MAXIMUM LIKELIHOOD ESTIMATION OF THE EQUITY PREMIUM

Efstathios Avdis
Jessica A. Wachter

Maximum likelihood estimation of the equity premium
Efstathios Avdis and Jessica A. Wachter
NBER Working Paper No. 19684
November 2013
JEL No. C32,C58,G11,G12

ABSTRACT

The equity premium is usually estimated by taking the sample average of returns. We propose an alternative estimator, based on maximum likelihood, that takes into account additional information contained in dividends and prices. Applying our method to the postwar sample leads to an economically significant reduction from the sample average of 6.4% to a maximum likelihood estimate of 5.1%. Using simulations, we show that our method is robust to mis-specification and is substantially less noisy than the sample average.

Efstathios Avdis
University of Alberta
3-30B Business
Edmonton, AB
Canada T6G 2R6
avdis@ualberta.ca

Jessica A. Wachter
Department of Finance
2300 SH-DH
The Wharton School
University of Pennsylvania
3620 Locust Walk
Philadelphia, PA 19104
and NBER
jwachter@wharton.upenn.edu

The equity premium, namely the expected return on equities less the risk-free 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 the challenge in estimating sample means, 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 (Blanchard, 1993; Fama and French, 2002; 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-year discount rates which are 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).

plicative.[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. However, our implementation is quite 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 log-normally 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 error of 0.089%, while our estimator has a standard error of only 0.050%.

---

[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 in what follows.

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.

1   Statistical Model and Estimation

1.1   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}) - $

$\log(1 + R_{f,t+1})$. Let x_t denote the log of the dividend-price ratio. 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, conditional on $(r_1, \ldots, r_t, x_0, \ldots, x_t)$, the vector of shocks $[u_{t+1}, v_{t+1}]^\top$ is normally distributed 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 x_t. While we focus on the case that the shocks are normally distributed, we also explore robustness to alternative distributional assumptions.

Equations (1a) and (1b) for the return and predictor processes are standard in the literature. Indeed, the equation for returns is equivalent to the ordinary least squares regression that has been a focus of measuring predictability in stock returns for almost 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. The stationary first-order autoregression for x_t is standard in settings where modeling x_t is necessary, e.g. understanding long-horizon returns or the statistical properties of estimators for β.[3] Indeed, most leading economic models imply that x_t is stationary (e.g. Bansal and Yaron, 2004; Campbell and Cochrane, 1999). A large and sophisticated literature uses this setting to explore the bias and size distortions in estimation of β, treating other parameters, including μ_r, as "nuisance" parameters.[4] Our

---

[3]See for example Campbell and Viceira (1999), Barberis (2000), Fama and French (2002), Lewellen (2004), Cochrane (2008), Van Binsbergen and Koijen (2010).

[4]See for example Bekaert, Hodrick, and Marshall (1997), Campbell and Yogo (2006),

work differs from this literature in that μ_r is not a nuisance parameter but rather the focus of our study.

1.2 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

$$p\left(r_1, \ldots, r_T; x_1, \ldots, x_T | \mu_r, \mu_x, \beta, \theta, \Sigma, x_0\right) =$$

$$|2\pi\Sigma|^{-\frac{T}{2}} \exp\left\{-\frac{1}{2}\left(\frac{\sigma_v^2}{|\Sigma|}\sum_{t=1}^{T}u_t^2 - 2\frac{\sigma_{uv}}{|\Sigma|}\sum_{t=1}^{T}u_t v_t + \frac{\sigma_u^2}{|\Sigma|}\sum_{t=1}^{T}v_t^2\right)\right\}. \quad (2)$$

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 u_t and v_t.

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

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

Nelson and Kim (1993), and Stambaugh (1999) for discussions on the bias in estimation of β and Cavanagh, Elliott, and Stock (1995), Elliott and Stock (1994), Jansson and Moreira (2006), Torous, Valkanov, and Yan (2004) and Ferson, Sarkissian, and Simin (2003) for discussion of size. Campbell (2006) surveys this literature. There is a connection between estimation of the mean and of the predictive coefficient, in that the bias in β arises from the bias in θ (Stambaugh, 1999), which ultimately arises from the need to estimate μ_x (Andrews, 1993).

(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. 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.[5] Other previous studies have focused on the effect of incorporating this first term (referred to as the initial condition) on unit root tests (Elliott, 1999; Müller and Elliott, 2003).

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 solution of every other parameter unravels easily. Because our method does not require numerical optimization, it is computationally expedient.

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

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

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}.[6]

1.3 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. Recall that our object of interest is r_t, the logarithm of the gross return in excess of the riskfree asset: $r_t = \log(1 + R_t) - \log(1 + R_t^f)$. We take R_t to be the 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 use the standard construction for the dividend-price ratio that eliminates seasonality, namely, we divide a monthly dividend series (constructed by summing over dividend payouts over the current month and previous eleven months) by the price. Our monthly data are from January 1927 to December 2011.

2 Results

2.1 Point estimates

Table 1 reports estimates of the parameters of our statistical model given in (1). We report estimates for the 1927-2011 sample and for the 1953-2011 postwar subsample. For the postwar subsample, the equity premium from MLE is 0.322% in monthly terms and 3.86% per annum. In contrast, the sample average (given under the column labeled "OLS") is 0.433% in monthly terms, or 5.20% per annum. The annualized difference is 133 basis points.

---

[6]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}.

Applying MLE to the 1927–2011 sample yields an estimated mean of 4.69% per annum, 88 basis points lower than the sample average.

Table 1 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 are nearly identical across the two methods, not surprisingly, because these can be estimated precisely in monthly data. Estimates for the average value of the predictor variable, the predictive coefficient and the autoregressive coefficient are noticeably different. 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 previously discussed, 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 1 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.

2.2   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 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 maximum likelihood estimates, generating a distribution of these estimates over the 10,000 paths.[7]

Table 2 reports the means, standard deviations, and the 5th, 50th, and 95th percentile values. Panel A shows the results of a simulation calibrated using the postwar sample. While the sample average of the excess return has a standard deviation of 0.089, the maximum likelihood estimate has a standard deviation of only 0.050 (unless stated otherwise, units are in monthly percentage terms). 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

---

[7]In every sample, both actual and artificial, we have been able to find a unique solution to the first order conditions such that θ is real and between -1 and 1. Given this value for θ, there is a unique solution for the other parameters. Figure 1 shows the histogram of the resulting values of θ from the postwar simulation. The distribution is well-behaved in that it falls as θ approaches the inadmissible value of 1. Further discussion of the polynomial for θ is contained in Appendix A.

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 estimate.

Table 2 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 estimate of the regression coefficient β also has a smaller standard deviation and confidence intervals than the OLS estimate, 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 system at least, the equity premium is unique in the improvement offered by maximum likelihood. This is in part due to the fact that estimation of first moments is more difficult than that of second moments in the time series (Merton, 1980). However, the result that the mean of returns is affected more than the mean of the predictor shows that this is not all that is going on. We return to this issue in Section 3.

Figure 2 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.[8] The data generating process is calibrated to the postwar period, assuming the parameter estimated using maximum likelihood (unless otherwise stated, all simulations that follow assume this calibration). 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 2, 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

---

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

themselves in the simulation. However, Table 3 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 2 and 3 replicate this result. For example, in the simulation in Table 2, the "true" value of the predictive coefficient β in the simulated data is 0.69, however, the mean OLS value from the simulated samples is 1.28. 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 mean OLS value is 0.987.[9] Maximum likelihood reduces the bias somewhat: the mean maximum likelihood estimate of β is 1.24 as opposed to 1.28, but it does not eliminate it entirely.[10]

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 the mean computed across the simulated samples matches the observed value in the data. The results are given in Panel A of Table 4. This adjustment lowers β and increases θ, but does not change the 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. The sample average has a standard deviation of

---

[9]These tables also show a downward bias in σ_u, the estimate of return shocks. We return to this issue in Section 3.

[10]The estimates of the equity premium are not biased however; the mean for both maximum likelihood and the sample average is close to the population value.

0.138, while the standard deviation of the maximum likelihood estimate if 0.072. Namely, after accounting for biases, maximum likelihood gives an equity premium estimate with standard deviation that is about half of the standard deviation of the sample mean excess return.

In Panel B of Table 4 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 degree of freedom ν. To estimate ν, we measure the kurtosis of the estimated residuals in the return and predictor regressions and take the average. 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 mean kurtosis of the residuals across our simulations by adjusting the ν parameter of the simulated t shocks.[11] 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.

2.3   The equity premium in levels

So far we have defined the equity premium in terms of log returns. However, our result is also indicative of a lower equity premium using return levels. For simplicity, assume that the log returns $\log{(1 + R_t)}$ are normally distributed. Then
$$E[R_t] = E\left[e^{\log(1+R_t)}\right] - 1 = e^{E[\log(1+R_t)] + \frac{1}{2}\mathrm{Var}(\log(1+R_t))} - 1.$$

___

[11]The value of ν that achieves this is 5.96, which corresponds to a population kurtosis of 6.05. The difference between the population number and the mean across our samples reflects the fact that kurtosis in downward biased.

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

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

Our maximum likelihood method provides an estimate of $E[r_t]$ and all other quantities above can be easily calculated using sample moments. Taking the sample 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.[12] 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.

2.4   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 return. Noting the following return identity:

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

and taking the expectation:

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

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 mean growth rates should be the same. They find that

---

[12]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%.

13

the resulting expected return is less than half the sample average, namely 4.74% rather than 9.62%.

While their argument seems intuitive, a closer look reveals a problem. Let $X_t = D_t/P_t$, and let lower-case letters denote natural logs. Then

$$d_{t+1} - d_t = x_{t+1} - x_t + p_{t+1} - p_t. \tag{4}$$

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

$$E[d_{t+1} - d_t] = E[p_{t+1} - p_t]. \tag{5}$$

However, exponentiating (4) and subtracting 1 implies

$$\frac{D_{t+1} - D_t}{D_t} = \frac{X_{t+1}}{X_t}\frac{P_{t+1}}{P_t} - 1. \tag{6}$$

Namely, stationarity of X_t implies (5) and (6), but not $E[(P_t - P_{t-1})/P_{t-1}] = E[(D_t - D_{t-1})/D_{t-1}]$ (namely it does not imply that the average level growth rates be equal). For expected growth rates to be equal in levels, (6) shows that it must be the case that $E[\frac{X_{t+1}}{X_t}\frac{P_{t+1}}{P_t}] = E[\frac{P_{t+1}}{P_t}]$. It is not clear under what conditions this equation should hold (except, of course, if the dividend-price ratio were constant). Even if we were to assume that

$$E\left[\frac{X_{t+1}}{X_t}\frac{P_{t+1}}{P_t}\right] \approx E\left[\frac{X_{t+1}}{X_t}\right] E\left[\frac{P_{t+1}}{P_t}\right],$$

which is unlikely to hold, we would still have $E[X_{t+1}/X_t] > 1$ because of Jensen's inequality and $E[\log(X_{t+1}/X_t)] = 0$.

Nonetheless, our results show that assuming cointegration of prices and dividends can be very informative for estimation of the mean return. Indeed, the intuition that we will develop in the section that follows is closely related 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.

3 Discussion

3.1 Source of the gain in efficiency

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 $\hat{u}_t$ and $\hat{v}_t$ as follows:

$$\hat{u}_t = r_t - \hat{\mu}_r - \hat{\beta}(x_{t-1} - \hat{\mu}_x), \tag{7a}$$

$$\hat{v}_t = x_t - \hat{\mu}_x - \hat{\theta}(x_{t-1} - \hat{\mu}_x). \tag{7b}$$

By definition, then,

$$\hat{\mu}_r = \frac{1}{T}\sum_{t=1}^{T} r_t - \frac{1}{T}\sum_{t=1}^{T} \hat{u}_t - \hat{\beta}\frac{1}{T}\sum_{t=1}^{T}(x_{t-1} - \hat{\mu}_x). \tag{8}$$

As (8) 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 shocks $\hat{u}_t$ may not average to zero over the sample. The second is that, if returns are predictable in the sense that $\beta \neq 0$, 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 quantitatively important for our sample. 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.

3.1.1 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

$$\hat{\mu}_x = \frac{(1+\theta)}{1+\theta+(1-\theta)T}x_0 + \frac{1}{(1+\theta)+(1-\theta)T}\sum_{t=1}^{T}(x_t - \theta x_{t-1}). \qquad (9)$$

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

$$\mu_x = \frac{1}{1-\theta}\frac{1}{T}\sum_{t=1}^{T}(x_t - \theta x_{t-1}) - \frac{1}{T(1-\theta)}\sum_{t=1}^{T}v_t, \qquad (10)$$

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 $1-\theta$. 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 (10) 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 (9) 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:

$$\hat{\mu}_x = \frac{1 - \theta^2}{1 - \theta^2 + (1 - \theta)^2 T} x_0 + \frac{(1 - \theta)^2 T}{1 - \theta^2 + (1 - \theta)^2 T} \hat{\mu}_x^c. \tag{11}$$

Equation 11 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

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

Also, because the shocks v_t are independent, we have that

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

Therefore $T(1 - \theta)^2$ can be viewed as proportional to the precision of the conditional maximum likelihood 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.[13]

While (11) rests on the assumption that θ is known, we can nevertheless 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. Namely, the values of the predictor variable for the last T observations are lower, on average, than they would be in a representative sample. It follows that the predictor variable must have declined over the sample period. Thus the shocks v_t do not average to zero,

---

[13]Note that we cannot interpret (11) 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).

as OLS (or conditional maximum likelihood) would imply, but rather, they average to a negative value.

Figure 3 shows the historical time series of the dividend-price ratio, with the starting value in bold, and a horizontal line representing the mean. Given the appearance of this figure, the conclusion that the dividend-price ratio has been subject to shocks that are, on average, negative does not seem surprising.

3.1.2 Estimation of the equity premium

We now return to the problem of estimating the equity premium. Equation 8 shows that the average shock $\frac{1}{T}\sum_{t=1}^{T} \hat{u}_t$ plays an important role in explaining the difference between the maximum likelihood estimate of the equity premium and the sample mean return. When these shocks are computed using the OLS estimates of the parameters, they must, by definition, average to zero.[14] When the shocks are computed using the maximum likelihood estimate, however, they will not.

To understand the properties of the average shocks to returns, we note that the first-order condition for estimation of $\hat{\mu}_r$ implies

$$\frac{1}{T}\sum_{t=1}^{T} \hat{u}_t = \frac{\hat{\sigma}_{uv}}{\hat{\sigma}_v^2}\frac{1}{T}\sum_{t=1}^{T} \hat{v}_t. \tag{12}$$

This is analogous to a result of Stambaugh (1999), in which the averages of the error terms are replaced by the deviation of β and of θ from the true means. Equation 12 implies a connection between the average value of the shocks to the predictor variable and the average value of the shocks to returns. As the previous section shows, MLE implies that the average shock to the predictor variable is negative in our sample. Because shocks to returns are negatively correlated with shocks to the predictor variable, the average shock to returns

---

[14]Note that the OLS estimate of μ_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.

is positive.[15] Note that this result operates purely through the correlation of the shocks, and is not related to predictability.[16]

Based on this intuition, we can label the terms in (8) as follows:

$$\hat{\mu}_r = \frac{1}{T}\sum_{t=1}^{T} r_t \quad - \quad \underbrace{\frac{1}{T}\sum_{t=1}^{T}\hat{u}_t}_{\text{Correlated shock term}} \quad - \quad \underbrace{\hat{\beta}\frac{1}{T}\sum_{t=1}^{T}(x_{t-1} - \hat{\mu}_x)}_{\text{Predictability term}}. \quad (13)$$

As discussed above, the correlated shock term accounts for more than 100% of the difference between the sample mean and the maximum likelihood estimate of the equity premium. It makes sense that these terms would have opposite signs: 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 4 shows that this result is typical in our simulated samples. This figure shows a scatter plot of the correlated shock term and the predictability term. The correlated shock term tends to be much larger in magnitude than the predictability term. Moreover, the two effects are clearly negatively correlated.

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 1 shows that the difference between the sample mean of excess returns and the maximum likelihood estimate of

---

[15]This point is related to the result that longer time series can help estimate parameters determined by shorter time series, as long as the shocks are correlated. See Stambaugh (1997) and Lynch and Wachter (2013). 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.

[16]Ultimately, however, there may be a connection in that variation in the equity premium is the main driver of variation in the dividend-price ratio and thus the reason why the shocks are negatively correlated.

the equity premium is many times that of the difference between the two estimates of the mean of the predictor variable. Moreover, Table 2 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 12 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.

3.2 Properties of the maximum likelihood estimator

In this section we investigate the properties of the maximum likelihood estimator, and, in particular, how the variance of the estimator 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.

3.2.1 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 5 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 4). For each value of θ, we simulate 10,000 samples.

Panel A shows the standard deviation of estimators of μ_x, the mean of the predictor variable. The standard deviation of both the sample mean and

20

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 of the equity premium. 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, an 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):

$$r_t = \beta x_{t-1} + u_t$$

$$r_{t+1} = \beta\theta x_{t-1} + \beta v_t + u_{t+1}$$

$$r_{t+2} = \beta\theta^2 x_{t-1} + \beta\theta v_t + \beta v_{t+1} + u_{t+2}$$

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

$$\text{Cov}\left(r_t, r_{t+k}\right) = \theta^k \beta^2 \text{Var}(x_t) + \theta^{k-1}\beta\sigma_{uv}, \tag{14}$$

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 through the term $\theta^k \beta^2 \text{Var}(x_t)$. However, because u and v are negatively correlated, the second term in (14), $\theta^{k-1}\beta\sigma_{uv}$ is also negative. In Appendix B, we show that this second term dominates the first for all positive values of θ up until a critical value, at which point the first comes to dominate.

Intuitively, if in a given sample there is a sequence of unusually high returns, this will tend to be followed by unusually low returns. Thus a sequence of unusually high observations or unusually low observations are less likely to dominate in any given sample, and so the sample average will be more stable than it would be if returns were iid.[17] Because the sample mean is simply the

---

[17]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 C). 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.

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 5 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 feature greater mean reversion, and a greater benefit of maximum likelihood.[18]

3.3   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. Like the previous section on the properties of the estimator of the mean return, this section illustrates how a small persistent component of returns can have a substantial impact not only on conditional moments but on unconditional ones as well.

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 pre-

---

[18]If we assume that returns are unpredictable, the standard deviation of the sample mean is constant as a function of θ because the marginal distribution of returns is iid. However, the negative correlation between the shocks u_t and v_t still leads to an improvement from maximum likelihood. A simulation exercise analogous to that reported in Figure 5 shows that the benefits of the maximum likelihood estimate grow as θ increases.

dictor 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 non-stationary 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   Conclusion

A large literature has grown up around the empirical quantity known as the equity premium, in part because of its significance for evaluating models in macro-finance (Mehra and Prescott (1985)) and in part because of its practical significance as indicated by discussions in popular classics on investing (e.g. Siegel (1994), Malkiel (2003)) and in undergraduate and masters' level textbooks.

Estimation of the equity premium is almost always accomplished by taking sample means. The implicit assumption is that the period in question con-

tains a representative sample of returns. We show that it is possible to relax this assumption, and obtain a better estimate of the premium, by bringing additional information to bear on the problem, specifically the information contained separately in prices and dividends.

We show that the time series behavior of prices, dividends and returns, suggests that shocks to returns have been unusually positive over the post-war period. Thus the sample average will overstate the equity premium. We show, surprisingly, that this intuition can be formalized with the standard econometric technique of maximum likelihood. Applying maximum likelihood rather than taking the sample average leads to an economically significant reduction in the equity premium of 1.3 percentage points from 6.4% to 5.1%. Furthermore, Monte Carlo experiments indicate that the small-sample noise is greatly reduced.

It is well-known that the dividend-price ratio contains information about the conditional mean of returns. Our study shows that the process for the dividend-price ratio also has implications for unconditional moments of returns. Besides showing that the average return overstates the equity premium in the post-war period, we also show that that the standard deviation of returns is biased downwards. The degree of these biases depend on the persistence of the dividend-price ratio. We have assumed that the dividend-price ratio follows a stationary auto-regressive process. To the degree there is uncertainty about the dividend-price ratio process itself, our results suggest that there may be considerably more uncertainty about the unconditional distribution of returns than reflected in conventional standard error measures. We look forward to exploring these issues in further work.

References

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

Bansal, Ravi, and Amir Yaron, 2004, Risks for the long-run: A potential resolution of asset pricing puzzles, *Journal of Finance* 59, 1481–1509.

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

Bekaert, Geert, Robert J. Hodrick, and David A. Marshall, 1997, On biases in tests of the expectations hypothesis of the term structure of interest rates, *Journal of Financial Economics* 44, 309–348.

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., 2006, Household Finance, *Journal of Finance* 61, 1553 – 1604.

Campbell, John Y., and John H. Cochrane, 1999, By force of habit: A consumption-based explanation of aggregate stock market behavior, *Journal of Political Economy* 107, 205–251.

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.

Campbell, John Y., and Motohiro Yogo, 2006, Efficient tests of stock return predictability, *Journal of Financial Economics* 81, 27–60.

Cavanagh, Christopher L., Graham Elliott, and James H. Stock, 1995, Inference in models with nearly integrated regressors, *Econometric Theory* 11, 1131–1147.

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.

Elliott, Graham, 1999, Efficient Tests for a Unit Root When the Initial Observation is Drawn from Its Unconditional Distribution, *International Economic Review* 40, 767–783.

Elliott, Graham, and James H Stock, 1994, Inference in Time Series Regression When the Order of Integration of a Regressor Is Unknown, *Econometric Theory* 10, 672–700.

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.

Ferson, Wayne E., Sergei Sarkissian, and Timothy T. Simin, 2003, Spurious regressions in financial economics?, *Journal of Finance* 58, 1393–1413.

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).

Jansson, Michael, and Marcelo J. Moreira, 2006, Optimal Inference in Regression Models with Nearly Integrated Regressors, *Econometrica* 74, 681–714.

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, 2013, Using Samples of Unequal Length in Generalized Method of Moments Estimation, *Journal of Financial and Quantitative Analysis* 48, 277–307.

Malkiel, Burton Gordon, 2003, *A random walk down Wall Street.* (W. W. Norton and Company, Inc. New York, NY).

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.

Müller, Ulrich K., and Graham Elliott, 2003, Tests for Unit Roots and the Initial Condition, *Econometrica* 71, 1269–1286.

Nelson, C. R., and M. J. Kim, 1993, Predictable stock returns: The role of small sample bias, *Journal of Finance* 48, 641–661.

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.

Siegel, Jeremy J., 1994, *Stocks for the long run: a guide to selecting markets for long-term growth.* (Irwin Burr Ridge, IL).

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.

Torous, Walter, Rossen Valkanov, and Shu Yan, 2004, On predicting stock returns with nearly integrated explanatory variables, *Journal of Business* 77, 937–966.

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, 2013, What is the chance that the equity premium varies over time? Evidence from regressions on the dividend-price ratio, NBER Working Paper No. 17334.

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

Table 1: 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

Notes: Estimates of

$$
\begin{aligned}
r_{t+1} - \mu_r &= \beta(x_t - \mu_x) + u_{t+1} \\
x_{t+1} - \mu_x &= \theta(x_t - \mu_x) + v_{t+1},
\end{aligned}
$$

where the shocks $[u_{t+1}, v_{t+1}]^\top$ are iid Gaussian over time with variance-covariance matrix

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

and where r_t is the continuously compounded return on the value-weighted CRSP portfolio in excess of the return on the 30-day Treasury Bill and x_t is the log of the dividend-price ratio. Data are monthly. Means and standard deviations of returns are in percentage terms. Under the OLS columns, parameters are estimated by ordinary least squares, except for μ_r and μ_x, which are equal to the sample averages of excess returns and the predictor variable respectively. Under the MLE columns, parameters are estimated using maximum likelihood.

Table 2: The distribution of estimators in simulations: Calibration to MLE

	True Value	Method	Mean	Std. Dev.	5 %	50 %	95 %
		Panel A: January 1953 to December 2011					
μ_r	0.322	Sample	0.322	0.089	0.175	0.322	0.467
		MLE	0.323	0.050	0.241	0.324	0.404
μ_x	-3.504	Sample	-3.508	0.231	-3.894	-3.507	-3.126
		MLE	-3.508	0.221	-3.875	-3.507	-3.145
β	0.686	OLS	1.284	0.699	0.420	1.145	2.639
		MLE	1.243	0.670	0.440	1.103	2.541
θ	0.993	OLS	0.987	0.007	0.973	0.988	0.996
		MLE	0.987	0.007	0.974	0.989	0.996
σ_u	4.416	OLS	4.408	0.119	4.213	4.408	4.603
		MLE	4.406	0.119	4.211	4.406	4.600
σ_v	0.046	OLS	0.046	0.001	0.044	0.046	0.048
		MLE	0.046	0.001	0.044	0.046	0.048
ρ_{uv}	-0.961	OLS	-0.961	0.003	-0.965	-0.961	-0.956
		MLE	-0.961	0.003	-0.965	-0.961	-0.956
		Panel B: January 1927 to December 2011					
μ_r	0.391	Sample	0.390	0.080	0.258	0.389	0.522
		MLE	0.391	0.058	0.295	0.390	0.485
μ_x	-3.383	Sample	-3.383	0.196	-3.710	-3.385	-3.063
		MLE	-3.384	0.190	-3.701	-3.384	-3.074
β	0.650	OLS	1.039	0.547	0.336	0.941	2.063
		MLE	1.018	0.530	0.345	0.923	2.007
θ	0.991	OLS	0.987	0.006	0.976	0.988	0.995
		MLE	0.987	0.006	0.977	0.989	0.994
σ_u	5.464	OLS	5.460	0.119	5.265	5.459	5.655
		MLE	5.458	0.119	5.263	5.458	5.653
σ_v	0.057	OLS	0.057	0.001	0.055	0.057	0.059
		MLE	0.057	0.001	0.055	0.057	0.059
ρ_{uv}	-0.953	OLS	-0.953	0.003	-0.958	-0.953	-0.948
		MLE	-0.953	0.003	-0.958	-0.953	-0.948

Notes: We simulate 10,000 monthly data samples from (1) with length and parameters as in the postwar data series (Panel A) and as in the long data series (Panel B). Parameters are set to their maximum likelihood estimates given in Table 1. We conduct maximum likelihood estimation (MLE) for each sample path. As a comparison, we take sample means to estimate μ_r and μ_x (Sample) and use ordinary least squares to estimate the slope coefficients and the variance and correlations of the residuals (OLS). The table reports the means, standard deviations, and 5th, 50th, and 95th percentile values across simulations.

Table 3: The distribution of estimators in simulations: Calibration to OLS estimates

	True Value	Method	Mean	Std. Dev.	5 %	50 %	95 %
		Panel A: January 1953 to December 2011					
μ_r	0.433	Sample	0.432	0.082	0.297	0.431	0.565
		MLE	0.432	0.049	0.352	0.432	0.513
μ_x	-3.545	Sample	-3.550	0.192	-3.865	-3.551	-3.232
		MLE	-3.550	0.184	-3.854	-3.552	-3.242
β	0.828	OLS	1.414	0.715	0.512	1.276	2.801
		MLE	1.372	0.689	0.515	1.241	2.675
θ	0.992	OLS	0.986	0.007	0.971	0.987	0.995
		MLE	0.986	0.007	0.972	0.988	0.995
σ_u	4.414	OLS	4.410	0.118	4.215	4.410	4.603
		MLE	4.408	0.118	4.214	4.408	4.601
σ_v	0.046	OLS	0.046	0.001	0.044	0.046	0.048
		MLE	0.046	0.001	0.044	0.046	0.048
ρ_{uv}	-0.961	OLS	-0.961	0.003	-0.965	-0.961	-0.956
		MLE	-0.961	0.003	-0.965	-0.961	-0.956
		Panel B: January 1927 to December 2011					
μ_r	0.464	Sample	0.463	0.082	0.326	0.462	0.596
		MLE	0.464	0.058	0.367	0.463	0.560
μ_x	-3.374	Sample	-3.373	0.200	-3.702	-3.373	-3.044
		MLE	-3.373	0.194	-3.690	-3.374	-3.054
β	0.623	OLS	1.019	0.543	0.322	0.925	2.051
		MLE	0.995	0.527	0.329	0.903	1.983
θ	0.992	OLS	0.987	0.006	0.976	0.988	0.995
		MLE	0.988	0.006	0.977	0.989	0.995
σ_u	5.466	OLS	5.465	0.121	5.269	5.463	5.668
		MLE	5.463	0.121	5.268	5.461	5.666
σ_v	0.057	OLS	0.057	0.001	0.055	0.057	0.059
		MLE	0.057	0.001	0.055	0.057	0.059
ρ_{uv}	-0.953	OLS	-0.953	0.003	-0.958	-0.953	-0.948
		MLE	-0.953	0.003	-0.958	-0.953	-0.948

Notes: We simulate 10,000 monthly data samples from (1) with length and parameters as in the postwar data series (Panel A) and as in the long data series (Panel B). Parameters are set to their OLS estimates given in Table 1. We conduct maximum likelihood estimation (MLE) for each sample path. As a comparison, we take sample means to estimate μ_r and μ_x (Sample) and use ordinary least squares to estimate the slope coefficients and the variance and correlations of the residuals (OLS). The table reports the means, standard deviations, and 5th, 50th, and 95th percentile values across simulations.

Table 4: The distribution of estimators in simulations: Bias-correction and fat-tailed shocks

	True Value	Method	Mean	Std. Dev.	5 %	50 %	95 %
		Panel A: Bias Correction					
μ_r	0.322	Sample	0.324	0.138	0.097	0.327	0.546
		MLE	0.322	0.072	0.205	0.323	0.441
μ_x	-3.504	Sample	-3.510	0.582	-4.464	-3.512	-2.567
		MLE	-3.510	0.557	-4.425	-3.506	-2.601
β	0.090	OLS	0.750	0.643	-0.009	0.610	1.989
		MLE	0.686	0.601	0.036	0.528	1.881
θ	0.998	OLS	0.991	0.007	0.978	0.992	0.999
		MLE	0.992	0.006	0.979	0.993	0.998
σ_u	4.424	OLS	4.417	0.118	4.223	4.416	4.611
		MLE	4.417	0.118	4.225	4.416	4.612
σ_v	0.046	OLS	0.046	0.001	0.044	0.046	0.048
		MLE	0.046	0.001	0.044	0.046	0.048
ρ_{uv}	-0.961	OLS	-0.961	0.003	-0.965	-0.961	-0.956
		MLE	-0.961	0.003	-0.965	-0.961	-0.956
		Panel B: Fat-Tailed Shocks					
μ_r	0.322	Sample	0.323	0.138	0.098	0.320	0.552
		MLE	0.322	0.072	0.204	0.322	0.440
μ_x	-3.504	Sample	-3.504	0.578	-4.454	-3.498	-2.543
		MLE	-3.504	0.549	-4.404	-3.498	-2.589
β	0.090	OLS	0.746	0.634	-0.007	0.601	1.947
		MLE	0.683	0.594	0.040	0.533	1.836
θ	0.998	OLS	0.991	0.007	0.978	0.993	0.999
		MLE	0.992	0.006	0.980	0.993	0.998
σ_u	4.430	OLS	4.419	0.185	4.136	4.411	4.727
		MLE	4.419	0.185	4.136	4.410	4.727
σ_v	0.046	OLS	0.046	0.002	0.043	0.045	0.049
		MLE	0.046	0.002	0.043	0.045	0.049
ρ_{uv}	-0.961	OLS	-0.961	0.004	-0.967	-0.961	-0.954
		MLE	-0.961	0.004	-0.967	-0.961	-0.954

Notes: We simulate 10,000 monthly data samples from (1) with length and parameters as in the postwar data series. Parameters are set to their maximum likelihood estimates from Table 1, adjusted so that the median maximum likelihood estimates of θ and β from the simulation match the data values. We conduct maximum likelihood estimation (MLE) for each sample path. As a comparison, we take sample means to estimate μ_r and μ_x (Sample) and use ordinary least squares to estimate the slope coefficients and the variance and correlations of the residuals (OLS). The table reports the means, standard deviations, and 5th, 50th, and 95th percentile values across simulations. Panel A assumes normally distributed shocks while Panel B uses t-distributed shocks with degrees of freedom parameter equal to 5.96.
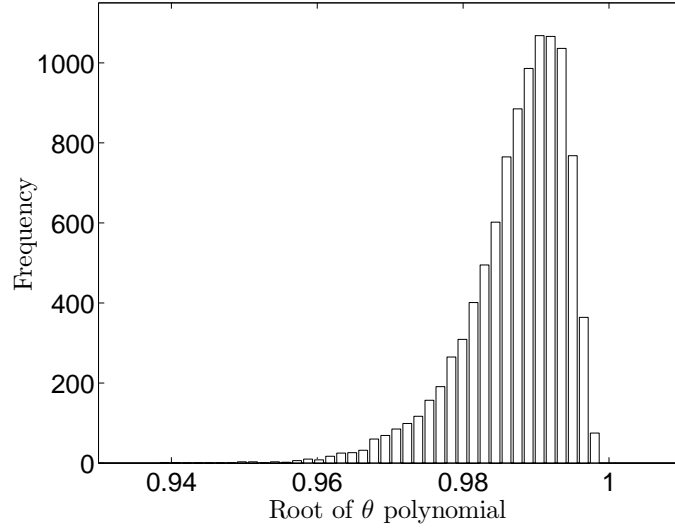
Figure 1: We simulate 10,000 monthly data samples from (1) with length and parameters as in the postwar data series. The figure reports the histogram of maximum likelihood estimates of θ, the autocorrelation of the dividend-price ratio.
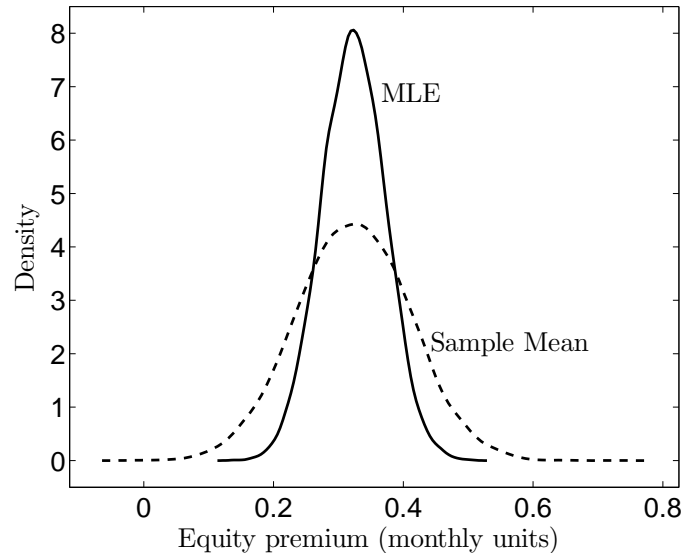

Figure 2: We simulate 10,000 monthly data samples from (1) with length and parameters as in the postwar data series. The figure reports densities of the estimators of the equity premium μ_r. The solid line shows the density of the maximum likelihood estimates while the dashed line shows the density of the sample means of excess returns.
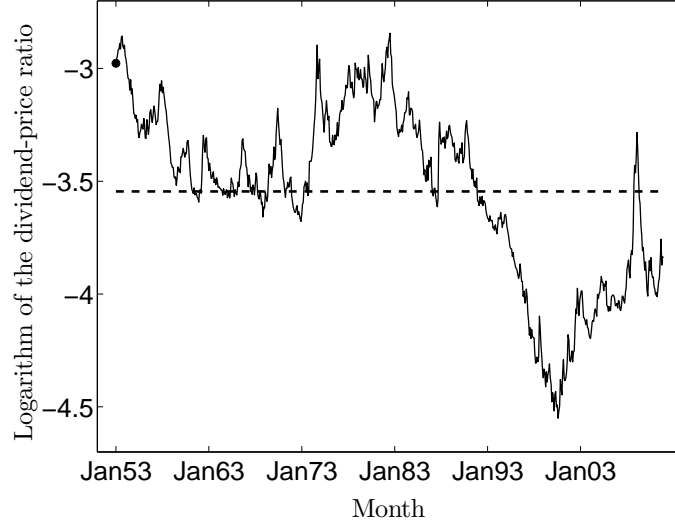
Figure 3: The logarithm of the dividend-price ratio over the period January 1953 to December 2011 (the postwar sample). The dotted line indicates the mean, and the black dot the initial value.
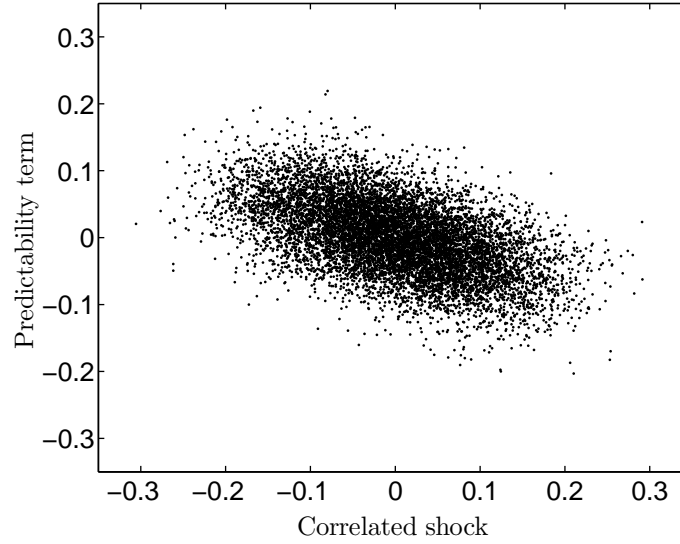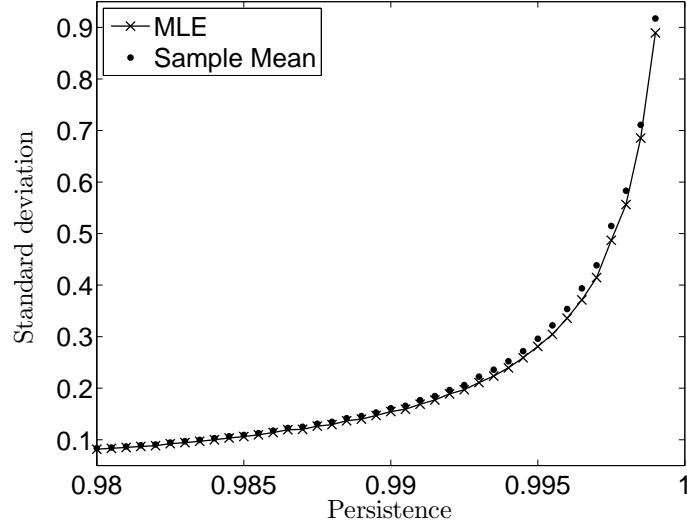

Figure 4: We simulate 10,000 monthly data samples from (1) with length and parameters as in the postwar data series. The figure shows the joint distribution of the predictability term $\hat{\beta}\frac{1}{T}\sum_{t=1}^{T}(x_{t-1} - \hat{\mu}_x)$ and the correlated shock term $\frac{1}{T}\sum_{t=1}^{T}\hat{u}_t$ that sum to the difference between the maximum likelihood estimate and the sample mean.

Panel A: Standard deviation of the mean of the predictor

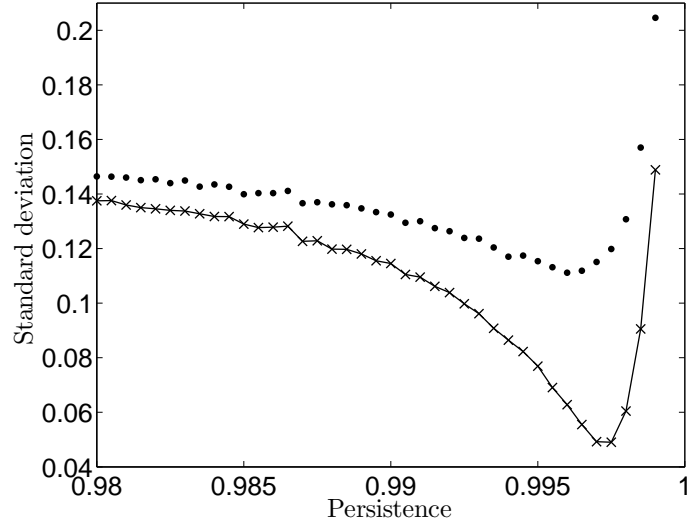Panel B: Standard deviation of the equity premium estimate

Figure 5: We simulate 10,000 monthly data samples from (1) with length and parameters as in the postwar data series. Panel A reports the standard deviations of the sample mean (dots) and of the maximum likelihood estimates of the mean (crosses) of the predictor variable. Panel B does the same for the excess return. The figure shows these standard deviations as functions of the autocorrelation θ. Other parameters are set to their maximum likelihood estimates, adjusted for biases.

Panel A: Standard deviation of the predictor
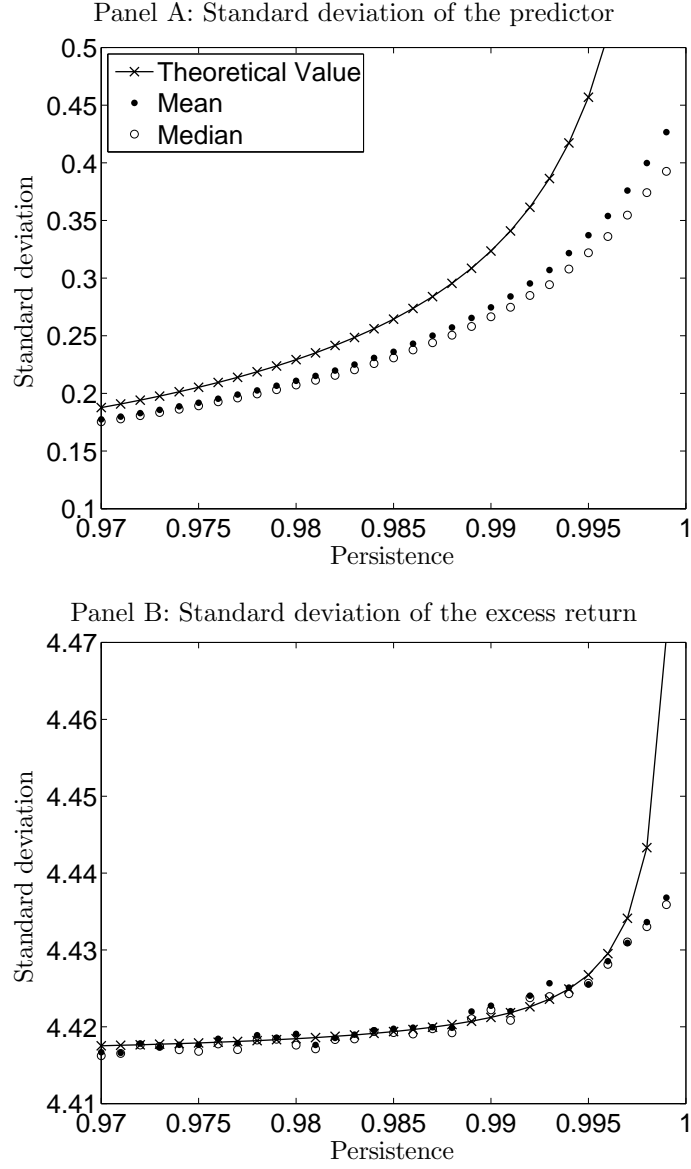
Panel B: Standard deviation of the excess return

Figure 6: We simulate 40,000 monthly data samples from (1) with length and parameters as in the postwar data series. Panel A reports the mean, the median, and the population value of the standard deviation of the predictor variable. Panel B does the same for the excess return. The figure shows these statistics as functions of the autocorrelation θ. Other parameters are set to their maximum likelihood estimates, adjusted for biases.

38

For Online Publication

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 $\mathcal{L}$:

$$\mathcal{L}(\beta, \theta, \mu_r, \mu_x, \sigma_{uv}, \sigma_u, \sigma_v) = \log(\sigma_v^2) - \log(1 - \theta^2) + \frac{1 - \theta^2}{\sigma_v^2}(x_0 - \mu_x)^2$$

$$+ T\log(|\Sigma|) + \frac{\sigma_v^2}{|\Sigma|}\sum_{t=1}^{T} u_t^2 - 2\frac{\sigma_{uv}}{|\Sigma|}\sum_{t=1}^{T} u_t v_t + \frac{\sigma_u^2}{|\Sigma|}\sum_{t=1}^{T} v_t^2,$$

where $|\Sigma| = \sigma_u^2 \sigma_v^2 - \sigma_{uv}^2$. The first-order conditions arise from setting the following partial derivatives of the likelihood function to zero:

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

$$0 = \frac{\partial}{\partial \theta}\mathcal{L} = \frac{\theta}{1 - \theta^2} - \theta\frac{(x_0 - \mu_x)^2}{\sigma_v^2}$$

$$- \frac{\sigma_{uv}}{|\Sigma|}\sum_{t=1}^{T} u_t(\mu_x - x_{t-1}) + \frac{\sigma_u^2}{|\Sigma|}\sum_{t=1}^{T} v_t(\mu_x - x_{t-1}) \tag{A.1b}$$

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

$$0 = \frac{\partial}{\partial \mu_x}\mathcal{L} = -\frac{1 - \theta^2}{\sigma_v^2}(x_0 - \mu_x)$$

$$+ \frac{\sigma_v^2}{|\Sigma|}\sum_{t=1}^{T} \beta u_t - \frac{\sigma_{uv}}{|\Sigma|}\sum_{t=1}^{T}(\beta v_t - (1 - \theta)u_t) - \frac{\sigma_u^2}{|\Sigma|}\sum_{t=1}^{T}(1 - \theta)v_t \tag{A.1d}$$

$$0 = \frac{\partial}{\partial \sigma_{uv}}\mathcal{L} = -T\frac{2\sigma_{uv}}{|\Sigma|}$$

$$+ 2\frac{\sigma_{uv}\sigma_v^2}{|\Sigma|^2}\sum_{t=1}^{T}u_t^2 - 2\frac{\sigma_u^2\sigma_v^2 + \sigma_{uv}^2}{|\Sigma|^2}\sum_{t=1}^{T}u_tv_t + 2\frac{\sigma_{uv}\sigma_u^2}{|\Sigma|^2}\sum_{t=1}^{T}v_t^2$$

(A.1e)

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

$$0 = \frac{\partial}{\partial \sigma_v^2}\mathcal{L} = \frac{1}{\sigma_v^2} + T\frac{\sigma_u^2}{|\Sigma|} - (1-\theta^2)(x_0 - \mu_x)^2\frac{1}{\sigma_v^4}$$

$$- \frac{\sigma_{uv}^2}{|\Sigma|^2}\sum_{t=1}^{T}u_t^2 + 2\frac{\sigma_{uv}\sigma_u^2}{|\Sigma|^2}\sum_{t=1}^{T}u_tv_t - \frac{\sigma_u^4}{|\Sigma|^2}\sum_{t=1}^{T}v_t^2.$$

(A.1g)

Define the residuals

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

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

We now outline the algebra that allows us to solve these first order equations.

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

Combining the first-order conditions (A.1c) and (A.1d) gives

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

which we can write as

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

Step 2: Express the covariance matrix in terms of $\hat{\mu}_x$, $\hat{\theta}$, $\hat{\mu}_r$, $\hat{\beta}$ and the data.

The first-order conditions (A.1e), (A.1f) and (A.1g) give the relations

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

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

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

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 (A.1a) and (A.1b) gives

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

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

$$0 = T\left(\hat{\theta}-1\right)\left((T+1)\left(1-\hat{\theta}^2\right)+2\hat{\theta}\right)\left(\sum_{t=0}^{T}x_t - \hat{\theta}\sum_{t=1}^{T-1}x_t\right)^2$$

$$+ \left((T+1)-(T-1)\hat{\theta}\right)\left(\hat{\theta}-1\right)\left(\sum_{t=0}^{T}x_t - \hat{\theta}\sum_{t=1}^{T-1}x_t\right)$$

$$\times \left[2T\hat{\theta}(1+\hat{\theta})\left(\sum_{t=1}^{T-1}x_t\right) - \left((T+1)+(T-1)\hat{\theta}\right)\left(\sum_{t=0}^{T}x_t + \sum_{t=1}^{T-1}x_t\right)\right]$$

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

41

$$\times \left[\hat{\theta} \left(\left(1 - \hat{\theta}^2 \right) T + 1 \right) \left(\sum_{t=1}^{T-1} x_t^2 \right) + \left(\hat{\theta}^2 (T-1) - (T+1) \right) \sum_{t=1}^{T} x_t x_{t-1} + \hat{\theta} \sum_{t=0}^{T} x_t^2 \right].$$

This is a fifth-order polynomial in $\hat{\theta}$ where the coefficients are determined by the sample. 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 (A.1c) 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 (A.1a) yields

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

where

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

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

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

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

$$\hat{\mu}_r = \frac{1}{J}\left[\frac{1}{T}\sum_{t=1}^{T}r_t - \left(\frac{1}{T}\sum_{t=1}^{T}x_t - \hat{\mu}_x\right)\frac{F - \beta^{\mathrm{OLS}}H}{1 + (\hat{\theta} - \theta^{\mathrm{OLS}})H}\right.$$

$$\left. - \left(\frac{1}{T}\sum_{t=1}^{T}x_{t-1} - \hat{\mu}_x\right)\frac{\beta^{\mathrm{OLS}}(1 + \hat{\theta}H) - \theta^{\mathrm{OLS}}F}{1 + (\hat{\theta} - \theta^{\mathrm{OLS}})H}\right] \tag{A.10}$$

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

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

where

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

$$F = \frac{\sum_{t=1}^{T}r_t\hat{v}_t}{\sum_{t=1}^{T}\hat{v}_t^2}$$

$$G = \frac{\sum_{t=1}^{T}\hat{v}_t}{\sum_{t=1}^{T}\hat{v}_t^2}$$

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

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 expressed 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 Autocovariance of Returns

It follows from (14) that

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

We sign this autocovariance under the assumptions $\theta > 0$, $\beta > 0$ and $\sigma_{uv} < 0$, as we estimate the case to be in our data. Substituting in $\text{Var}(x_t) = \sigma_v^2/(1 - \theta^2)$, multiplying by $(1 - \theta^2) > 0$ and dividing through by $\theta^{k-1}\beta > 0$ shows that the autocovariance of returns is negative whenever

$$-\sigma_{uv}\theta^2 + \beta\sigma_v^2\theta + \sigma_{uv} < 0.$$

The left-hand side is a quadratic polynomial in θ with a positive leading coefficient. As a result, whenever this polynomial has two real roots in θ, the entire expression is negative if and only if θ lies in between those roots. Indeed, the polynomial has two real roots because its discriminant equals $\beta^2\sigma_v^4 + 4\sigma_{uv}^2 > 0$. Let θ_1 be the smaller of the two roots and let θ_2 be the larger one, that is,

$$\theta_2 = \frac{-\beta\sigma_v^2 + \sqrt{\beta^2\sigma_v^4 + 4\sigma_{uv}^2}}{-2\sigma_{uv}}.$$

Under our assumptions it is straightforward to prove that $\theta_1 < -1$ and $-1 < \theta_2 < 1$, so the only possible change of sign of the return autocovariance happens at θ_2. In particular, $\text{Cov}\left(r_t, r_{t+k}\right) < 0$ whenever $\theta < \theta_2$ and $\text{Cov}\left(r_t, r_{t+k}\right) > 0$ whenever $\theta > \theta_2$.

C   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,

$$\mathrm{Var}\left(\frac{1}{T}\sum_{t=1}^{T} u_t\right) = \sigma_u^2\frac{1}{T},$$

$$\mathrm{Var}\left(\frac{1}{T}\sum_{t=1}^{T} x_{t-1}\right) = \frac{\sigma_v^2}{1-\theta^2}\left[\frac{1}{T}\left(1+2\frac{\theta}{1-\theta}\right) + \frac{2}{T^2}\frac{\theta(\theta^T-1)}{(1-\theta)^2}\right],$$

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

so that

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