Are Stocks Really Less Volatile in the Long Run?

by*

Ľuboš Pástor

and

Robert F. Stambaugh

First Draft: April 22, 2008
This revision: September 6, 2008

Abstract

Stocks are more volatile over long horizons than over short horizons from an investor's perspective. This perspective recognizes that observable predictors imperfectly deliver the conditional expected return and that parameters are uncertain, even with two centuries of data. Stocks are often considered less volatile over long horizons due to mean reversion induced by predictability. However, mean reversion's negative contribution to long-horizon variance is more than offset by uncertainty about future expected return, combined with effects of predictor imperfection and parameter uncertainty. Using a predictive system to capture these effects, we find 30-year variance is 21 to 53 percent higher per year than 1-year variance.

---

*Graduate School of Business, University of Chicago, NBER, and CEPR (Pástor) and the Wharton School, University of Pennsylvania, and NBER (Stambaugh). The authors are grateful for comments from John Campbell, Gene Fama, and workshop participants at George Mason University.

1. Introduction

Stock returns are often thought to be less volatile over longer investment horizons. Various empirical estimates are consistent with such a view. For example, using over two centuries of U.S equity returns, Siegel (2008) reports that variances realized over investment horizons of several decades are substantially lower than short-horizon variances on a per-year basis. Such evidence pertains to unconditional variance, but a similar message is delivered by studies that condition variance on information useful in predicting returns. Campbell and Viceira (2002, 2005), for example, report estimates of conditional variances that generally decrease with the investment horizon. The long-run volatility of stocks is no doubt of interest to investors. Evidence of lower long-horizon variance is cited in support of higher equity allocations for long-run investors (e.g, Siegel, 2008) as well as the increasingly popular "life-cycle" mutual funds that allocate less to equity as investors grow older (e.g., Gordon and Stockton, 2006, Greer, 2004, and Viceira, 2008).

We find that stocks are actually *more* volatile over long horizons. At a 30-year horizon, for example, we find return variance per year to be 21 to 53 percent higher than the variance at a 1-year horizon. This conclusion stems from the fact that we assess variance from the perspective of investors who condition on available information but realize their knowledge is limited in two key respects. First, even after observing 206 years of data (1802–2007), investors do not know the values of the parameters that govern the processes generating returns and observable "predictors" used to forecast returns. Second, investors recognize that, even if those parameter values were known, the predictors could deliver only an imperfect proxy for the conditional expected return.

Under the traditional random-walk assumption that returns are distributed independently and identically (i.i.d.) through time, return variance per period is equal at all investment horizons. Explanations for lower variance at long horizons commonly focus on "mean reversion," whereby a negative shock to the current return is offset by positive shocks to future returns, and vice versa. Define the conditional expected return μ_t in the equation

$$r_{t+1} = \mu_t + u_{t+1}, \tag{1}$$

where r_{t+1} denotes the continuously compounded return from time t to time $t+1$, and u_{t+1} has zero mean conditional on all information at time t. With mean reversion, the unexpected return u_{t+1} is negatively correlated with future values of μ_t. If μ_t follows an AR(1) process,

$$\mu_{t+1} = (1-\beta)E_r + \beta\mu_t + w_{t+1}, \tag{2}$$

mean reversion is equivalent to a negative correlation between the innovations u_{t+1} and w_{t+1}, or $\rho_{uw} < 0$. If fluctuations in μ_t are fairly persistent as well (i.e., high β), then a negative shock in

u_{t+1} is accompanied by offsetting positive shifts in the μ_t's for multiple future periods, resulting in a stronger negative contribution to the variance of long-horizon returns.

Our conclusion that stocks are more volatile in the long run obtains despite the presence of mean reversion. We show that mean reversion is only one of five components of long-run variance:

(i) i.i.d. uncertainty

(ii) mean reversion

(iii) uncertainty about future expected returns

(iv) uncertainty about current expected return

(v) estimation risk.

Whereas the mean-reversion component is strongly negative, the other components are all positive, and their combined effect outweighs that of mean reversion.

Of the four components contributing positively, the one making the largest contribution at the 30-year horizon reflects uncertainty about future expected returns. This component (iii) is often neglected in discussions of how return predictability affects long-horizon return variance. Such discussions typically highlight mean reversion, but mean reversion—and predictability more generally—require variance in the conditional expected return μ_t. That variance makes the future values of μ_t uncertain, especially in the more distant future periods, thereby contributing to overall uncertainty about future returns. The greater the true degree of predictability (i.e., the higher the R^2 in equation (1)), the larger is the variance of μ_t and thus the greater is the relative contribution of uncertainty about future expected returns to long-horizon return variance.

Three additional components also make significant positive contributions to long-horizon variance. One is simply the variance attributable to the unexpected return u_{t+1}. Under an i.i.d. assumption for u_{t+1}, the variance of u_{t+1} makes a constant contribution to variance per period at all investment horizons. At the 30-year horizon, this component (i), though quite important, is actually smaller in magnitude than both components (ii) and (iii) discussed above.

Another component of long-horizon variance reflects uncertainty about the current μ_t. Components (i), (ii), and (iii) all condition on the current value of μ_t. Conditioning on the current expected return is standard in long-horizon variance calculations using a vector autoregression (VAR), such as Campbell (1991) and Campbell, Chan, and Viceira (2003). In reality, though, an investor does not observe μ_t but instead observes a vector of predictors, x_t, capable of producing only an imperfect proxy for μ_t. Pástor and Stambaugh (2008) introduce a predictive system to deal with imperfect predictors, and we use that framework to assess long-horizon variance and capture component (iv). When expected returns are persistent (high β), this component grows with the horizon. Uncertainty about the current μ_t then contributes to uncertainty about μ_t in multiple future periods, on top of the uncertainty about future μ_t's discussed earlier.

The fifth and last component adding to long-horizon variance, also positively, is one we label "estimation risk," following common usage of that term. This component reflects the fact that, after observing the available data, an investor remains uncertain about the parameters of the joint process generating return r_{t+1}, expected return μ_t, and the observed predictors x_t. That parameter uncertainty adds to the overall variance of returns assessed by an investor. If the investor knew the parameter values, this estimation-risk component would be zero.

Parameter uncertainty also enters long-horizon variance more pervasively. Unlike the fifth component, the first four components are non-zero even if the parameters are known to an investor. At the same time, those four components can be affected significantly by parameter uncertainty. Each component is an expectation of a function of the parameters, with the expectation evaluated over the distribution characterizing an investor's parameter uncertainty. We find that Bayesian posterior distributions of these functions are often skewed, so that less likely parameter values exert a significant influence on the posterior means, and thus on long-horizon variance.

Variance that incorporates parameter uncertainty is known as *predictive* variance in a Bayesian setting. In contrast, *true* variance excludes parameter uncertainty and is defined by setting parameters equal to their true values. True variance is the more common focus of statistical inference; the usual sample variance, for example, is an estimate of true unconditional variance. We compare long- and short-horizon predictive variances, which are relevant from an investor's perspective. Our objective is thus different from trying to infer whether true return variances per period differ across long and short horizons. The latter inference problem is the focus of an extensive literature that uses variance ratios and other statistics to test whether true return variances differ across horizons.[1] The variance of interest in that hypothesis is generally unconditional, as opposed to being conditioned on current information, but even ignoring that distinction leaves the results of such exercises less relevant to investors. Investors might well infer from the data that the true variance, whether conditional or unconditional, is probably lower at long horizons. At the same time, investors remain uncertain about the values of the parameters, enough so that they assess the relevant variance from their perspective to be higher at long horizons.

The effects of parameter uncertainty on the variance of long-horizon returns are analyzed in previous studies, such as Stambaugh (1999), Barberis (2000), and Hoevenaars et al (2007). Barberis discusses how parameter uncertainty essentially compounds across periods and exerts

[1] A partial list of such studies includes Fama and French (1988), Poterba and Summers (1988), Lo and MacKinlay (1988, 1989), Richardson and Stock (1989), Kim, Nelson, and Startz (1991), and Richardson (1993).

stronger effects at long horizons. All three studies find that the Bayesian predictive variance is substantially higher than variance estimates that ignore parameter uncertainty. However, all studies find that long-horizon predictive variance is lower than short-horizon variance for the horizons those studies consider—up to 10 years in Barberis (2000), up to 20 years in Stambaugh (1999), and up to 50 years in Hoevenaars et al (2007).[2] In contrast, we find that predictive variance even at a 10-year horizon is higher than at a 1-year horizon.

A key difference between our analysis and the above studies is our inclusion of uncertainty about the current expected return μ_t. The above studies employ VAR approaches in which observed predictors perfectly capture μ_t, whereas we consider predictors to be imperfect, as explained earlier. We compare predictive variances under perfect versus imperfect predictors, and find that long-run variance is substantially higher when predictors are imperfect. Predictor imperfection increases long-run variance both directly and indirectly. The direct effect, component (iv) of predictive variance, is large enough at a 10-year horizon that subtracting it from predictive variance leaves the remaining portion lower than the 1-year variance. The indirect effect is even larger. It stems from the fact that once predictor imperfection is admitted, parameter uncertainty is more important in general. That is, when μ_t is not observed, learning about its persistence (β) and its predictive ability (R^2) is more difficult than when μ_t is assumed to be given by observed predictors. The effects of parameter uncertainty pervade all components of long-horizon returns, as noted earlier. The greater parameter uncertainty accompanying predictor imperfection further widens the gap between our analysis and the previous studies.[3]

The remainder of the paper proceeds as follows. Section 2 derives expressions for the five components of long-horizon variance discussed above and analyzes their theoretical properties. The effects of parameter uncertainty on long-horizon variance are first explored in Section 3 using a simplified setting. Section 4 then presents our empirical analysis. We use a predictive system, with 206 years of data, to examine the effects of parameter uncertainty on long-horizon predictive variance and its components. Section 5 compares predictive variances computed using a predictive system to those computed using a "perfect-predictor" framework that excludes uncertainty about the current expected return. Section 6 returns to the above discussion of the distinction between an investor's problem and inference about true variance. Section 7 summarizes our conclusions.

[2]Instead of predictive variances, Barberis reports asset allocations for buy-and-hold, power-utility investors. His allocations for the 10-year horizon exceed those for short horizons, even when parameter uncertainty is incorporated.

[3]Schotman, Tschernig, and Budek (2008) find that if the predictors are fractionally integrated, long-horizon variance of stock returns can exceed short-horizon variance. With stationary predictors, though, they find long-horizon variance is smaller than short-horizon variance. By incorporating predictor imperfection as well as parameter uncertainty, we find that long-horizon variance exceeds short-horizon variance even when predictors are stationary.

2. Components of long-horizon variance

Define the k-period return from period $T+1$ through period $T+k$,

$$r_{T,T+k} = r_{T+1} + r_{T+2} + \ldots + r_{T+k}. \tag{3}$$

An investor assessing the variance of $r_{T,T+k}$ uses D_T, a subset of all information at time T. As noted earlier, D_T typically reveals neither the value of μ_T in (1) nor the values of the parameters governing the joint dynamics of r_{t+1}, μ_{t+1}, and the predictors that investors use in forecasting returns. Let ϕ denote the vector containing those parameter values.

In computing the desired variance $\text{Var}(r_{T,T+k}|D_T)$, a useful building block is the conditional variance $\text{Var}(r_{T,T+k}|\mu_T, \phi, D_T)$. We assume throughout, for simplicity, that μ_t follows the AR(1) process in (2), and that the conditional covariance matrix of $[u_{t+1}\ w_{t+1}]$ is constant.[4] These assumptions imply that $\text{Var}(r_{T,T+k}|\mu_T, \phi, D_T) = \text{Var}(r_{T,T+k}|\mu_T, \phi)$. The Appendix shows that

$$\text{Var}(r_{T,T+k}|\mu_T, \phi) \;=\; k\sigma_u^2 \left[1 + 2\bar{d}\rho_{uw}A(k) + \bar{d}^2 B(k)\right], \tag{4}$$

where

$$A(k) \;=\; 1 + \frac{1}{k}\left(-1 - \beta\frac{1-\beta^{k-1}}{1-\beta}\right) \tag{5}$$

$$B(k) \;=\; 1 + \frac{1}{k}\left(-1 - 2\beta\frac{1-\beta^{k-1}}{1-\beta} + \beta^2\frac{1-\beta^{2(k-1)}}{1-\beta^2}\right) \tag{6}$$

$$\bar{d} \;=\; \left[\frac{1+\beta}{1-\beta}\,\frac{R^2}{1-R^2}\right]^{1/2}. \tag{7}$$

(Recall that ρ_{uw} is the correlation between u_t and w_t, and that R^2 is the true predictive R-squared—the ratio of the variance of μ_t to the variance of r_{t+1}, based on equation (1).)

The conditional variance in (4) consists of three terms. The first term, $k\sigma_u^2$, captures the well-known feature of i.i.d. returns—the variance of k-period returns increases linearly with k. The second term, containing $A(k)$, reflects mean reversion in returns arising from the likely negative correlation between realized returns and expected future returns ($\rho_{uw} < 0$), and it contributes negatively to long-horizon variance. The third term, containing $B(k)$, reflects the uncertainty about future values of μ_t, and it contributes positively to long-horizon variance. When returns are unpredictable, only the first term is present (because $R^2 = 0$ implies $\bar{d} = 0$, so the terms

[4] Our stationary AR(1) process for μ_t nests a popular model in which the stock price is the sum of a random walk and a positively autocorrelated stationary AR(1) component (e.g., Summers, 1986, Fama and French, 1988). In that special case, ρ_{uw} as well as return autocorrelations at all lags are negative. See the Appendix.

involving $A(k)$ and $B(k)$ are zero). Now suppose that returns are predictable, so that $R^2 > 0$ and $\bar{d} > 0$. When $k = 1$, the first term is still the only one standing, because $A(1) = B(1) = 0$. As k increases, though, the terms involving $A(k)$ and $B(k)$ become increasingly important, because both $A(k)$ and $B(k)$ increase monotonically from 0 to 1 as k goes from 1 to infinity.

Figure 1 plots the variance in (4) on a per-period basis (i.e., divided by k), as a function of the investment horizon k. Also shown are the terms containing $A(k)$ and $B(k)$. It can be verified that $A(k)$ converges to 1 faster than $B(k)$. (See Appendix.) As a result, the conditional variance in Figure 1 is U-shaped: as k increases, mean reversion exerts a stronger effect initially, but uncertainty about future expected returns dominates eventually.[5] The contribution of the mean reversion term, and thus the extent of the U-shape, is stronger when ρ_{uw} takes larger negative values. This effect is illustrated in Figure 1. The contributions of mean reversion and uncertainty about future μ_{T+i}'s both become stronger as predictability increases. These effects are illustrated in Figure 2, which plots the same quantities as Figure 1, but for three different R^2 values.

The key insight arising from Figures 1 and 2 is that, although mean reversion can significantly reduce long-horizon variance, that reduction can be more than offset by uncertainty about future expected returns. Both effects become stronger as R^2 increases, since R^2 enters the variance in (4) via $\bar{d}$ in (7), and $\bar{d}$ is increasing in R^2. Note, though, that $\bar{d}$ is squared in the $B(k)$ term, which captures uncertainty about future expected returns, but $\bar{d}$ is not squared in the $A(k)$ term, which captures mean reversion. As a result, mean reversion can be stronger when R^2 is low while uncertainty about future expected returns prevails when R^2 is high.

The persistence in expected return also plays an important role in multiperiod variance, albeit in a more complicated fashion, since β appears in $\bar{d}$ as well as in $A(k)$ and $B(k)$. Figure 3 illustrates effects of β, ρ_{uw} and R^2 by plotting the ratio of per-period conditional variances,

$$V_c(k) = \frac{(1/k)\mathrm{Var}(r_{T,T+k}|\mu_T, \phi)}{\mathrm{Var}(r_{T+1}|\mu_T, \phi)}, \tag{8}$$

for $k = 20$ years. Note that $V_c(20)$ is generally not monotonic in β. At lower values of R^2 and larger negative values of ρ_{uw}, $V_c(20)$ is higher at $\beta = 0.99$ than at the two lower β values. At higher R^2 values, however, $V_c(20)$ is higher at $\beta = 0.85$ than at both the higher and lower β values. At larger negative values of ρ_{uw}, $V_c(20)$ exhibits a U-shape with respect to R^2.

As observed above, uncertainty about future expected returns can cause the long-horizon variance per period to exceed the short-horizon variance, even in the presence of strong mean re-

[5] Campbell and Viceira (2002, pp. 95–96) also model expected return as an AR(1) process, but they conclude that variance per period cannot increase with k when $\rho_{uw} < 0$. They appear to equate conditional variances of single-period returns across future periods, which would omit the uncertainty about future expected return.

version. Importantly, the long-horizon variance can be larger even without including uncertainty about parameters ϕ and the current μ_T. That additional uncertainty exerts a greater effect at longer horizons, further increasing the long-horizon variance relative to the short-horizon variance.

To incorporate the uncertainty about μ_T and ϕ, observe that the variance of $r_{T,T+k}$ conditional on an investor's information D_T can be decomposed as

$$\mathrm{Var}(r_{T,T+k}|D_T) = \mathrm{E}\{\mathrm{Var}(r_{T,T+k}|\mu_T, \phi, D_T)|D_T\} + \mathrm{Var}\{\mathrm{E}(r_{T,T+k}|\mu_T, \phi, D_T)|D_T\}. \tag{9}$$

The first term on the right is the expectation of the conditional variance in (4). Each of the three terms in (4) is now replaced by its expectation with respect to ϕ. (We need not take the expectation with respect to μ_T, since μ_T does not appear on the right in (4).) The interpretations of these terms are the same as before, except that now each term also incorporates parameter uncertainty.

The second term on the right in equation (9) is the variance of the true conditional expected return. This variance is taken with respect to ϕ and μ_T. It can be decomposed into two components: one reflecting uncertainty about the current μ_T, or predictor imperfection, and the other reflecting uncertainty about ϕ, or "estimation risk." (See the Appendix.) Let b_T and q_T denote the conditional mean and variance of the unobservable expected return μ_T:

$$\begin{aligned} b_T &= \mathrm{E}(\mu_T|\phi, D_T) & (10) \\ q_T &= \mathrm{Var}(\mu_T|\phi, D_T). & (11) \end{aligned}$$

The right-hand side of equation (9) can then be expressed as the sum of five components:

$$\begin{aligned} \mathrm{Var}(r_{T,T+k}|D_T) \; = \; & \\ & \underbrace{\mathrm{E}\left\{k\sigma_u^2|D_T\right\}}_{\text{i.i.d. uncertainty}} + \underbrace{\mathrm{E}\left\{2k\sigma_u^2\bar{d}\rho_{uw}A(k)|D_T\right\}}_{\text{mean reversion}} + \underbrace{\mathrm{E}\left\{k\sigma_u^2\bar{d}^2B(k)|D_T\right\}}_{\text{future } \mu_{T+i} \text{ uncertainty}} \\ & + \underbrace{\mathrm{E}\left\{\left(\frac{1-\beta^k}{1-\beta}\right)^2 q_T|D_T\right\}}_{\text{current } \mu_T \text{ uncertainty}} + \underbrace{\mathrm{Var}\left\{kE_r + \frac{1-\beta^k}{1-\beta}(b_T - E_r)|D_T\right\}}_{\text{estimation risk}}. \qquad (12) \end{aligned}$$

Parameter uncertainty plays a role in all five components in equation (12). The first four components are expected values of quantities that are viewed as random due to uncertainty about ϕ, the parameters governing the joint dynamics of returns and predictors. (If the values of these parameters were known to the investor, the expectation operators could be removed from those four components.) Parameter uncertainty can exert a non-trivial effect on the first four components, in that the expectations can be influenced by parameter values that are unlikely but cannot be ruled out. The fifth component in equation (12) is the variance of a quantity whose randomness is also due to parameter uncertainty. In the absence of such uncertainty, the fifth component is zero, which is why we assign it the interpretation of estimation risk.

The estimation risk term includes the variance of kE_r, where E_r denotes the unconditional mean return. This variance equals $k^2\text{Var}(E_r|D_T)$, so dividing by k leaves the per-period variance $\text{Var}(r_{T,T+k}|D_T)/k$ increasing at rate k. As a result, if E_r is unknown, then the per-period variance grows without bounds as k goes to infinity. For finite horizons k that are typically of interest to investors, however, the fifth component in equation (12) can nevertheless be smaller in magnitude than the other four components. In general, the k-period variance ratio, defined as

$$V(k) = \frac{(1/k)\text{Var}(r_{T,T+k}|D_T)}{\text{Var}(r_{T+1}|D_T)}, \tag{13}$$

can exhibit a variety of patterns as k increases. Whether or not $V(k) > 1$ at various horizons k is an empirical question.

3. Parameter uncertainty: A simple illustration

In Section 4, we compute $\text{Var}(r_{T,T+k}|D_T)$ and its components empirically, incorporating parameter uncertainty via Bayesian posterior distributions. Before turning to that analysis, we use a simpler setting to illustrate the effects of parameter uncertainty on multiperiod return variance.

Let $\rho_{\mu b}$ denote the correlation between μ_T and b_T, conditional on all other parameters. If the observed predictors capture μ_T perfectly, then $\rho_{\mu b} = 1$; otherwise $\rho_{\mu b} < 1$. We then have

$$\text{Var}(\mu_T|\phi, D_T) = (1-\rho_{\mu b}^2)\sigma_\mu^2 = (1-\rho_{\mu b}^2)R^2\sigma_r^2 \tag{14}$$

$$\text{Var}(b_T|\phi, D_T) = \rho_{\mu b}^2\sigma_\mu^2 = \rho_{\mu b}^2 R^2\sigma_r^2, \tag{15}$$

where σ_μ^2 and σ_r^2 are the unconditional variances of μ_t and r_{t+1}, respectively. The parameter vector is $\phi = [\beta, R^2, \rho_{uw}, E_r, \sigma_r, \rho_{\mu b}]$. We assume for this simple illustration that the elements of ϕ are distributed independently of each other, conditional on D_T. (This is generally not true in the Bayesian posteriors in the next section.) We define τ such that

$$\text{Var}(E_r|D_T) = \tau\text{E}(\sigma_r^2) \tag{16}$$

and set $\tau = 1/200$, so that the uncertainty about the unconditional mean return E_r corresponds to the imprecision in a 200-year sample mean. With the above independence assumption, equations

(14) through (16), and the fact that $\sigma_u^2 = (1 - R^2)\sigma_r^2$, it is easily verified that $\mathrm{E}(\sigma_r^2)$ can be factored from each component in $\mathrm{Var}(r_{T,T+k}|D_T)$ and thus does not enter the variance ratio in (13). The uncertainty for the remaining parameters is specified by the probability densities displayed in Figure 4, whose medians are 0.86 for β, 0.12 for R^2, -0.66 for ρ_{uw}, and 0.70 for $\rho_{\mu b}$.

Table 1 displays the 20-year variance ratio, $V(20)$, under different specifications of uncertainty about the parameters. In the first row, β, R^2, ρ_{uw}, and E_r are held fixed, by setting the first three parameters equal to their medians and by setting $\tau = 0$ in (16). Successive rows then specify one or more of those parameters as uncertain, by drawing from the densities in Figure 4 (for β, R^2, and ρ_{uw}) or setting $\tau = 0$ (for E_r). For each row, $\rho_{\mu b}$ is either fixed at one of the values 0, 0.70 (its median), and 1, or it is drawn from its density in Figure 4. Note that the return variances are unconditional when $\rho_{\mu b} = 0$ and conditional on full knowledge of μ_T when $\rho_{\mu b} = 1$.

Table 1 shows that when all parameters are fixed, $V(20) < 1$ at all levels of conditioning (all values of $\rho_{\mu b}$). That is, in the absence of parameter uncertainty, the values in the first row range from 0.95 at the unconditional level to 0.77 when μ_T is fully known. Thus, this fixed-parameter specification is consistent with mean reversion playing a dominant role, causing the return variance (per period) to be lower at the long horizon. Rows 2 through 5 specify one of the parameters β, R^2, ρ_{uw}, and E_r as uncertain. Uncertainty about β exerts the strongest effect, raising $V(20)$ by 17% to 26% (depending on $\rho_{\mu b}$), but uncertainty about any one of these parameters raises $V(20)$. In the last row of Table 1, all parameters are uncertain, and the values of $V(20)$ substantially exceed 1, ranging from 1.17 (when $\rho_{\mu b} = 1$) to 1.45 (when $\rho_{\mu b} = 0$). Even though the density for ρ_{uw} in Figure 4 has almost all of its mass below 0, so that returns almost certainly exhibit mean reversion, parameter uncertainty causes the long-run variance to exceed the short-run variance.

As noted earlier, uncertainty about E_r implies $V(k) \to \infty$ as $k \to \infty$. We can see from Table 1 that uncertainty about E_r contributes nontrivially to $V(20)$, but somewhat less than uncertainty about β or R^2 and only slightly more than uncertainty about ρ_{uw}. With uncertainty about only the latter three parameters, $V(20)$ is still well above 1, especially when $\rho_{\mu b} < 1$. Thus, although uncertainty about E_r must eventually dominate variance at sufficiently long horizons, it does not do so here at the 20-year horizon.

The variance ratios in Table 1 increase as $\rho_{\mu b}$ decreases. In other words, less knowledge about μ_T makes long-run variance greater relative to short-run variance. We also see that drawing $\rho_{\mu b}$ from its density in Figure 4 produces the same values of $V(20)$ as fixing $\rho_{\mu b}$ at its median.

4. Long-horizon predictive variance: Empirical results

This section takes a Bayesian empirical approach to assess long-horizon return variance from an investor's perspective. After describing the data and the empirical framework, we specify prior distributions for the parameters and analyze the resulting posteriors. Those posterior distributions characterize the remaining parameter uncertainty faced by an investor who conditions on essentially the entire history of U.S. equity returns. That uncertainty is incorporated in the Bayesian predictive variance, which we then analyze along with its five components.

4.1. Empirical framework: Predictive system

As discussed previously, the return variance faced by an investor is higher when observable predictors at time t do not perfectly capture the expected return μ_t. To incorporate the likely presence of predictor imperfection, we employ the predictive system of Pástor and Stambaugh (2008), which consists of equations (1) and (2) along with a model characterizing the dynamics of the predictors, x_t. We follow that study in modeling x_t as a first-order vector autoregression,

$$x_{t+1} = \theta + A x_t + v_{t+1}. \tag{17}$$

The vector of residuals in the system, $[u_t \;\; v_t' \;\; w_t]$, are assumed to be normally distributed, independently across t, with a constant covariance matrix Σ. We also assume that $0 < \beta < 1$ and that the eigenvalues of A lie inside the unit circle.

Our data consist of annual observations for the 206-year period from 1802 through 2007, as compiled by Siegel (1992, 2008).[6] The return r_t is the annual real log return on the U.S. equity market, and x_t contains three predictors: the dividend yield on U.S equity, the first difference in the long-term high-grade bond yield, and the difference between the long-term bond yield and the short-term interest rate. We refer to these quantities as the "dividend yield," the "bond yield," and the "term spread," respectively. These three predictors seem reasonable choices given the various predictors used in previous studies and the information available in Siegel's dataset. Dividend yield and the term spread have long been entertained as return predictors (e.g., Fama and French, 1989). Using post-war quarterly data, Pástor and Stambaugh (2008) find that the long-term bond yield, relative to its recent levels, exhibits significant predictive ability in predictive regressions. That evidence motivates our choice of the bond-yield variable used here.

[6] We are grateful to Jeremy Siegel for supplying these data.

Table 2 reports properties of the three predictors in the standard predictive regression,

$$r_{t+1} = a + b'x_t + e_{t+1}. \tag{18}$$

The first three regressions in Table 2 contain just one predictor, while the fourth contains all three. When all predictors are included, each one exhibits significant ability to predict returns, and the overall R^2 is 5.6%. The estimated correlation between e_{t+1} and the estimated innovation in expected return, $b'v_{t+1}$, is negative. Pástor and Stambaugh (2008) suggest this correlation as a diagnostic in predictive regressions, with a negative value being what one would hope to see for predictors able to deliver a reasonable proxy for expected return. Table 2 also reports the OLS t-statistics and the bootstrapped p-values associated with these t-statistics as well as with the R^2.[7]

For each of the three key parameters that affect multiperiod variance—ρ_{uw}, β, and R^2—we implement the Bayesian empirical framework under three different prior distributions, displayed in Figure 5. The priors are assumed to be independent across parameters. For each parameter, we specify a "benchmark" prior as well as two priors that depart from the benchmark in opposite directions but seem at least somewhat plausible as alternative specifications. When we depart from the benchmark prior for one of the parameters, we hold the priors for the other two parameters at their benchmarks, obtaining a total of seven different specifications of the joint prior for ρ_{uw}, β, and R^2. We estimate the predictive system under each specification, to explore the extent to which a Bayesian investor's assessment of long-horizon variance is sensitive to prior beliefs.

The benchmark prior for ρ_{uw}, the correlation between expected and unexpected returns, has 97% of its mass below 0. This prior follows the reasoning of Pástor and Stambaugh (2008), who suggest that, a priori, the correlation between unexpected return and the innovation in expected return is likely to be negative. The more informative prior concentrates toward larger negative values, whereas the less informative prior essentially spreads evenly over the range from -1 to 1. The benchmark prior for β, the first-order autocorrelation in the annual expected return μ_t, has a median of 0.83 and assigns a low (2%) probability to β values less than 0.4. The two alternative priors then assign higher probability to either more persistence or less persistence. The benchmark prior for R^2, the fraction of variance in annual returns explained by μ_t, has 63% of its mass below 0.1 and relatively little (17%) above 0.2. The alternative priors are then either more concentrated or less concentrated on low values. These priors on the true R^2 are shown in Panel C of Figure 5.

[7] In the bootstrap, we repeat the following procedure 20,000 times: (i) Resample T pairs of $(\hat{v}_t, \hat{e}_t)$, with replacement, from the set of OLS residuals from regressions (17) and (18); (ii) Build up the time series of x_t, starting from the unconditional mean and iterating forward on equation (17), using the OLS estimates $(\hat{\theta}, \hat{A})$ and the resampled values of $\hat{v}_t$; (iii) Construct the time series of returns, r_t, by adding the resampled values of $\hat{e}_t$ to the sample mean (i.e., under the null that returns are not predictable); (iv) Use the resulting series of x_t and r_t to estimate regressions (17) and (18) by OLS. The bootstrapped p-value associated with the reported t-statistic (or R^2) is the relative frequency with which the reported quantity is smaller than its 20,000 counterparts bootstrapped under the null of no predictability.

Panel D displays the corresponding implied priors on the "observed" R^2—the fraction of variance in annual real returns explained by the predictors. Each of the three priors in Panel D is implied by those in Panel C, while holding the priors for ρ_{uw} and β at their benchmarks and specifying non-informative priors for the degree of imperfection in the predictors. Observe that the benchmark prior for the observed R^2 has much of its mass below 0.05.

We compute posterior distributions for the parameters using the Markov Chain Monte Carlo (MCMC) method discussed in Pástor and Stambaugh (2008). Figure 6 plots posterior distributions computed under the benchmark priors. These posteriors characterize the parameter uncertainty faced by an investor after updating the priors using the 206-year history of equity returns and predictors. Panel B displays the posterior of the true R^2. The posterior lies to the right of the benchmark prior, in the direction of greater predictability. The prior mode for R^2 is less than 0.05, while the posterior mode is nearly 0.1. The posterior of β, shown in Panel C, lies to the right of the prior, in the direction of higher persistence. The benchmark prior essentially admits values of β down to about 0.4, while the posterior ranges only to about 0.7 and has a mode around 0.9.

Compared to the benchmark prior, the posterior for ρ_{uw} is substantially more concentrated toward larger negative values, even to a greater degree than the more concentrated prior. Very similar posteriors for ρ_{uw} are also obtained under the two alternative priors for ρ_{uw} in Figure 5. These results are consistent with observed autocorrelations of annual real returns and the posteriors of R^2 and β discussed above. Equations (1) and (2) imply that the autocovariances of returns are given by

$$\mathrm{Cov}(r_t, r_{t-k}) = \beta^{k-1}\left(\beta\sigma_\mu^2 + \sigma_{uw}\right), \quad k = 1, 2, \dots \quad , \tag{19}$$

where $\sigma_\mu^2 = \sigma_w^2/(1-\beta^2)$. From (19) we can also obtain the autocorrelations of returns,

$$\mathrm{Corr}(r_t, r_{t-k}) = \beta^{k-1}\left(\beta R^2 + \rho_{uw}\sqrt{(1-R^2)R^2(1-\beta^2)}\right), \quad k = 1, 2, \dots \quad , \tag{20}$$

by noting that $\sigma_\mu^2 = R^2\sigma_r^2$ and that $\sigma_u^2 = (1-R^2)\sigma_r^2$. The posterior mode of ρ_{uw} in Figure 6 is about -0.9, and the posterior modes of R^2 and β are about 0.1 and 0.9, as observed earlier. Evaluating (20) at those values gives autocorrelations starting at -0.028 for $k = 1$ and then increasing gradually toward 0 as k increases. Such values seem consistent with observed autocorrelations that are typically near or below zero. For example, the first five autocorrelations of annual real returns in our 206-year sample are 0.02, -0.17, -0.04, 0.01, and -0.10.

Panel A of Figure 6 plots the posterior for the R^2 in a regression of the conditional expected return μ_t on the three predictors in x_t. This R^2 quantifies the degree of imperfection in the predictors ($R^2 = 1$ if and only if the predictors are perfect). Recall from the earlier discussion that predictor imperfection—incomplete knowledge of μ_t—gives rise to the fourth component of return variance in equation (12). The posterior for this R^2 indicates a substantial degree of predictor imperfection, in that the density's mode is about 0.3, and values above 0.8 have near-zero probability.

Further perspective on the predictive abilities of the individual predictors is provided by Figure 7, which plots the posterior densities of the partial correlation coefficients between μ_t and each predictor. Dividend yield exhibits the strongest relation to expected return, with the posterior for its partial correlation ranging between 0 and 0.9 and having a mode around 0.6. Most of the posterior mass for the term spread's partial correlation lies above zero, but there is little posterior mass above 0.5. The bond yield's marginal contribution is the weakest, with much of the posterior density lying between -0.2 and 0.2. In the multiple regression reported in the last row of Table 2, all three variables (rescaled to have unit variances) have comparable slope coefficients and t-statistics. When compared to those estimates, the posterior distributions in Figure 7 indicate that dividend yield is more attractive as a predictor but that bond yield is less attractive. These differences are consistent with the predictors' autocorrelations and the fact that the posterior distribution of β, the autocorrelation of expected return μ_t, centers around 0.9. The autocorrelations for the three predictors are 0.92 for dividend yield, 0.65 for the term spread, and -0.04 for the bond yield. The bond yield's low autocorrelation makes it look less correlated with μ_t, whereas dividend yield's higher autocorrelation makes it look more like μ_t.

4.2. Multiperiod predictive variance and its components

Each of the five components of multiperiod return variance in equation (12) is a moment of a quantity evaluated with respect to the distribution of the parameters ϕ, conditional on the information D_T available to an investor at time T. In our Bayesian empirical setting, D_T consists of the 206-year history of returns and predictors, and the distribution of parameters is the posterior density given that sample. Draws of ϕ from this density are obtained via the MCMC procedure and then used to evaluate the required moments of each of the components in equation (12). The sum of those components, $\mathrm{Var}(r_{T,T+k}|D_T)$, is the Bayesian predictive variance of $r_{T,T+k}$.

Figure 8 displays the predictive variance and its five components for horizons of $k = 1$ through $k = 30$ years, computed under the benchmark priors. The values are stated on a per-year basis (i.e., divided by k). The predictive variance (Panel A) increases significantly with the investment horizon, with the per-year variance exceeding the one-year variance by about 8% at a 10-year horizon and about 45% at a 30-year horizon. This is the main result of the paper.

The five variance components, displayed in Panel B of Figure 8, reveal the sources of the greater predictive variance at long horizons. Over a one-year horizon ($k = 1$), virtually all of the variance is due to the i.i.d. uncertainty in returns, with uncertainty about the current μ_T and parameter uncertainty also making small contributions. Mean reversion and uncertainty about future μ_t's make no contribution for $k = 1$, but they become quite important for larger k. Mean reversion contributes negatively at all horizons, consistent with $\rho_{uw} < 0$ in the posterior (cf. Figure 6), and the magnitude of this contribution increases with the horizon. Nearly offsetting the negative mean reversion component is the positive component due to uncertainty about future μ_t's. At longer horizons, the magnitudes of both components exceed the i.i.d. component, which is flat across horizons. At a 10-year horizon, the mean reversion component is nearly equal in magnitude to the i.i.d. component. At a 30-year horizon, both mean reversion and future-μ_t uncertainty are substantially larger in magnitude than the i.i.d. component. In fact, the mean reversion component is larger in magnitude than the overall predictive variance.

Both estimation risk and uncertainty about the current μ_T make stronger positive contributions to predictive variance as the investment horizon lengthens. At the 30-year horizon, the contribution of estimation risk is about two thirds of the contribution of the i.i.d. component. Uncertainty about the current μ_T, arising from predictor imperfection, makes the smallest contribution among the five components at long horizons, but it still accounts for almost a quarter of the total predictive variance at the 30-year horizon.

Table 3 reports the predictive variance at horizons of 15 and 30 years under various prior distributions for ρ_{uw}, β, and R^2. For each of the three parameters, the prior for that parameter is specified as one of the three alternatives displayed in Figure 5, while the prior distributions for the other two parameters are maintained at their benchmarks. Also reported in Table 3 is the ratio of the long-horizon predictive variance to the one-year variance, as well as the contribution of each of the five components to the long-horizon predictive variance.

Across the different priors in Table 3, the 15-year variance ratio ranges from 1.03 to 1.20, and the 30-year variance ratio ranges from 1.21 to 1.53. The variance ratios exhibit the greatest sensitivity to prior beliefs about R^2. The "loose" prior beliefs that assign higher probability to larger R^2 values produce the lowest variance ratios. When returns are more predictable, mean reversion makes a stronger negative contribution to variance, but uncertainty about future μ_t's makes a stronger positive contribution. The contributions of these two components offset to a large degree as the prior on R^2 moves from tight to loose. At both horizons, the decline in predictive variance as the R^2 prior moves from tight to loose is accompanied by a decline of similar magnitude in estimation risk. The reason why greater predictability implies lower estimation risk involves β. The estimation-risk term in equation (12) contains the expression $(1-\beta^k)/(1-\beta)$ inside the variance operator, so we can roughly gauge the relative effects of changing β by squaring that expression.

As the prior for R^2 moves from tight to loose, the posterior median of β declines from 0.90 to 0.83, and the squared value of $(1-\beta^k)/(1-\beta)$ declines by 30% for $k = 15$ and by 39% for $k = 30$. These drops are comparable to those in the estimation-risk component: 39% for $k = 15$ and 49% for $k = 30$. To then understand why making *higher* R^2 more likely also makes *lower* β more likely, we turn again to the return autocorrelations in (20). Recall that the posterior for ρ_{uw} is concentrated around -0.9 and is relatively insensitive to prior beliefs. Holding ρ_{uw} roughly fixed, therefore, an increase in R^2 requires a decrease in β in order to maintain the same return autocorrelations (for R^2 within the range relevant here). Since the sample is relatively informative about such autocorrelations, the prior (and posterior) that makes higher R^2 more likely is thus accompanied by a posterior that makes lower β more likely.

As the prior for R^2 becomes looser, we also see a smaller positive contribution from i.i.d. uncertainty, which is the posterior mean of $k\sigma_u^2$. This result is expected, as greater posterior density on high values of R^2 necessitates less density on high values of $\sigma_u^2 = (1-R^2)\sigma_r^2$, given that the sample is informative about the unconditional return variance σ_r^2. Finally, prior beliefs about ρ_{uw} and β have a smaller effect on the predictive variance and its components.[8]

In sum, when viewed by an investor whose prior beliefs lie within the wide range of priors considered here, stocks are considerably more volatile at longer horizons. The greater volatility obtains despite the presence of a large negative contribution from mean reversion.

4.3. Robustness

Our main empirical result—that long-run predictive variance of stock returns is larger than short-run variance—is robust to various sample changes. We describe these changes below, along with the corresponding results. We do not tabulate the results to save space.

First, we conduct subperiod analysis. We split the 1802–2007 sample in half and estimate the predictive variances separately at the ends of both subperiods. In the first subperiod, the predictive variance per period rises monotonically with the horizon, under the benchmark priors. In the second subperiod, the predictive variance exhibits a U-shape with respect to the horizon: it initially decreases, reaching its minimum at the horizon of 7 years, but it increases afterwards, rising above the 1-year variance at the horizon of 18 years. That is, the negative effect of mean reversion

[8]This relative insensitivity to prior beliefs about ρ_{uw} and β appears to be specific to the long sample of real equity returns. Greater sensitivity to prior beliefs appears if returns in excess of the short-term interest rate are used instead, or if quarterly returns on a shorter and more recent sample period are used. In all of these alternative samples, we obtain variance results that lead to the same qualitative conclusions.

prevails at short horizons, but the combined positive effects of estimation risk and uncertainty about current and future μ_t's prevail at long horizons. For both subperiods, the 30-year predictive variance exceeds the 1-year variance across all prior specifications. The 30-year predictive variance ratios, which correspond to the ratios reported in the first row of Panel B in Table 3, range from 1.03 to 1.67 across the 14 specifications (seven prior specifications times two subperiods).

Second, we analyze excess returns instead of real returns. We compute annual excess stock returns in 1802–2007 by subtracting the short-term interest rate from the realized stock return. The predictive variance again exhibits a U-shape under the benchmark priors: it slightly decreases before reaching the bottom at the horizon of 3 years, but it quickly rises thereafter. The 30-year predictive variance ratios range from 1.18 to 1.35 across the seven prior specifications.

Third, instead of using three predictors, we use only one, the dividend yield. The predictive variance is U-shaped again in the benchmark case, and the 30-year predictive variance ratio is 1.09. Across the seven prior specifications, the variance ratios range from 0.92 to 1.25.

Finally, we replace our annual 1802–2007 data by quarterly 1952Q1–2006Q4 data. In the postwar period, the data quality is higher, and the available predictors of stock returns have more predictive power. We use the same three predictors as Pástor and Stambaugh (2008): dividend yield, CAY, and bond yield.[9] The R^2 from the predictive regression of quarterly real stock returns on the three predictors is 11.1%, twice as large as the corresponding R^2 in our annual 206-year sample. We adjust the prior distributions to reflect the different data frequency: we shift the priors for R^2 and ρ_{uw} to the left and for β to the right. We find that the results in this quarterly sample are even stronger than the results in our annual sample. Using our benchmark priors, the 15-year predictive variance is 44% larger than the 1-year variance, and the 30-year predictive variance is more than double the 1-year variance. Across our seven prior specifications, the 30-year predictive variance ratios range from 2.02 to 2.44. In short, our empirical results seem robust.

In our baseline estimation, we assume that all parameters of the predictive system are constant over 206 years. This strong assumption seems conservative in that it minimizes parameter uncertainty. As discussed earlier, parameter uncertainty increases long-horizon variance by more than short-horizon variance. If we allowed the unknown parameters to vary over time, an investor's uncertainty about the current parameter values would most likely increase, and the larger parameter uncertainty would then further increase the long-horizon predictive variance ratios.

[9] See that paper for more detailed descriptions of the predictors. Our quarterly sample ends in 2006Q4 because the 2007 data on CAY are not yet available as of this writing. Our quarterly sample begins in 1952Q1, after the 1951 Treasury-Fed accord that made possible the independent conduct of monetary policy.

Time variation in the parameters, if present, need not change our algebraic results. For example, suppose there is time variation in the conditional covariance matrix of the residuals in the predictive system, $\eta_{t+1} \equiv [u_{t+1} \ v_{t+1} \ w_{t+1}]'$. Let Σ_t denote this conditional covariance matrix, and let $\Sigma = \mathrm{E}(\Sigma_t)$ denote the unconditional covariance matrix. It seems plausible to assume that, if $\Sigma_t = \Sigma$ at a given time t, then $\mathrm{E}_t\left(\eta_{t+k}\eta'_{t+k}\right) = \Sigma$ for all $k > 0$.[10] Under this assumption, the conditional variance of the k-period return in equation (4) is unchanged, provided we interpret it as $\mathrm{Var}(r_{T,T+k}|\mu_T, \phi, \Sigma_T = \Sigma)$. The introduction of parameter uncertainty is also unchanged, under the interpretation that Σ is uncertain but that, whatever it is, it also equals Σ_T. Setting $\Sigma_T = \Sigma$ removes horizon effects due to mean-reversion in Σ_T. If instead Σ_T were low relative to Σ, for example, then the reversion of future Σ_{T+i}'s to Σ could also contribute to volatility that is higher over longer horizons. Setting $\Sigma_T = \Sigma$ excludes such a contribution to higher long-run volatility.

5. Perfect predictors versus imperfect predictors

When predictors are imperfect, the current expected return remains uncertain even if the parameters of the processes generating returns and predictors are known. Incorporating predictor imperfection is a key difference between our analysis and earlier studies by Stambaugh (1999) and Barberis (2000) that investigate the effects of parameter uncertainty on long-run equity volatility. Those studies model expected return as $\mu_t = a + b'x_t$, so that the observed predictors deliver expected return perfectly if the parameters a and b are known. The latter "perfect-predictor" assumption implies the predictive regression in (18). Combining that equation with the VAR for x_t in (17) delivers implications for long-run variance, as in Stambaugh (1999) and Barberis (2000).

To assess the importance of recognizing predictor imperfection, we compute long-run predictive variances in the above perfect-predictor framework and compare them to predictive variances obtained using our (imperfect-predictor) predictive system. We conduct this comparison using non-informative priors for both settings, noting that Stambaugh and Barberis use non-informative priors as well.[11] Panel A of Figure 9 displays results based on annual data for the 1802–2007 period, while Panel B displays results based on quarterly data for the 1952Q1–2006Q4 period. All results are based on real returns and three predictors (described earlier).

For the 1802–2007 period, predictive variance computed using the predictive system (solid line) increases with horizon to produce a variance ratio of 1.70 at 30 years. This value, obtained

[10] Such a property is satisfied, for example, by a stationary first-order multivariate GARCH process, $\mathrm{vech}(\Sigma_{t+1}) = c_0 + C_1\mathrm{vech}(\eta_{t+1}\eta'_{t+1}) + C_2\mathrm{vech}(\Sigma_t)$, where vech(.) stacks the columns of the lower triangular part of its argument.

[11] Details of the calculations in the perfect-predictor case are provided in the Appendix.

with non-informative priors, is even higher than the 30-year ratio of 1.45 obtained with the benchmark informative priors (see Figure 8 and Table 3). The results under both priors deliver the same message: volatility is substantially higher in the long run when predictors are imperfect. In contrast, predictive variance computed using the perfect-predictor framework (dashed line) is much flatter across horizons, with a 30-year variance ratio of 1.08.

Stambaugh (1999) and Barberis (2000) use data beginning in 1952 to investigate the effects of parameter uncertainty at longer horizons. For a similar post-1951 period (Panel B of Figure 9), we find that the effect of predictor imperfection is especially large. Consistent with Stambaugh and Barberis, we find that predictive variance under perfect predictors is substantially lower at a 120-quarter horizon than a 1-quarter horizon, with a variance ratio of 0.45. In dramatic contrast, this variance ratio is 3.73 when predictor imperfection is incorporated. That is, accounting for predictor imperfection increases the 30-year variance ratio from well below 1 to well above 1, flipping around the answer to the question whether stocks are less volatile in the long run.

We also see that the findings of Stambaugh and Barberis, indicating stocks are less volatile at longer horizons even after incorporating parameter uncertainty, do not obtain over the longer 206-year period. As noted earlier, the predictive variances in Panel A are actually slightly higher at longer horizons in the perfect-predictor case. For both sample periods, however, we see that predictor imperfection produces long-run variances that substantially exceed not only short-run variance but also long-run variances computed under a perfect-predictor assumption.

6. Predictive variance versus true variance

We have thus far analyzed multiperiod return variance from the perspective of an investor who conditions on the historical data but remains uncertain about the true values of the parameters. One can instead conduct inference about the true multiperiod variance implied by those parameters. In that inference setting, a commonly employed statistic is the sample long-horizon variance ratio. Values of such ratios are often less than 1 for stocks, suggesting lower unconditional variance per period at long horizons. Figure 10 plots sample variance ratios for horizons of 2 through 30 years computed with the 206-year sample of annual real log stock returns analyzed above. The calculations use overlapping returns and unbiased variance estimates.[12] Also plotted are percentiles of the variance ratio's Monte Carlo sampling distribution under the null hypothesis that returns are i.i.d. normal. That distribution exhibits significant positive skewness and has nearly 60% of its

[12]Each ratio is computed as $\overline{VR}(q)$ in equation (2.4.37) of Campbell, Lo, and MacKinlay (1997).

mass below 1. The realized value of 0.28 at the 30-year horizon corresponds to a Monte Carlo p-value of about 0.01, supporting the inference that the true 30-year variance ratio lies below 1 (setting aside the multiple-comparison issues of selecting one horizon from many). Panel A of Figure 11 plots the posterior distribution of the 30-year ratio for true unconditional variance, based on the benchmark priors. The posterior probability that this ratio lies below 1 is 63%. We thus see that the variance ratio statistic in a frequentist setting and the posterior distribution in a Bayesian setting both favor the inference that the true unconditional variance ratio is below 1.

Inference about unconditional variance ratios is of limited relevance to investors. One reason is that conditional variance, rather than unconditional variance, is the more relevant quantity when returns are predictable. The ratio of true unconditional variances can be less than 1 while the ratio of true conditional variances exceeds 1, or vice versa. At a horizon of $k = 30$ years, for example, parameter values of $\beta = 0.60$, $R^2 = 0.30$, and $\rho_{uw} = -0.55$ imply a ratio of 0.90 for unconditional variances but 1.20 for conditional variances given μ_T.[13] Even if the parameters and the conditional mean were known, the unconditional variance would not be the appropriate measure from an investor's perspective.

The larger point is that inference about true variance, conditional or unconditional, is distinct from assessing the predictive variance perceived by an investor who does not know the true parameters. This distinction can be drawn clearly in the context of the variance decomposition,

$$\mathrm{Var}(r_{T,T+k}|D_T) = \mathrm{E}\left\{\mathrm{Var}(r_{T,T+k}|\phi, D_T)|D_T\right\} + \mathrm{Var}\left\{\mathrm{E}(r_{T,T+k}|\phi, D_T)|D_T\right\}. \tag{21}$$

The variance on the left-hand side of (21) is the predictive variance. The quantity inside the expectation in the first term, $\mathrm{Var}(r_{T,T+k}|\phi, D_T)$, is the true variance, relevant only to an investor who knows the true parameter vector ϕ (but not μ_T, thus maintaining predictor imperfection). That true variance is the sum of the first four components in equation (12) with the expectations operators removed. The data can imply that this *true* variance is probably *lower* at long horizons than at short horizons while also implying that the *predictive* variance is *higher* at long horizons. In other words, investors who observe D_T can infer that if they were told the true parameter values, they would probably assess 30-year variance to be less than 1-year variance. These investors realize, however, that they do not know the true parameters. As a consequence, they evaluate the posterior mean of the true variance, the first term in (21). That posterior mean can exceed the most likely values of the true variance, because the posterior distribution of the true variance can be skewed (we return to this point below). Moreover, investors must add to that posterior mean the posterior

[13] The relation between the ratios of conditional and unconditional variances is derived in the Appendix. Campbell and Viceira (2002, p. 96) state that the unconditional variance ratio is always greater than the conditional ratio, but it appears they equate single-period conditional and unconditional variances in reaching that conclusion.

variance of the true conditional mean, the second term in (21), which is the same as the estimation-risk term in equation (12). In a sense, investors do conduct inference about true variance—they compute its posterior mean—but they realize that estimate is only part of predictive variance.

The results based on our 206-year sample illustrate how predictive variance can be higher at long horizons while true variance is inferred to be most likely higher at short horizons. Panel B of Figure 11 plots the posterior distribution (using benchmark priors) of the variance ratio

$$V^*(k) = \frac{(1/k)\mathrm{Var}(r_{T,T+k}|\phi, D_T)}{\mathrm{Var}(r_{T+1}|\phi, D_T)}, \tag{22}$$

for $k = 30$ years. The posterior probability that this ratio of true variances lies below 1 is 76%, and the posterior mode is below 0.5. In contrast, recall that 30-year predictive variance is substantially greater than 1-year variance, as shown earlier in Figure 8 and Table 3.

As noted above, the true variance $\mathrm{Var}(r_{T,T+k}|\phi, D_T)$ is the sum of four quantities, the first four components in equation (12) with the expectations removed. The posterior distributions of those four quantities are displayed in Figure 12, again using benchmark priors. Three of the four distributions exhibit significant asymmetry. As a result, less likely values of these quantities exert a disproportionate effect on the posterior means and, therefore, on the first term of the predictive variance in (21). The components reflecting uncertainty about current and future μ_t are positively skewed, so their contributions to predictive variance exceed what they would be if evaluated at the most likely parameter values. This feature of parameter uncertainty also helps drive predictive variance above what one would infer true variance is most likely to be.

7. Conclusions

We use a predictive system and 206 years of data to compute long-horizon "predictive" variance of annual real stock returns from the perspective of an investor who recognizes that parameters are uncertain and predictors are imperfect. Mean reversion reduces long-horizon variance considerably, but it is more than offset by other effects. As a result, long-horizon variance substantially exceeds short-horizon variance on a per-year basis. A major contributor to higher long-horizon variance is uncertainty about future expected returns, a component of variance that is inherent to return predictability, especially when expected return is persistent. Estimation risk is another important component of predictive variance that is higher at longer horizons. Uncertainty about current expected return, arising from predictor imperfection, also adds considerably to long-horizon variance. We show that accounting for predictor imperfection is key in reaching the conclusion that stocks are substantially more volatile in the long run. Overall, our results show that long-horizon stock investors face more volatility than short-horizon investors, in contrast to previous research.

In computing predictive variance, we assume that the parameters of the predictive system remain constant over the 206-year sample period. While such an assumption is certainly strong, it also allows us to be conservative in concluding that stocks are more volatile at long horizons. Parameter uncertainty, which already contributes substantially to that conclusion, would generally be even greater under alternative scenarios in which investors would effectively have less information about the current values of the parameters.

Although we find that stock volatility is greater at long horizons than at short horizons, this finding does not necessarily imply that long-horizon investors should hold less stock than short-horizon investors. Volatility is only one key ingredient in a problem that no doubt involves other considerations of first-order importance, such as human capital, that are beyond the scope of this study.[14] Investigating asset-allocation decisions while allowing the higher long-run volatility to enter the problem offers an interesting direction for future research.

[14] See Benzoni et al. (2007) for a recent analysis of the role of human capital in portfolio choice.

Appendix

A.1. Derivation of Var($r_{T,T+k} | \mu_T, \phi$)

We can rewrite the AR(1) process for μ_t in equation (2) as an MA(∞) process

$$\mu_t \;=\; \mathrm{E}_r + \sum_{i=0}^{\infty} \beta^i w_{t-i}, \tag{A1}$$

given our assumption that $0 < \beta < 1$. From (1) and (A1), the return k periods ahead is equal to

$$r_{T+k} \;=\; (1-\beta^{k-1})\mathrm{E}_r + \beta^{k-1}\mu_T + \sum_{i=1}^{k-1} \beta^{k-1-i} w_{T+i} + u_{T+k}. \tag{A2}$$

The multiperiod return from period $T+1$ through period $T+k$ is then

$$r_{T,T+k} = \sum_{i=1}^{k} r_{T+i} \;=\; k\mathrm{E}_r + \frac{1-\beta^k}{1-\beta}(\mu_T - \mathrm{E}_r) + \sum_{i=1}^{k-1} \frac{1-\beta^{k-i}}{1-\beta} w_{T+i} + \sum_{i=1}^{k} u_{T+i}. \tag{A3}$$

The conditional variance of the k-period return can be obtained from equation (A3) as

$$\begin{aligned}\mathrm{Var}\,(r_{T,T+k} | \, \mu_T, \phi) \;&=\; k\sigma_u^2 + \frac{\sigma_w^2}{(1-\beta)^2}\left[k - 1 - 2\beta\frac{1-\beta^{k-1}}{1-\beta} + \beta^2\frac{1-\beta^{2(k-1)}}{1-\beta^2}\right] \\ &\quad + \frac{2\sigma_{uw}}{1-\beta}\left[k - 1 - \beta\frac{1-\beta^{k-1}}{1-\beta}\right].\end{aligned} \tag{A4}$$

Equation (A4) can then be written as in equations (4) to (7), where $\bar{d}$ arises from the relation

$$\sigma_w^2 = \sigma_\mu^2(1-\beta^2) = \sigma_r^2 R^2(1-\beta^2) = (\sigma_u^2/(1-R^2))R^2(1-\beta^2). \tag{A5}$$

A.2. Properties of $A(k)$ and $B(k)$

1. $A(1) = 0, \quad B(1) = 0$
2. $A(k) \to 1$ as $k \to \infty, \quad B(k) \to 1$ as $k \to \infty$
3. $A(k+1) > A(k)\ \forall k, \quad B(k+1) > B(k)\ \forall k$
4. $A(k) \geq B(k)\ \forall k$, with a strict inequality for all $k > 1$
5. $0 \leq A(k) < 1, \quad 0 \leq B(k) < 1$
6. $A(k)$ converges to one more quickly than $B(k)$

Properties 1 and 2 are obvious. Properties 3 and 4 are proved below. Property 5 follows from Properties 1–3. Property 6 follows from Properties 1–4.

Proof that $A(k+1) > A(k)\ \forall k$:

$$\begin{aligned}
A(k+1) &= 1 + \frac{1}{k+1}\left[-1-\beta(1+\beta+\ldots+\beta^{k-2}+\beta^{k-1})\right] \\
&= 1 + \frac{k}{k+1}\frac{1}{k}\left[-1-\beta(1+\beta+\ldots+\beta^{k-2}+\beta^{k-1})\right] \\
&= 1 + \frac{k}{k+1}\left[A(k)-1-\frac{\beta^k}{k}\right],
\end{aligned}$$

which exceeds $A(k)$ if and only if $A(k) < 1-\beta^k$. This is indeed true because

$$A(k) = 1-\frac{1}{k}-\frac{1}{k}\left[\beta^1+\ldots+\beta^{k-1}\right] = 1-\frac{1}{k}\left[\beta^0+\beta^1+\ldots+\beta^{k-1}\right] < 1-\frac{1}{k}\left[k\beta^k\right] = 1-\beta^k.$$

Proof that $B(k+1) > B(k)\ \forall k$:

$$\begin{aligned}
&B(k+1) \\
&= 1 + \frac{1}{k+1}\left[-1-2\beta(1+\beta+\ldots+\beta^{k-2}+\beta^{k-1})+\beta^2(1+\beta^2+\ldots+(\beta^2)^{k-2}+(\beta^2)^{k-1})\right] \\
&= 1 + \frac{k}{k+1}\frac{1}{k}\left[\left\{-1-2\beta(1+\beta+\ldots+\beta^{k-2})+\beta^2(1+\beta^2+\ldots+(\beta^2)^{k-2})\right\}-2\beta^k+\beta^{2k}\right] \\
&= 1 + \frac{k}{k+1}\left[B(k)-1+\frac{1}{k}\left(-2\beta^k+\beta^{2k}\right)\right],
\end{aligned}$$

which exceeds $B(k)$ if and only if $B(k) < 1+\beta^{2k}-2\beta^k$. This is indeed true because

$$\begin{aligned}
B(k) &= 1-2\frac{1}{k}+\frac{1}{k}-2\frac{1}{k}\left(\beta+\ldots+\beta^{k-2}+\beta^{k-1}\right)+\frac{1}{k}\left(\beta^2+\ldots+(\beta^2)^{k-2}+(\beta^2)^{k-1}\right) \\
&= 1+\frac{1}{k}\left[\left((\beta^2)^0-2\beta^0\right)+\left((\beta^2)^1-2\beta^1\right)+\ldots+\left((\beta^2)^{k-1}-2\beta^{k-1}\right)\right] \\
&< 1+\frac{1}{k}\left[k\left((\beta^2)^k-2\beta^k\right)\right] \\
&= 1+\beta^{2k}-2\beta^k,
\end{aligned}$$

where the inequality follows from the fact that the function $f(x) = (\beta^2)^x - 2\beta^x$ is increasing in x (because $f'(x) = 2(\ln\beta)\beta^x(\beta^x-1) > 0$, for $0 < \beta < 1$).

Proof that $A(k) > B(k)\ \forall k > 1$:

$$\begin{aligned}
B(k)-A(k) &= \frac{1}{k}\left[\beta^2\frac{1-\beta^{2(k-1)}}{1-\beta^2}-\beta\frac{1-\beta^{k-1}}{1-\beta}\right] = \frac{1}{k}\left[\beta^2+\ldots+(\beta^2)^{k-1}-\left(\beta+\ldots+\beta^{k-1}\right)\right] \\
&= \frac{1}{k}\sum_{i=1}^{k-1}\left(\beta^{2i}-\beta^i\right) = \frac{1}{k}\sum_{i=1}^{k-1}\beta^i\left(\beta^i-1\right) < 0.
\end{aligned}$$

A.3. Decomposition of $\mathrm{Var}\{\mathrm{E}(r_{T,T+k}|\mu_T, \phi, D_T)|D_T\}$

Let $E_{T,k} = \mathrm{E}(r_{T,T+k}|\mu_T, \phi, D_T)$. The variance of $E_{T,k}$ given D_T can be decomposed as

$$\mathrm{Var}\{E_{T,k}|D_T\} = \mathrm{E}\{\mathrm{Var}[E_{T,k}|\phi, D_T]|D_T\} + \mathrm{Var}\{\mathrm{E}[E_{T,k}|\phi, D_T]|D_T\}. \tag{A6}$$

To simplify each term on the right-hand side, observe from equations (1), (2), and (3), that

$$\begin{aligned} E_{T,k} &= \mathrm{E}(r_{T+1} + r_{T+2} + \ldots + r_{T+k}|\mu_T, \phi, D_T) \\ &= \mathrm{E}(\mu_T + \mu_{T+1} + \ldots + \mu_{T+k-1}|\mu_T, \phi) \\ &= kE_r + \frac{1-\beta^k}{1-\beta}(\mu_T - E_r). \end{aligned} \tag{A7}$$

Taking the first and second moments of (A7), using (10) and (11), then gives

$$\mathrm{E}[E_{T,k}|\phi, D_T] = kE_r + \frac{1-\beta^k}{1-\beta}(b_T - E_r) \tag{A8}$$

$$\mathrm{Var}[E_{T,k}|\phi, D_T] = \left(\frac{1-\beta^k}{1-\beta}\right)^2 q_T. \tag{A9}$$

Substituting (A8) and (A9) into (A6) then gives the fourth and fifth terms in (12), using (9).

A.4. Perfect predictors

This paper focuses on the realistic scenario in which the observable predictors x_t are imperfect, in that they do not perfectly capture the conditional expected return μ_t. The Bayesian analysis of the model with imperfect predictors is provided in Pástor and Stambaugh (2008). In this section, we discuss the Bayesian analysis of the model with perfect predictors, for which $\mu_t = a + b'x_t$. In that case, the predictive system is replaced by a model consisting of equations (17) and (18), combined with the following distributional assumption on the residuals:

$$\begin{bmatrix} e_t \\ v_t \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_e^2 & \sigma_{ev} \\ \sigma_{ve} & \Sigma_{vv} \end{bmatrix}\right). \tag{A10}$$

Let δ denote the full set of parameters in equations (17), (18), and (A10). Let Ω denote the covariance matrix in (A10), let B denote the matrix of the slope coefficients in (17) and (18),

$$B = \begin{bmatrix} a & \theta' \\ b & A' \end{bmatrix},$$

and let $c = \mathrm{vec}\,(B)$. Note that δ consists of the elements of c and Ω.

A.4.1. Posterior distributions under perfect predictors

We specify the prior distribution on δ as $p(\delta) = p(c)p(\Omega)$. The priors on c and Ω are non-informative, except for the restriction that the spectral radius of A, $\rho(A)$, is less than 1. The prior on c is $p(c) \propto I[\rho(A) < 1]$, where $I[\,]$ denotes the indicator function. The prior on Ω is $p(\Omega) \propto |\Omega|^{-(m+1)/2}$, where m is the number of rows in Ω (i.e., x_t is $(m-1) \times 1$).

To obtain the posterior distribution of δ, the prior $p(\delta)$ is combined with the normal likelihood function $p(D_T|\delta)$ implied by equation (A10). The posterior draws of δ can be obtained by applying standard results from the multivariate regression model (e.g., Zellner, 1971). Define the following notation: $r = [r_1\; r_2\; \cdots\; r_T]'$, $Q^+ = [x_1\; x_2\; \cdots\; x_T]'$, $Q = [x_0\; x_1\; \cdots\; x_{T-1}]'$, $X = [\iota_T\; Q]$, where ι_T denotes a $T \times 1$ vector of ones, $Y = \left[r\; Q^+\right]$, $\hat{B} = (X'X)^{-1}X'Y$, and $S = (Y - X\hat{B})'(Y - X\hat{B})$. We first draw Ω^{-1} from a Wishart distribution with $T-m$ degrees of freedom and parameter matrix S^{-1}. Given that draw of Ω^{-1}, we then draw c from a normal distribution with mean $\hat{c} = \mathrm{vec}\,(\hat{B})$ and covariance matrix $\Omega \otimes (X'X)^{-1}$. That draw of δ is retained as a draw from $p(\delta|D_T)$ if $\rho(A) < 1$.

A.4.2. Predictive variance under perfect predictors

The conditional moments of the k-period return $r_{T,T+k}$ are given by

$$\mathrm{E}(r_{T,T+k}|D_T,\delta) = ka + b'\Psi_{k-1}\theta + b'\Lambda_k x_T \tag{A11}$$

$$\mathrm{Var}(r_{T,T+k}|D_T,\delta) = k\sigma_e^2 + 2b'\Psi_{k-1}\sigma_{ve} + b'\left(\sum_{i=1}^{k-1}\Lambda_i\Sigma_{vv}\Lambda_i'\right)b, \tag{A12}$$

where

$$\Lambda_i = I + A + \cdots + A^{i-1} = (I-A)^{-1}(I-A^i) \tag{A13}$$

$$\Psi_{k-1} = \Lambda_1 + \Lambda_2 + \cdots + \Lambda_{k-1} = (I-A)^{-1}[kI - (I-A)^{-1}(I-A^k)]. \tag{A14}$$

The first term in (A12) reflects i.i.d. uncertainty. The second term reflects correlation between unexpected returns and innovations in future x_{T+i}'s, which deliver innovations in future μ_{T+i}'s. That term can be positive or negative and captures any mean reversion. The third term, always positive, reflects uncertainty about future x_{T+i}'s, and thus uncertainty about future μ_{T+i}'s. This third term, which contains a summation, can also be written without the summation as

$$b'\left(\sum_{i=1}^{k-1}\Lambda_i\Sigma_{vv}\Lambda_i'\right)b = \left(b'\otimes b'\right)\left[(I-A)^{-1}\otimes(I-A)^{-1}\right]\Big[kI - \Lambda_k\otimes I - I\otimes\Lambda_k + (I - A\otimes A)^{-1}(I-(A\otimes A)^k)\Big]\mathrm{vec}\,(\Sigma_{vv})\,.$$

Applying the decomposition in equation (9), the predictive variance $\mathrm{Var}(r_{T,T+k}|D_T)$ can be computed as the sum of the posterior mean of the right-hand side of equation (A12) and the posterior variance of the right-hand side of equation (A11). These posterior moments are computed from the posterior draws of δ, which are described in the previous subsection.

A.5. Relation between conditional and unconditional variance ratios

The unconditional variance (which does not condition on μ_T) is given by

$$\begin{aligned}
\mathrm{Var}(r_{T,T+k}|\phi) &= \mathrm{E}[\mathrm{Var}(r_{T,T+k}|\mu_T,\phi,D_T)|\phi] + \mathrm{Var}[\mathrm{E}(r_{T,T+k}|\mu_T,\phi,D_T)|\phi] \\
&= \mathrm{Var}(r_{T,T+k}|\mu_T,\phi) + \left(\frac{1-\beta^k}{1-\beta}\right)^2 \mathrm{Var}(\mu_T|\phi) \\
&= \mathrm{Var}(r_{T,T+k}|\mu_T,\phi) + \left(\frac{1-\beta^k}{1-\beta}\right)^2 \sigma_u^2 \left(\frac{R^2}{1-R^2}\right),
\end{aligned} \tag{A15}$$

using equation (A7). It follows from equation (4) that

$$\mathrm{Var}(r_{T,T+1}|\mu_T,\phi) = \sigma_u^2. \tag{A16}$$

Combining equations (A15) and (A16) for $k = 1$ gives

$$\mathrm{Var}(r_{T,T+1}|\phi) = \mathrm{Var}(r_{T,T+1}|\mu_T,\phi) + \frac{\sigma_u^2 R^2}{1-R^2} = \frac{\sigma_u^2}{1-R^2} = \frac{\mathrm{Var}(r_{T,T+1}|\mu_T,\phi)}{1-R^2}. \tag{A17}$$

The unconditional variance ratio $V_u(k)$ and the conditional variance ratio $V_c(k)$ can then be related as follows, combining (A15), (A17), and (8):

$$\begin{aligned}
V_u(k) &= \frac{(1/k)\mathrm{Var}(r_{T,T+k}|\phi)}{\mathrm{Var}(r_{T,T+1}|\phi)} \\
&= \frac{(1/k)\mathrm{Var}(r_{T,T+k}|\phi)(1-R^2)}{\mathrm{Var}(r_{T,T+1}|\mu_T,\phi)} \\
&= \frac{(1/k)\mathrm{Var}(r_{T,T+k}|\mu_T,\phi)(1-R^2)}{\mathrm{Var}(r_{T,T+1}|\mu_T,\phi)} + \frac{1}{k}\left(\frac{1-\beta^k}{1-\beta}\right)^2 R^2 \\
&= (1-R^2)V_c(k) + \frac{1}{k}\left(\frac{1-\beta^k}{1-\beta}\right)^2 R^2.
\end{aligned} \tag{A18}$$

A.6. Permanent and temporary price components in our setting

Fama and French (1988), Summers (1986), and others employ a model in which the log stock price p_t is the sum of a random walk s_t and a stationary component y_t that follows an AR(1) process:

$$\begin{aligned} p_t &= s_t + y_t && \text{(A19)} \\ s_t &= \mu + s_{t-1} + \epsilon_t && \text{(A20)} \\ y_t &= b y_{t-1} + e_t, && \text{(A21)} \end{aligned}$$

where e_t and ϵ_t are mean-zero variables independent of each other, and $|b| < 1$. Noting that $r_{t+1} = p_{t+1} - p_t$, it is easy to verify that equations (A19) through (A21) deliver a special case of our model in equations (1) and (2), in which

$$\begin{aligned} E_r &= \mu && \text{(A22)} \\ \beta &= b && \text{(A23)} \\ \mu_t &= \mu - (1-b) y_t && \text{(A24)} \\ u_{t+1} &= \epsilon_{t+1} + e_{t+1} && \text{(A25)} \\ w_{t+1} &= -(1-b) e_{t+1}. && \text{(A26)} \end{aligned}$$

This special case has the property

$$\sigma_{uw} = \text{Cov}(u_{t+1}, w_{t+1}) = -(1-b)\sigma_e^2 < 0, \tag{A27}$$

implying the presence of mean reversion. We also see

$$\sigma_\mu^2 \;=\; \text{Var}(\mu_t) = (1-b)^2 \sigma_y^2 = (1-b)^2 \frac{\sigma_e^2}{1-b^2} = \frac{1-b}{1+b}\sigma_e^2 \tag{A28}$$

and, therefore, using (19),

$$\text{Cov}(r_{t+1}, r_t) \;=\; \beta\sigma_\mu^2 + \sigma_{uw} = \frac{b(1-b)}{1+b}\sigma_e^2 - (1-b)\sigma_e^2 = -\frac{1-b}{1+b}\sigma_e^2 < 0. \tag{A29}$$

Thus, under (A19) through (A21) with $b > 0$, all autocovariances in (19) are negative and all unconditional variance ratios are less than 1.

Table 1

Effects of Parameter Uncertainty on 20-Year Variance Ratio

The table displays the ratio $(1/20)\text{Var}(r_{T,T+20}|D_T)/\text{Var}(r_{T+1}|D_T)$, where D_T is information used by an investor at time T. The value of the ratio is computed under various parametric scenarios for β (autocorrelation of the conditional expected return μ_t), R^2 (fraction of variance in r_{t+1} explained by μ_t), ρ_{uw} (correlation between unexpected returns and innovations in expected returns), $\rho_{\mu b}$ (correlation between μ_T and its best available estimate given D_T), and E_r (the unconditional mean return). For β, R^2, ρ_{uw}, and $\rho_{\mu b}$, each parameter is either drawn from its density in Figure 4 when uncertain or set to a fixed value. The parameters β, R^2, and ρ_{uw} are set to their medians when held fixed, while $\rho_{\mu b}$ is fixed at its median as well as 0 and 1. The medians are 0.86 for β, 0.12 for R^2, -0.66 for ρ_{uw}, and 0.70 for $\rho_{\mu b}$. The variance of E_r given D_T is either 0 (when fixed) or 1/200 times the expected variance of one-year returns (when uncertain).

fixed (F) or uncertain (U)				$\rho_{\mu b}$ fixed at			$\rho_{\mu b}$
β	R^2	ρ_{uw}	E_r	0	0.70	1	uncertain
F	F	F	F	0.95	0.87	0.77	0.87
U	F	F	F	1.20	1.06	0.90	1.06
F	U	F	F	1.05	0.97	0.87	0.97
F	F	U	F	1.02	0.94	0.84	0.94
F	F	F	U	1.05	0.97	0.88	0.97
U	U	U	F	1.36	1.22	1.06	1.22
U	U	U	U	1.45	1.32	1.17	1.32

Table 2
Predictive Regressions
1802–2007

This table summarizes the results from predictive regressions $r_t = a + b'x_{t-1} + e_t$, where r_t denotes annual real log stock market return and x_{t-1} contains the predictors (listed in the column headings) lagged by one year. Innovations in expected returns are constructed as $b'v_t$, where v_t contains the disturbances estimated in a vector autoregression for the predictors, $x_t = \theta + Ax_{t-1} + v_t$. The table reports the estimated slope coefficients $\hat{b}$, the correlation Corr($e_t, b'v_t$) between unexpected returns and innovations in expected returns, and the (unadjusted) R^2 from the predictive regression. The independent variables are rescaled to have unit variance. The correlations and R^2's are reported in percent (i.e., $\times 100$). The OLS t-statistics are given in parentheses "( )". The t-statistic of Corr($e_t, b'v_t$) is computed as the t-statistic of the slope from the regression of the sample residuals $\hat{e}_t$ on $\hat{b}\hat{v}_t$. The p-values associated with all t-statistics and R^2's are computed by bootstrapping and reported in brackets "[ ]".

Dividend Yield	Term Spread	Bond Yield	Corr($e_t, b'v_t$)	R^2
0.023			-56.515	1.714
(1.891)			(-9.808)	[0.070]
[0.057]			[1.000]	
	0.008		22.445	0.232
	(0.690)		(3.298)	[0.498]
	[0.236]		[0.000]	
		0.025	-19.231	2.163
		(2.129)	(-2.806)	[0.034]
		[0.018]	[0.997]	
0.031	0.028	0.028	-13.754	5.558
(2.383)	(2.137)	(2.373)	(-1.988)	[0.013]
[0.021]	[0.017]	[0.010]	[0.973]	

Table 3
Variance Ratios and Components of Long-Horizon Variance

The first row of each panel reports the ratio $(1/k)\mathrm{Var}(r_{T,T+k}|D_T)/\mathrm{Var}(r_{T+1}|D_T)$, where $\mathrm{Var}(r_{T,T+k}|D_T)$ is the predictive variance of the k-year return based on 206 years of annual data for real equity returns and the three predictors over the 1802–2007 period. The second row reports $\mathrm{Var}(r_{T,T+k}|D_T)$, multiplied by 100. The remaining rows report the five components of $\mathrm{Var}(r_{T,T+k}|D_T)$, also multiplied by 100 (they add up to total variance). Panel A contains results for $k = 15$ years, and Panel B contains results for $k = 30$ years. Results are reported under each of three priors for ρ_{uw}, R^2, and β. As the prior for one of the parameters departs from the benchmark, the priors on the other two parameters are held at the benchmark priors. The "tight" priors, as compared to the benchmarks, are more concentrated towards -1 for ρ_{uw}, 0 for R^2, and 1 for β; the "loose" priors are less concentrated in those directions.

	ρ_{uw}			R^2			β		
Prior	Tight	Bench	Loose	Tight	Bench	Loose	Tight	Bench	Loose
Panel A. Investment Horizon $k = 15$ years									
Variance Ratio	1.13	1.17	1.10	1.15	1.17	1.03	1.20	1.17	1.15
Predictive Variance	3.30	3.43	3.24	3.45	3.43	2.94	3.52	3.43	3.39
IID Component	2.59	2.60	2.59	2.75	2.60	2.43	2.58	2.60	2.60
Mean Reversion	-3.05	-2.96	-3.04	-2.15	-2.96	-3.51	-3.13	-2.96	-2.94
Uncertain Future μ	1.60	1.57	1.58	0.86	1.57	2.16	1.70	1.57	1.55
Uncertain Current μ	0.89	0.89	0.88	0.60	0.89	1.00	1.07	0.89	0.86
Estimation Risk	1.26	1.34	1.22	1.39	1.34	0.85	1.30	1.34	1.33
Panel B. Investment Horizon $k = 30$ years									
Variance Ratio	1.40	1.45	1.33	1.39	1.45	1.21	1.53	1.45	1.43
Predictive Variance	4.09	4.27	3.91	4.16	4.27	3.46	4.49	4.27	4.20
IID Component	2.59	2.60	2.59	2.75	2.60	2.43	2.58	2.60	2.60
Mean Reversion	-4.51	-4.38	-4.47	-3.37	-4.38	-4.84	-4.70	-4.38	-4.33
Uncertain Future μ	3.49	3.43	3.38	2.11	3.43	4.03	3.80	3.43	3.34
Uncertain Current μ	0.97	0.97	0.95	0.80	0.97	0.88	1.18	0.97	0.94
Estimation Risk	1.54	1.65	1.46	1.87	1.65	0.95	1.64	1.65	1.65

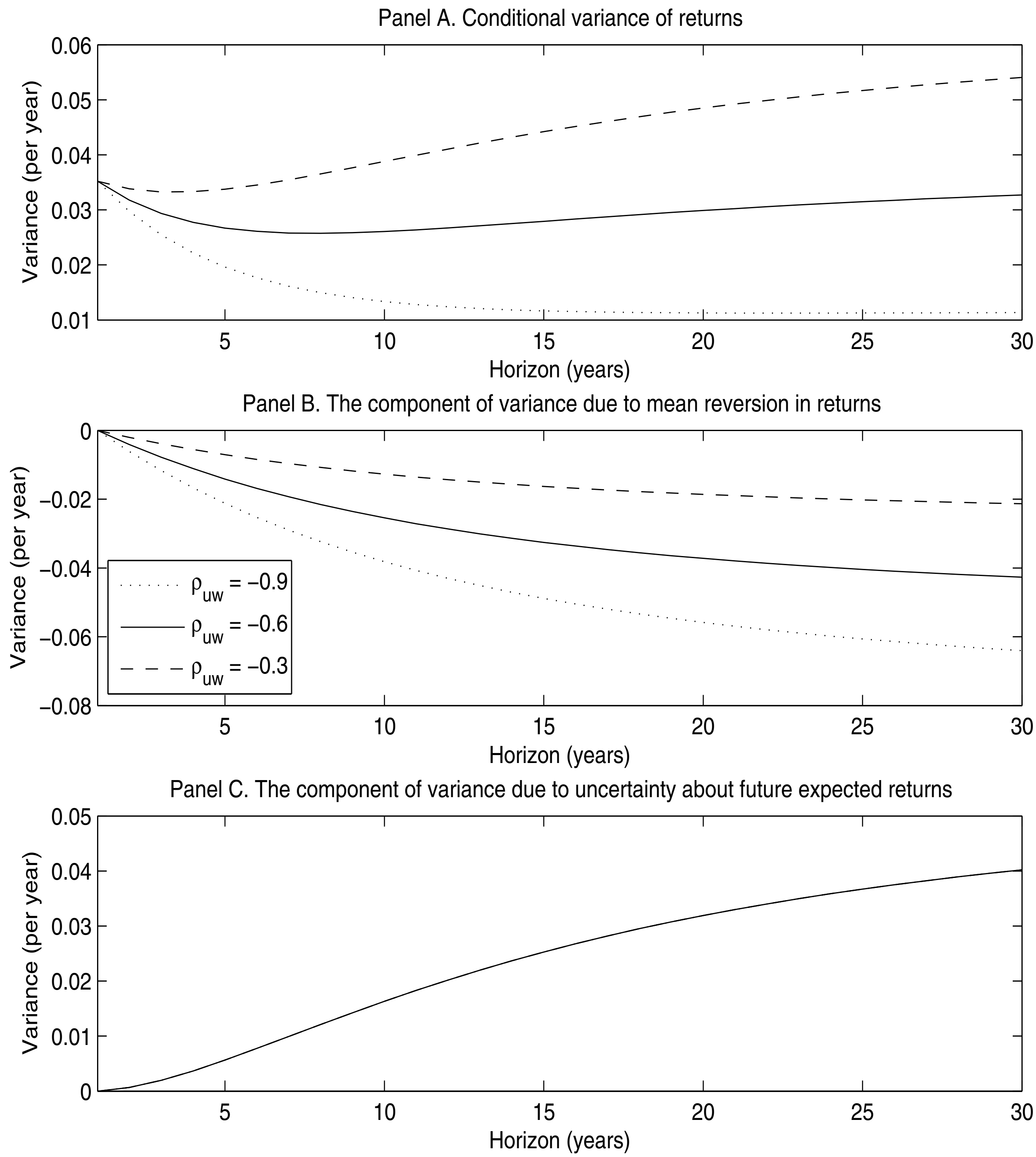

Figure 1. Conditional multiperiod variance and its components for different values of ρ_{uw}. Panel A plots the conditional per-period variance of multiperiod returns from equation (4), $\mathrm{Var}(r_{T,T+k}|\mu_T,\phi)/k$, as a function of the investment horizon k, for three different values of ρ_{uw}. Panel B plots the component of the variance that is due to mean reversion in returns, $\sigma_u^2 2\bar{d}\rho_{uw}A(k)$. Panel C plots the component of this variance that is due to uncertainty about future values of the expected return, $\sigma_u^2\bar{d}^2B(k)$. For all three values of ρ_{uw}, variances are computed with $\beta = 0.85$, $R^2 = 0.12$, and an unconditional standard deviation of returns of 20% per year.

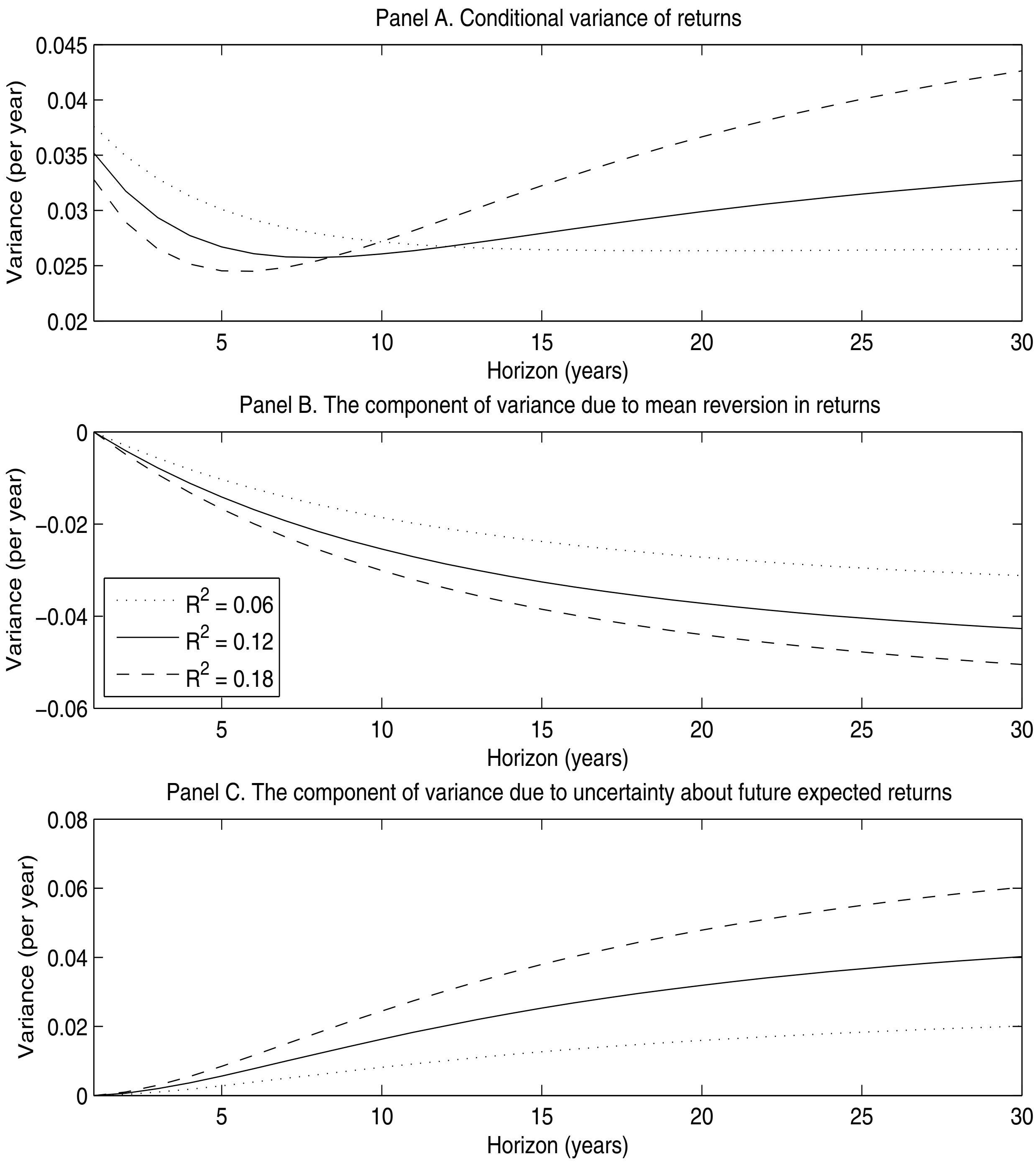

Figure 2. Conditional multiperiod variance and its components for different values of R^2. Panel A plots the conditional per-period variance of multiperiod returns from equation (4), $\mathrm{Var}(r_{T,T+k}|\mu_T,\phi)/k$, as a function of the investment horizon k, for three different values of R^2. Panel B plots the component of the variance that is due to mean reversion in returns, $\sigma_u^2 2\bar{d}\rho_{uw}A(k)$. Panel C plots the component of this variance that is due to uncertainty about future values of the expected return, $\sigma_u^2\bar{d}^2B(k)$. For all three values of R^2, variances are computed with $\beta = 0.85$, $\rho_{uw} = -0.6$, and an unconditional standard deviation of returns of 20% per year.

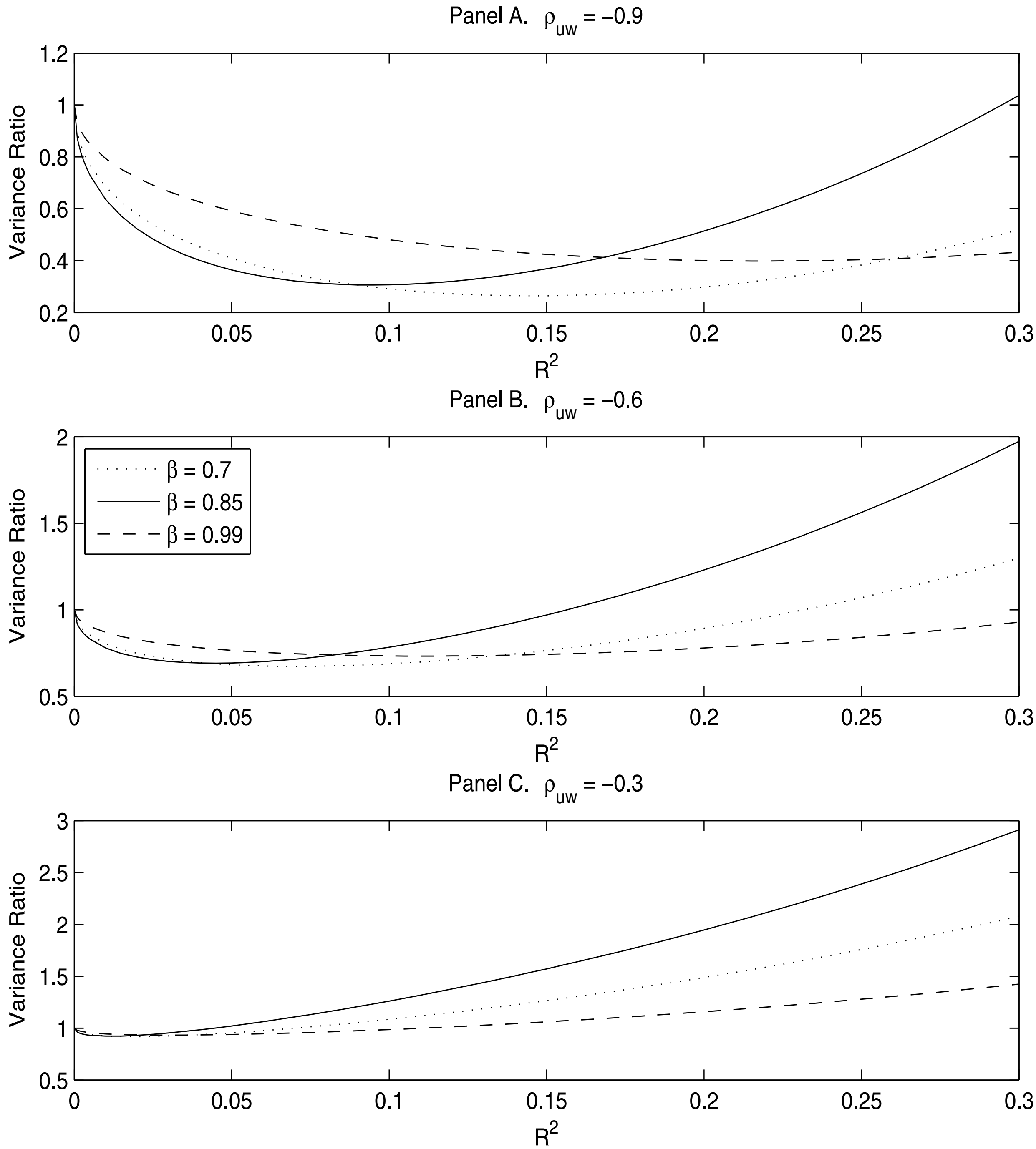

Figure 3. Variance ratios at the 20-year horizon. This figure plots $V_c(k)$ for $k = 20$ years, where $V_c(k)$ denotes the ratio of the conditional variance of k-period returns to the conditional variance of 1-period returns. This ratio is formally defined in equation (8).

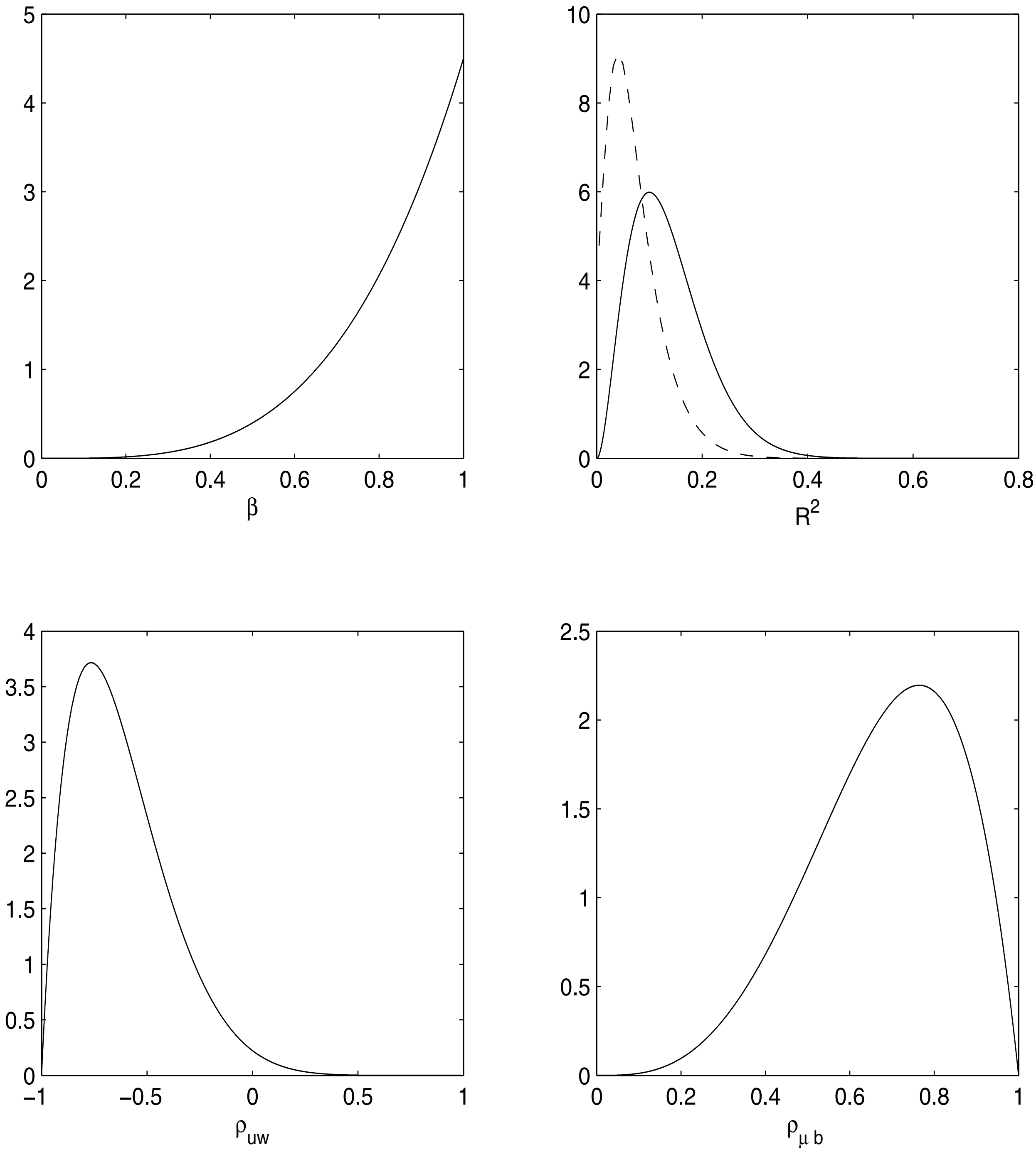

Figure 4. Distributions for uncertain parameters The plots display the probability densities used to illustrate the effects of parameter uncertainty on long-run variance. In the R^2 panel, the solid line plots the density of the true R^2 (predictability given μ_T), and the dashed line plots the implied density of the R-squared in a regression of returns on b_T. The dashed line incorporates the uncertainty about $\rho_{\mu b}$.

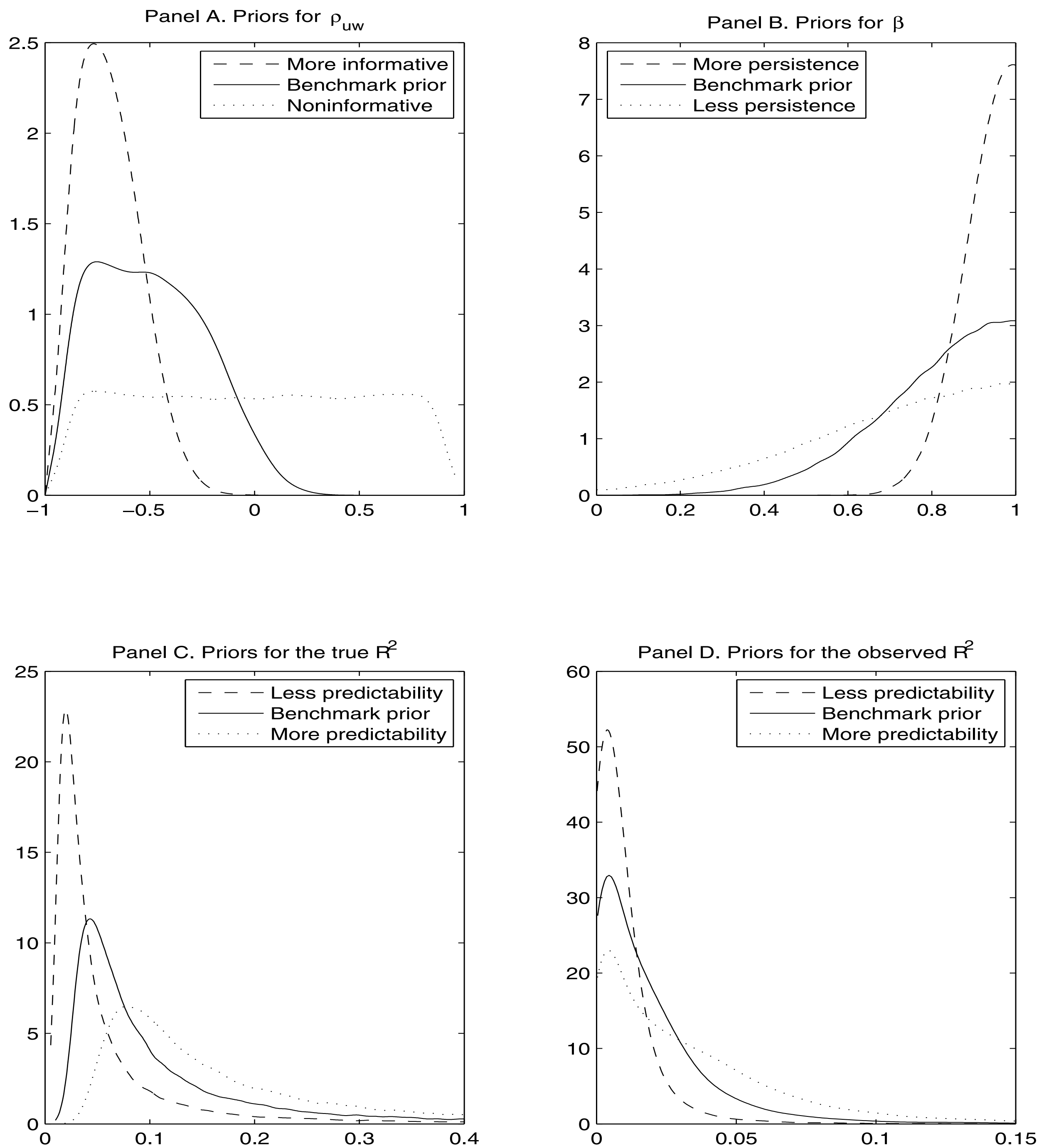

Figure 5. Prior distributions of parameters. The plots display the prior distributions for β, ρ_{uw}, the true R^2 (fraction of variance in the return r_{t+1} explained by the conditional mean μ_t), and the "observed" R^2 (fraction of variance in r_{t+1} explained by the observed predictors x_t). The priors shown for the observed R^2 correspond to the three priors for the true R^2 and the benchmark priors for β and ρ_{uw}.

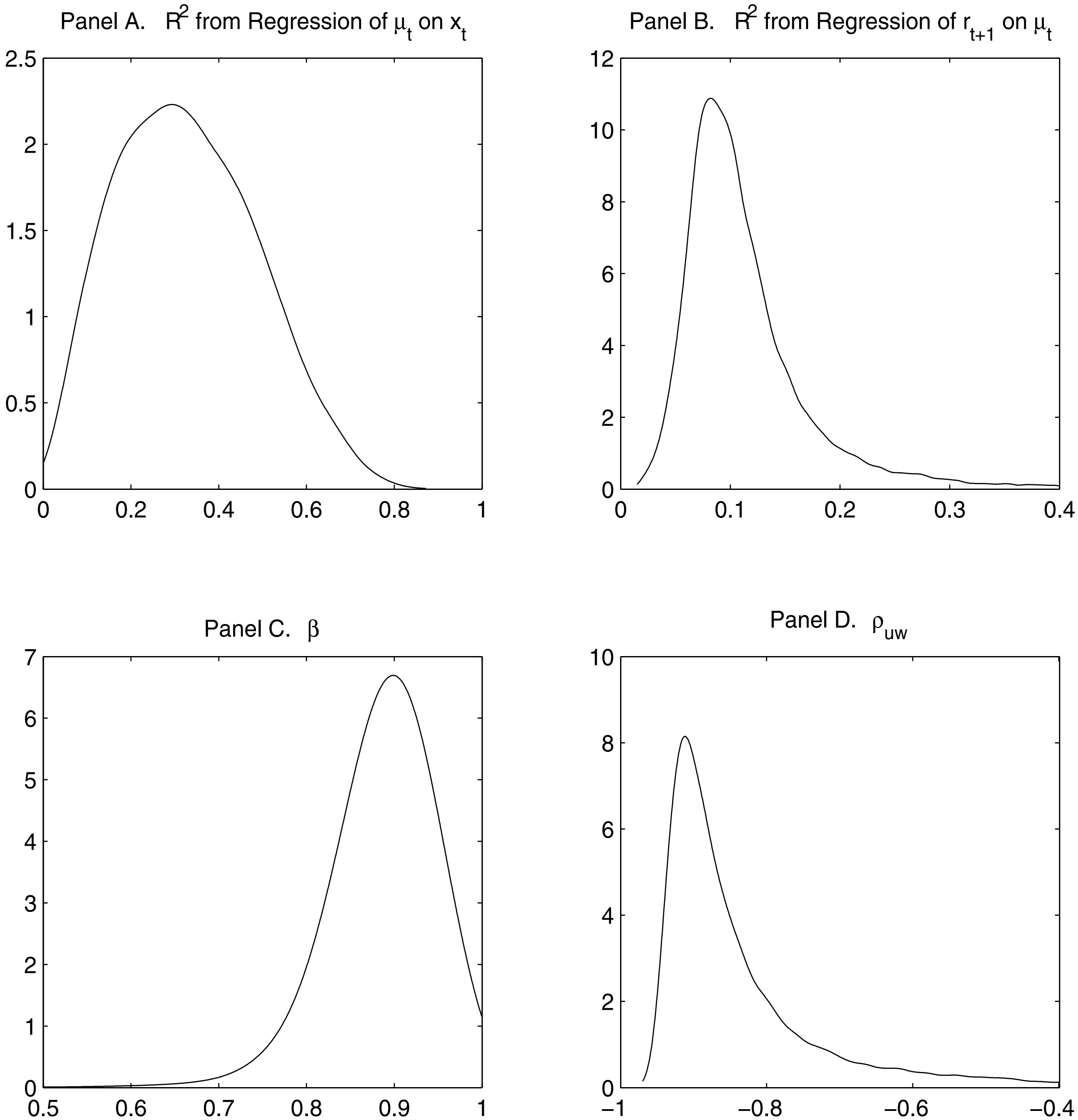

Figure 6. Posterior distributions of parameters. Panel A plots the posterior of the fraction of variance in the conditional expected return μ_t that can be explained by the predictors. Panel B plots the posterior of the true R^2 (fraction of variance in the return r_{t+1} explained by μ_t). Panel C plots the posterior of β, and Panel D plots the posterior of ρ_{uw}. These posteriors are obtained under the benchmark priors for β, ρ_{uw}, and R^2. The results are obtained by estimating the predictive system on annual real U.S. stock market returns in 1802-2007. Three predictors are used: the dividend yield, the bond yield, and the term spread.

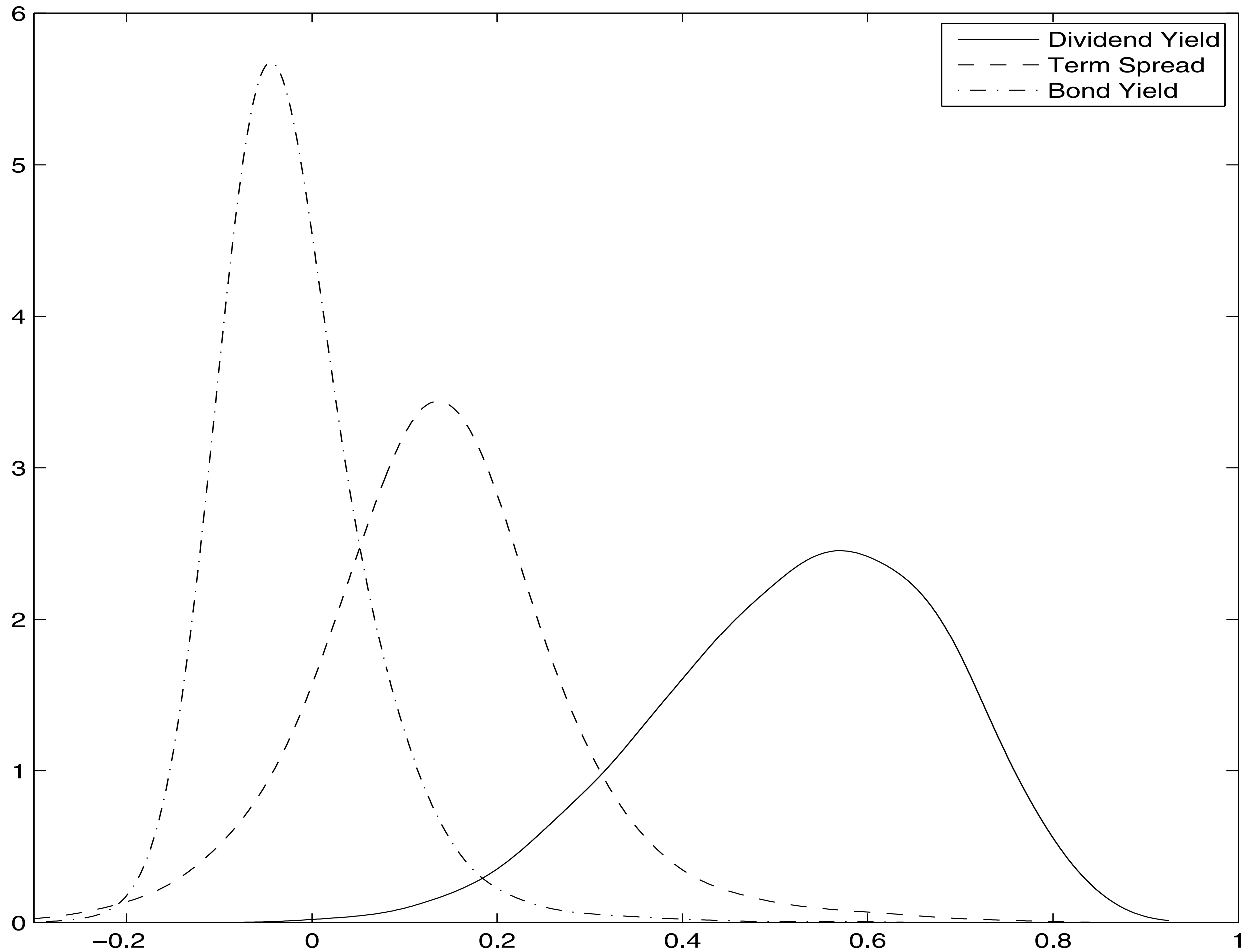

Figure 7. Posterior distributions of partial correlations between each of the three predictors and the conditional expected return μ_t. The results are obtained by estimating the predictive system on annual real U.S. stock market returns in 1802-2007. Three predictors are used: the dividend yield, the bond yield, and the term spread.

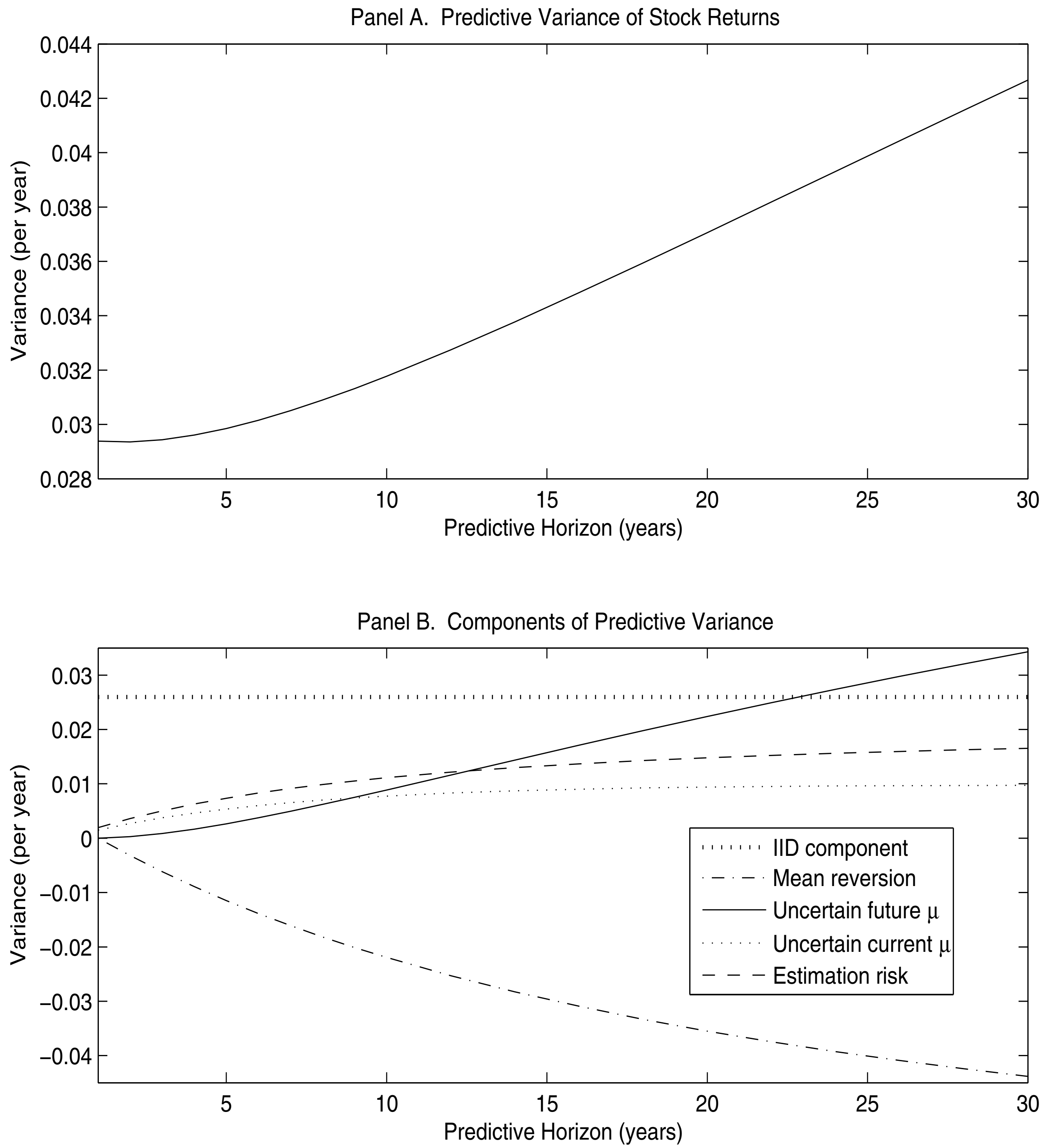

Figure 8. Predictive variance of multiperiod return and its components. Panel A plots the variance of the predictive distribution of long-horizon returns, $\mathrm{Var}(r_{T,T+k}|D_T)$. Panel B plots the five components of the predictive variance. All quantities are divided by k, the number of periods in the return horizon. The results are obtained by estimating the predictive system on annual real U.S. stock market returns in 1802-2007. Three predictors are used: the dividend yield, the bond yield, and the term spread.

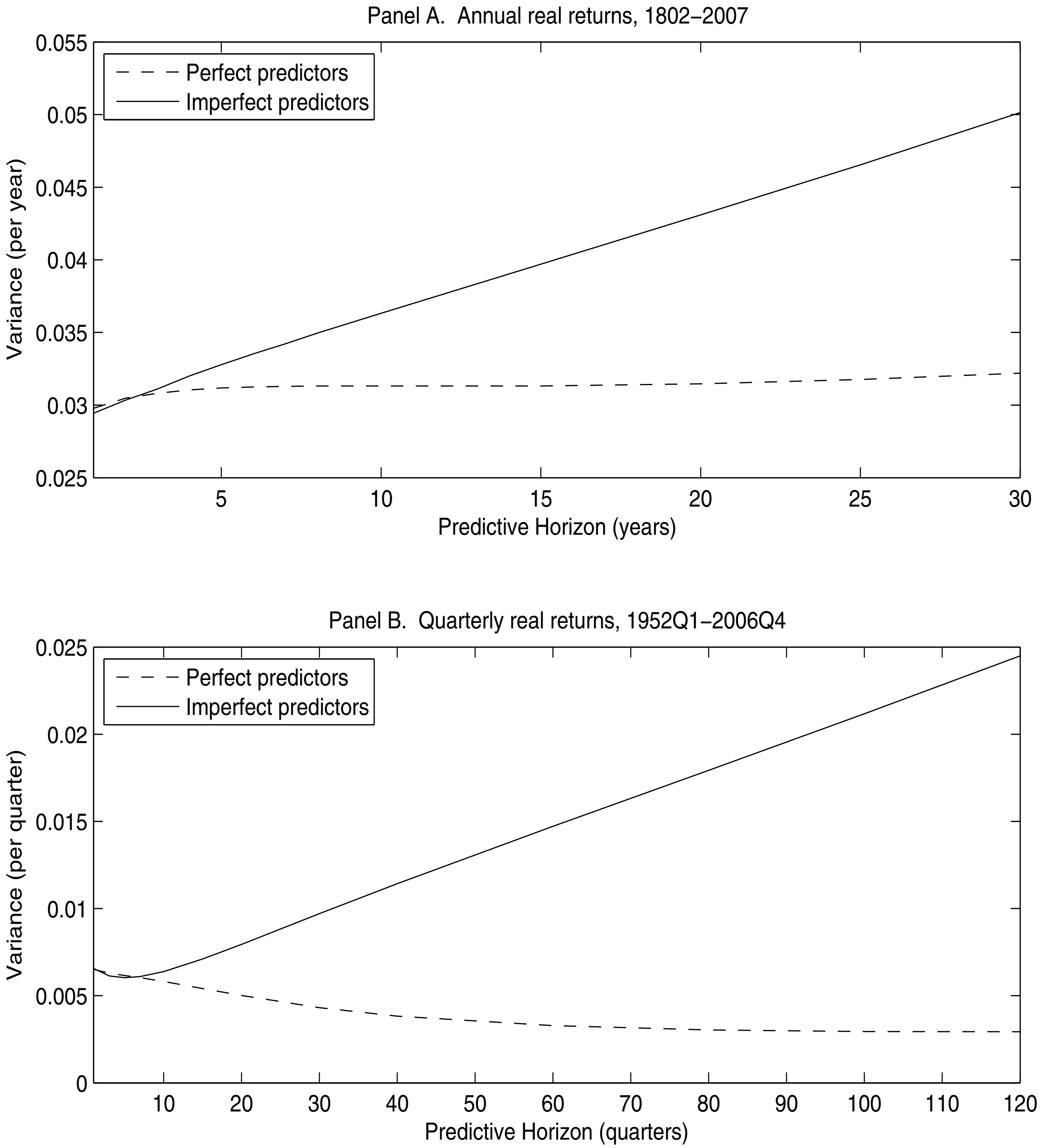

Figure 9. Predictive variance with perfect predictors versus imperfect predictors. The plots display the predictive variance computed under two alternative frameworks: a predictive system allowing imperfect predictors (solid line) and a predictive regression/VAR assuming perfect predictors (dashed line). Panel A is based on annual data from the 1802–2007 period for real U.S. stock returns and three predictors: the dividend yield, the bond yield, and the term spread. Panel B is based on quarterly data from the 1952Q1–2006Q4 period for real returns and three predictors: the dividend yield, CAY, and the bond yield.

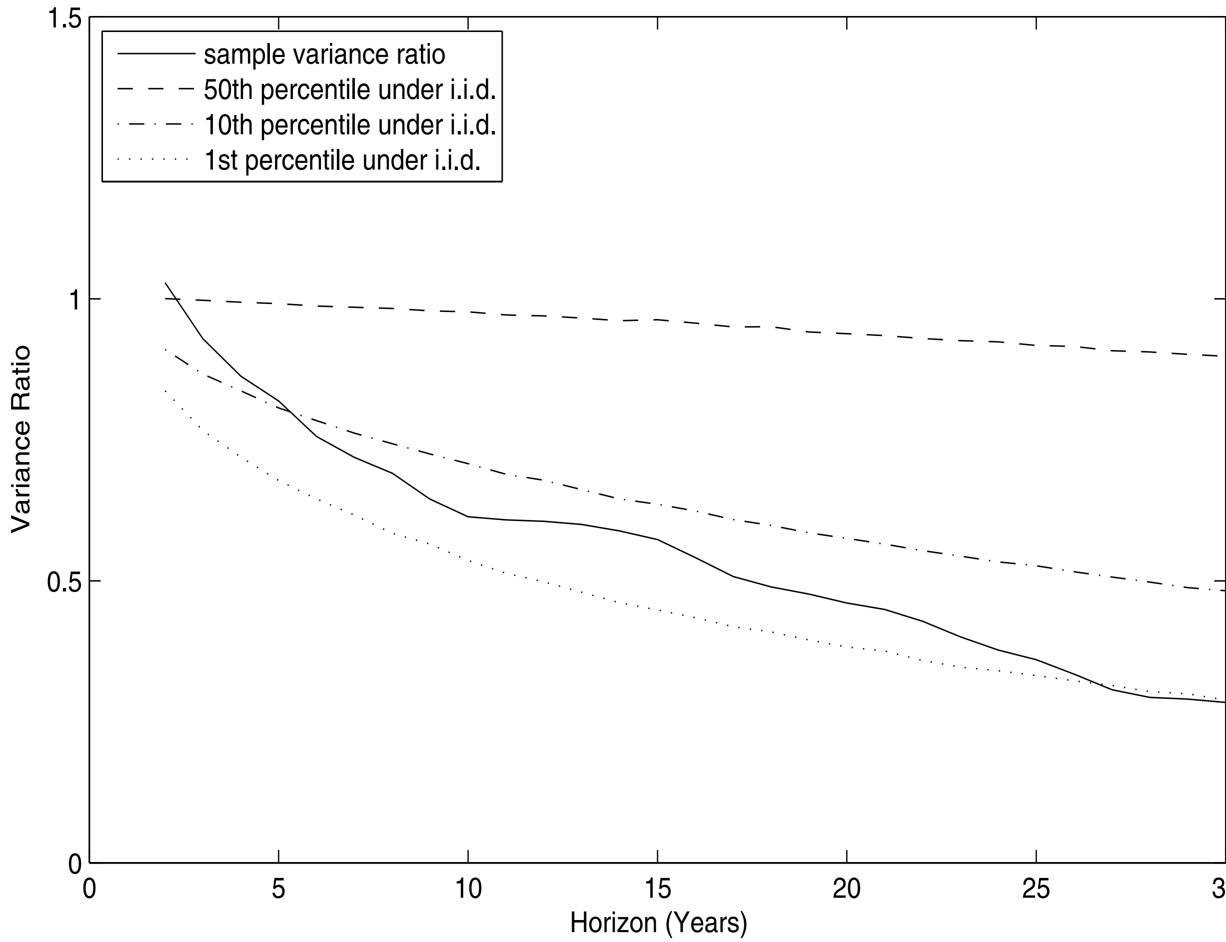

Figure 10. Sample variance ratios of annual real equity returns, 1802–2007. The plot displays the sample variance ratio $\hat{V}(k) = \hat{\text{Var}}(r_{t,t+k})/(k\hat{\text{Var}}(r_{t,t+1}))$, where $\hat{\text{Var}}(r_{t,t+k})$ is the unbiased sample variance of k-year log returns, computed at an overlapping annual frequency. Also shown are the 1st, 10th, and 50th percentiles of the Monte Carlo sampling distribution of $\hat{V}(k)$ under the hypothesis that annual log returns are independently and identically distributed as normal.

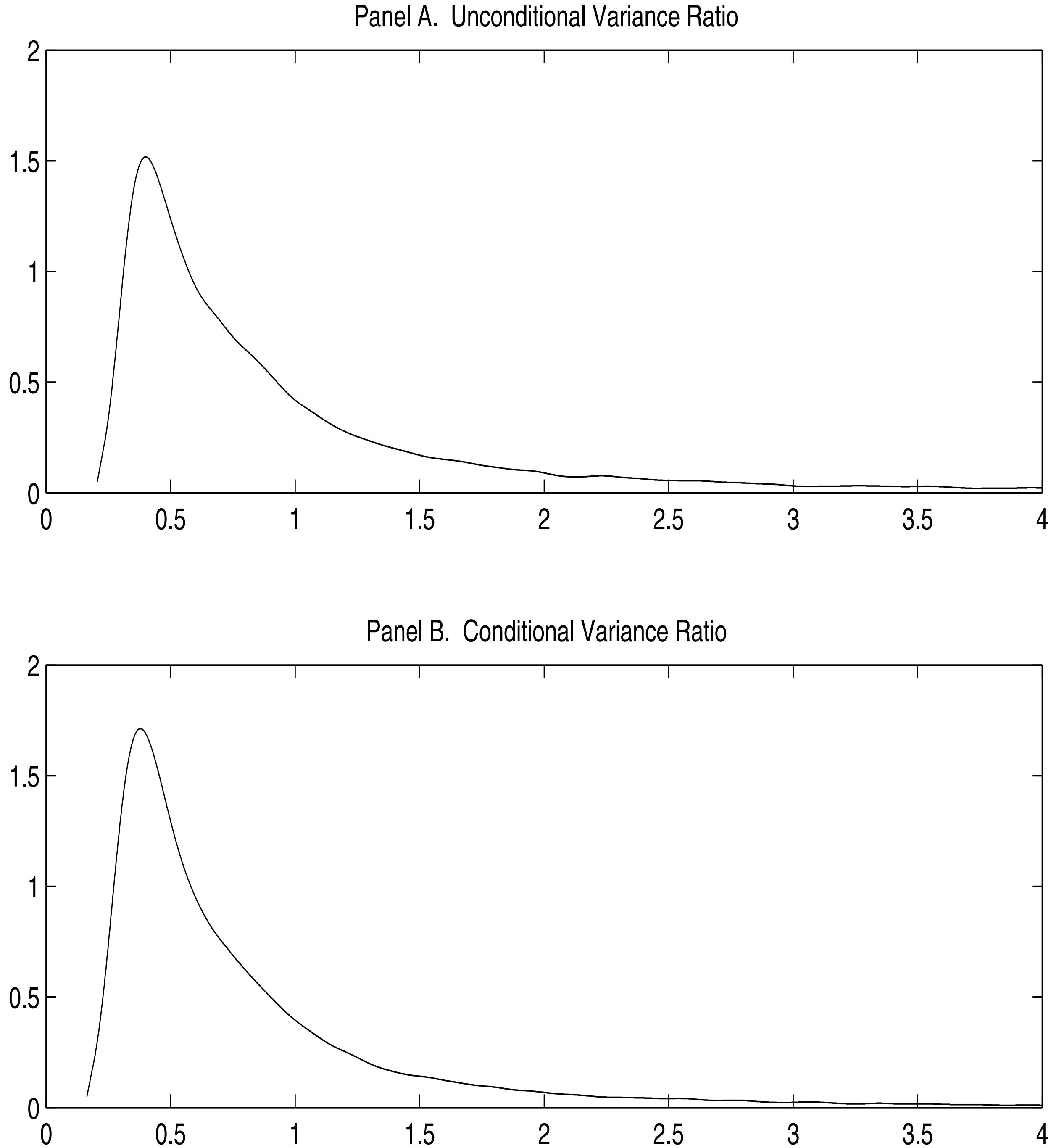

Figure 11. Posterior distributions for 30-year variance ratios. Panel A plots the posterior distribution of the unconditional variance of 30-year stock market returns, $\text{Var}(r_{T,T+30}|\phi)$, divided by 30 times the unconditional variance of one-year returns, $\text{Var}(r_{T+1}|\phi)$. Panel B plots the analogous ratio for the conditional variance, $\text{Var}(r_{T,T+30}|D_T, \phi)$. (The posterior mean of that variance is the first term of the predictive variance in equation (21).) The results are obtained by estimating the predictive system on annual real U.S. stock market returns in 1802-2007. Three predictors are used: the dividend yield, the bond yield, and the term spread.

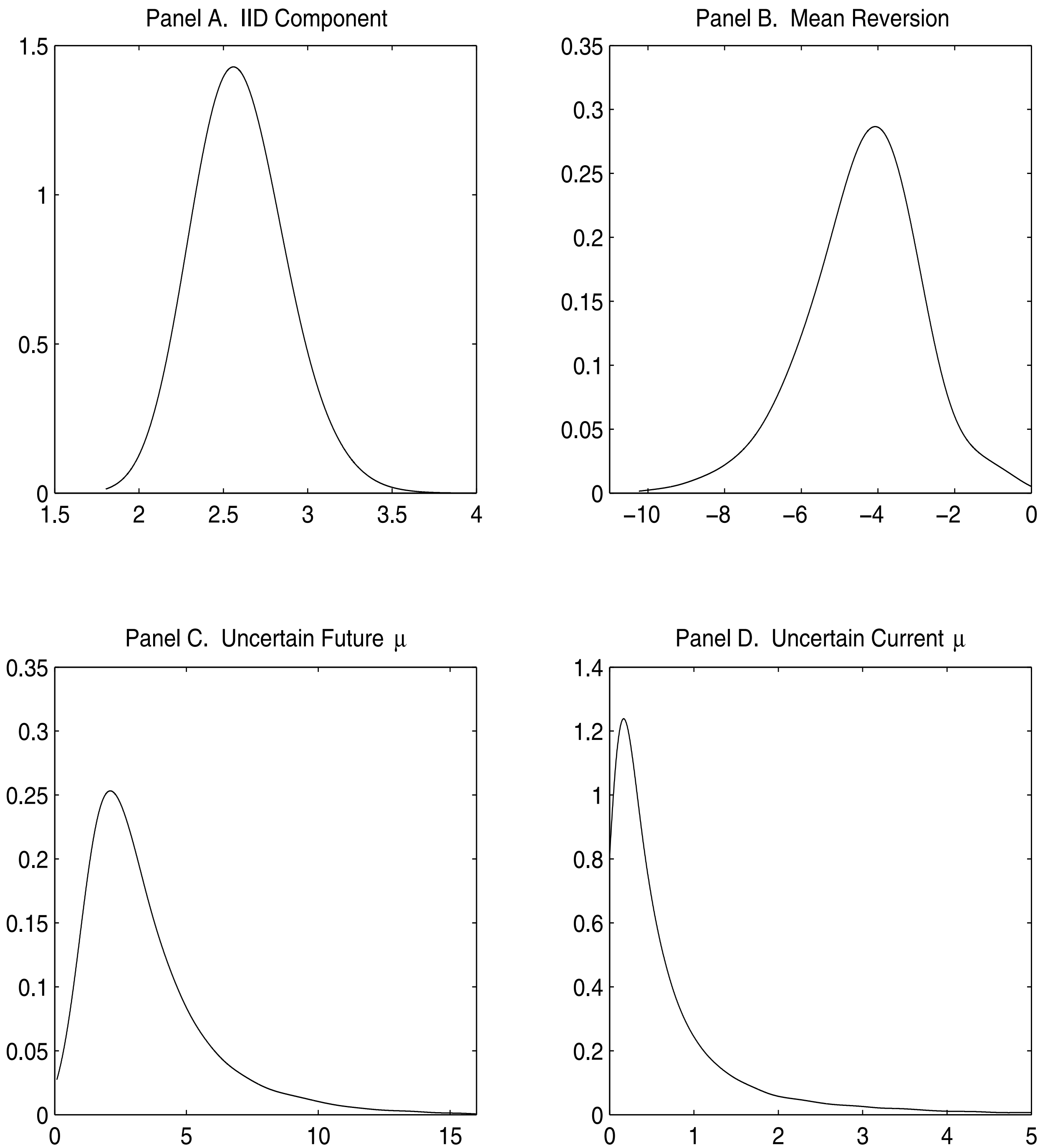

Figure 12. Posterior distributions for the first four components of 30-year predictive variance. The results are obtained by estimating the predictive system on annual real U.S. stock market returns in 1802-2007. Three predictors are used: the dividend yield, the bond yield, and the term spread.

References

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

Benzoni, Luca, Pierre Collin-Dufresne, and Robert S. Goldstein, 2007, Portfolio choice over the life-cycle when the stock and labor markets are cointegrated, *Journal of Finance* 62, 2123–2167.

Campbell, John Y., 1991, A variance decomposition for stock returns, *The Economic Journal* 101, 157–179.

Campbell, John Y., Yeung Lewis Chan and Luis M. Viceira, 2003, A multivariate model of strategic asset allocation, *Journal of Financial Economics* 67, 41–80.

Campbell, John Y., Andrew W. Lo, and A. Craig MacKinlay, 1997, *The Econometrics of Financial Markets* (Princeton University Press, Princeton, NJ).

Campbell, John Y., and Luis M. Viceira, 2002, *Strategic Asset Allocation: Portfolio Choice for Long-Term Investors* (Oxford University Press).

Campbell, John Y., and Luis M. Viceira, 2005, The term structure of the risk-return tradeoff, *Financial Analysts Journal* 61.

Fama, Eugene F., and Kenneth R. French, 1988, Permanent and temporary components of stock prices, *Journal of Political Economy* 96, 246–273.

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

Gordon, Katherine D. and Kimberly A. Stockton, 2006, *Funds for Retirement: The 'Life-Cycle' Approach* (The Vanguard Group, Valley Forge, PA).

Greer, Boyce, 2004, *The case for age-based life-cycle investing* (Fidelity Investments, Boston, MA).

Hoevenaars, Roy, Roderick Molenaar, Peter C. Schotman, and Tom Steenkamp, 2007, Strategic asset allocation for long-term investors: Parameter uncertainty and prior information, Working paper, Maastricht University.

Kim, M., Nelson, C. and Startz, R., 1991, Mean reversion in stock prices? A reappraisal of the empirical evidence, *Review of Economic Studies* 58, 515-28.

Lo, Andrew and MacKinlay, Craig A., 1988, Stock market prices do not follow random walks: Evidence from a simple specification test, *Review of Financial Studies 1*, 41-66.

Lo, Andrew and MacKinlay, Craig A., 1989, The size and power of the variance ratio test in finite samples, *Journal of Econometrics* 40, 203–238.

Pástor, Ľuboš, and Robert F. Stambaugh, 2008, Predictive systems: Living with imperfect predictors, *Journal of Finance*, forthcoming.

Poterba, James M. and Lawrence H. Summers, 1988, Mean reversion in stock returns: Evidence and implications, *Journal of Financial Economics* 22, 27–60.

Richardson, Matthew, 1993, Temporary components of stock prices: a skeptics view, *Journal of Business and Economics Statistics* 11, 199-207.

Richardson, Matthew, and James H. Stock, 1989, Drawing inferences from statistics based on multiyear asset returns, *Journal of Financial Economics* 25, 323–348.

Schotman, Peter C., Rolf Tschernig, and Jan Budek, 2008, Long memory and the term structure of risk, *Journal of Financial Econometrics*, forthcoming.

Siegel, Jeremy J., 1992, The real rate of interest from 1800–1990: A study of the U.S. and the U.K., *Journal of Monetary Economics* 29, 227–252.

Siegel, Jeremy J., 2008, *Stocks for the Long Run*, 4th ed. (McGraw Hill, New York, NY).

Summers, Lawrence H., 1996, Does the stock market rationally reflect fundamental values? *Journal of Finance* 41, 591–601.

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

Viceira, Luis M., 2008, Life-cycle funds, forthcoming in Annamaria Lusardi, ed., *Overcoming the Saving Slump: How to Increase the Effectiveness of Financial Education and Saving Programs* (University of Chicago Press, Chicago, IL).