u^2$, $\hat{\sigma}_v^2$ and $\hat{\sigma}_{uv}$. Given these, we report $\sqrt{\hat{\sigma}_u^2}$, $\sqrt{\hat{\sigma}_v^2}$ and $\hat{\sigma}_{uv}/\sqrt{\hat{\sigma}_u^2\hat{\sigma}_v^2}$ as estimates of σ_u, σ_v and the correlation ρ_{uv}.

ble 1, where we report monthly quantities. The estimated equity premium in levels using the sample average is 7.38% per annum for the period January 1927 to December 2011. For the postwar period January 1953 to December 2011 it is 6.38% per annum. For log returns these sample averages are 5.57% per annum for the long sample and 5.20% per annum for the post-war sample.

3 Results

3.1 Point estimates

Our first main result is that the maximum likelihood estimate for the equity premium is substantially lower than the sample average. We summarize this in Table 2, where we report estimates of the parameters in our predictive regression system for log returns at a monthly frequency. For the postwar period of 1953 to 2011, the estimated mean excess monthly return using maximum likelihood is 3.86% in annualized terms. This implies that our estimate of the equity premium is 133 basis points lower per annum than the sample average. Using our maximum likelihood method on the 1927–2011 sample yields an estimated mean of 4.69% per annum, 88 basis points lower than the sample average.

Table 2 also reports results for maximum likelihood estimation of the predictive coefficient β, the autoregressive coefficient θ, and the standard deviations and correlation between the shocks. The estimation of the standard deviations and correlation ρ are nearly identical across the two methods, not surprisingly, because these tend to be estimated precisely in monthly data. Estimates for the average value of the predictor variable, the predictive coefficient and the autoregressive coefficient exhibit are noticeably different between the methods, however. The estimate for the average of the predictor variable is lower for maximum likelihood

estimation (MLE) than for OLS in both samples. The difference in the postwar data is 4 basis points, an order of magnitude smaller than the difference in the estimate of the equity premium. Nonetheless, the two results are closely related, as we will discuss in what follows

As mentioned in the introduction, the estimation of the predictive coefficient β and its relation to the autoregressive coefficient θ is itself the subject of a large literature, and is not the focus of our manuscript. Table 2 shows that maximum likelihood implies a postwar estimate of β of 0.69, lower than the OLS value of 0.83. Because OLS is biased upward, the fact that our method generates a lower value for β is intriguing; however the result is sample-dependent. In the longer sample, the estimate for β generated by maximum likelihood is in fact slightly higher than the OLS estimate. The estimates for θ vary in the opposite direction to the estimates for β: in the postwar sample the estimate for θ is (slightly) higher, while in the longer sample it is (slightly) lower.

3.2 Efficiency

We next evaluate efficiency. Asymptotically, maximum likelihood is known to be the most efficient estimation method, and so in large samples (assuming that the specification is correct), our method is guaranteed to be more efficient than taking the sample average. However, because our method requires a nonlinear optimization, it is possible that this asymptotic result does not extend to small samples. The asymptotic result may also not be robust to a reasonable degree of mis-specification. We investigate both of these issues.

We simulate 10,000 samples of excess returns and predictor variables, each of length equal to the data. Namely, we simulate from (1), setting parameter values equal to their maximum likelihood estimates, and, for each sample, initializing

x using a draw from the stationary distribution. For each simulated sample, we calculate sample averages, OLS estimates and our maximum likelihood estimates, generating a distribution of these estimates over the 10,000 paths. Table 3 reports the standard deviations, medians, and 5th and 95th percentile values. Panel A shows the results of the exercise designed to capture the postwar sample, while Panel B shows the results for the longer sample. Panel A shows that the sample average has a standard deviation of 0.087. In contrast, the maximum likelihood estimate has a standard deviation of only 0.05.[8] Panel B shows an economically significant decline in standard deviation for the long sample as well: the standard deviation falls from 0.080 to 0.058. It is noteworthy that our results still hold in the longer sample, indicating that our method has value even when there is a large amount of data available to estimate the sample mean. Besides lower standard deviations, the maximum likelihood estimates also have a tighter distribution. For example, the 95th percentile value for the sample mean of returns is 0.47 , while the 95th percentile value for the maximum likelihood estimate is 0.40 (in monthly terms, the value of the maximum likelihood estimate is 0.32). The 5th percentile is 0.18 for the sample average but 0.24 for the maximum likelihood estimates.

Table 3 shows that the maximum likelihood estimate of the mean of the predictor also has a lower standard deviation and tighter confidence intervals than the sample average, though the difference is much less pronounced. Similarly, the maximum likelihood estimates of the regression coefficients θ and ρ also have smaller standard deviations and confidence intervals than the OLS estimates, though again, the differences for these parameters between MLE and OLS are not large. The results in this table show that, in terms of the parameters of this

[8]Note that this implies that the sample mean and the mean from MLE differ by more than two standard deviations in the postwar sample.

system at least, the equity premium is unique in the potential improvement offered by maximum likelihood. This is no doubt in part because estimation of first moments is more difficult than that of second moments in the time series (Merton, 1980). This doesn't explain the difference between the results for returns and for the predictor variable however. We return to this issue in Section 4.

Figure 1 provides another view of the difference between the sample mean and the maximum likelihood estimate of the equity premium. The solid line shows the probability density of the maximum likelihood estimates while the dashed line shows the probability density of the sample mean.[9] Results are shown for the postwar period. The distribution of the maximum likelihood estimate is visibly more concentrated around the true value of the equity premium, and the tails of this distribution fall well under the tails of the distribution of sample means.

In Table 3, we used coefficients estimated by maximum likelihood to evaluate whether MLE is more efficient than OLS. Perhaps it is not surprising that MLE delivers better estimates, if we use the maximum likelihood estimates themselves in the simulation. However, Table 4 shows nearly identical results from setting the parameters equal to their sample means and OLS estimates.

It is well known that OLS estimates of predictive coefficients can be severely biased (Stambaugh, 1999). Tables 3 and 4 replicate this result. For example, in the simulation in Table 3, the "true" value of the predictive coefficient β in the simulated data is 0.69, however, the median OLS value from the simulated samples is 1.16. That is, OLS estimates the predictive coefficient to be much higher than the true value, and thus the predictive relation to be stronger. The bias in the predictive coefficient is associated with bias in the autoregressive coefficient on the dividend yield. The true value of θ in the simulated data is 0.993, but the

[9]Both densities are computed non-parametrically and smoothed by a normal kernel.

median OLS value is 0.988.[10]. Maximum likelihood reduces the bias somewhat: the median maximum likelihood estimate of β is 1.10 as opposed to 1.16, but it does not eliminate it entirely.[11]

These results suggest that 0.69 is probably not a good estimate of β, and likewise, 0.993 is likely not to be a good estimate of θ. Does the superior performance of maximum likelihood continue to hold if these estimates are corrected for bias? We turn to this question next. We repeat the exercise described above, but instead of using the maximum likelihood estimates, we adjust the values of β and θ so that median computed across the simulated samples matches the observed value in the data. The results are gigen in Panel A of Table 5. This adjustment lowers β and increases θ, but does not change the median maximum likelihood estimate of the equity premium. If anything, adjusting for biases shows that we are being conservative in how much more efficient our method of estimating the equity premium is in comparison to using the sample average. After accounting for biases, maximum likelihood gives an equity premium estimate with standard deviation about 40% of that of the standard deviation of the sample mean excess return for the postwar sample.

In Panel B of Table 5 we conduct an additional robustness check. Here, we check the impact of fat-tailed shocks on the efficiency of our method. We simulate system (1) under the assumption that the shocks u_t and v_t are distributed as a bivariate Student's t distribution with a common degree of freedom ν. We set the true value of ν in our simulations by using the feature that the excess kurtosis of a t random variable equals $6/(\nu - 4)$ to back out an estimated ν. In particular,

[10]These tables also show a downward bias in σ_u, the estimate of return shocks. We return to this issue in Section 4

[11]The estimates of the equity premium are not biased however; the median for both OLS and the sample average is close to the mean.

we measure the kurtosis of the estimated residuals in the return and predictor regressions from our maximum likelihood estimation and take the average of the two numbers to be the common kurtosis of u_t and v_t. In the postwar sample the kurtosis of the residual to the return regression is 5.76 and the kurtosis of the residual to the predictor regression is 5.43, giving an estimated kurtosis of 5.60. We match this number to the median kurtosis of the residuals in our simulations by adjusting the ν parameter of the simulated t shocks.[12] In addition, the true values of the parameters we use in our simulations have been adjusted to account for estimation biases as above. Our results show that the efficiency gain of our MLE method is virtually unchanged by the fat tails in the shocks.

3.3 The equity premium in levels

So far we have used log returns to provide an estimate of the equity premium. In what sense is our lower estimate representative of the equity premium using return levels? If we assume that the log returns $\log(1+R_t)$ are normally distributed we can write

$$\mathbb{E}[R_t] = \mathbb{E}\left[e^{\log(1+R_t)}\right] - 1 = e^{\mathbb{E}[\log(1+R_t)]+\frac{1}{2}\mathrm{Var}(\log(1+R_t))} - 1.$$

Using the definition of the excess log return, $\mathbb{E}[\log(1+R_t)] = \mathbb{E}[r_t]+\mathbb{E}[\log(1+R_t^f)]$, so the above implies that

$$\mathbb{E}[R_t - R_t^f] = e^{\mathbb{E}[r_t]}e^{\mathbb{E}\left[\log\left(1+R_t^f\right)\right]+\frac{1}{2}\mathrm{Var}(\log(1+R_t))} - 1 - \mathbb{E}[R_t^f].$$

Our maximum likelihood method provides an estimate of $\mathbb{E}[r_t]$ and all other quantities above can be easily calculated using sample averages. Taking the sample

[12] The value of ν that achieves this is 5.12, which corresponds to a population kurtosis of 8.35. This shows that the kurtosis statistic is severely downward biased.

mean of the series $R_t - R_t^f$ for the period 1953-2011 yields a risk premium that is 0.530% per month, or 6.37% per annum. On the other hand, using the above calculation and our maximum likelihood estimate of the mean of r_t gives an estimate of $\mathbb{E}[R_t - R_t^f]$ of 0.422% per month, or 5.06% per annum[13]. Thus our estimate of the risk premium in return levels is 131 basis lower than taking the sample average, in line with our results for log returns.

4 Discussion

4.1 Source of the gain in efficiency

4.1.1 Predictability or shocks?

What determines the difference between the maximum likelihood estimate of the equity premium and the sample average of excess returns? Let $\hat{\mu}_r$ denote the maximum likelihood estimate of the equity premium. Given the maximum likelihood estimates, we can define a time series of shocks u_t and v_t as follows:

$$\hat{u}_t = r_t - \hat{\mu}_r - \hat{\beta}(x_{t-1} - \hat{\mu}_x), \tag{4a}$$

$$\hat{v}_t = x_t - \hat{\mu}_x - \hat{\theta}(x_{t-1} - \hat{\mu}_x). \tag{4b}$$

By definition, then,

$$\hat{\mu}_r = \frac{1}{T}\sum_{t=1}^{T} r_t - \frac{1}{T}\sum_{t=1}^{T} \hat{u}_t - \hat{\beta}\frac{1}{T}\sum_{t=1}^{T}(x_{t-1} - \hat{\mu}_x). \tag{5}$$

As (5) shows, there are two reasons why the maximum likelihood estimate of the mean, $\hat{\mu}_r$, might differ from the sample mean $\frac{1}{T}\sum_{t=1}^{T} r_t$. The first is that the

[13]In the data, in monthly terms for the period 1953-2011, the sample mean of R_t is 0.918%, the sample mean of R_t^f is 0.387%, the sample mean of $\log(1 + R_t^f)$ is 0.386% and the variance of $\log(1 + R_t)$ is 0.194%.

shocks $\hat{u}_t$ may not average to zero over the sample. The second is that, if returns are determined in part by the value of the predictor variable, and if the average value of the predictor variable is not equal to its mean over the sample, then the average return will not equal the mean.

It turns out that only the first of these effects is of any importance in explaining our result. For the period January 1953 to December 2001, the sample average $\frac{1}{T}\sum_{t=1}^{T}\hat{u}_t$ is equal to 0.1382% per month, while $\hat{\beta}\frac{1}{T}\sum_{t=1}^{T}(x_{t-1}-\hat{\mu}_x)$ is -0.0278% per month. The difference in the maximum likelihood estimate and the sample mean thus ultimately comes down to the interpretation of the shocks $\hat{u}_t$. To understand the behavior of these shocks, we will argue it is necessary to understand the behavior of the shocks $\hat{v}_t$. And, to understand $\hat{v}_t$, it is necessary to understand why the maximum likelihood estimate of the mean of x differs from the sample mean.

4.1.2 Estimation of the mean of the predictor variable

To build intuition, we consider a simpler problem in which the true value of the autocorrelation coefficient θ is known. Appendix A shows that the first-order condition in the exact likelihood function with respect to μ_x implies

$$\hat{\mu}_x = \frac{(1+\theta)}{1+\theta+(1-\theta)T}x_0 + \frac{1}{(1+\theta)+(1-\theta)T}\sum_{t=1}^{T}(x_t-\theta x_{t-1}). \tag{6}$$

We can rearrange (1b) as follows:

$$x_{t+1}-\theta x_t = (1-\theta)\mu_x + v_{t+1}.$$

Summing over t and solving for μ_x implies that

$$\mu_x = \frac{1}{1-\theta}\frac{1}{T}\sum_{t=1}^{T}(x_t-\theta x_{t-1}) - \frac{1}{T(1-\theta)}\sum_{t=1}^{T}v_t \tag{7}$$

where the shocks v_t are defined using the mean μ_x and the autocorrelation θ.

Consider the conditional maximum likelihood estimate of μ_x, the estimate that arises from maximizing the conditional likelihood (2). We will call this $\hat{\mu}_x^c$. Note that this is also equal to the OLS estimate of μ_x, which arises from estimating the intercept $(1-\theta)\mu_x$ in the regression equation

$$x_{t+1} = (1-\theta)\mu_x + \theta x_t + v_{t+1}$$

and dividing by $1-\theta$. Because it is the same as OLS, the conditional maximum likelihood estimate of μ_x is determined by the requirement that the shocks v_t average to zero. Therefore, it follows from (7) that

$$\hat{\mu}_x^c = \frac{1}{1-\theta}\frac{1}{T}\sum_{t=1}^{T}(x_t - \theta x_{t-1}).$$

Substituting back into (6) implies

$$\hat{\mu}_x = \frac{(1+\theta)}{1+\theta+(1-\theta)T}x_0 + \frac{(1-\theta)T}{(1+\theta)+(1-\theta)T}\hat{\mu}_x^c.$$

Multiplying and dividing by $1-\theta$ implies a more intuitive formula:

$$\hat{\mu}_x = \frac{1-\theta^2}{1-\theta^2+(1-\theta)^2T}x_0 + \frac{(1-\theta)^2T}{1-\theta^2+(1-\theta)^2T}\hat{\mu}_x^c. \tag{8}$$

Equation 8 shows that the exact maximum likelihood estimate is a weighted average of the first observation and the conditional maximum likelihood estimate. The weights are determined by the precision of each estimate. Recall that the assumption of stationarity implies that

$$x_0 \sim \mathcal{N}\left(0, \frac{\sigma_v^2}{1-\theta^2}\right).$$

Because the shocks v_t are independent we have that

$$\frac{1}{T(1-\theta)}\sum_{t=1}^{T} v_t \sim \mathcal{N}\left(0, \frac{\sigma_v^2}{T(1-\theta)^2}\right),$$

so $(1-\theta)^2 T$ can be viewed as proportional to the precision of this estimate, just as $1-\theta^2$ can be viewed as proportional to the precision of x_0. Note that when $\theta = 0$, there is no persistence and the weight on x_0 is $1/(T+1)$, its appropriate weight if all the observations were independent. At the other extreme, if $\theta = 1$, no information is conveyed by the shocks v_t. In this case, the "estimate" of $\hat{\mu}_x$ is simply equal to x_0.[14]

While (8) rests on the assumption that θ is known, we can use it to qualitatively understand the effect of including the first observation. Because of the information contained in x_0, we can conclude that the last T observations of the predictor variable are not entirely representative of values of the predictor variable in population. Specifically, the average predictor over the last T of $T+1$ observations are lower than a true representative sample. It follows that the predictor variable must, on average, have declined over the sample period. Thus the shocks v_t do not average to zero, as OLS (or conditional maximum likelihood) would imply, but rather, they average to a negative value. The time series of the predictor variable, shown in Figure 2, shows that the dividend-price ratio does appear to be declining over the sample period. This figure suggests why a form of estimation

[14]Our MLE of θ is 0.997 after correcting for biases for the period January 1953 to December 2011 (see Table 5). Using this value as the true value of θ means that the weight that the exact maximum likelihood estimate of μ_x places on the initial observation x_0 is 48.49%. This highlights the substantial role of the initial observation x_0 in producing a good estimate of μ_x: if we had the conditional maximum likelihood estimate μ_x^c and wanted an approximate value for the exact maximum likelihood estimate $\hat{\mu}_x$, we would need to adjust μ_x^c by weighting it roughly equally with the initial observation x_0. However, we cannot interpret (8) as precisely giving our maximum likelihood estimate, because θ is not known (more precisely, the conditional and exact maximum likelihood estimates of θ will differ). However, x_0 clearly plays a substantial role. In postwar data, $\mu_x^c = -3.695$, the initial condition is $x_0 = -2.977$ and the exact estimate is $\hat{\mu}_x = -3.504$.

that makes use of the initial condition would lead to the conclusion that shocks to the dividend-price ratio were, on average, negative.

4.1.3 Estimation of the equity premium

We now return to the problem of estimating the equity premium. Equation 5 shows the role of average shocks $\frac{1}{T}\sum_{t=1}^{T}\hat{u}_t$ in explaining the difference between the maximum likelihood estimate $\hat{\mu}_r$ and the sample mean. The OLS estimate of (1a), by definition, sets the average of these shocks to equal zero. The maximum likelihood estimate, however does not.[15]

Why then does maximum likelihood set the average value of the shocks to something other than zero? As shown in Appendix A, the first-order condition for estimation of $\hat{\mu}_r$ implies

$$\frac{1}{T}\sum_{t=1}^{T}\hat{u}_t = \frac{\hat{\sigma}_{uv}}{\hat{\sigma}_v^2}\frac{1}{T}\sum_{t=1}^{T}\hat{v}_t. \tag{9}$$

This is analogous to a result of Stambaugh (1999), in in which the error terms are replaced by the deviation of β and of θ from the true means. Equation 9 implies a connection between the average value of the shocks to the predictor variable and the average value of the shocks to returns. OLS would imply that $\frac{1}{T}\sum_{t=1}^{T}\hat{v}_t = 0$. However, as we described above, MLE does not imply that the average shocks to the predictor variable equals zero. To the extent that incorporating the first observation leads these shocks to not equal zero on average, then, because v_t is correlated with u_t, shocks to returns should also not be expected to equal zero on average.[16] Note that this result operates purely through the correlation of the

[15]Note that the OLS estimate of $\hat{\mu}_r$ is not the same as the sample average, though they will be close. The reason is that OLS adjusts the intercept in (1a) for the difference between the average of the first T observations of the predictor variable and the OLS estimate of μ_x.

[16]This point is related to the result that longer time series can help estimate parameters determined by shorter time series, as long as the errors are correlated. See Stambaugh (1997)

errors, and is not related to predictability.[17] Based on this intuition, we can label the terms in (5) as follows:

$$\hat{\mu}_r = \frac{1}{T}\sum_{t=1}^{T} r_t \quad - \quad \underbrace{\frac{1}{T}\sum_{t=1}^{T}\hat{u}_t}_{\text{Correlated error term}} \quad - \quad \underbrace{\hat{\beta}\frac{1}{T}\sum_{t=1}^{T}(x_{t-1}-\hat{\mu}_x)}_{\text{Predictability term}}. \tag{10}$$

As discussed above, the correlated error term accounts for more than 100% of our results, and is an order of magnitude larger than the predictability term. It makes sense that these would go in opposite directions: If the shocks to the dividend-price ratio were negative over the sample (as is consistent with the positive shocks to returns), then the earlier observations of x_t would tend to be above the estimated mean. Indeed, Figure 3, where we plot a scatter of the correlated error term and the predictability term, shows that this result is typical in our simulated samples. This figure shows a scatter plot of the correlated error term and the predictability term. The correlated error term tends to be much larger in magnitude then the predictability term. Moreover, the two effects are clearly negatively correlated.

Another way to see this result is to use first order conditions derived in Appendix A (Equations A.2, A.8) to rewrite the sum of errors as follows:

$$\sum_{t=1}^{T}\hat{u}_t = \left(1+\hat{\theta}\right)\frac{\hat{\sigma}_{uv}}{\hat{\sigma}_v^2}\left(\hat{\mu}_x - x_0\right).$$

This furthermore yields the relationship

$$\hat{\mu}_r = \frac{1}{T}\sum_{t=1}^{T} r_t - \frac{1+\hat{\theta}}{T}\frac{\hat{\sigma}_{uv}}{\hat{\sigma}_v^2}\left(\hat{\mu}_x - x_0\right) - \hat{\beta}\left(\frac{1}{T}\sum_{t=1}^{T} x_{t-1} - \hat{\mu}_x\right).$$

and Lynch and Wachter (2011). Here, the time series for the predictor is slightly longer than the time series of the return. Despite the small difference in the lengths of the data, the structure of the problem implies that the effect of including the full predictor variable series is very strong.

[17] Ultimately, however, the two may be related in that an accounting identity links changes in the dividend-price ratio to variation in the equity premium.

This relation between the sample mean return and the MLE of the mean return explains the difference between these two estimates in our sample. Suppose that the predictability term is small in terms of the correlated errors term, as is true in our data. Then, a small sample in a predictive regression that features negatively correlated shocks between the returns and the predictor will also have the feature

$$\hat{\mu}_r > \frac{1}{T}\sum_{t=1}^{T} r_t$$

whenever $\hat{\mu}_x > x_0$ and

$$\hat{\mu}_r < \frac{1}{T}\sum_{t=1}^{T} r_t$$

whenever $\hat{\mu}_x < x_0$. If the correlation between the return shocks and the predictor shocks was positive, the inequalities between the sample mean return and the maximum likelihood estimate of the mean return would flip. In this manner, MLE incorporates information from the path of the predictor in a way that OLS cannot. For example, if $\sigma_{uv}/\sigma_v^2 < 0$, then in small samples where the predictor has been decreasing, in which case $\hat{\mu}_x > x_0$, estimating the equity premium using the sample mean produces estimates that are higher than the true value. This is exactly what happens in the data: dividend-price ratios have been decreasing over time, and incorporating this information into the equity premium leads to a downward revision in the estimates.

This section explains the difference between the sample mean and the maximum likelihood estimate of the equity premium by appealing to the difference between the sample mean and the maximum likelihood estimate of the mean of the predictor variable. However, Table 2 shows that the difference between the sample mean of excess returns and the maximum likelihood estimate of the equity premium is many times that of the difference between the two estimates of the mean of the predictor variable. Moreover, Table 3 shows that the difference in efficiency

for returns is also much greater than the difference in efficiency for the predictor variable. How is it then that the difference in the estimates for the mean of the predictor variable could be driving the results? Equation 9 offers an explanation. Shocks to returns are far more volatile than shocks to the predictor variable. The term $\hat{\sigma}_{uv}/\hat{\sigma}_v^2$ is about -100 in the data. What seems like only a small increase in information concerning the shocks to the predictor variable translates to quite a lot of information concerning returns.

4.2 Properties of the maximum likelihood estimator

In this section we investigate the properties of the maximum likelihood estimator, and, in particular, how the variance depends on the persistence of the predictor variable and the amount of predictability. We also return to the issue of the finite-sample properties of the variance of returns themselves.

4.2.1 Variance of the estimator as a function of the persistence

The theoretical discussion in the previous section suggests that the persistence θ is the main determinant of the increase in efficiency from maximum likelihood. Figure 4 shows the standard deviation of estimators of the mean of the predictor variable (μ_x) and of estimators of the equity premium (μ_r) as functions of θ. Other parameters are set equal to their estimates from the postwar data, adjusted for bias (see Table 5). For each value of θ, we simulate 10,000 samples.

Panel A shows the standard deviation of estimators of μ_x. Both the standard deviation of both the sample mean and the maximum likelihood estimate are increasing as functions of θ. This is not surprising; holding all else equal, an increase in the persistence of θ makes the observations on the predictor variable more alike, thus decreasing their information content. The standard deviation of

the sample mean is larger than the standard deviation of the maximum likelihood estimate, indicating that our results above do not depend on a specific value of θ. Moreover, the improvement in efficiency increases as θ grows larger.

Panel B shows the standard deviation of estimators of μ_r. In contrast to the result of μ_x, the relation between the standard deviation and θ is non-monotonic for both the sample mean of excess returns and the maximum likelihood estimate of the equity premium. For values of θ below about 0.998, the standard deviations of the estimates are decreasing in θ, while for values of θ above this number they are increasing. This result is surprising given the result in Panel A. As θ increases, any given sample contains less information about the predictor variable, and thus about returns. One might expect that the standard deviation of estimators of the mean return would follow the same pattern as in Panel A. Indeed, this is the case for part of the parameter space, namely when the persistence of the predictor variable is very close to one.

However, an increase in θ has two opposing effects on the variance of the estimators. On the one hand, an increase in θ decreases the information content of the predictor variable series, and thus of the return series, as described above. On the other hand, for a given β, and increase in θ raises the R^2 in the return regression, namely it increases the relative amount of return variance that can be predicted. Moreover, innovations to the predictable part of returns are negatively correlated with innovations to the unpredictable part of returns. That is, an increase in θ increases mean reversion.

To see this, consider the effect of a series of shocks on excess returns (in this calculation, we will assume, for expositional reasons, that the mean excess return

is zero):

$$\begin{aligned} r_t &= \beta x_{t-1} + u_t \\ r_{t+1} &= \beta\theta x_{t-1} + \beta v_t + u_{t+1} \\ r_{t+2} &= \beta\theta^2 x_{t-1} + \beta\theta v_t + \beta v_{t+1} + u_{t+2} \end{aligned}$$

and so on. Thus, for $k \geq 1$, the autocovariance of returns is given by

$$\text{Cov}\left(r_t, r_{t+k}\right) = \theta^k \beta^2 \text{Var}(x_t) + \theta^{k-1}\beta\sigma_{uv}, \tag{11}$$

where $\text{Var}(x_t) = \sigma_v^2/(1-\theta^2)$. An increase in θ increases the variance of the predictor variable. In the absence of covariance between the shocks u and v, this effect would increase the autocovariance of returns. However, because u and v are negatively correlated, there is a second term. This second term is negative, and, unless θ is very high, it dominates the first term.

Because returns are mean-reverting, the sample average will be less variable. If there are returns that are unusually high compared to the mean, greater mean-reversion implies that they will be followed by returns that are unusually low compared to the mean.[18] A sequence of unusually high observations or unusually low observations are less likely to dominate in any given sample, and so the sample

[18]We can see this directly by computing the variance of the sample mean:

$$\text{Var}\left(\frac{1}{T}\sum_{t=1}^{T} r_t\right) = \frac{1}{T}\left(\sigma_u^2 + \beta^2\frac{\sigma_v^2}{1-\theta^2} + 2\beta\frac{\sigma_{uv}}{1-\theta}\right) + O\left(\frac{1}{T^2}\right)$$

(see Appendix B). The term $\sigma_u^2 + \beta^2\sigma_v^2/(1-\theta^2)$ measures the contribution of the return shocks and the predictor to the variability of the sample-mean return. The term $\beta\sigma_{uv}/(1-\theta)$ measures the contribution of the covariance of the return shocks and the predictor shocks to the variability of the sample-mean return. The former term increases as θ increases, which says that the sample-mean return is more variable because the predictor is more variable. At the same time, the latter term becomes more negative as θ increases, so that in fact the overall variability of the sample-mean return can decrease.

average will be more stable. Because the sample mean is simply the scaled long-horizon return, our result is related to the fact that mean reversion reduces the variability of long-horizon returns relative to short-horizon returns. Of course, for θ sufficiently large, the reduction in information from autocorrelation in the price-dividend ratio dominates, and both the sample mean and the maximum likelihood estimate increase. In the limit, as θ approaches one, returns become non-stationary and the sample mean has infinite variance.

Panel B of Figure 4 also shows that MLE is more efficient than the sample mean for any value of θ. The benefit of using the maximum estimate increases with θ. Indeed, while the standard deviation of the sample mean falls from 0.14 to 0.12 as θ goes from 0.980 to 0.995, the maximum likelihood estimate falls from 0.14 to 0.06. The effects appear to reinforce each other, perhaps because the samples that tend to feature a large degree of negative correlation between shocks to u and v both lead to greater mean reversion, and a greater benefit of maximum likelihood.

If we assume that returns are unpredictable, but maintain the correlation between shocks to the dividend-price ratio and shocks to returns, we have neither effect mentioned above.[19] That is, the standard deviation of the sample mean neither falls because of mean reversion, nor does it rise as the autocorrelation approaches one. Rather it remains constant, which makes sense, because the marginal distribution of returns is iid. This can be seen in Figure 5, which repeats the exercise of Figure 4, except setting the regression coefficient equal to zero. On the other hand, the benefits of the maximum likelihood estimate grow as θ increases.

This difference in the behavior of the two statistics comes from the kind of

[19] As discussed previously, the economic plausibility of this exercise is questionable. If the dividend-price ratio is variable, it seems likely that at least some of that variability would be due to change in the equity premium, in which case returns would be predictable. We consider this exercise merely to show the theoretical effect of predictability versus correlation of the shocks.

information that is incorporated in these two different measures of the mean return. The sample-mean return is simply the sum of all return observations divided by the length of the series. Unlike the sample-mean statistic, maximum likelihood estimation uses not just the return time series but also the dividend-price time series to produce a mean-return estimate. Since the shocks for the two series are correlated, so are the two time series, and thus the history of the dividend-price ratio contains return information that can be used to decrease the variability of the mean-return estimate. As we explain above, a key feature of the MLE is that it adjusts the mean-return estimate to capture stochastic properties of the predictor series. As θ increases, the specific time series of shocks matters more. Therefore, because MLE can adjust for such shocks, its variability relative to that of the sample-mean improves as θ increases.

4.2.2 Variance of the return

We conclude by using our framework to examine the properties of the variance of returns as the persistence increases to one. This is related to the discussion above because, as the return variance increases, in the absence of serial correlation at least, the variance of the mean estimators will increase as well. Our simulation setting also provides a convenient way of addressing questions of bias in the standard deviation raised in previously in the manuscript.

Figure 6 shows the standard deviation of the predictor variable (Panel A) and the excess return (Panel B) as functions of θ. The solid line shows the true value of the standard deviations. Both the standard deviation of the predictor variable and of returns increase in θ. However, the patterns differ: the standard deviation of the predictor variable increases steadily as θ approaches one, while the standard deviation of returns stays stable, and then increases relatively quickly for very high

values of θ. The reason is that the standard deviation of returns is mainly driven by the unexpected portion of returns u_t, unless θ is very high. For high values of θ there are two reinforcing effects: a greater percentage of the variance of returns is driven by the predictor, and the predictor is also more volatile.

Figure 6 also shows the mean and median values of the standard deviations, computed across the simulated samples. The figure shows a clear downward bias in the standard deviations, both for the predictor variable and for returns. Intuitively, as θ approaches one, the distribution for the return becomes nonstationary and the true variance is infinite. However, in a finite sample, it is always possible to compute a number for the variance. This bias may be especially pernicious in the case of returns, where it has little effect unless θ is very large, at which point the effect becomes dramatic. Thus, when returns have even a small persistent component, the standard deviation computed in any one sample can be severely biased downward.

4.3 Comparison with Fama and French (2002)

Fama and French (2002) also propose an estimator that takes the time series of the dividend-price ratio into account in estimating the mean. Noting the following return identity:

$$R_t = \frac{D_t}{P_{t-1}} + \frac{P_t - P_{t-1}}{P_{t-1}},$$

and rewriting in terms of expectations:

$$E[R_t] = E\left[\frac{D_t}{P_{t-1}}\right] + E\left[\frac{P_t - P_{t-1}}{P_{t-1}}\right],$$

they propose replacing the capital gain term $E[(P_t - P_{t-1})/P_{t-1}]$ with dividend growth $E[(D_t - D_{t-1})/D_{t-1}]$. They argue that, because prices and dividends are cointegrated, their means should be the same. They find that the resulting ex-

pected return is less than half the sample average: It is 4.74% rather than 9.62%, the value over their sample.

At first glance, the two findings seem like they should be closely related, since they both are based on the cointegration of dividends and prices. However, cointegration need not imply that the growth rates of dividends and prices be the same. The reason is Jensen's inequality. Note that

$$p_{t+1} - p_t = -(x_{t+1} - x_t) + (d_{t+1} - d_t).$$

Because $E[x_{t+1} - x_t] = 0$, then it is indeed the case that

$$E[p_{t+1} - p_t] = E[d_{t+1} - d_t].$$

However, exponentiating (4.3) and subtracting 1 implies

$$\frac{P_{t+1} - P_t}{P_t} = e^{-(x_{t+1} - x_t)} \frac{D_{t+1}}{D_t} - 1.$$

We can replace average growth in prices with average growth in dividends if $E[e^{-(x_{t+1} - x_t)} \frac{D_{t+1}}{D_t}] = E[\frac{D_{t+1}}{D_t}]$, a relation that is unlikely to hold exactly. Assuming for the moment that dividend growth and changes to the price-dividend ratio are independent (or nearly independent), the less volatile are shocks to the price-dividend ratio, the closer the two means will be.

Because of Jensen's inequality, replacing price growth with dividend growth need not lead to estimates of the mean return. However, that does not mean that cointegration of the price-dividend ratio cannot be used to help with return estimation. Indeed, the intuition given in Section 4.1 for adjusting the shocks is identical to that conjectured by Fama and French (2002): The sample average of the return is "too high", because shocks to discount rates were positive on average over the sample period. It is the behavior of the price-dividend ratio that allows us to identify that indeed, these shocks were positive on average.

5 Conclusion

In this paper, we have shown that, using maximum likelihood, it is possible to infer that shocks to the dividend-price ratio have been on average negative over the post-war period. It follows that shocks to realized returns are likely to be positive on average. This simple point can be used to construct a less noisy estimate of the equity premium than one given by the sample average alone. It also implies that, in annual terms, the equity premium is more than one percentage point lower than the sample average would imply.

We also show that our method is more reliable than the sample average in finite samples, and across a range of specifications. Our result more generally points to implications of including a small but highly persistent process in returns. One such implication is that the standard deviation of returns may be biased downward to a significant extent.

References

Andrews, Donald W. K., 1993, Exactly median-unbiased estimation of first order autoregressive/unit root models, *Econometrica* 61, 139–165.

Barberis, Nicholas, 2000, Investing for the long run when returns are predictable, *Journal of Finance* 55, 225–264.

Blanchard, Olivier J., 1993, Movements in the Equity Premium, *Brookings Papers on Economic Activity* 1993, 75–138.

Box, George E.P., and George C. Tiao, 1973, *Bayesian Inference in Statistical Analysis*. (Addison-Wesley Pub. Co. Reading, MA).

Campbell, John Y., 2003, Consumption-based asset pricing, in G. Constantinides, M. Harris, and R. Stulz, eds.: *Handbook of the Economics of Finance, vol. 1b* (Elsevier Science, North-Holland).

Campbell, John Y., and Robert J. Shiller, 1988, The dividend-price ratio and expectations of future dividends and discount factors, *Review of Financial Studies* 1, 195–228.

Campbell, John Y., and Luis M. Viceira, 1999, Consumption and portfolio decisions when expected returns are time-varying, *Quarterly Journal of Economics* 114, 433–495.

Claus, James, and Jacob Thomas, 2001, Equity Premia as Low as Three Percent? Evidence from Analysts' Earnings Forecasts for Domestic and International Stock Markets, *The Journal of Finance* 56, 1629–1666.

Cochrane, John H., 2008, The Dog That Did Not Bark: A Defense of Return Predictability, *The Review of Financial Studies* 21, 1533–1575.

DeLong, J. Bradford, and Konstantin Magin, 2009, The U.S. Equity Return Premium: Past, Present, and Future, *The Journal of Economic Perspectives* 23, 193–208.

Donaldson, R. Glen, Mark J. Kamstra, and Lisa A. Kramer, 2010, Estimating the equity premium, *Journal of Financial and Quantitative Analysis* 45, 813–846.

Fama, Eugene F., and Kenneth R. French, 1989, Business conditions and expected returns on stocks and bonds, *Journal of Financial Economics* 25, 23–49.

Fama, Eugene F., and Kenneth R. French, 2002, The Equity Premium, *The Journal of Finance* 57, pp. 637–659.

Graham, John R., and Campbell R. Harvey, 2005, The long-run equity risk premium, *Finance Research Letters* 2, 185–194.

Hamilton, J. D., 1994, *Time Series Analysis*. (Oxford University Press Princeton, NJ).

Kandel, Shmuel, and Robert F. Stambaugh, 1996, On the predictability of stock returns: An asset allocation perspective, *Journal of Finance* 51, 385–424.

Keim, Donald B., and Robert F. Stambaugh, 1986, Predicting returns in the stock and bond markets, *Journal of Financial Economics* 17, 357–390.

Kocherlakota, Narayana R., 1996, The Equity Premium: It's Still a Puzzle, *Journal of Economic Literature* 34, 42–71.

Lewellen, Jonathan, 2004, Predicting returns with financial ratios, *Journal of Financial Economics* 74, 209–235.

Lynch, Anthony W., and Jessica A. Wachter, 2011, Using samples of unequal length in generalized method of moments estimation, Working paper, New York University and the University of Pennsylvania.

Mehra, Rajnish, and Edward Prescott, 1985, The equity premium puzzle, *Journal of Monetary Economics* 15, 145–161.

Mehra, Rajnish, and Edward C. Prescott, 2003, The equity premium in retrospect, in G. M. Constantinides, M. Harris, and R. M. Stulz, eds.: *Handbook of the Economics of Finance* (Elsevier, North-Holland).

Merton, Robert C., 1980, On estimating the expected return on the market: An exploratory investigation, *Journal of Financial Economics* 8, 323–361.

Poirier, Dale J., 1978, The effect of the first observation in regression models with first-order autoregressive disturbances, *Journal of the Royal Statistical Society, Series C, Applied Statistics* 27, 67–68.

Polk, Christopher, Samuel Thompson, and Tuomo Vuolteenaho, 2006, Cross-sectional forecasts of the equity premium, *Journal of Financial Economics* 81, 101–141.

Stambaugh, Robert F., 1997, Analyzing investments whose histories differ in length, *Journal of Financial Economics* 45, 285–331.

Stambaugh, Robert F., 1999, Predictive regressions, *Journal of Financial Economics* 54, 375–421.

Van Binsbergen, Jules H., and Ralph S. J. Koijen, 2010, Predictive regressions: A present-value approach, *The Journal of Finance* 65, 1439–1471.

Wachter, Jessica A., and Missaka Warusawitharana, 2009, Predictable returns and asset allocation: Should a skeptical investor time the market?, *Journal of Econometrics* 148, 162–178.

Wachter, Jessica A., and Missaka Warusawitharana, 2011, What is the chance that the equity premium varies over time? Evidence from predictive regressions, Working paper, Board of Governors of the Federal Reserve and University of Pennsylvania.

Welch, Ivo, 2000, Views of Financial Economists on the Equity Premium and on Professional Controversies, *The Journal of Business* 73, 501–537.

Appendix

A Derivation of the Maximum Likelihood Estimators

We denote the maximum likelihood estimate of parameter q as $\hat{q}$. Here we derive the estimators for μ_r, μ_x, β, θ, σ_u^2, σ_v^2 and σ_{uv}. We note in particular that $\hat{\sigma}_u^2$ is the estimator of σ_u^2, not the square of the estimator of σ_u, and similarly for $\hat{\sigma}_v^2$. Maximizing the exact log likelihood function is the same as minimizing the function $\mathcal{L}$:

$$\begin{aligned}\mathcal{L}(\beta,\theta,\mu_r,\mu_x,\sigma_{uv},\sigma_u,\sigma_v) = \log(\sigma_v^2) - \log(1-\theta^2) \\ + T\log(|\Sigma|) + \frac{1-\theta^2}{\sigma_v^2}(x_0-\mu_x)^2 \\ + \frac{\sigma_v^2}{|\Sigma|}\sum_{t=1}^{T} u_t^2 - 2\frac{\sigma_{uv}}{|\Sigma|}\sum_{t=1}^{T} u_t v_t + \frac{\sigma_u^2}{|\Sigma|}\sum_{t=1}^{T} v_t^2\end{aligned}$$

where $|\Sigma| = \sigma_u^2\sigma_v^2 - \sigma_{uv}^2$. The first-order conditions are

$$0 = \frac{\partial}{\partial\beta}\mathcal{L} = \frac{\sigma_v^2}{|\Sigma|}\sum_{t=1}^{T} u_t(\mu_x - x_{t-1}) - \frac{\sigma_{uv}}{|\Sigma|}\sum_{t=1}^{T}(\mu_x - x_{t-1})v_t \tag{A.1a}$$

$$\begin{aligned}0 = \frac{\partial}{\partial\theta}\mathcal{L} = \frac{\theta}{1-\theta^2} - \theta\frac{(x_0-\mu_x)^2}{\sigma_v^2} \\ - \frac{\sigma_{uv}}{|\Sigma|}\sum_{t=1}^{T} u_t(\mu_x - x_{t-1}) + \frac{\sigma_u^2}{|\Sigma|}\sum_{t=1}^{T} v_t(\mu_x - x_{t-1})\end{aligned} \tag{A.1b}$$

$$0 = \frac{\partial}{\partial\mu_r}\mathcal{L} = -\frac{\sigma_v^2}{|\Sigma|}\sum_{t=1}^{T} u_t + \frac{\sigma_{uv}}{|\Sigma|}\sum_{t=1}^{T} v_t \tag{A.1c}$$

$$\begin{aligned}0 = \frac{\partial}{\partial\mu_x}\mathcal{L} = -\frac{1-\theta^2}{\sigma_v^2}(x_0-\mu_x) + \frac{\sigma_v^2}{|\Sigma|}\sum_{t=1}^{T}\beta u_t \\ - \frac{\sigma_{uv}}{|\Sigma|}\sum_{t=1}^{T}(\beta v_t - (1-\theta)u_t) - \frac{\sigma_u^2}{|\Sigma|}\sum_{t=1}^{T}(1-\theta)v_t\end{aligned} \tag{A.1d}$$

$$\begin{aligned}0 = \frac{\partial}{\partial\sigma_{uv}}\mathcal{L} = -T\frac{2\sigma_{uv}}{|\Sigma|} \\ + 2\frac{\sigma_{uv}\sigma_v^2}{|\Sigma|^2}\sum_{t=1}^{T} u_t^2 - 2\frac{\sigma_u^2\sigma_v^2 + \sigma_{uv}^2}{|\Sigma|^2}\sum_{t=1}^{T} u_t v_t + 2\frac{\sigma_{uv}\sigma_u^2}{|\Sigma|^2}\sum_{t=1}^{T} v_t^2\end{aligned} \tag{A.1e}$$

$$0 = \frac{\partial}{\partial\sigma_u^2}\mathcal{L} = T\frac{\sigma_v^2}{|\Sigma|} - \frac{\sigma_v^4}{|\Sigma|^2}\sum_{t=1}^{T} u_t^2 + 2\frac{\sigma_{uv}\sigma_v^2}{|\Sigma|^2}\sum_{t=1}^{T} u_t v_t - \frac{\sigma_{uv}^2}{|\Sigma|^2}\sum_{t=1}^{T} v_t^2 \tag{A.1f}$$

$$0 = \frac{\partial}{\partial \sigma_v^2}\mathcal{L} = \frac{1}{\sigma_v^2} + T\frac{\sigma_u^2}{|\Sigma|} - (1-\theta^2)(x_0-\mu_x)^2\frac{1}{\sigma_v^4} - \frac{\sigma_{uv}^2}{|\Sigma|^2}\sum_{t=1}^{T} u_t^2 + 2\frac{\sigma_{uv}\sigma_u^2}{|\Sigma|^2}\sum_{t=1}^{T} u_t v_t - \frac{\sigma_u^4}{|\Sigma|^2}\sum_{t=1}^{T} v_t^2 \tag{A.1g}$$

Define the residuals

$$\hat{u}_t = r_t - \hat{\mu}_r - \hat{\beta}(x_{t-1} - \hat{\mu}_x)$$

$$\hat{v}_t = x_t - \hat{\mu}_x - \hat{\theta}(x_{t-1} - \hat{\mu}_x).$$

Step 1: Solve for $\hat{\mu}_x$ in terms of $\hat{\theta}$ and the data.

Combining the first-order conditions with respect to $\hat{\mu}_r$ and $\hat{\mu}_x$ gives

$$\sum_{t=1}^{T} \hat{v}_t = \left(1+\hat{\theta}\right)(\hat{\mu}_x - x_0) \tag{A.2}$$

which we can write as

$$\hat{\mu}_x = \frac{\left(1+\hat{\theta}\right)x_0 + \sum_{t=1}^{T}\left(x_t - \hat{\theta}x_{t-1}\right)}{\left(1+\hat{\theta}\right) + \left(1-\hat{\theta}\right)T}. \tag{A.3}$$

Step 2: Relations for the covariance matrix.

The first-order conditions with respect to $\hat{\sigma}_u^2$, $\hat{\sigma}_v^2$ and $\hat{\sigma}_{uv}$ give the relations

$$T\hat{\sigma}_u^2 = -\frac{\hat{\sigma}_{uv}}{\hat{\sigma}_v^2}\hat{\sigma}_{uv} + (1-\hat{\theta}^2)(x_0-\hat{\mu}_x)^2\left(\frac{\hat{\sigma}_{uv}}{\hat{\sigma}_v^2}\right)^2 + \sum_{t=1}^{T}\hat{u}_t^2 \tag{A.4}$$

$$(T+1)\hat{\sigma}_v^2 = (1-\hat{\theta}^2)(x_0-\hat{\mu}_x)^2 + \sum_{t=1}^{T}\hat{v}_t^2 \tag{A.5}$$

$$\frac{\hat{\sigma}_{uv}}{\hat{\sigma}_v^2} = \frac{\sum_{t=1}^{T}\hat{u}_t\hat{v}_t}{\sum_{t=1}^{T}\hat{v}_t^2} \tag{A.6}$$

Step 3: Solve for $\hat{\theta}$ in terms of the data. This also gives $\hat{\mu}_x$ and $\hat{\sigma}_v^2$ in terms of the data.

Combining the first-order conditions with respect to $\hat{\beta}$ and $\hat{\theta}$ gives

$$0 = \sum_{t=1}^{T}(\hat{\mu}_x - x_{t-1})\hat{v}_t + \hat{\sigma}_v^2\frac{\hat{\theta}}{1-\hat{\theta}^2} - \hat{\theta}(x_0-\hat{\mu}_x)^2 \tag{A.7}$$

Here $\hat{\mu}_x$ and $\hat{v}_t$ are functions of only $\hat{\theta}$ and the data, so if we combine (A.7) and (A.5) we can get an equation for $\hat{\theta}$:

$$0 = (T+1)\sum_{t=1}^{T}(\hat{\mu}_x - x_{t-1})\hat{v}_t + \frac{\hat{\theta}}{1-\hat{\theta}^2}\sum_{t=1}^{T}\hat{v}_t^2 - T\hat{\theta}(x_0 - \hat{\mu}_x)^2$$

Because we require that $-1 < \hat{\theta} < 1$, we can multiply this by

$$\left[(T+1) - (T-1)\hat{\theta}\right]^2\left(1-\hat{\theta}^2\right)$$

and rearrange to obtain

$$\begin{aligned}
0 = T&\left[\sum_{t=0}^{T}x_t - \hat{\theta}\sum_{t=1}^{T-1}x_t\right]^2\left(\hat{\theta}-1\right)\left[(T+1)\left(1-\hat{\theta}^2\right)+2\hat{\theta}\right] \\
&- \left[\sum_{t=0}^{T}x_t - \hat{\theta}\sum_{t=1}^{T-1}x_t\right]\left[(T+1)-(T-1)\hat{\theta}\right]\left(1-\hat{\theta}\right) \\
&\cdot\left\{2T\left(\sum_{t=1}^{T-1}x_t\right)\hat{\theta}(1+\hat{\theta}) - \left(\sum_{t=0}^{T}x_t + \sum_{t=1}^{T-1}x_t\right)\left[(T+1)+(T-1)\hat{\theta}\right]\right\} \\
&+ \left[(T+1)-(T-1)\hat{\theta}\right]^2 \\
&\cdot\left\{\hat{\theta}\left[\left(1-\hat{\theta}^2\right)T+1\right]\left(\sum_{t=1}^{T-1}x_t^2\right) + \left[\hat{\theta}^2(T-1)-(T+1)\right]\sum_{t=1}^{T}x_t x_{t-1} + \hat{\theta}\sum_{t=0}^{T}x_t^2\right\}
\end{aligned}$$

This is a fifth-order polynomial in $\hat{\theta}$ where the coefficients are random variables. As a consequence, it is very hard to establish analytical results on existence and uniqueness of solutions that would be accepted as estimators of θ. Nevertheless, in lengthy experimentation and simulation runs we have always found that this polynomial only has one root within the unit circle of the complex plane and that this root is real. Therefore this root is a valid MLE of θ. Given this solution for $\hat{\theta}$, (A.3) gives the estimator for μ_x and (A.5) gives the estimator for σ_v^2.

Step 4: Solve for $\hat{\mu}_r$ and $\hat{\beta}$ in terms of the data. This also gives the solution for $\hat{\sigma}_{uv}$ and $\hat{\sigma}_u^2$.

The first-order condition with respect to $\hat{\mu}_r$ gives

$$\sum_{t=1}^{T} \hat{u}_t = \frac{\hat{\sigma}_{uv}}{\hat{\sigma}_v^2} \sum_{t=1}^{T} \hat{v}_t \tag{A.8}$$

Combining this with the first-order condition with respect to $\hat{\beta}$ yields

$$\hat{\beta} = \beta^{\text{OLS}} + \frac{\hat{\sigma}_{uv}}{\hat{\sigma}_v^2} \left(\hat{\theta} - \theta^{\text{OLS}} \right) \tag{A.9}$$

where

$$\theta^{\text{OLS}} = \frac{1}{\frac{1}{T}\sum_{t=1}^{T} x_{t-1}^2 - \left(\frac{1}{T}\sum_{t=1}^{T} x_{t-1}\right)^2} \left[\frac{1}{T}\sum_{t=1}^{T} x_{t-1}x_t - \left(\frac{1}{T}\sum_{t=1}^{T} x_{t-1}\right)\left(\frac{1}{T}\sum_{s=1}^{T} x_s\right) \right]$$

is the OLS coefficient of regressing x_t on x_{t-1} and

$$\beta^{\text{OLS}} = \frac{1}{\frac{1}{T}\sum_{t=1}^{T} x_{t-1}^2 - \left(\frac{1}{T}\sum_{t=1}^{T} x_{t-1}\right)^2} \left[\frac{1}{T}\sum_{t=1}^{T} x_{t-1}r_t - \left(\frac{1}{T}\sum_{t=1}^{T} x_{t-1}\right)\left(-\frac{1}{T}\sum_{s=1}^{T} r_s\right) \right]$$

is the OLS coefficient of regressing r_t on x_{t-1}.

Equations (A.6), (A.8) and (A.9) constitute a system of three equations in the three unknowns $\hat{\mu}_r$, $\hat{\beta}$ and $\frac{\hat{\sigma}_{uv}}{\hat{\sigma}_v^2}$. The solution is

$$\begin{aligned} \hat{\mu}_r = \frac{1}{J}\Bigg[\frac{1}{T}\sum_{t=1}^{T} r_t - \left(\frac{1}{T}\sum_{t=1}^{T} x_t - \hat{\mu}_x\right) \frac{F - \beta^{\text{OLS}} H}{1 + (\hat{\theta} - \theta^{\text{OLS}})H} \\ - \left(\frac{1}{T}\sum_{t=1}^{T} x_{t-1} - \hat{\mu}_x\right) \frac{\beta^{\text{OLS}}(1 + \hat{\theta}H) - \theta^{\text{OLS}} F}{1 + (\hat{\theta} - \theta^{\text{OLS}})H} \Bigg] \end{aligned} \tag{A.10}$$

$$\hat{\beta} = \frac{\beta^{\text{OLS}} + (\hat{\theta} - \theta^{\text{OLS}})F}{1 + (\hat{\theta} - \theta^{\text{OLS}})H} - \frac{(\hat{\theta} - \theta^{\text{OLS}})G}{1 + (\hat{\theta} - \theta^{\text{OLS}})H}\hat{\mu}_r. \tag{A.11}$$

$$\frac{\hat{\sigma}_{uv}}{\hat{\sigma}_v^2} = \frac{F - \beta^{\text{OLS}} H}{1 + (\hat{\theta} - \theta^{\text{OLS}})H} - \frac{G}{1 + (\hat{\theta} - \theta^{\text{OLS}})H}\hat{\mu}_r. \tag{A.12}$$

where

$$J = 1 - \frac{G}{1 + (\hat{\theta} - \theta^{\text{OLS}})H} \left[\frac{1}{T}\sum_{t=1}^{T} x_t - \hat{\mu}_x - \theta^{\text{OLS}} \left(\frac{1}{T}\sum_{t=1}^{T} x_{t-1} - \hat{\mu}_x \right) \right],$$

$$F = \frac{\sum_{t=1}^{T} r_t \hat{v}_t}{\sum_{t=1}^{T} \hat{v}_t^2},$$
$$G = \frac{\sum_{t=1}^{T} \hat{v}_t}{\sum_{t=1}^{T} \hat{v}_t^2},$$
$$H = \frac{\sum_{t=1}^{T} (x_{t-1} - \hat{\mu}_x)\hat{v}_t}{\sum_{t=1}^{T} \hat{v}_t^2}.$$

Expressions (A.10) and (A.11) provide the estimators for μ_r and β because they depend only on the data and $\hat{\mu}_x$ and $\hat{\theta}$, which we have already solved in terms of the data. Finally, (A.12) gives the estimator the estimator of σ_{uv} via (A.5), which further yields the estimator of σ_u^2 via (A.4).

B The Effect of θ on the Variance of the Sample Mean Return

By definition

$$\frac{1}{T}\sum_{t=1}^{T} r_t = \mu_r + \beta\left(\frac{1}{T}\sum_{t=1}^{T} x_{t-1} - \mu_x\right) + \frac{1}{T}\sum_{t=1}^{T} u_t$$

thus

$$\mathrm{Var}\left(\frac{1}{T}\sum_{t=1}^{T} r_t\right) = \beta^2\mathrm{Var}\left(\frac{1}{T}\sum_{t=1}^{T} x_{t-1}\right) + \mathrm{Var}\left(\frac{1}{T}\sum_{t=1}^{T} u_t\right) + 2\beta\mathrm{Cov}\left(\frac{1}{T}\sum_{t=1}^{T} x_{t-1}, \frac{1}{T}\sum_{t=1}^{T} u_t\right)$$

The variance of the average predictor is available and it depends on θ. The variance of the average residual does not depend on θ. Finally, the covariance of the average predictor and the average predictor depends on θ and ρ_{uv}. It is not a trivial quantity because even though u_t is uncorrelated with x_{t-1}, it is correlated with x_t via v_t whenever $\rho_{uv} \neq 0$ and thus it is also correlated with $x_{t+1}, x_{t+2}, \ldots, x_{T-1}$ whenever $\theta \neq 0$.

In particular,

$$\begin{aligned}
\mathrm{Var}\left(\frac{1}{T}\sum_{t=1}^{T} u_t\right) &= \sigma_u^2\frac{1}{T}, \\
\mathrm{Var}\left(\frac{1}{T}\sum_{t=1}^{T} x_{t-1}\right) &= \frac{\sigma_v^2}{1-\theta^2}\left[\frac{1}{T}\left(1 + 2\frac{\theta}{1-\theta}\right) + \frac{2}{T^2}\frac{\theta(\theta^T - 1)}{(1-\theta)^2}\right], \\
\mathrm{Cov}\left(\frac{1}{T}\sum_{t=1}^{T} x_{t-1}, \frac{1}{T}\sum_{t=1}^{T} u_t\right) &= \sigma_{uv}\left[\frac{1}{T}\frac{1}{1-\theta} + \frac{1}{T^2}\frac{\theta^T - 1}{(1-\theta)^2}\right],
\end{aligned}$$

so that

$$\mathrm{Var}\left(\frac{1}{T}\sum_{t=1}^{T} r_t\right) = \frac{1}{T}\left(\sigma_u^2 + 2\beta\frac{\sigma_{uv}}{1-\theta} + \beta^2\frac{\sigma_v^2}{1-\theta^2}\right) \quad - \frac{1}{T^2}2\beta\frac{1-\theta^T}{(1-\theta)^2}\left(\beta\theta\frac{\sigma_v^2}{1-\theta^2} + \sigma_{uv}\right)$$

C Conditional Maximum Likelihood Estimation

We denote our conditional maximum likelihood estimator of parameter q as q^c. Here we derive the estimators for μ_r, μ_x, β, θ, σ_u^2, σ_v^2 and σ_{uv} treating the first observation of the predictor, x_0, as fixed. Again, we note that σ_u^{2c} is the estimator of σ_u^2, not the square of the estimator of σ_u, and similarly for σ_v^{2c}. Maximizing the log-likelihood function conditional on x_0 is the same as minimizing the function $\mathcal{L}^c$:

$$\mathcal{L}^c(\beta, \theta, \mu_r, \mu_x, \sigma_{uv}, \sigma_u, \sigma_v) = T\log(|\Sigma|) + \frac{\sigma_v^2}{|\Sigma|}\sum_{t=1}^{T} u_t^2 - 2\frac{\sigma_{uv}}{|\Sigma|}\sum_{t=1}^{T} u_t v_t + \frac{\sigma_u^2}{|\Sigma|}\sum_{t=1}^{T} v_t^2$$

where $|\Sigma| = \sigma_u^2\sigma_v^2 - \sigma_{uv}^2$. Define

$$\begin{aligned} r_t^c &= r_t - \mu_r^c, \\ x_t^c &= x_t - \mu_x^c, \\ u_t^c &= r_t^c - \beta^c x_{t-1}^c, \\ v_t^c &= x_t^c - \theta^c x_{t-1}^c. \end{aligned}$$

Taking first-order conditions and solving them yields the following estimators:

$$\begin{aligned} \mu_x^c &= \frac{\left(\frac{1}{T}\sum_{t=1}^{T} x_t x_{t-1}\right)\left(\frac{1}{T}\sum_{t=1}^{T} x_{t-1}\right) - \left(\frac{1}{T}\sum_{t=1}^{T} x_{t-1}^2\right)\left(\frac{1}{T}\sum_{t=1}^{T} x_t\right)}{\frac{1}{T}\sum_{t=1}^{T} x_t x_{t-1} - \left(\frac{1}{T}\sum_{t=1}^{T} x_{t-1}\right)\left(\frac{1}{T}\sum_{t=1}^{T} x_t\right) + \left(\frac{1}{T}\sum_{t=1}^{T} x_{t-1}\right)^2 - \frac{1}{T}\sum_{t=1}^{T} x_{t-1}^2} \\ \mu_r^c &= \frac{\left[\frac{1}{T}\sum_{t=1}^{T} r_t (x_{t-1}^c)\right]\left[\frac{1}{T}\sum_{t=1}^{T} x_{t-1}^c\right] - \left[\frac{1}{T}\sum_{t=1}^{T} r_t\right]\left[\frac{1}{T}\sum_{t=1}^{T} (x_{t-1}^c)^2\right]}{\left[\frac{1}{T}\sum_{t=1}^{T} x_{t-1}^c\right]^2 - \frac{1}{T}\sum_{t=1}^{T} (x_{t-1}^c)^2} \\ \beta^c &= \frac{\frac{1}{T}\sum_{t=1}^{T} r_t^c x_{t-1}^c}{\frac{1}{T}\sum_{t=1}^{T} (x_{t-1}^c)^2} \\ \theta^c &= \frac{\frac{1}{T}\sum_{t=1}^{T} x_t^c x_{t-1}^c}{\frac{1}{T}\sum_{t=1}^{T} (x_{t-1}^c)^2} \\ \sigma_u^{2c} &= \frac{1}{T}\sum_{t=1}^{T} (u_t^c)^2 \\ \sigma_v^{2c} &= \frac{1}{T}\sum_{t=1}^{T} (v_t^c)^2 \\ \sigma_{uv}^c &= \frac{1}{T}\sum_{t=1}^{T} u_t^c v_t^c \end{aligned}$$

Table 1: Summary Statistics

	1953–2011			1927–2011		
	dp	log returns	return levels	dp	log returns	return levels
mean	−3.545	0.433	0.532	−3.374	0.464	0.615
variance	0.149	19.587	19.337	0.193	29.957	29.975
skewness	−0.570	−0.792	−0.534	−0.389	−0.526	0.170
kurtosis	2.553	5.781	4.915	2.961	9.473	10.340

Sample moments computed at a monthly frequency.The returns are measured as the cum-dividend returns of the value-weighted portfolio from CRSP in excess of the return of the 30-day Treasury Bill.

Table 2: Estimates

	1953–2011		1927–2011	
	OLS	MLE	OLS	MLE
μ_r	0.433	0.322	0.464	0.391
μ_x	−3.545	−3.504	−3.374	−3.383
β	0.828	0.686	0.623	0.650
θ	0.992	0.993	0.992	0.991
σ_u	4.414	4.416	5.466	5.464
σ_v	0.046	0.046	0.057	0.057
ρ_{uv}	−0.961	−0.961	−0.953	−0.953

Estimates at monthly frequency. The predictor is the logarithm of the dividend-price ratio. The return is the logarithm of the cum-dividend return of the value-weighted CRSP portfolio without reinvestment in excess of the logarithm of the return of the 30-day Treasury Bill. Means and standard deviations of returns are in percentage terms. Under the column labeled "OLS", the entries for μ_r and μ_x are the sample averages of the mean return and the mean predictor, whereas the entries for β and θ are the OLS estimates of those parameters. The entries under the columns labeled "MLE" for μ_r, μ_x, β and θ are maximum likelihood estimates.

Table 3: The distribution of estimators in simulations

	True Value	Method	St. Dev.	5 %	50 %	95 %
		Panel A: January 1953 to December 2011				
μ_r	0.322	Sample	0.087	0.177	0.322	0.466
		MLE	0.050	0.241	0.323	0.402
μ_x	−3.504	Sample	0.232	−3.892	−3.503	−3.124
		MLE	0.221	−3.871	−3.504	−3.144
β	0.686	OLS	0.702	0.429	1.157	2.614
		MLE	0.667	0.438	1.101	2.517
θ	0.993	OLS	0.007	0.973	0.988	0.996
		MLE	0.007	0.974	0.989	0.996
σ_u	4.416	OLS	0.119	4.216	4.408	4.607
		MLE	0.118	4.215	4.407	4.604
σ_v	0.046	OLS	0.001	0.044	0.046	0.048
		MLE	0.001	0.044	0.046	0.048
ρ_{uv}	−0.961	OLS	0.003	−0.965	−0.961	−0.956
		MLE	0.003	−0.965	−0.961	−0.956
		Panel B: January 1927 to December 2011				
μ_r	0.391	Sample	0.080	0.257	0.391	0.521
		MLE	0.058	0.294	0.390	0.486
μ_x	−3.383	Sample	0.198	−3.706	−3.386	−3.054
		MLE	0.190	−3.695	−3.384	−3.067
β	0.650	OLS	0.542	0.346	0.944	2.082
		MLE	0.526	0.350	0.919	2.022
θ	0.991	OLS	0.006	0.976	0.988	0.995
		MLE	0.006	0.977	0.989	0.994
σ_u	5.464	OLS	0.122	5.261	5.461	5.664
		MLE	0.122	5.260	5.459	5.662
σ_v	0.057	OLS	0.001	0.055	0.057	0.059
		MLE	0.001	0.055	0.057	0.059
ρ_{uv}	−0.953	OLS	0.003	−0.958	−0.953	−0.948
		MLE	0.003	−0.958	−0.953	−0.948

Standard deviations and percentiles of estimates using the sample mean (Sample), OLS, and maximum likelihood (MLE) with $N = 10,000$ time-series paths matched to two different data samples at monthly frequency. Each panel reports results obtained from first simulating (1) and then estimating the underlying parameters in each simulation run. We set the true value of each parameter to its maximum likelihood estimate in the data. For μ_r and μ_x we report the sample average and the maximum likelihood estimate. For β, θ, σ_u, σ_v and ρ_{uv} we report the OLS estimate and the maximum likelihood estimate.

Table 4: The distribution of estimators in simulations with OLS and sample-mean estimates

	True Value	Method	St. Dev.	5 %	50 %	95 %
			Panel A: January 1953 to December 2011			
μ_r	0.433	Sample	0.082	0.297	0.433	0.568
		MLE	0.049	0.351	0.432	0.511
μ_x	−3.545	Sample	0.189	−3.854	−3.547	−3.239
		MLE	0.180	−3.840	−3.545	−3.249
β	0.828	OLS	0.720	0.517	1.276	2.765
		MLE	0.692	0.519	1.231	2.682
θ	0.992	OLS	0.008	0.971	0.987	0.995
		MLE	0.007	0.972	0.987	0.995
σ_u	4.414	OLS	0.117	4.215	4.409	4.599
		MLE	0.117	4.214	4.407	4.598
σ_v	0.046	OLS	0.001	0.044	0.046	0.048
		MLE	0.001	0.044	0.046	0.047
ρ_{uv}	−0.961	OLS	0.003	−0.965	−0.961	−0.956
		MLE	0.003	−0.965	−0.961	−0.956
			Panel B: January 1927 to December 2011			
μ_r	0.464	Sample	0.081	0.330	0.464	0.598
		MLE	0.058	0.369	0.466	0.561
μ_x	−3.374	Sample	0.203	−3.702	−3.378	−3.038
		MLE	0.196	−3.694	−3.377	−3.049
β	0.623	OLS	0.542	0.323	0.920	2.055
		MLE	0.524	0.326	0.902	1.994
θ	0.992	OLS	0.006	0.977	0.988	0.995
		MLE	0.006	0.977	0.989	0.995
σ_u	5.466	OLS	0.121	5.261	5.460	5.665
		MLE	0.121	5.259	5.457	5.663
σ_v	0.057	OLS	0.001	0.055	0.057	0.059
		MLE	0.001	0.055	0.057	0.059
ρ_{uv}	−0.953	OLS	0.003	−0.958	−0.953	−0.948
		MLE	0.003	−0.958	−0.953	−0.948

Standard deviations and percentiles of estimates using the sample mean (Sample), OLS, and maximum likelihood (MLE) with $N = 10,000$ time-series paths matched to two different data samples at monthly frequency. Each panel reports results obtained from first simulating (1) and then estimating the underlying parameters in each simulation run. We set the true values of β, θ, σ_u, σ_v and ρ_{uv} to their OLS estimate in the data and the true value of μ_r and μ_x to the sample average in the data. For μ_r and μ_x we report the sample average and the maximum likelihood estimate. For β, θ, σ_u, σ_v and ρ_{uv} we report the OLS estimate and the maximum likelihood estimate.

Table 5: The distribution of estimators in simulations with bias correction and fat tails

	True Value	Method	St. Dev.	5 %	50 %	95 %
			Panel A: Bias Correction			
μ_r	0.322	Sample	0.118	0.132	0.32	0.515
		MLE	0.049	0.241	0.321	0.402
μ_x	−3.504	Sample	0.489	−4.305	−3.501	−2.696
		MLE	0.465	−4.267	−3.505	−2.738
β	0.250	OLS	0.644	0.135	0.764	2.160
		MLE	0.606	0.172	0.688	2.057
θ	0.997	OLS	0.007	0.977	0.992	0.998
		MLE	0.006	0.978	0.993	0.998
σ_u	4.426	OLS	0.118	4.224	4.418	4.613
		MLE	0.118	4.223	4.417	4.611
σ_v	0.046	OLS	0.001	0.044	0.046	0.048
		MLE	0.001	0.044	0.046	0.048
ρ_{uv}	−0.961	OLS	0.003	−0.965	−0.961	−0.956
		MLE	0.003	−0.965	−0.961	−0.956
			Panel B: Fat-Tailed Shocks			
μ_r	0.322	Sample	0.119	0.130	0.322	0.519
		MLE	0.049	0.243	0.323	0.402
μ_x	−3.504	Sample	0.488	−4.309	−3.502	−2.698
		MLE	0.463	−4.263	−3.503	−2.742
β	0.250	OLS	0.621	0.134	0.750	2.061
		MLE	0.586	0.166	0.683	1.957
θ	0.997	OLS	0.006	0.978	0.992	0.998
		MLE	0.006	0.980	0.993	0.998
σ_u	4.446	OLS	0.224	4.123	4.420	4.800
		MLE	0.224	4.122	4.419	4.801
σ_v	0.046	OLS	0.002	0.042	0.045	0.049
		MLE	0.002	0.042	0.045	0.049
ρ_{uv}	−0.961	OLS	0.005	−0.967	−0.961	−0.953
		MLE	0.005	−0.968	−0.961	−0.953

Standard deviations and percentiles of estimates using the sample mean (Sample), OLS, and maximum likelihood (MLE) with $N = 10,000$ time-series paths matched to two different data samples at monthly frequency. Each panel reports results obtained from first simulating (1) and then estimating the underlying parameters in each simulation run. We set the true value of each parameter to its maximum likelihood estimate in the data, adjusted so as to match the median values of $\hat{\beta}$ and $\hat{\theta}$ to those we observe in the data. Panel A assumes normally distributed errors while Panel B assumes t-distributed errors. For μ_r and μ_x we report the sample average and the maximum likelihood estimate. For β, θ, σ_u, σ_v and ρ_{uv} we report the OLS estimate and the maximum likelihood estimate.

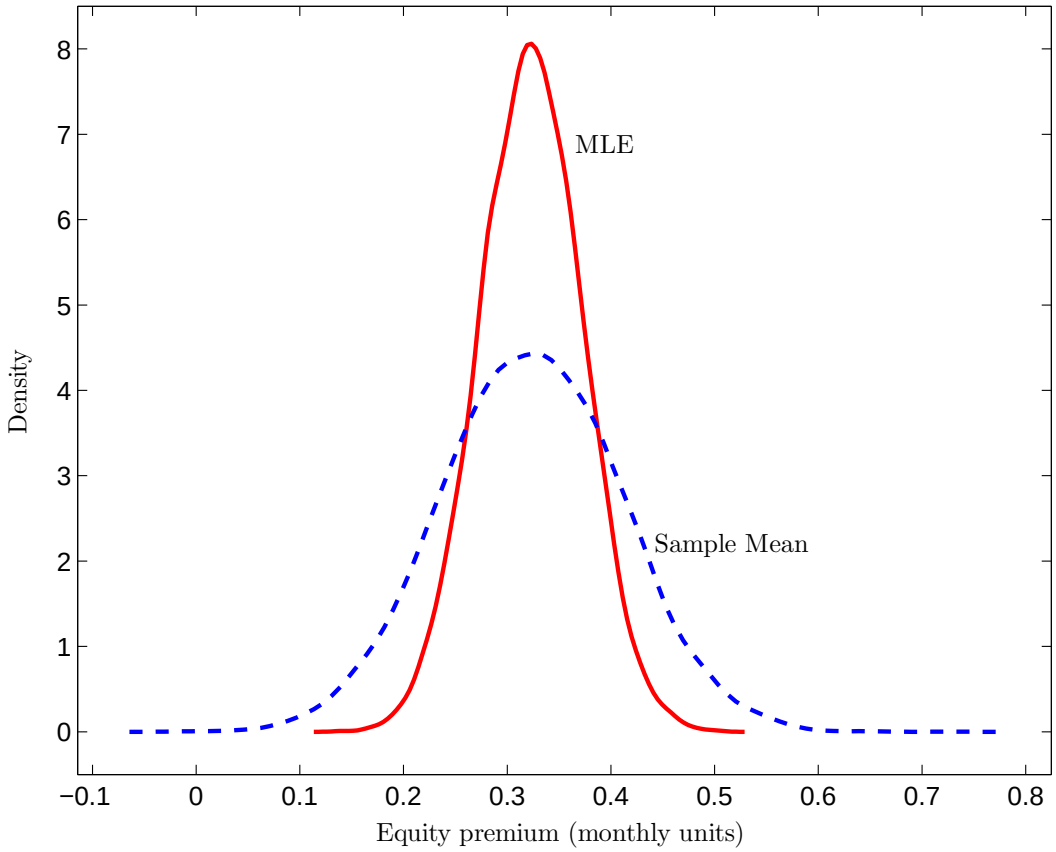


Figure 1: Densities of the estimators of the equity premium. We show estimated densities of the sample mean (dashed line) and the maximum likelihood estimate (solid line) from $N = 10,000$ samples of length $T = 707$ by simulating system (1) with all the parameters fixed to their maximum likelihood estimates in the data for 1953-2011. We have estimated these densities by smoothing out histograms of the estimators with a normal kernel.

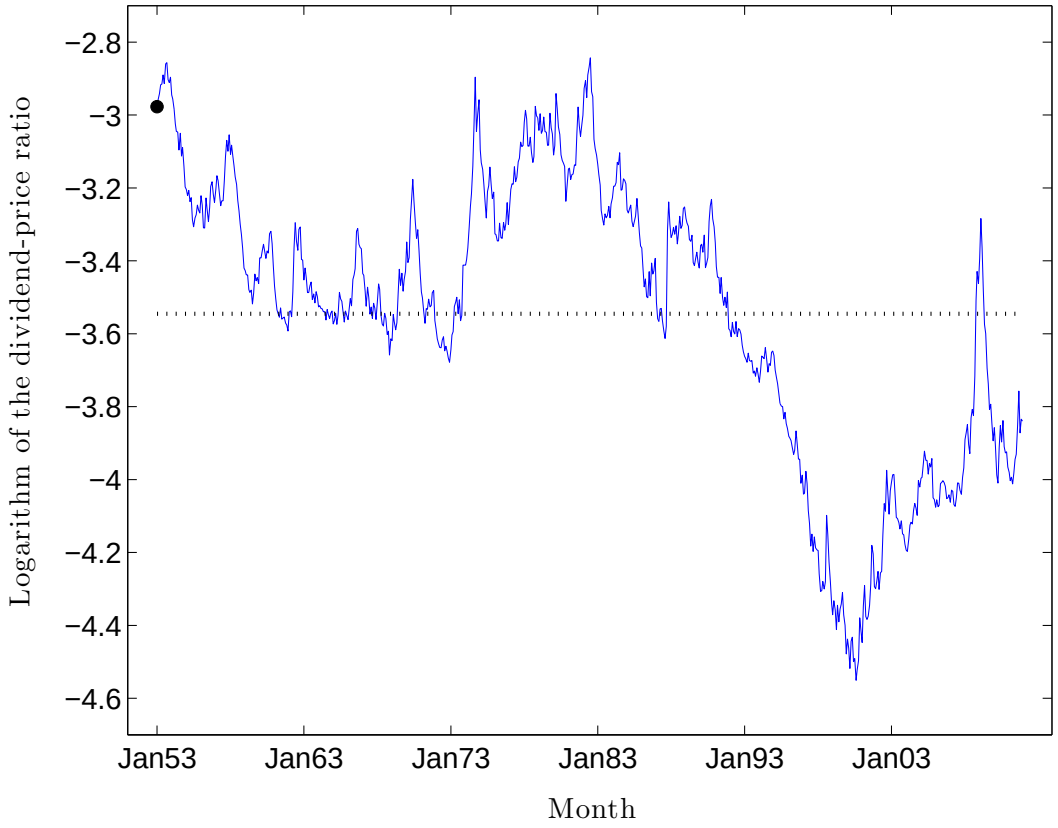


Figure 2: The logarithm of the dividend-price ratio over the period January 1953 to December 2011. The dotted line indicates the mean, and the black dot the initial value.

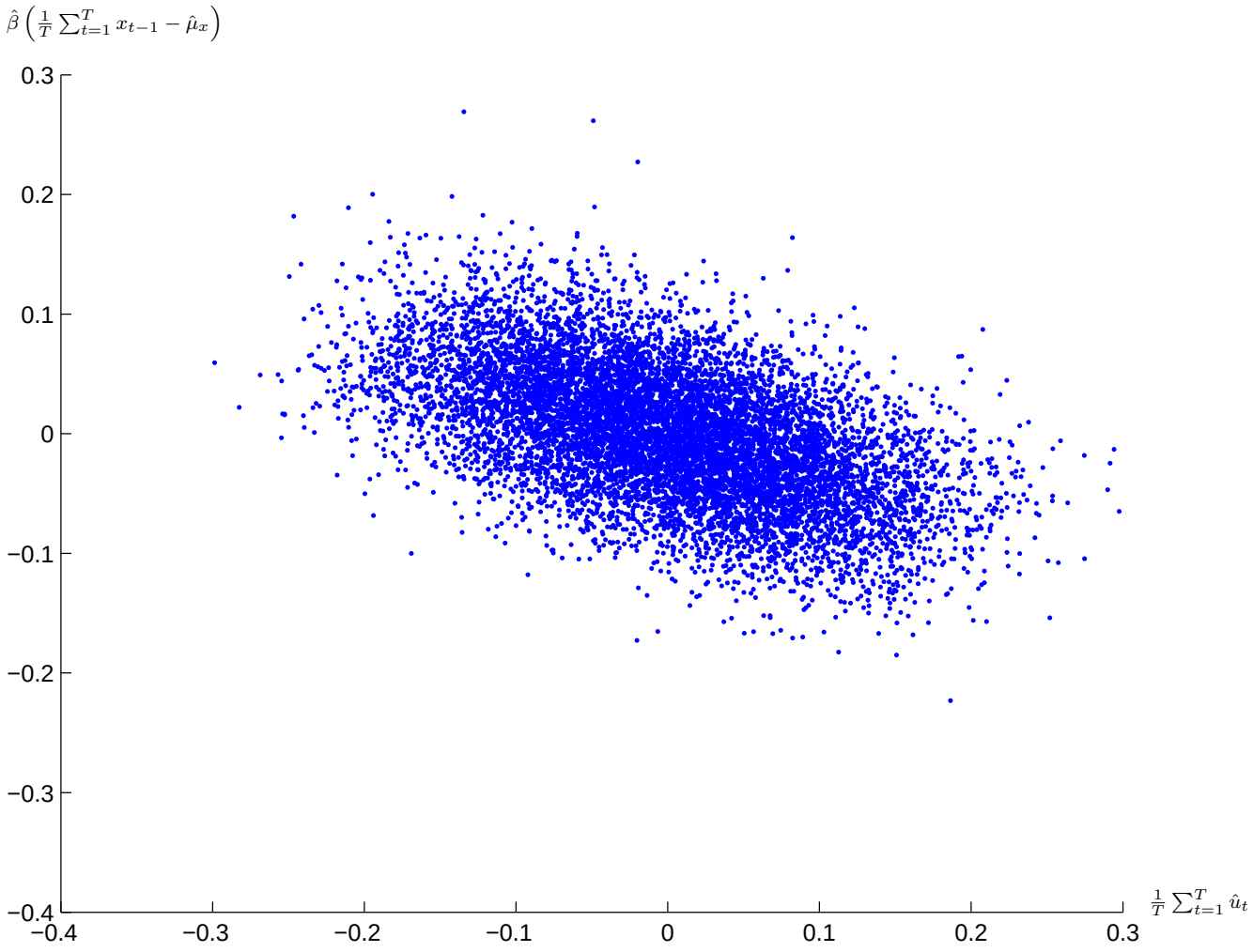
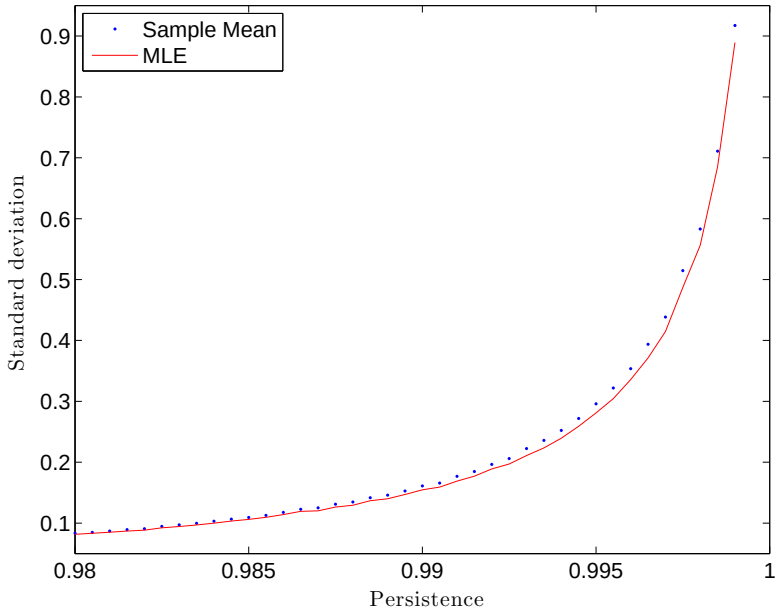
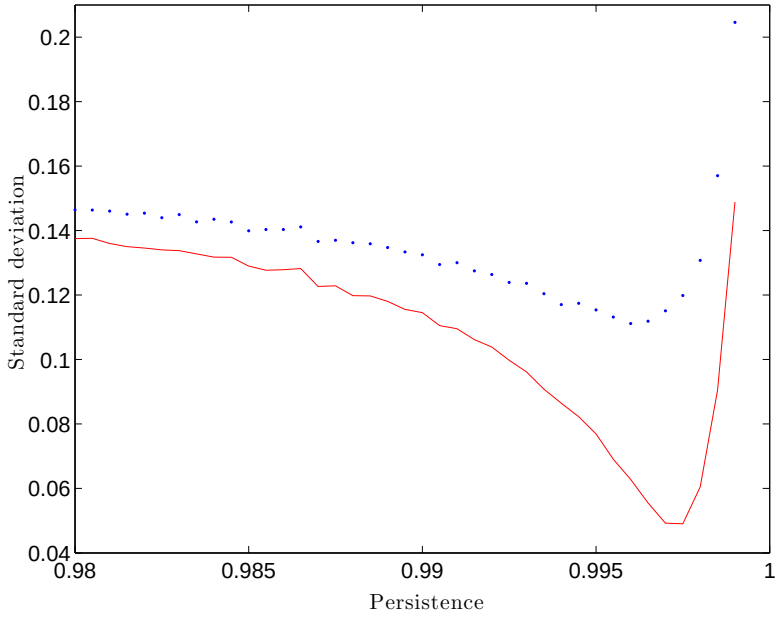


Figure 3: The joint distribution of the error terms in the decomposition of the sample mean. We use $N = 10,000$ samples of length $T = 707$ with parameters set at their maximum likelihood estimates for the 1953-2011 sample.

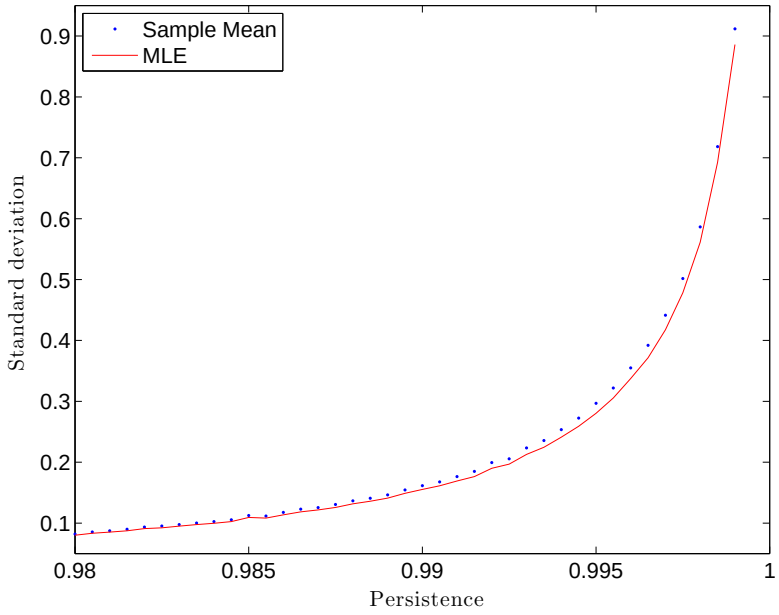


(a) The standard deviation of $\bar{\mu}_x$ (dots) and $\hat{\mu}_x$ (curve) as θ varies.

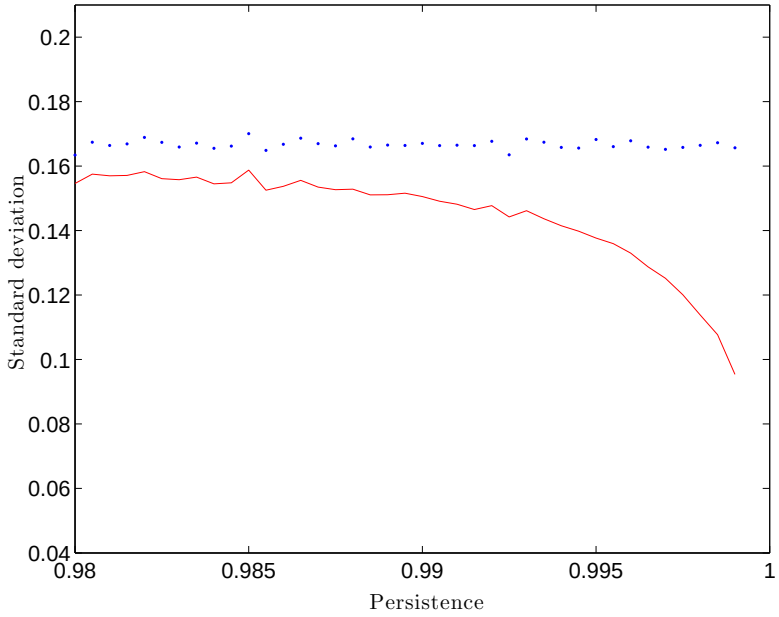


(b) The standard deviation of $\bar{\mu}_r$ (dots) and $\hat{\mu}_r$ (curve) as θ varies.

Figure 4: The standard deviation of estimating the mean of a time series using the sample mean and our maximum likelihood estimate as θ varies. We show these standard deviations for the predictor in (a) and for the return in (b). We use $N = 10,000$ samples of length $T = 707$ with all the parameters other than θ fixed to their maximum likelihood estimates in the data for the 1953-2011 sample, adjusted for biases.

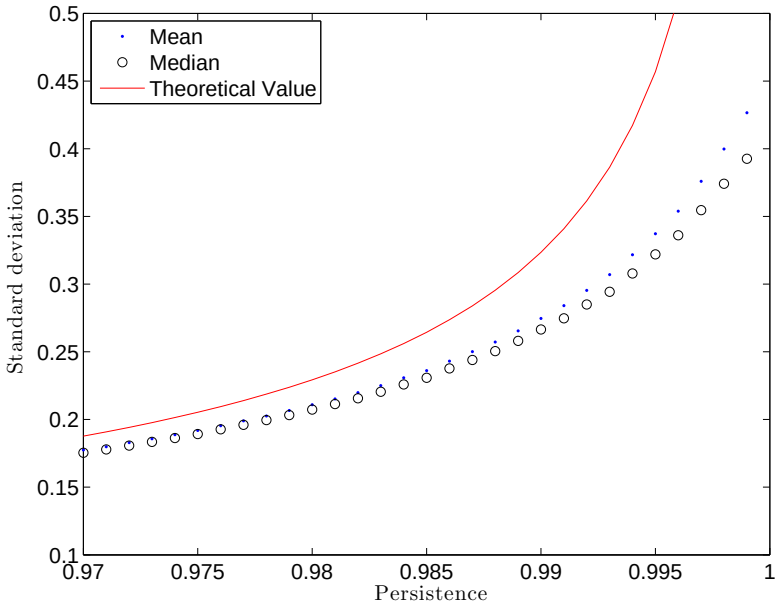


(a) The standard deviation of $\bar{\mu}_x$ (dots) and $\hat{\mu}_x$ (curve) as θ varies.

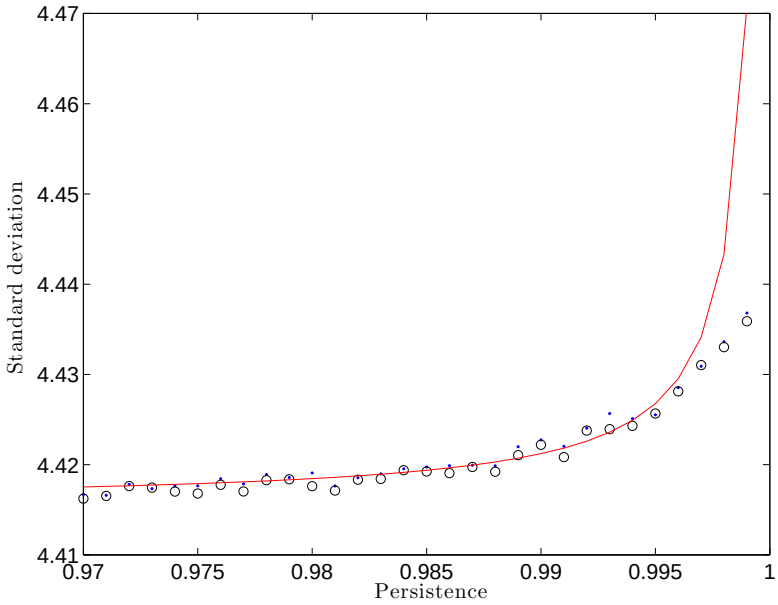


(b) The standard deviation of $\bar{\mu}_r$ (dots) and $\hat{\mu}_r$ (curve) as θ varies.

Figure 5: The standard deviation of estimating the mean of a time series using the sample mean and our maximum likelihood estimate as θ varies in the absence of predictability. We show these standard deviations for the predictor in (a) and for the return in (b). We use $N = 10,000$ samples of length $T = 707$ with $\beta = 0$ and the remaining parameters other than θ fixed to their maximum likelihood estimates in the data for the 1953-2011 sample, adjusted for biases.



(a) The standard deviation of x_t as θ varies: the theoretical value (curve), the mean (dots), and the median (circles).



(b) The standard deviation of r_t as θ varies: the theoretical value (curve), the mean (dots), and the median (circles).

Figure 6: Measurements of the standard deviation of a time series as θ varies. We show the theoretical value, the mean and the median standard deviation for the predictor in (a) and for the return in (b). We use $N = 40,000$ samples of length $T = 707$ with all parameters other than θ fixed to their maximum likelihood estimates in the data for the 1953-2011 sample.