Internet Appendix to

Are Stocks Really Less Volatile in the Long Run?

by

Ľuboš Pástor

and

Robert F. Stambaugh

January 10, 2011

B1. Roadmap

This Appendix is organized as follows. A simple illustration demonstrating the effects of parameter uncertainty on long-horizon predictive variance is provided in Section B2. The Bayesian empirical methodology for System 2 is presented in Section B3. Additional empirical evidence that complements the evidence presented in the paper is reported in Section B4. That section first presents the estimates of predictive regressions for both annual and quarterly data, followed by various robustness results regarding the long-horizon predictive variance of stock returns.

B2. Parameter uncertainty: A simple illustration

In the paper, we compute $\mathrm{Var}(r_{T,T+k}|D_T)$ and its components empirically, incorporating parameter uncertainty via Bayesian posterior distributions. Here we use a simpler setting to illustrate the effects of parameter uncertainty on multiperiod return variance. All the notation that is not defined here is defined in Section 2 of the paper.

The elements of the parameter vector ϕ are viewed as random, given that they are unknown to an investor. For the purpose of this illustration, we make several assumptions related to ϕ with the objective of simplifying the variance ratio calculations in the presence of parameter uncertainty. Our first assumption is that, given data D_T, the elements of ϕ are distributed independently of each other. We also make two assumptions about the deviations of the conditional mean from the unconditional mean: given D_T, $\mu_T - E_r$ is uncorrelated with E_r, and the quantity $z_T \equiv b_T - E_r$ is fixed across ϕ. Next, let $\rho_{\mu b}$ denote the true unconditional correlation between μ_T and b_T, $\rho_{\mu b} \equiv \mathrm{Corr}(\mu_T, b_T|\phi)$ (this correlation is "true" in that it depends on the true parameter values ϕ, and it is "unconditional" in that it does not condition on D_T). If the observed predictors capture μ_T perfectly, then $\rho_{\mu b} = 1$; otherwise $\rho_{\mu b} < 1$. We make the "homoskedasticity" assumption that q_T is constant across D_T, which implies that

$$q_T = (1 - \rho_{\mu b}^2)\sigma_\mu^2 = (1 - \rho_{\mu b}^2)R^2\sigma_r^2, \tag{B1}$$

where σ_μ^2 and σ_r^2 are the true unconditional variances of μ_t and r_{t+1}, respectively (i.e., $\sigma_\mu^2 \equiv \mathrm{Var}(\mu_t|\phi)$ and $\sigma_r^2 \equiv \mathrm{Var}(r_{t+1}|\phi)$). The parameter vector is $\phi = [\beta, R^2, \rho_{uw}, E_r, \sigma_r, \rho_{\mu b}]$. Finally, we specify τ such that

$$\mathrm{Var}(E_r|D_T) = \tau \mathrm{E}(\sigma_r^2|D_T)\ , \tag{B2}$$

so that the uncertainty about the unconditional mean return E_r is as large as the imprecision in a sample mean computed over a sample of length $1/\tau$.

None of the above simplifying assumptions hold in the Bayesian empirical framework in the paper; they are made for the purpose of this simple illustration only. Given these assumptions, we are able to characterize the variance ratio in a parsimonious fashion that does not depend on the unconditional variance of single-period returns.

The variance of the k-period return can be decomposed as follows:

$$\begin{aligned}
\text{Var}(r_{T,T+k}|D_T) &= \text{E}\left\{\text{Var}(r_{T,T+k}|\phi,\mu_T,D_T)|D_T\right\} + \text{Var}\left\{\text{E}(r_{T,T+k}|\phi,\mu_T,D_T)|D_T\right\} \\
&= \text{E}\left\{\text{Var}(r_{T,T+k}|\phi,\mu_T)|D_T\right\} + \text{Var}\left\{\text{E}(r_{T,T+k}|\phi,\mu_T)|D_T\right\} \\
&= \text{E}\left\{k\sigma_r^2(1-R^2)[1+2\bar{d}\rho_{uw}A(k)+\bar{d}^2B(k)]|D_T\right\} \\
&\quad +\text{Var}\left\{kE_r + \frac{1-\beta^k}{1-\beta}(\mu_T - E_r)|D_T\right\} \\
&= k\text{E}(\sigma_r^2|D_T)\text{E}\left\{(1-R^2)[1+2\bar{d}\rho_{uw}A(k)+\bar{d}^2B(k)]\Big|\,D_T\right\} \\
&\quad +k^2\text{Var}(E_r|D_T) + \text{Var}\left\{\frac{1-\beta^k}{1-\beta}(\mu_T - E_r)|D_T\right\} \\
&= k\text{E}(\sigma_r^2|D_T)\text{E}\left\{(1-R^2)[1+2\bar{d}\rho_{uw}A(k)+\bar{d}^2B(k)]|D_T\right\} \\
&\quad +k^2\tau\text{E}(\sigma_r^2|D_T) + \text{Var}\left\{\frac{1-\beta^k}{1-\beta}(\mu_T - E_r)|D_T\right\}\ .
\end{aligned} \tag{B3}$$

The next to last equality uses the property that E_r is uncorrelated with $\mu_T - E_r$, and the last equality uses (B2). Next observe that

$$\begin{aligned}
\text{Var}\left\{\frac{1-\beta^k}{1-\beta}(\mu_T - E_r)|D_T\right\} &= \text{E}\left\{\text{Var}\left[\frac{1-\beta^k}{1-\beta}(\mu_T - E_r)|\phi, D_T\right]|D_T\right\} \\
&\quad +\text{Var}\left\{\text{E}\left[\frac{1-\beta^k}{1-\beta}(\mu_T - E_r)|\phi, D_T\right]|D_T\right\} \\
&= \text{E}\left\{\left(\frac{1-\beta^k}{1-\beta}\right)^2 \text{Var}(\mu_T|\phi, D_T)|D_T\right\} \\
&\quad +\text{Var}\left\{\frac{1-\beta^k}{1-\beta}\text{E}\left(\mu_T - E_r|\phi, D_T\right)|D_T\right\} \\
&= \text{E}\left\{\left(\frac{1-\beta^k}{1-\beta}\right)^2 q_T|D_T\right\} \\
&\quad +\text{Var}\left\{\frac{1-\beta^k}{1-\beta}(b_T - E_r)|D_T\right\} \\
&= \text{E}\left\{\left(\frac{1-\beta^k}{1-\beta}\right)^2 \sigma_r^2R^2(1-\rho_{\mu b}^2)|D_T\right\} \\
&\quad +\text{Var}\left\{\frac{1-\beta^k}{1-\beta}z_T|D_T\right\}
\end{aligned}$$

$$
\begin{aligned}
&= \mathrm{E}\left\{\left(\frac{1-\beta^k}{1-\beta}\right)^2 |D_T\right\} \mathrm{E}(\sigma_r^2|D_T)\mathrm{E}\left\{R^2(1-\rho_{\mu b}^2)|D_T\right\} \\
&\quad + z_T^2 \mathrm{Var}\left\{\frac{1-\beta^k}{1-\beta}|D_T\right\}. \qquad \text{(B4)}
\end{aligned}
$$

Substituting the right-hand side of (B4) for the last term in (B3) then gives

$$
\begin{aligned}
\mathrm{Var}(r_{T,T+k}|D_T) &= k\mathrm{E}(\sigma_r^2|D_T))\mathrm{E}\left\{(1-R^2)[1+2\bar{d}\rho_{uw}A(k)+\bar{d}^2B(k)]|D_T\right\} \\
&\quad + k^2\tau\mathrm{E}(\sigma_r^2|D_T) + \mathrm{E}(\sigma_r^2|D_T)\mathrm{E}\left\{\left(\frac{1-\beta^k}{1-\beta}\right)^2 |D_T\right\} \mathrm{E}\left\{R^2(1-\rho_{\mu b}^2)\Big| D_T\right\} \\
&\quad + z_T^2 \mathrm{Var}\left\{\frac{1-\beta^k}{1-\beta}|D_T\right\} . \qquad \text{(B5)}
\end{aligned}
$$

When $k = 1$, equation (B5) simplifies to

$$
\mathrm{Var}(r_{T,T+1}|D_T) = \mathrm{E}(\sigma_r^2|D_T)\left[1+\tau-\mathrm{E}(R^2|D_T)\mathrm{E}(\rho_{\mu b}^2|D_T)\right] . \qquad \text{(B6)}
$$

Observe that the k-period variance in (B5) depends on the value of z_T^2, which enters the last term multiplied by the variance of $(1-\beta^k)/(1-\beta)$. This dependence makes sense: when μ_T is estimated to be farther from the unconditional mean E_r, so that the absolute value of z_T is large, uncertainty about the speed with which μ_T reverts to E_r is more important. To achieve a further algebraic simplification, we evaluate the variance in (B5) by setting z_T^2 equal to the posterior mean of the true unconditional mean of z_T^2:

$$
z_T^2 = \mathrm{E}[\mathrm{E}(z_T^2|\phi)|D_T]. \qquad \text{(B7)}
$$

To evaluate the right-hand side of (B7), first note that $\mathrm{Var}(z_T|\phi) = \mathrm{Var}(b_T|\phi)$, since $z_T = b_T - E_r$ and E_r is in ϕ. Also observe that $\mathrm{E}(z_T|\phi) = 0$, since across samples (D_T's), the expected conditional mean b_T is the unconditional mean E_r. Therefore, $\mathrm{E}(z_T^2|\phi) = \mathrm{Var}(z_T|\phi) = \mathrm{Var}(b_T|\phi)$. We thus have

$$
\begin{aligned}
\mathrm{E}[\mathrm{E}(z_T^2|\phi)|D_T] &= \mathrm{E}[\mathrm{Var}(b_T|\phi)|D_T] \\
&= \mathrm{E}[\rho_{\mu b}^2\mathrm{Var}(\mu_T|\phi)|D_T] \\
&= \mathrm{E}[\rho_{\mu b}^2\sigma_r^2R^2|D_T] \\
&= \mathrm{E}(\rho_{\mu b}^2|D_T)\mathrm{E}(\sigma_r^2|D_T)\mathrm{E}(R^2|D_T). \qquad \text{(B8)}
\end{aligned}
$$

We compute the k-period variance ratio,

$$
V(k) = \frac{\overline{\mathrm{Var}(r_{T,T+k}|D_T)}}{k \cdot \mathrm{Var}(r_{T,T+1}|D_T)} , \qquad \text{(B9)}
$$

where $\mathrm{Var}(r_{T,T+1}|D_T)$ is given in equation (B6) and $\overline{\mathrm{Var}(r_{T,T+k}|D_T)}$ denotes the value of (B5) obtained by substituting the right-hand side of (B8) for z_T^2. Note that the variance ratio in equation (B9) does not depend on $\mathrm{E}(\sigma_r^2|D_T)$.

We set $\tau = 1/200$ in equation (B2), so that the uncertainty about the unconditional mean return E_r corresponds to the imprecision in a 200-year sample mean. To specify the uncertainty for the remaining parameters, we choose the probability densities displayed in Figure A1, 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 A1 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 (B2). Successive rows then specify one or more of those parameters as uncertain, by drawing from the densities in Figure A1 (for β, R^2, and ρ_{uw}) or setting $\tau = 1/200$ (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 A1. 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 A1 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 A1, 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 A1 has almost all of its mass below 0, so that mean reversion is almost certainly present, parameter uncertainty causes the long-run variance to exceed the short-run variance.

As discussed in the paper, uncertainty about E_r implies $V(k) \to \infty$ as $k \to \infty$. We can see from Table A1 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 A1 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 A1 produces the same values of $V(20)$ as fixing $\rho_{\mu b}$ at its median.

B3. Predictive System 2

System 2 is implemented in the paper for one asset, but it can be specified for n assets, in which case r_{t+1} and π_{t+1} are $n \times 1$ vectors. We begin the analysis with that more general form and then later restrict it to the single-asset setting. With multiple assets, System 2 is given by

$$\begin{bmatrix} r_{t+1} \\ x_{t+1} \\ \pi_{t+1} \end{bmatrix} = \begin{bmatrix} a \\ \theta \\ 0 \end{bmatrix} + \begin{bmatrix} 0 & A_{12} & I \\ 0 & A_{22} & 0 \\ 0 & 0 & A_{33} \end{bmatrix} \begin{bmatrix} r_t \\ x_t \\ \pi_t \end{bmatrix} + \begin{bmatrix} u_{t+1} \\ v_{t+1} \\ \eta_{t+1} \end{bmatrix}, \tag{B10}$$

which can also be written as

$$\begin{bmatrix} r_{t+1} - E_r \\ x_{t+1} - E_x \\ \pi_{t+1} - E_\pi \end{bmatrix} = \begin{bmatrix} 0 & A_{12} & I \\ 0 & A_{22} & 0 \\ 0 & 0 & A_{33} \end{bmatrix} \begin{bmatrix} r_t - E_r \\ x_t - E_x \\ \pi_t - E_\pi \end{bmatrix} + \begin{bmatrix} u_{t+1} \\ v_{t+1} \\ \eta_{t+1} \end{bmatrix}, \tag{B11}$$

where $E_x = (I - A_{22})^{-1}\theta$, $E_r = a + A_{12}E_x$, and without loss of generality $E_\pi = 0$. We begin working with multiple assets, so that not only x_t but also r_t and π_t are vectors. We assume the errors in (B11) are i.i.d. across $t = 1, \ldots, T$:

$$\begin{bmatrix} u_t \\ v_t \\ \eta_t \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma_{uu} & \Sigma_{uv} & \Sigma_{u\eta} \\ \Sigma_{vu} & \Sigma_{vv} & \Sigma_{v\eta} \\ \Sigma_{\eta u} & \Sigma_{\eta v} & \Sigma_{\eta\eta} \end{bmatrix} \right). \tag{B12}$$

We assume the eigenvalues of both A_{22} and A_{33} lie inside the unit circle. As the elements of $\Sigma_{\eta\eta}$ approach zero, the above model approaches the perfect-predictor specification,

$$\begin{bmatrix} r_{t+1} - E_r \\ x_{t+1} - E_x \end{bmatrix} = \begin{bmatrix} 0 & A_{12} \\ 0 & A_{22} \end{bmatrix} \begin{bmatrix} r_t - E_r \\ x_t - E_x \end{bmatrix} + \begin{bmatrix} u_{t+1} \\ v_{t+1} \end{bmatrix}. \tag{B13}$$

Let $\bar{A}$ denote the entire coefficient matrix in (B11), and let Σ denote the entire covariance matrix in (B12). Define the vector

$$\zeta_t = \begin{bmatrix} r_t \\ x_t \\ \pi_t \end{bmatrix}, \tag{B14}$$

and let $V_{\zeta\zeta}$ denote its unconditional covariance matrix. Then

$$V_{\zeta\zeta} = \begin{bmatrix} V_{rr} & V_{rx} & V_{r\pi} \\ V_{xr} & V_{xx} & V_{x\pi} \\ V_{\pi r} & V_{\pi x} & V_{\pi\pi} \end{bmatrix} = \bar{A} V_{\zeta\zeta} \bar{A}' + \Sigma, \tag{B15}$$

which can be solved as

$$\text{vec}\,(V_{\zeta\zeta}) = [I - (\bar{A} \otimes \bar{A})]^{-1}\text{vec}\,(\Sigma), \tag{B16}$$

using the well known identity vec $(DFG) = (G' \otimes D)$vec (F).

Let z_t denote the vector of the observed data at time t,

$$z_t = \begin{bmatrix} r_t \\ x_t \end{bmatrix}.$$

Denote the data we observe through time t as $D_t = (z_1, \ldots, z_t)$, and note that our complete data consist of D_T. Also define

$$E_z = \begin{bmatrix} E_r \\ E_x \end{bmatrix}, \quad V_{zz} = \begin{bmatrix} V_{rr} & V_{rx} \\ V_{xr} & V_{xx} \end{bmatrix}, \quad V_{z\pi} = \begin{bmatrix} V_{r\pi} \\ V_{x\pi} \end{bmatrix}. \tag{B17}$$

Let ϕ denote the full set of parameters in the model, $\phi \equiv (\bar{A}, \Sigma, E_z)$, and let π denote the full time series of π_t, $t = 1, \ldots, T$. To obtain the joint posterior distribution of ϕ and π, denoted by $p(\phi, \pi|D_T)$, we use an MCMC procedure in which we alternate between drawing π from the conditional posterior $p(\pi|\phi, D_T)$ and drawing ϕ from the conditional posterior $p(\phi|\pi, D_T)$. The procedure for drawing π from $p(\pi|\phi, D_T)$ is described in Section B3.1. The procedure for drawing ϕ from $p(\phi|\pi, D_T) \propto p(\phi)p(D_T, \pi|\phi)$ is described in Section B3.2.

B3.1. Drawing π given the parameters

To draw the time series of the unobservable values of π_t conditional on the current parameter draws, we apply the *forward filtering, backward sampling* (FFBS) approach developed by Carter and Kohn (1994) and Frühwirth-Schnatter (1994). See also West and Harrison (1997, chapter 15).

B3.1.1. Filtering

The first stage follows the standard methodology of Kalman filtering. Define

$$a_t = \text{E}(\pi_t|D_{t-1}) \qquad b_t = \text{E}(\pi_t|D_t) \qquad e_t = \text{E}(z_t|\pi_t, D_{t-1}) \tag{B18}$$

$$f_t = \text{E}(z_t|D_{t-1}) \qquad P_t = \text{Var}(\pi_t|D_{t-1}) \qquad Q_t = \text{Var}(\pi_t|D_t) \tag{B19}$$

$$R_t = \text{Var}(z_t|\pi_t, D_{t-1}) \qquad S_t = \text{Var}(z_t|D_{t-1}) \qquad G_t = \text{Cov}(z_t, \pi_t'|D_{t-1}) \tag{B20}$$

Conditioning on ϕ is assumed throughout this section but suppressed in the notation for convenience. First observe that

$$\pi_1|D_0 \sim N(a_1, P_1), \tag{B21}$$

where D_0 denotes the null information set, so that the unconditional moments of π_0 are given by $a_1 = E_\pi = 0$ and $P_1 = V_{\pi\pi}$. Also,

$$z_1|D_0 \sim N(f_1, S_1), \tag{B22}$$

where $f_1 = E_z$ and $S_1 = V_{zz}$. Note that

$$G_1 = V_{z\pi} \tag{B23}$$

and that

$$z_1|\pi_1, D_0 \sim N(e_1, R_1), \tag{B24}$$

where

$$e_1 = f_1 + G_1 P_1^{-1}(\pi_1 - a_1) \tag{B25}$$

$$R_1 = S_1 - G_1 P_1^{-1} G_1'. \tag{B26}$$

Combining this density with equation (B21) using Bayes rule gives

$$\pi_1|D_1 \sim N(b_1, Q_1), \tag{B27}$$

where

$$b_1 = a_1 + P_1(P_1 + G_1' R_1^{-1} G_1)^{-1} G_1' R_1^{-1}(z_1 - f_1) \tag{B28}$$

$$Q_1 = P_1(P_1 + G_1' R_1^{-1} G_1)^{-1} P_1. \tag{B29}$$

Continuing in this fashion, we find that all conditional densities are normally distributed, and we obtain all the required moments for $t = 2, \ldots, T$:

$$a_t = A_{33} b_{t-1} \tag{B30}$$

$$f_t = \begin{bmatrix} E_r + A_{12}(x_{t-1} - E_x) + b_{t-1} \\ E_x + A_{22}(x_{t-1} - E_x) \end{bmatrix} \tag{B31}$$

$$S_t = \begin{bmatrix} Q_{t-1} & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} \Sigma_{uu} & \Sigma_{uv} \\ \Sigma_{vu} & \Sigma_{vv} \end{bmatrix} \tag{B32}$$

$$G_t = \begin{bmatrix} Q_{t-1} A_{33}' \\ 0 \end{bmatrix} + \begin{bmatrix} \Sigma_{u\eta} \\ \Sigma_{v\eta} \end{bmatrix} \tag{B33}$$

$$P_t = A_{33} Q_{t-1} A_{33}' + \Sigma_{\eta\eta} \tag{B34}$$

$$e_t = f_t + G_t P_t^{-1}(\pi_t - a_t) \tag{B35}$$

$$R_t = S_t - G_t P_t^{-1} G_t' \tag{B36}$$

$$b_t = a_t + P_t(P_t + G_t' R_t^{-1} G_t)^{-1} G_t' R_t^{-1}(z_t - f_t) \tag{B37}$$

$$= a_t + G_t' S_t^{-1}(z_t - f_t) \tag{B38}$$

$$Q_t = P_t(P_t + G_t' R_t^{-1} G_t)^{-1} P_t. \tag{B39}$$

The values of $\{a_t, b_t, Q_t, S_t, G_t, P_t\}$ for $t = 1, \ldots, T$ are retained for the next stage. Equations (B32) through (B34) are derived as

$$\begin{aligned}
\begin{bmatrix} S_t & G_t \\ G_t' & P_t \end{bmatrix} &= \text{Var}(\zeta_t|D_{t-1}) \\
&= \bar{A}\text{Var}(\zeta_{t-1}|D_{t-1})\bar{A}' + \Sigma \\
&= \bar{A}\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & Q_{t-1} \end{bmatrix}\bar{A}' + \Sigma \\
&= \begin{bmatrix} Q_{t-1} & 0 & Q_{t-1}A_{33}' \\ 0 & 0 & 0 \\ A_{33}Q_{t-1} & 0 & A_{33}Q_{t-1}A_{33}' \end{bmatrix} + \begin{bmatrix} \Sigma_{uu} & \Sigma_{uv} & \Sigma_{u\eta} \\ \Sigma_{vu} & \Sigma_{vv} & \Sigma_{v\eta} \\ \Sigma_{\eta u} & \Sigma_{\eta v} & \Sigma_{\eta\eta} \end{bmatrix}.
\end{aligned}$$

B3.1.2. Sampling—drawing π

We wish to draw $(\pi_1, \ldots, \pi_T)$ conditional on D_T (and the parameters, ϕ). The backward-sampling approach relies on the Markov property of the evolution of ζ_t and the resulting identity,

$$p(\zeta_1, \ldots, \zeta_T|D_T) = p(\zeta_T|D_T)p(\zeta_{T-1}|\zeta_T, D_{T-1}) \cdots p(\zeta_1|\zeta_2, D_1). \tag{B40}$$

We first sample π_T from $p(\pi_T|D_T)$, the normal density obtained in the last step of the filtering. Then, for $t = T-1, T-2, \ldots, 1$, we sample π_t from the conditional density $p(\zeta_t|\zeta_{t+1}, D_t)$. (Note that the first two subvectors of ζ_t are already observed and thus need not be sampled.) To obtain that conditional density, first note that

$$\zeta_{t+1}|D_t \sim N\left(\begin{bmatrix} f_{t+1} \\ a_{t+1} \end{bmatrix}, \begin{bmatrix} S_{t+1} & G_{t+1} \\ G_{t+1}' & P_{t+1} \end{bmatrix}\right), \tag{B41}$$

$$\zeta_t|D_t \sim N\left(\begin{bmatrix} r_t \\ x_t \\ b_t \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & Q_t \end{bmatrix}\right), \tag{B42}$$

and

$$\begin{aligned}
\text{Cov}(\zeta_t, \zeta_{t+1}'|D_t) &= \text{Var}(\zeta_t|D_t)\bar{A}' \\
&= \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & Q_t \end{bmatrix}\begin{bmatrix} 0 & 0 & 0 \\ A_{12}' & A_{22}' & 0 \\ I & 0 & A_{33}' \end{bmatrix} \\
&= \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ Q_t & 0 & Q_tA_{33}' \end{bmatrix}.
\end{aligned} \tag{B43}$$

Therefore,

$$\zeta_t|\zeta_{t+1}, D_t \sim N(h_t, H_t), \tag{B44}$$

where

$$\begin{aligned}
h_t &= \mathrm{E}(\zeta_t|D_t) + \left[\mathrm{Cov}(\zeta_t, \zeta'_{t+1}|D_t)\right] \left[\mathrm{Var}(\zeta_{t+1}|D_t)\right]^{-1} \left[\zeta_{t+1} - \mathrm{E}(\zeta_{t+1}|D_t)\right] \\
&= \begin{bmatrix} r_t \\ x_t \\ b_t \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ Q_t & 0 & Q_t A'_{33} \end{bmatrix} \begin{bmatrix} S_{t+1} & G_{t+1} \\ G'_{t+1} & P_{t+1} \end{bmatrix}^{-1} \begin{bmatrix} z_{t+1} - f_{t+1} \\ \pi_{t+1} - a_{t+1} \end{bmatrix}
\end{aligned}$$

and

$$\begin{aligned}
H_t &= \mathrm{Var}(\zeta_t|D_t) - \left[\mathrm{Cov}(\zeta_t, \zeta'_{t+1}|D_t)\right] \left[\mathrm{Var}(\zeta_{t+1}|D_t)\right]^{-1} \left[\mathrm{Cov}(\zeta_t, \zeta'_{t+1}|D_t)\right]' \\
&= \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & Q_t \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ Q_t & 0 & Q_t A'_{33} \end{bmatrix} \begin{bmatrix} S_{t+1} & G_{t+1} \\ G'_{t+1} & P_{t+1} \end{bmatrix}^{-1} \begin{bmatrix} 0 & 0 & Q_t \\ 0 & 0 & 0 \\ 0 & 0 & A_{33} Q_t \end{bmatrix}
\end{aligned}$$

The mean and covariance matrix of π_t are taken as the relevant elements of h_t and H_t.

In the rest of the Appendix, we discuss the special case (implemented in the paper) in which r_t and π_t are scalars. The dimensions of $\bar{A}$ and Σ are then $(K+2) \times (K+2)$, where K is the number of predictors in the vector x_t. In this case, A'_{12} is a $K \times 1$ vector, which we denote as b (distinct from b_t, defined previously), and A_{33} is a scalar that we denote as δ. We also denote A_{22} as simply A (distinct from $\bar{A}$, defined previously).

B3.2. Drawing the parameters given π

B3.2.1. Prior distributions

With r_t and π_t being scalars,

$$\Sigma = \begin{bmatrix} \sigma_u^2 & \sigma_{uv} & \sigma_{u\eta} \\ \sigma_{vu} & \Sigma_{vv} & \sigma_{v\eta} \\ \sigma_{\eta u} & \sigma_{\eta v} & \sigma_\eta^2 \end{bmatrix} = \begin{bmatrix} \Sigma_{\xi\xi} & \sigma_{\xi\eta} \\ \sigma'_{\xi\eta} & \sigma_\eta^2 \end{bmatrix},$$

where $\xi'_t \equiv [u_t \; v'_t]$. We wish to be informative about σ_η^2, to varying degrees, while being noninformative about $\Sigma_{\xi\xi}$ and $\sigma_{\xi\eta}$. This objective is accomplished by specifying a prior for Σ equal to the posterior that obtains when a diffuse prior for Σ is combined with a hypothetical sample of T_0 observations of η_t and S_0 observations of ξ_t, where $S_0 \leq T_0$ and the shorter period is a subset of the longer. This posterior is given by Stambaugh (1997), except that we impose the additional restriction that the population means of η_t and ξ_t are zero. The posterior in Stambaugh (1997) relies on a change of variables, which we adopt here as well. Specifically, $c = (1/\sigma_\eta^2)\sigma_{\xi\eta}$, and $\Omega = \Sigma_{\xi\xi} - \sigma_\eta^2 cc'$. The hypothetical T_0 observations of η_t produce the sample variance $\sigma_{\eta,0}^2$. The S_0 observations of ξ_t and η_t produce the sample covariance matrix $\hat{\Sigma}_0$, giving $\hat{\sigma}_{\eta,0}^2$, $\hat{c}_0$ and $\hat{\Omega}_0$. (All hypothetical second moments are non-central; equivalently the hypothetical sample means are zero.) With the change in variables from Σ to $(\sigma_\eta^2, c, \Omega)$, the latter quantities have the posterior, taken here as the prior, given by

$$p(\sigma_\eta^2, c, \Omega) = p(\sigma_\eta^2)p(c|\Omega)p(\Omega),$$

where[1]

$$\sigma_\eta^2 \sim \frac{T_0\sigma_{\eta,0}^2}{\chi_{T_0-K-1}^2} \tag{B45}$$

$$c|\Omega \sim N\left(\hat{c}_0, \frac{1}{S_0\hat{\sigma}_{\eta,0}^2}\Omega\right) \tag{B46}$$

$$\Omega \sim IW(S_0\hat{\Omega}_0, S_0 - K). \tag{B47}$$

Be specifying different values of $\sigma_{\eta,0}^2$ and T_0, we vary the degree to which the prior imposes a belief that σ_η^2 is close to zero, where the limiting case of $\sigma_\eta^2 = 0$ corresponds to perfect predictors. The larger we make T_0, the greater is the prior information we supply about σ_η^2. We set $S_0 = K+1$, where K is the number of predictors, which makes the prior on c and Ω essentially noninformative (as informative as a sample of only $K+1$ observations, where $K = 1$ or 3). The specification of $\hat{\Sigma}_0$ is thus inconsequential, and we simply set it to a scalar times the identity matrix. For the cases with imperfect predictors, we set $T_0 = K + 2$ and then set $\sigma_{\eta,0}$ equal to either 0.005 or 0.01 when using annual data and either 0.001 or 0.003 when using quarterly data.

We also wish to be informative about δ, the autocorrelation of π_t. Here we specify priors for δ identical to those specified for the autocorrelation (β) of the conditional mean μ_t in the predictive system analyzed previously. Specifically, the prior for δ is a truncated Normal distribution, with the mean and standard deviation of the corresponding non-truncated distribution denoted as $\bar{\delta}$ and

[1] The densities are given as follows:

$$p(\sigma_\eta^2) \propto \sigma_\eta^{-(T_0-K+1)} \exp\left\{-\frac{T_0\sigma_{\eta,0}^2}{2\sigma_\eta^2}\right\} \quad \left(\Rightarrow p(\sigma_\eta) \propto \sigma_\eta^{-(T_0-K)} \exp\left\{-\frac{T_0\sigma_{\eta,0}^2}{2\sigma_\eta^2}\right\}\right)$$

$$p(c|\Omega) \propto |\Omega|^{-\frac{(K+1)}{2}} \exp\left\{-\frac{1}{2}(c-\hat{c}_0)'\left(\frac{1}{S_0\hat{\sigma}_{\eta,0}^2}\Omega\right)^{-1}(c-\hat{c}_0)\right\}$$

$$p(\Omega) \propto |\Omega|^{-\frac{(S_0+2)}{2}} \exp\left\{-\frac{1}{2}\mathrm{tr}\,(S_0\hat{\Omega}_0)\Omega^{-1}\right\}.$$

σ_δ. The distribution is then truncated to satisfy the stationarity requirement $|\delta| < 1$. We set σ_δ to 0.25 with annual data and to 0.15 with quarterly data, and we set $\bar{\delta}$ to 0.99 in both cases.

The priors on the remaining parameters are non-informative, except for the condition required for stationary of x_t that $\rho(A)$, the spectral radius of A, be less than 1. Specifically, define

$$\psi = \begin{bmatrix} \text{vec}\begin{bmatrix} a & \theta' \\ b & A' \end{bmatrix} \\ \delta \end{bmatrix}. \tag{B48}$$

Then the prior for ψ is given by

$$p(\psi) \propto \exp\left\{-\frac{1}{2}(\psi - \bar{\psi})' V_\psi^{-1}(\psi - \bar{\psi})\right\} \times 1_S, \tag{B49}$$

where

$$\bar{\psi} = \begin{bmatrix} 0 \\ \bar{\delta} \end{bmatrix}, \quad V_\psi = \begin{bmatrix} \sigma_\psi^2 I_{(K+1)^2} & 0 \\ 0 & \sigma_\delta^2 \end{bmatrix}, \tag{B50}$$

σ_ψ^2 is set to a large number, 1_S equals 1 under the conditions that $\rho(A) < 1$ and $|\delta| < 1$, and 1_S equals 0 otherwise. We also assume that the priors for Σ and ψ are independent.

B3.2.2. Posterior distributions

Define

$$y_{t+1} = \begin{bmatrix} r_{t+1} - \pi_t \\ x_{t+1} \\ \pi_{t+1} \end{bmatrix}, Y_t = \begin{bmatrix} I_{K+1} \otimes [1\ x_t'] & 0 \\ 0 & \pi_t \end{bmatrix}, \upsilon_{t+1} = \begin{bmatrix} u_{t+1} \\ v_{t+1} \\ \eta_{t+1} \end{bmatrix},$$

$$y = \begin{bmatrix} y_2 \\ \vdots \\ y_T \end{bmatrix}, Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_{T-1} \end{bmatrix}, \upsilon = \begin{bmatrix} \upsilon_2 \\ \vdots \\ \upsilon_T \end{bmatrix}.$$

The sample representation of the predictive system in (B10) is then

$$y = Y\psi + \upsilon,$$

for ψ defined in (B48). For tractability we employ the "conditional" likelihood, which treats values in period $t = 1$ as non-stochastic. We can then apply Gibbs sampling to draw ψ and Σ. This simplification seems reasonable, given that $T = 206$ for our main results.[2]

[2] An alternative would be to employ the "exact" likelihood that includes the density of the values at $t = 1$ and then use the Metropolis-Hastings algorithm, with the conditional posteriors given here used as proposal densities.

B3.2.3. Drawing ψ

Given the prior for ψ in (B49), we can apply standard results from the multivariate regression model (e.g., Zellner, 1971) to obtain the full conditional posterior for ψ as

$$\psi|\cdot \sim N(\tilde{\psi}, \tilde{V}_\psi) \times 1_S \tag{B51}$$

where

$$\tilde{\psi} = \tilde{V}_\psi \left[V_\psi^{-1}\bar{\psi} + Y'(I_{T-1} \otimes \Sigma^{-1})y \right]$$

and

$$\tilde{V}_\psi = \left[V_\psi^{-1} + Y'(I_{T-1} \otimes \Sigma^{-1})Y \right]^{-1}$$

.

B3.2.4. Drawing Σ

Define

$$\hat{\Sigma} = \frac{1}{T-1}\sum_{t=1}^{T-1}(y_{t+1} - Y_t\psi)(y_{t+1} - Y_t\psi)' = \begin{bmatrix} \hat{\Sigma}_{\xi\xi} & \hat{\sigma}_{\xi\eta} \\ \hat{\sigma}'_{\xi\eta} & \hat{\sigma}^2_\eta \end{bmatrix}$$

$$\bar{\Sigma} = \begin{bmatrix} \bar{\Sigma}_{\xi\xi} & \bar{\sigma}_{\xi\eta} \\ \bar{\sigma}'_{\xi\eta} & \bar{\sigma}^2_\eta \end{bmatrix} = \frac{1}{\bar{S}}\left(S_0\hat{\Sigma}_0 + (T-1)\hat{\Sigma}\right), \bar{S} = S_0 + (T-1)$$

$$\bar{c} = (1/\bar{\sigma}^2_\eta)\bar{\sigma}_{\xi\eta}, \bar{\Omega} = \bar{\Sigma}_{\xi\xi} - \bar{\sigma}^2_\eta\bar{c}\bar{c}',$$

$$\tilde{\sigma}^2_\eta = \frac{1}{\tilde{T}}\left(T_0\sigma^2_{\eta,0} + (T-1)\hat{\sigma}^2_\eta\right), \tilde{T} = T_0 + (T-1).$$

A draw of Σ is constructed by drawing $(\sigma^2_\eta, c, \Omega)$. The joint full conditional of the latter is given by

$$p(\sigma^2_\eta, c, \Omega|\cdot) = p(\sigma^2_\eta|\cdot)p(c|\Omega,\cdot)p(\Omega|\cdot),$$

where the posteriors on the right-hand side are of the same form as (B45) through (B47):

$$\sigma^2_\eta|\cdot \sim \frac{\tilde{T}\tilde{\sigma}^2_\eta}{\chi^2_{\tilde{T}-K-1}} \tag{B52}$$

$$c|\Omega,\cdot \sim N\left(\bar{c}, \frac{1}{\bar{S}\bar{\sigma}^2_\eta}\Omega\right) \tag{B53}$$

$$\Omega|\cdot \sim IW(\bar{S}\bar{\Omega}, \bar{S} - K). \tag{B54}$$

The draw of Σ is then obtained by computing $\Sigma_{\xi\xi} = \Omega + \sigma^2_\eta cc'$ and $\sigma_{\xi\eta} = \sigma^2_\eta c$.

Our results based on 25,000 draws from the posterior distribution. First, we generate a sequence of 76,000 draws. We discard the first 1,000 draws as a "burn-in" and take every third draw from the rest to obtain a series of 25,000 draws that exhibit little serial correlation.

B3.3. Predictive variance

In addition to the notation from equations (B11), (B12), and (B14), define also

$$E_\zeta = \begin{bmatrix} E_r \\ E_x \\ 0 \end{bmatrix} \quad , \quad \epsilon_t = \begin{bmatrix} u_t \\ v_t \\ \eta_t \end{bmatrix}, \tag{B55}$$

and

$$\zeta_t^e = \zeta_t - E_\zeta. \tag{B56}$$

Equation (B11) can then be written as

$$\zeta_{t+1}^e = \bar{A}\zeta_t^e + \epsilon_{t+1}. \tag{B57}$$

For $i > 1$, successive substitution using (B57) gives

$$\zeta_{t+i}^e = \bar{A}^i\zeta_t^e + \bar{A}^{i-1}\epsilon_{t+1} + \bar{A}^{i-2}\epsilon_{t+2} + \cdots + \epsilon_{t+i}. \tag{B58}$$

Define

$$\begin{aligned} \zeta_{T,T+k} &= \sum_{i=1}^{k} \zeta_{T+i}, \\ \zeta_{T,T+k}^e &= \sum_{i=1}^{k} \zeta_{T+i}^e = \zeta_{T,T+k} - kE_\zeta. \end{aligned}$$

Summing (B57) over k periods then gives

$$\begin{aligned} \zeta_{T,T+k}^e &= \left(\sum_{i=1}^{k} \bar{A}^i\right)\zeta_t^e + \left(I + \bar{A} + \cdots + \bar{A}^{k-1}\right)\epsilon_{t+1} \\ &\quad + \left(I + \bar{A} + \cdots + \bar{A}^{k-2}\right)\epsilon_{t+2} + \cdots + \epsilon_{t+k} \\ &= (\bar{\Lambda}_{k+1} - I)\zeta_t^e + \bar{\Lambda}_k\epsilon_{t+1} + \bar{\Lambda}_{k-1}\epsilon_{t+2} + \cdots + \epsilon_{t+k}, \end{aligned} \tag{B59}$$

where

$$\begin{aligned} \bar{\Lambda}_i &= I + \bar{A} + \cdots + \bar{A}^{i-1} \\ &= (I - \bar{A})^{-1}(I - \bar{A}^i). \end{aligned} \tag{B60}$$

It then follows that

$$\begin{aligned} \mathrm{E}\left(\zeta_{T,T+k}^e | D_T, \phi, \pi_T\right) &= (\bar{\Lambda}_{k+1} - I)\zeta_T^e, \\ \mathrm{E}\left(\zeta_{T,T+k} | D_T, \phi, \pi_T\right) &= (\bar{\Lambda}_{k+1} - I)\zeta_T^e + kE_\zeta, \end{aligned} \tag{B61}$$

and

$$
\begin{aligned}
\text{Var}\left(\zeta_{T,T+k}|D_T,\phi,\pi_T\right) &= \text{Var}\left(\zeta^e_{T,T+k}|D_T,\phi,\pi_T\right) \\
&= \sum_{i=1}^{k}\bar{\Lambda}_i\Sigma\bar{\Lambda}_i'.
\end{aligned} \tag{B62}
$$

The first and second moments of (B61) given D_T and ϕ are given by

$$
\text{E}\left(\zeta_{T,T+k}|D_T,\phi\right) = \left(\bar{\Lambda}_{k+1}-I\right)\begin{bmatrix} r_T - E_r \\ x_T - E_x \\ b_T \end{bmatrix} + kE_\zeta \tag{B63}
$$

and

$$
\text{Var}\left[\text{E}\left(\zeta_{T,T+k}|D_T,\phi,\pi_T\right)|D_T,\phi\right] = \left(\bar{\Lambda}_{k+1}-I\right)\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & Q_T \end{bmatrix}\left(\bar{\Lambda}_{k+1}-I\right)'. \tag{B64}
$$

Combining (B62) and (B64) gives

$$
\begin{aligned}
\text{Var}\left(\zeta_{T,T+k}|D_T,\phi\right) &= \text{E}\left[\text{Var}\left(\zeta_{T,T+k}|D_T,\phi,\pi_T\right)|D_T,\phi\right] + \text{Var}\left[\text{E}\left(\zeta_{T,T+k}|D_T,\phi,\pi_T\right)|D_T,\phi\right] \\
&= \sum_{i=1}^{k}\bar{\Lambda}_i\Sigma\bar{\Lambda}_i' + \left(\bar{\Lambda}_{k+1}-I\right)\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & Q_T \end{bmatrix}\left(\bar{\Lambda}_{k+1}-I\right)'.
\end{aligned} \tag{B65}
$$

By evaluating (B63) and (B65) for repeated draws of ϕ from its posterior, the predictive variance of $\zeta_{T,T+k}$ can be computed using the decomposition,

$$
\text{Var}(\zeta_{T,T+k}|D_T) = \text{E}\left\{\text{Var}(\zeta_{T,T+k}|\phi,D_T)|D_T\right\} + \text{Var}\left\{\text{E}(\zeta_{T,T+k}|\phi,D_T)|D_T\right\}. \tag{B66}
$$

Finally, the predictive variance of $r_{T,T+k}$ is the (1,1) element of $\text{Var}(\zeta_{T,T+k}|D_T)$.

B3.4. Perfect predictors

If the predictors are perfect, then π_t is absent from the first equation in (B10) and the third equation simply drops out. In that case the model consists of the two equations,

$$
r_{t+1} = a + b'x_t + u_{t+1} \tag{B67}
$$

$$
x_{t+1} = \theta + Ax_t + v_{t+1}, \tag{B68}
$$

combined with the distributional assumption on the residuals,

$$
\begin{bmatrix} u_t \\ v_t \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{vu} & \Sigma_{vv} \end{bmatrix}\right). \tag{B69}
$$

This perfect-predictor model obtains as the limiting case of the above imperfect-predictor setting as $\sigma_\eta^2 \to 0$. Our perfect-predictor results can be computed in that manner, by setting T_0 to a large number and $\sigma_{\eta,0}$ to a small number. For completeness, however, we also provide here the simplified calculations that arise in that special case.

In this case ϕ denotes the full set of parameters in equations (B75), (B68), and (B69), and Σ denotes the covariance matrix in (B69). Let B denote the matrix of coefficients in (B75) and (B68),

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

and let $\psi = \text{vec}\,(B)$.

B3.4.1. Posterior distributions under perfect predictors

We specify the prior distribution on ϕ as $p(\phi) = p(\psi)p(\Sigma)$. The prior on ψ is $p(\psi) \propto 1_S$, where 1_S equals 1 if $\rho(A) < 1$ and equals 0 otherwise. The prior on Σ is $p(\Sigma) \propto |\Sigma|^{-(K+2)/2}$. Define the following notation: $r = [r_2\, r_3\, \cdots\, r_T]'$, $Q^+ = [x_2\, x_3\, \cdots\, x_T]'$, $Q = [x_1\, x_2\, \cdots\, x_{T-1}]'$, $X = [\iota_{T-1}\, Q]$, where ι_{T-1} denotes a $(T-1) \times 1$ vector of ones, $Y = [r\, Q^+]$, $\hat{B} = (X'X)^{-1}X'Y$, and $S = (Y - X\hat{B})'(Y - X\hat{B})$. We first draw Σ^{-1} from a Wishart distribution with $T - K - 2$ degrees of freedom and parameter matrix S^{-1}. Given that draw of Σ^{-1}, we then draw ψ from a normal distribution with mean $\hat{\psi} = \text{vec}\,(\hat{B})$ and covariance matrix $\Sigma \otimes (X'X)^{-1}$. That draw of ϕ is retained as a draw from $p(\phi|D_T)$ if $\rho(A) < 1$.

B3.4.2. Predictive variance under perfect predictors

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

$$\begin{aligned} \text{E}(r_{T,T+k}|D_T, \phi) &= ka + b'\Psi_{k-1}\theta + b'\Lambda_k x_T && \text{(B70)} \\ \text{Var}(r_{T,T+k}|D_T, \phi) &= k\sigma_u^2 + 2b'\Psi_{k-1}\sigma_{vu} + b'\left(\sum_{i=1}^{k-1} \Lambda_i \Sigma_{vv} \Lambda_i'\right) b, && \text{(B71)} \end{aligned}$$

where

$$\begin{aligned} \Lambda_i &= I + A + \cdots + A^{i-1} = (I - A)^{-1}(I - A^i) && \text{(B72)} \\ \Psi_{k-1} &= \Lambda_1 + \Lambda_2 + \cdots + \Lambda_{k-1} = (I - A)^{-1}[kI - (I - A)^{-1}(I - A^k)]. && \text{(B73)} \end{aligned}$$

The first term in (B71) 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

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

Applying the standard variance decomposition

$$\mathrm{Var}(r_{T,T+k}|D_T) = \mathrm{E}\{\mathrm{Var}(r_{T,T+k}|D_T,\phi)|D_T\} + \mathrm{Var}\{\mathrm{E}(r_{T,T+k}|D_T,\phi)|D_T\}, \tag{B74}$$

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 (B71) and the posterior variance of the right-hand side of equation (B70). These posterior moments are computed from the posterior draws of ϕ, which are described in Section B3.4.1.

B4. Additional empirical results

B4.1. Predictive regressions

This section reports the results from standard predictive regressions,

$$r_{t+1} = a + b'x_t + e_{t+1}\ , \tag{B75}$$

for various combinations of the three predictors used in the paper. The results, obtained by OLS, are reported in Table A2. Panel A reports the results based on annual data; the results based on quarterly data are reported in Panel B. In both panels, the first three regressions contain just one predictor, while the fourth regression contains all three. The table reports the estimated coefficients as well as the t-statistics, along with the bootstrapped p-values associated with these t-statistics as well as with the R^2.[3]

[3] 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 (B68) and (B75); (ii) Build up the time series of x_t, starting from the unconditional mean and iterating forward on equation (B68), 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 (B68) and (B75) 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.

Table A2 shows that all the predictors exhibit significant ability to predict returns, especially in the multivariate regressions that involve all three predictors. In those regressions, the estimated correlation between e_{t+1} and the estimated innovation in expected return, $b'v_{t+1}$, is negative. Pástor and Stambaugh (2009) 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.

B4.2. Long-horizon predictive variance

This section provides detailed robustness results on the long-horizon predictive variance, expanding the evidence reported in the paper for both predictive systems. We examine the following cases:

Annual data: Baseline case reported in the paper
- System 1: Table A3; Figures A2, A3
- System 2: Figure A18, left column

Annual data: First subperiod
- System 1: Table A4; Figures A4, A5
- System 2: Figure A19, left column

Annual data: Second subperiod
- System 1: Table A5; Figures A6, A7
- System 2: Figure A19, right column

Annual data: One instead of three predictors
- System 1: Table A6; Figures A8, A9
- System 2: Figure A20, left column

Annual data: Excess instead of real returns
- System 1: Table A7; Figures A10, A11
- System 2: Figure A21, left column

Quarterly data: Baseline case reported in the paper
- System 1: Table A8; Figures A12, A13
- System 2: Figure A18, right column

Quarterly data: One instead of three predictors
- System 1: Table A9; Figures A14, A15
- System 2: Figure A20, right column

Quarterly data: Excess instead of real returns
- System 1: Table A10; Figures A16, A17

System 2: Figure A21, right column

We do not report subperiod results for quarterly data because the (post-war) quarterly sample is already rather short compared to the 206-year annual sample. Parameter uncertainty generally plays a larger role in shorter samples.

Table A1
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 A1 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 A2
Predictive Regressions

This table summarizes the results from predictive regressions $r_t = a + b'x_{t-1} + e_t$, where r_t denotes 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 "[ ]". Each p-value is the relative frequency with which the reported quantity is smaller than its 20,000 counterparts bootstrapped under the null of no predictability. See footnote 3 of this document for more details on the bootstrapping procedure.

Panel A. Annual data (1802–2007)				
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]	

Panel B. Quarterly data (1952Q1–2006Q4)				
Bond Yield	Dividend Yield	CAY	Corr$(e_t, b'v_t)$	R^2
0.018			22.801	4.733
(3.299)			(3.466)	[0.002]
[0.001]			[0.000]	
	0.011		-91.573	1.708
	(1.951)		(-33.729)	[0.095]
	[0.091]		[1.000]	
		0.021	-51.474	6.390
		(3.866)	(-8.885)	[0.000]
		[0.000]	[1.000]	
0.017	0.010	0.017	-27.520	11.121
(3.145)	(1.876)	(3.011)	(-4.236)	[0.000]
[0.002]	[0.067]	[0.004]	[1.000]	

Table A3 (same as Table 1 in the paper)
Variance Ratios and Components of Long-Horizon Variance

The first row of each panel reports the ratio $(1/k)\text{Var}(r_{T,T+k}|D_T)/\text{Var}(r_{T+1}|D_T)$, where $\text{Var}(r_{T,T+k}|D_T)$ 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 $\text{Var}(r_{T,T+k}|D_T)$, multiplied by 100. The remaining rows report the five components of $\text{Var}(r_{T,T+k}|D_T)$, also multiplied by 100 (they add up to total variance). Panel A contains results for $k = 25$ years, and Panel B contains results for $k = 50$ 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 $k = 25$ years					
Variance Ratio	1.30	1.36	1.26	1.31	1.36	1.15	1.42	1.36	1.34
Predictive Variance	3.82	3.99	3.68	3.92	3.99	3.28	4.17	3.99	3.93
IID Component	2.59	2.60	2.59	2.75	2.60	2.43	2.58	2.60	2.60
Mean Reversion	-4.13	-4.01	-4.10	-3.04	-4.01	-4.51	-4.28	-4.01	-3.97
Uncertain Future μ	2.91	2.86	2.84	1.70	2.86	3.51	3.14	2.86	2.79
Uncertain Current μ	0.97	0.96	0.94	0.75	0.96	0.92	1.17	0.96	0.93
Estimation Risk	1.48	1.58	1.41	1.75	1.58	0.93	1.56	1.58	1.57
				Panel B. Investment Horizon $k = 50$ years					
Variance Ratio	1.76	1.82	1.64	1.72	1.82	1.45	1.96	1.82	1.79
Predictive Variance	5.14	5.34	4.79	5.14	5.34	4.13	5.75	5.34	5.27
IID Component	2.59	2.60	2.59	2.75	2.60	2.43	2.58	2.60	2.60
Mean Reversion	-5.52	-5.36	-5.42	-4.32	-5.36	-5.61	-5.80	-5.36	-5.28
Uncertain Future μ	5.40	5.31	5.13	3.60	5.31	5.54	5.97	5.31	5.16
Uncertain Current μ	0.95	0.94	0.91	0.90	0.94	0.73	1.16	0.94	0.92
Estimation Risk	1.72	1.85	1.59	2.21	1.85	1.03	1.85	1.85	1.87

Table A4
Variance Ratios and Components of Long-Horizon Variance
First subperiod (1802–1904)

Counterpart of Table A3

	ρ_{uw}			R^2			β		
Prior	Tight	Bench	Loose	Tight	Bench	Loose	Tight	Bench	Loose
Panel A. Investment Horizon $k = 25$ years									
Variance Ratio	1.28	1.35	1.38	1.46	1.35	1.22	1.27	1.35	1.55
Predictive Variance	2.86	3.00	3.05	3.30	3.00	2.65	2.80	3.00	3.48
IID Component	2.01	2.01	2.02	2.11	2.01	1.92	1.96	2.01	2.04
Mean Reversion	-1.99	-1.71	-1.44	-1.32	-1.71	-1.92	-2.37	-1.71	-1.39
Uncertain Future μ	1.17	1.13	1.09	0.66	1.13	1.49	1.69	1.13	0.92
Uncertain Current μ	0.36	0.34	0.28	0.30	0.34	0.28	0.57	0.34	0.28
Estimation Risk	1.31	1.23	1.09	1.56	1.23	0.87	0.95	1.23	1.64
Panel B. Investment Horizon $k = 50$ years									
Variance Ratio	1.70	1.78	1.79	2.00	1.78	1.55	1.66	1.78	2.16
Predictive Variance	3.80	3.96	3.97	4.54	3.96	3.35	3.64	3.96	4.86
IID Component	2.01	2.01	2.02	2.11	2.01	1.92	1.96	2.01	2.04
Mean Reversion	-2.49	-2.15	-1.81	-1.76	-2.15	-2.31	-3.06	-2.15	-1.74
Uncertain Future μ	2.04	2.00	1.91	1.35	2.00	2.34	3.02	2.00	1.64
Uncertain Current μ	0.40	0.37	0.31	0.39	0.37	0.25	0.59	0.37	0.32
Estimation Risk	1.84	1.73	1.54	2.44	1.73	1.16	1.13	1.73	2.60

Table A5
Variance Ratios and Components of Long-Horizon Variance
Second subperiod (1905–2007)

Counterpart of Table A3

	ρ_{uw}			R^2			β		
Prior	Tight	Bench	Loose	Tight	Bench	Loose	Tight	Bench	Loose
				Panel A. Investment Horizon $k = 25$ years					
Variance Ratio	1.07	1.08	1.11	0.98	1.08	1.29	1.23	1.08	1.06
Predictive Variance	3.82	3.89	3.99	3.60	3.89	4.52	4.40	3.89	3.80
IID Component	3.28	3.29	3.29	3.48	3.29	3.10	3.24	3.29	3.30
Mean Reversion	-4.60	-4.50	-4.37	-3.19	-4.50	-5.33	-4.98	-4.50	-4.29
Uncertain Future μ	3.02	3.03	3.00	1.56	3.03	4.46	3.69	3.03	2.86
Uncertain Current μ	0.86	0.83	0.83	0.58	0.83	1.09	1.16	0.83	0.76
Estimation Risk	1.26	1.25	1.24	1.17	1.25	1.20	1.29	1.25	1.18
				Panel B. Investment Horizon $k = 50$ years					
Variance Ratio	1.45	1.47	1.51	1.25	1.47	1.86	1.78	1.47	1.42
Predictive Variance	5.19	5.29	5.42	4.57	5.29	6.51	6.37	5.29	5.09
IID Component	3.28	3.29	3.29	3.48	3.29	3.10	3.24	3.29	3.30
Mean Reversion	-5.98	-5.86	-5.71	-4.31	-5.86	-6.81	-6.71	-5.86	-5.55
Uncertain Future μ	5.48	5.52	5.50	3.13	5.52	7.78	7.06	5.52	5.15
Uncertain Current μ	0.93	0.90	0.87	0.72	0.90	1.04	1.23	0.90	0.79
Estimation Risk	1.48	1.45	1.46	1.55	1.45	1.39	1.54	1.45	1.40

Table A6
Variance Ratios and Components of Long-Horizon Variance
One instead of three predictors (Dividend yield)

Counterpart of Table A3

	ρ_{uw}			R^2			β		
Prior	Tight	Bench	Loose	Tight	Bench	Loose	Tight	Bench	Loose
Panel A. Investment Horizon $k = 25$ years									
Variance Ratio	1.08	1.05	1.05	1.05	1.05	0.90	1.18	1.05	1.09
Predictive Variance	3.14	3.06	3.06	3.11	3.06	2.55	3.44	3.06	3.19
IID Component	2.63	2.63	2.63	2.77	2.63	2.48	2.61	2.63	2.63
Mean Reversion	-3.61	-3.49	-3.47	-2.63	-3.49	-3.79	-3.83	-3.49	-3.47
Uncertain Future μ	2.33	2.25	2.24	1.35	2.25	2.68	2.68	2.25	2.25
Uncertain Current μ	0.74	0.70	0.69	0.50	0.70	0.64	0.92	0.70	0.67
Estimation Risk	1.05	0.97	0.97	1.12	0.97	0.55	1.06	0.97	1.11
Panel B. Investment Horizon $k = 50$ years									
Variance Ratio	1.30	1.26	1.25	1.25	1.26	1.02	1.49	1.26	1.32
Predictive Variance	3.81	3.68	3.64	3.69	3.68	2.90	4.35	3.68	3.86
IID Component	2.63	2.63	2.63	2.77	2.63	2.48	2.61	2.63	2.63
Mean Reversion	-4.62	-4.44	-4.42	-3.52	-4.44	-4.55	-5.02	-4.44	-4.43
Uncertain Future μ	3.94	3.75	3.72	2.52	3.75	3.86	4.69	3.75	3.79
Uncertain Current μ	0.66	0.63	0.61	0.53	0.63	0.47	0.83	0.63	0.61
Estimation Risk	1.21	1.10	1.10	1.40	1.10	0.63	1.24	1.10	1.27

Table A7
Variance Ratios and Components of Long-Horizon Variance
Excess instead of real returns

Counterpart of Table A3

	ρ_{uw}			R^2			β		
Prior	Tight	Bench	Loose	Tight	Bench	Loose	Tight	Bench	Loose
Panel A. Investment Horizon $k = 25$ years									
Variance Ratio	1.13	1.17	1.26	1.29	1.17	1.15	1.19	1.17	1.19
Predictive Variance	2.94	3.04	3.24	3.48	3.04	2.87	3.07	3.04	3.08
IID Component	2.41	2.43	2.45	2.58	2.43	2.29	2.39	2.43	2.44
Mean Reversion	-2.44	-1.78	-1.12	-1.09	-1.78	-2.30	-2.44	-1.78	-1.49
Uncertain Future μ	1.73	1.44	1.20	0.74	1.44	2.07	2.01	1.44	1.25
Uncertain Current μ	0.32	0.22	0.17	0.15	0.22	0.32	0.42	0.22	0.19
Estimation Risk	0.90	0.73	0.53	1.11	0.73	0.49	0.69	0.73	0.68
Panel B. Investment Horizon $k = 50$ years									
Variance Ratio	1.36	1.39	1.44	1.60	1.39	1.35	1.43	1.39	1.40
Predictive Variance	3.56	3.59	3.70	4.32	3.59	3.37	3.72	3.59	3.62
IID Component	2.41	2.43	2.45	2.58	2.43	2.29	2.39	2.43	2.44
Mean Reversion	-3.07	-2.21	-1.41	-1.40	-2.21	-2.78	-3.13	-2.21	-1.84
Uncertain Future μ	2.82	2.25	1.80	1.25	2.25	3.02	3.31	2.25	1.91
Uncertain Current μ	0.28	0.18	0.15	0.16	0.18	0.24	0.36	0.18	0.16
Estimation Risk	1.11	0.94	0.72	1.73	0.94	0.61	0.79	0.94	0.95

Table A8
Variance Ratios and Components of Long-Horizon Variance
Quarterly data (1952Q1–2006Q4)

Counterpart of Table A3

	ρ_{uw}			R^2			β		
Prior	Tight	Bench	Loose	Tight	Bench	Loose	Tight	Bench	Loose
Panel A. Investment Horizon $k = 25$ years									
Variance Ratio	1.73	1.92	1.95	1.89	1.92	2.05	2.17	1.92	1.93
Predictive Variance	1.13	1.25	1.27	1.26	1.25	1.31	1.41	1.25	1.26
IID Component	0.64	0.64	0.64	0.65	0.64	0.62	0.64	0.64	0.64
Mean Reversion	-1.57	-1.14	-1.11	-0.88	-1.14	-1.29	-1.26	-1.14	-1.12
Uncertain Future μ	1.59	1.33	1.33	1.06	1.33	1.54	1.58	1.33	1.33
Uncertain Current μ	0.08	0.05	0.04	0.05	0.05	0.05	0.06	0.05	0.04
Estimation Risk	0.39	0.38	0.38	0.38	0.38	0.39	0.39	0.38	0.38
Panel B. Investment Horizon $k = 50$ years									
Variance Ratio	3.12	2.88	2.94	2.77	2.88	3.10	3.38	2.88	2.93
Predictive Variance	2.05	1.88	1.92	1.85	1.88	1.99	2.21	1.88	1.91
IID Component	0.64	0.64	0.64	0.65	0.64	0.62	0.64	0.64	0.64
Mean Reversion	-1.84	-1.27	-1.24	-1.01	-1.27	-1.42	-1.43	-1.27	-1.25
Uncertain Future μ	2.34	1.67	1.68	1.40	1.67	1.91	2.10	1.67	1.69
Uncertain Current μ	0.07	0.04	0.03	0.04	0.04	0.04	0.05	0.04	0.03
Estimation Risk	0.83	0.80	0.81	0.76	0.80	0.84	0.85	0.80	0.80

Table A9
Variance Ratios and Components of Long-Horizon Variance
Quarterly data (1952Q1–2006Q4)
One predictor instead of three (dividend yield)

Counterpart of Table A3

	ρ_{uw}			R^2			β		
Prior	Tight	Bench	Loose	Tight	Bench	Loose	Tight	Bench	Loose
Panel A. Investment Horizon $k = 25$ years									
Variance Ratio	3.54	2.20	2.30	1.97	2.20	2.24	2.67	2.20	2.15
Predictive Variance	2.36	1.50	1.57	1.36	1.50	1.51	1.82	1.50	1.47
IID Component	0.64	0.66	0.67	0.67	0.66	0.65	0.66	0.66	0.66
Mean Reversion	-2.04	-0.76	-0.55	-0.79	-0.76	-0.73	-0.90	-0.76	-0.75
Uncertain Future μ	2.98	1.16	1.06	1.01	1.16	1.18	1.54	1.16	1.13
Uncertain Current μ	0.40	0.12	0.11	0.15	0.12	0.10	0.18	0.12	0.11
Estimation Risk	0.38	0.31	0.29	0.31	0.31	0.31	0.33	0.31	0.31
Panel B. Investment Horizon $k = 50$ years									
Variance Ratio	8.07	3.48	3.49	3.33	3.48	3.36	4.46	3.48	3.37
Predictive Variance	5.38	2.37	2.39	2.29	2.37	2.27	3.04	2.37	2.30
IID Component	0.64	0.66	0.67	0.67	0.66	0.65	0.66	0.66	0.66
Mean Reversion	-2.76	-0.90	-0.65	-0.99	-0.90	-0.83	-1.09	-0.90	-0.88
Uncertain Future μ	6.19	1.80	1.62	1.78	1.80	1.67	2.53	1.80	1.73
Uncertain Current μ	0.42	0.11	0.09	0.15	0.11	0.08	0.16	0.11	0.09
Estimation Risk	0.88	0.71	0.66	0.68	0.71	0.70	0.78	0.71	0.70

Table A10
Variance Ratios and Components of Long-Horizon Variance
Quarterly data (1952Q1–2006Q4)
Excess instead of real returns

Counterpart of Table A3

	ρ_{uw}			R^2			β		
Prior	Tight	Bench	Loose	Tight	Bench	Loose	Tight	Bench	Loose
	Panel A. Investment Horizon $k = 25$ years								
Variance Ratio	1.35	1.88	1.89	1.85	1.88	1.95	2.11	1.88	1.85
Predictive Variance	0.86	1.19	1.19	1.20	1.19	1.21	1.33	1.19	1.17
IID Component	0.62	0.62	0.62	0.63	0.62	0.61	0.62	0.62	0.62
Mean Reversion	-1.33	-1.00	-1.00	-0.80	-1.00	-1.16	-1.16	-1.00	-1.00
Uncertain Future μ	1.18	1.18	1.19	0.97	1.18	1.36	1.44	1.18	1.17
Uncertain Current μ	0.05	0.03	0.04	0.04	0.03	0.04	0.06	0.03	0.03
Estimation Risk	0.34	0.36	0.35	0.35	0.36	0.37	0.37	0.36	0.35
	Panel B. Investment Horizon $k = 50$ years								
Variance Ratio	2.35	2.78	2.78	2.69	2.78	2.92	3.25	2.78	2.72
Predictive Variance	1.49	1.75	1.76	1.74	1.75	1.81	2.06	1.75	1.72
IID Component	0.62	0.62	0.62	0.63	0.62	0.61	0.62	0.62	0.62
Mean Reversion	-1.50	-1.10	-1.10	-0.90	-1.10	-1.26	-1.30	-1.10	-1.10
Uncertain Future μ	1.60	1.45	1.46	1.25	1.45	1.65	1.88	1.45	1.42
Uncertain Current μ	0.04	0.03	0.03	0.03	0.03	0.03	0.05	0.03	0.03
Estimation Risk	0.73	0.76	0.75	0.72	0.76	0.80	0.81	0.76	0.75

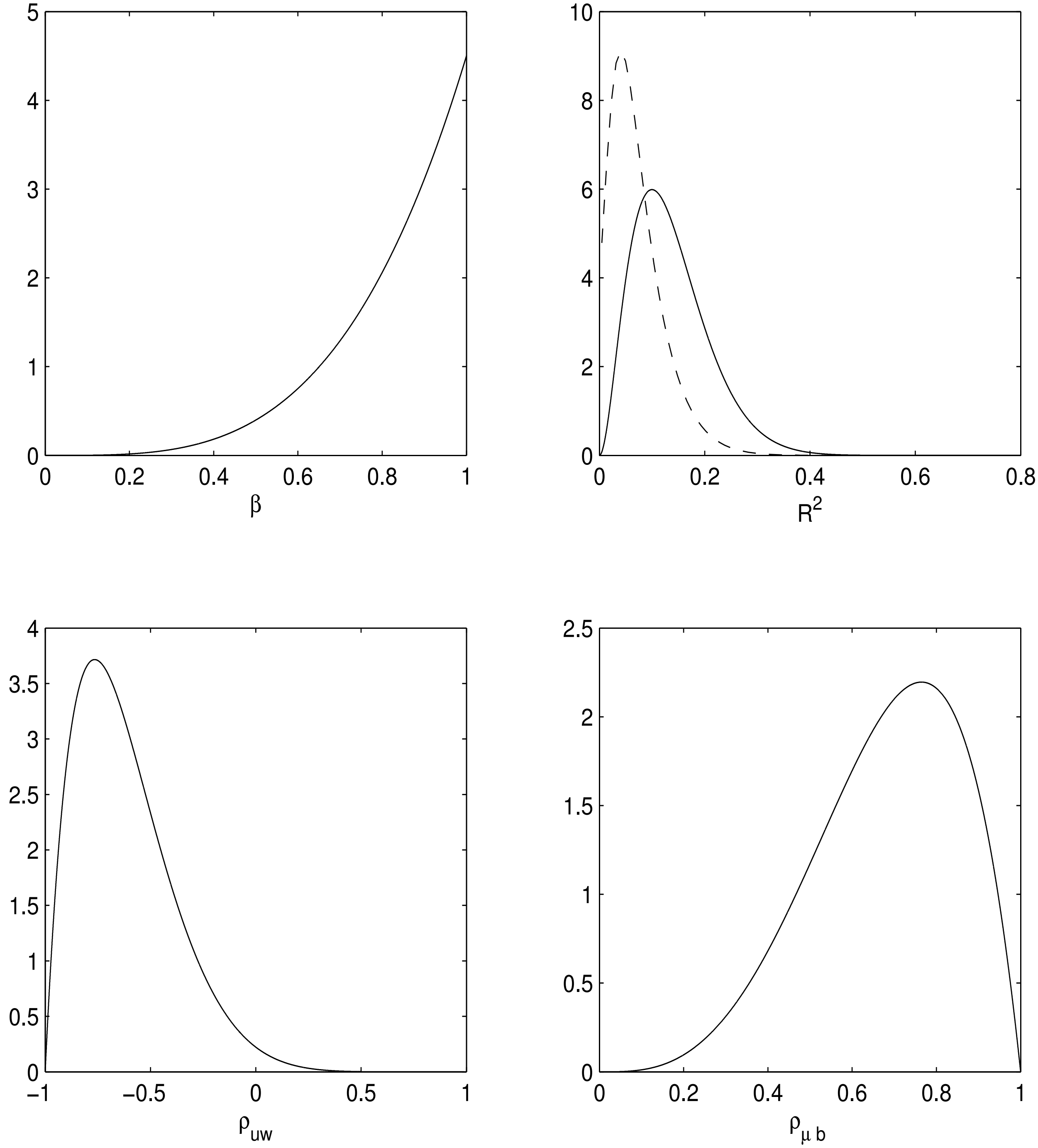


Figure A1. Distributions for uncertain parameters The plots display the probability densities used to illustrate the effects of parameter uncertainty on long-run variance. In the R^2 panel, the solid line plots the density of the true R^2 (predictability given μ_T), and the dashed line plots the implied density of the R-squared in a regression of returns on b_T. The dashed line incorporates the uncertainty about $\rho_{\mu b}$.

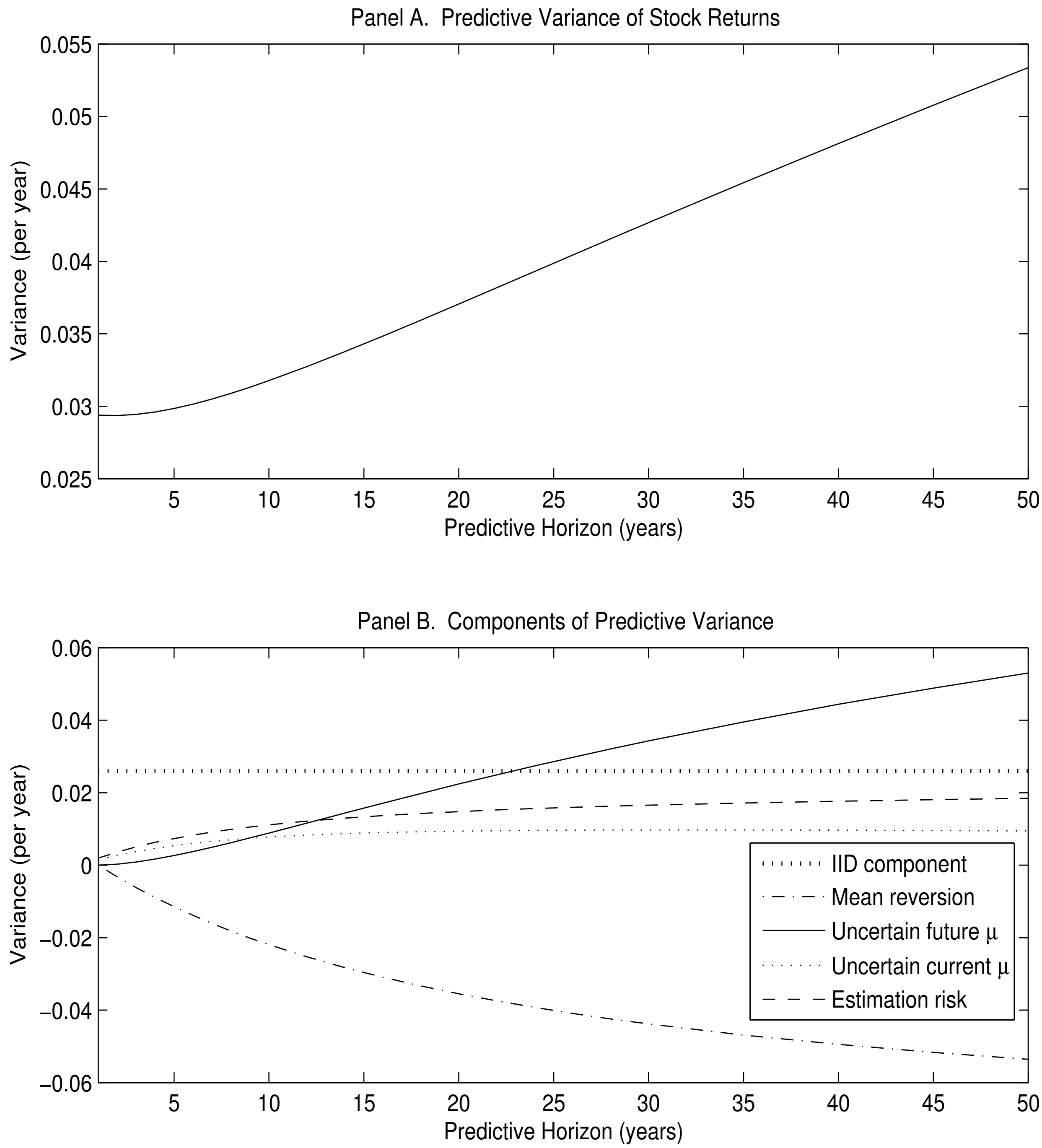


Figure A2 (same as Figure 6 in the paper). Predictive variance of multiperiod return and its components. Panel A plots the variance of the predictive distribution of long-horizon returns, Var$(r_{T,T+k}|D_T)$. Panel B plots the five components of the predictive variance. All quantities are divided by 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.

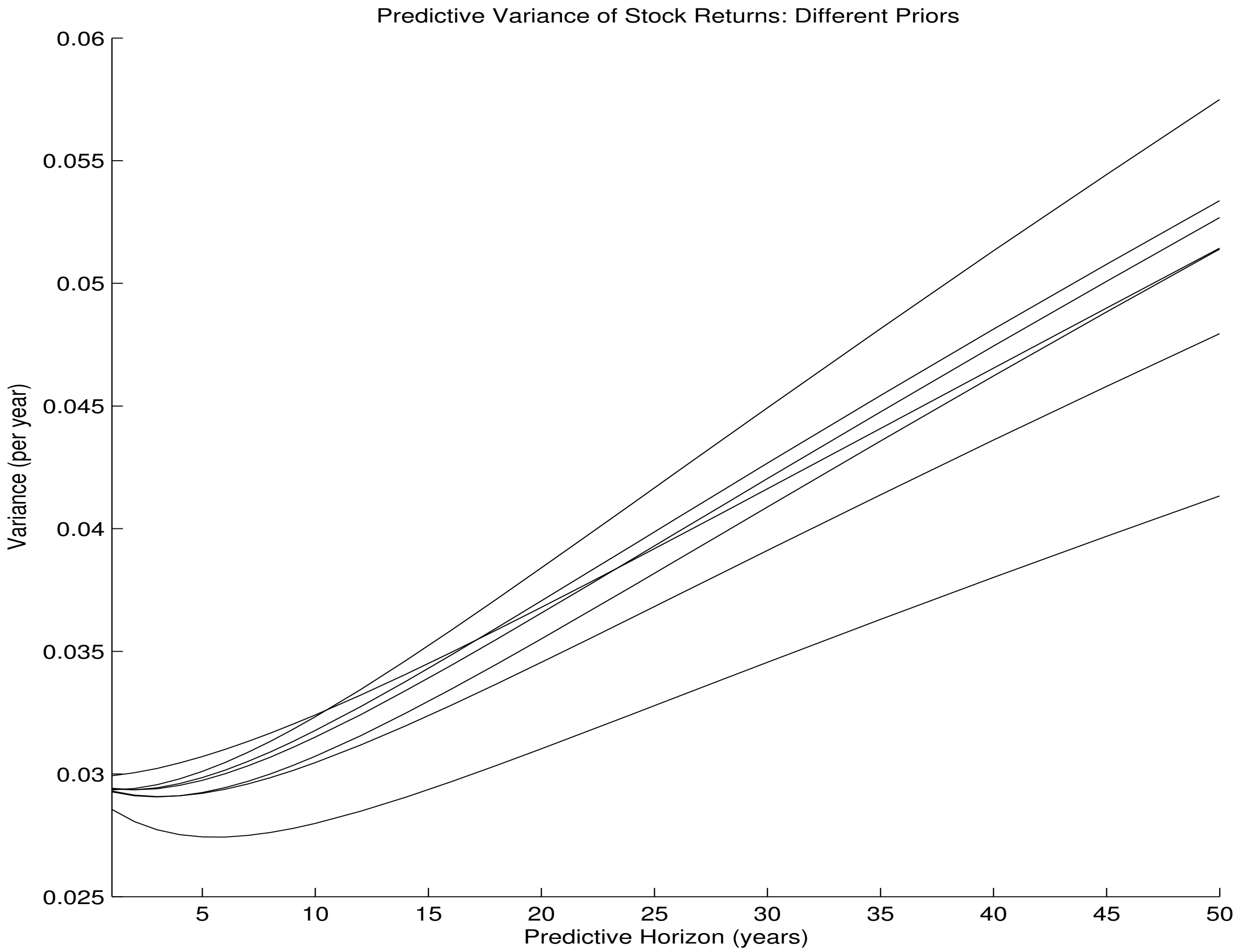


Figure A3. Predictive variance for seven different priors. The figure plots the variance of the predictive distribution of long-horizon returns, $\mathrm{Var}(r_{T,T+k}|D_T)$, as a function of the investment horison k. The variance is divided by k, the number of periods in the return horizon. The results 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.

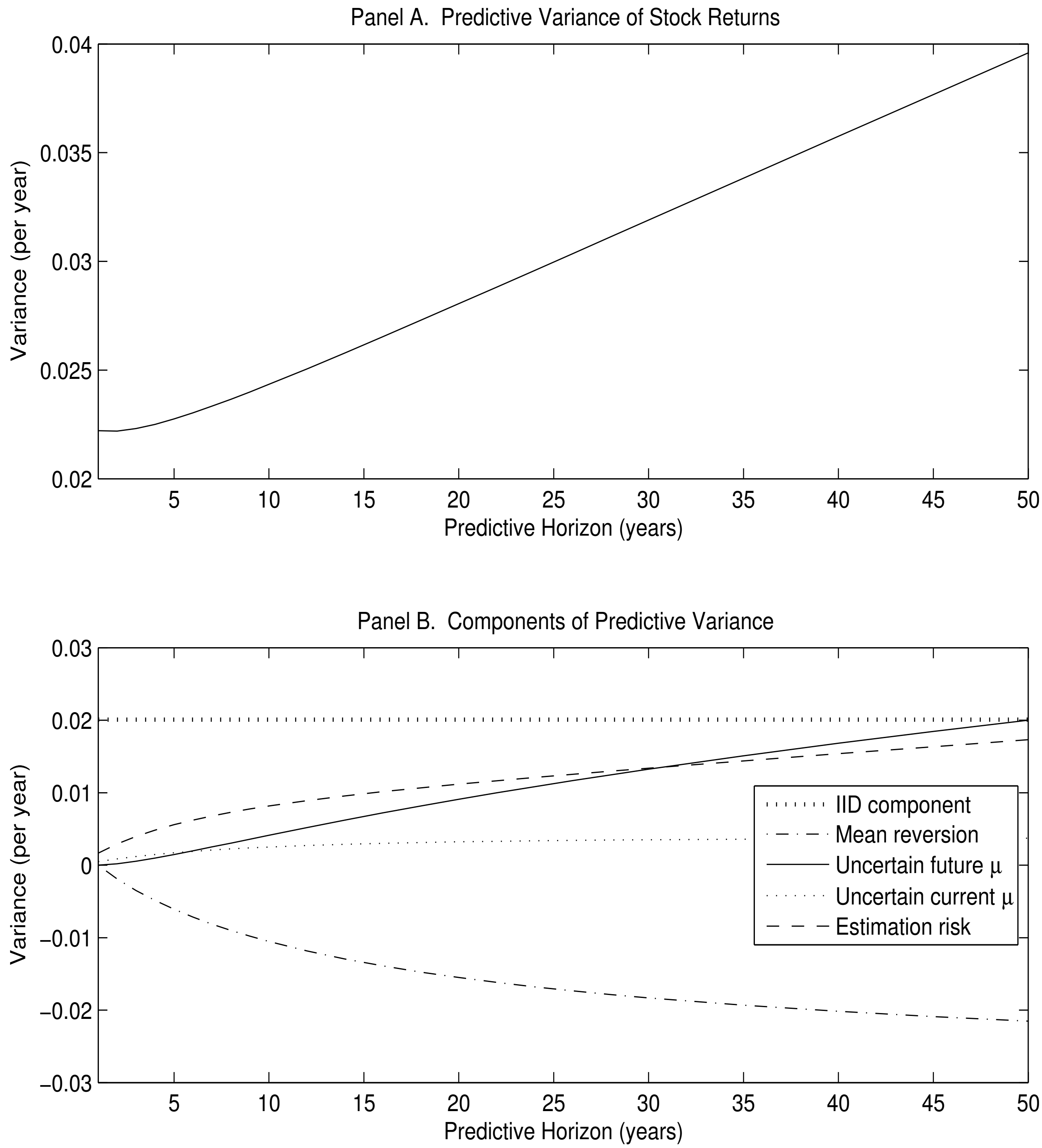


Figure A4. Predictive variance of multiperiod return and its components.
First subperiod (1802–1904). Counterpart of Figure A2.

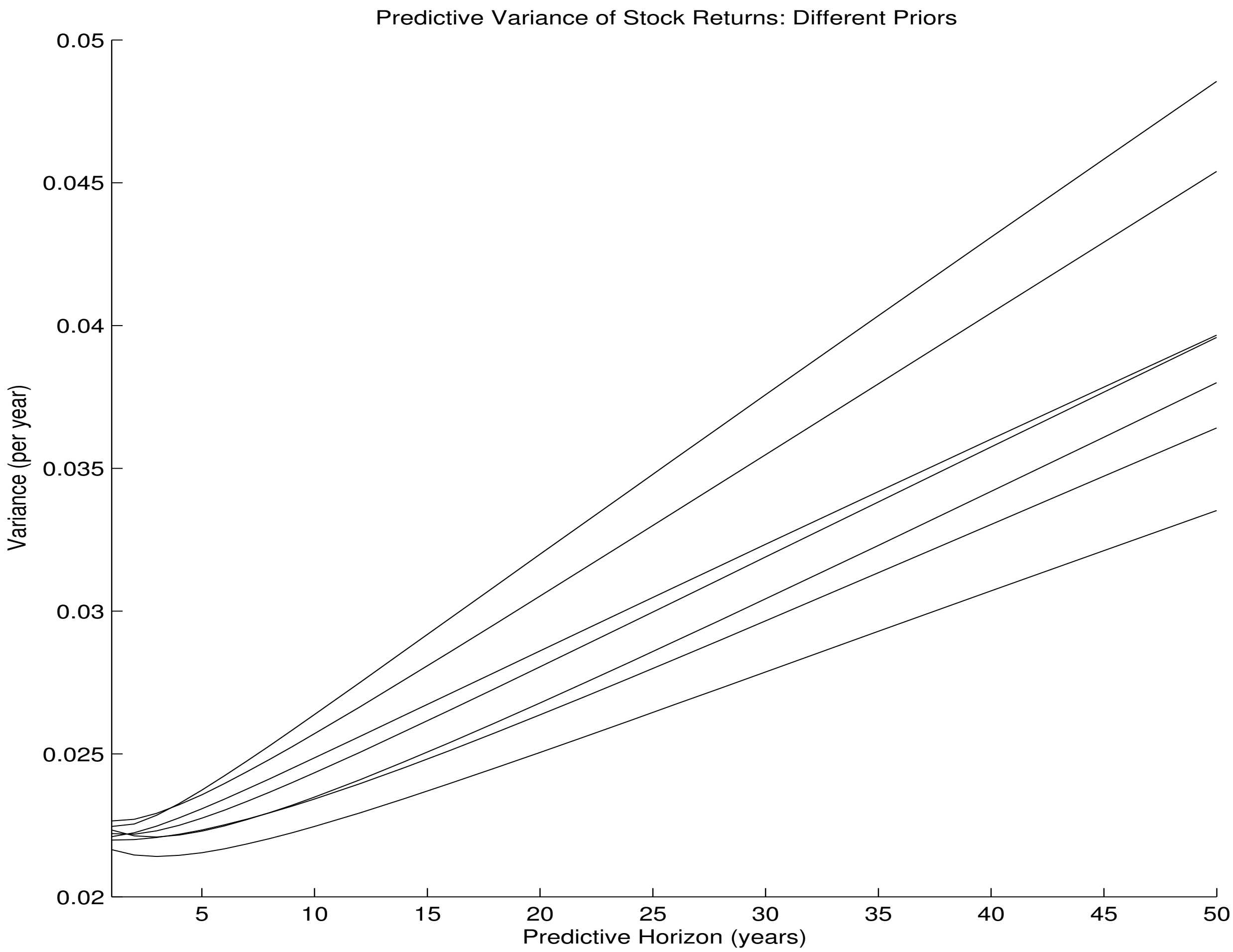


Figure A5. Predictive variance for seven different priors.
First subperiod (1802–1904). Counterpart of Figure A3.

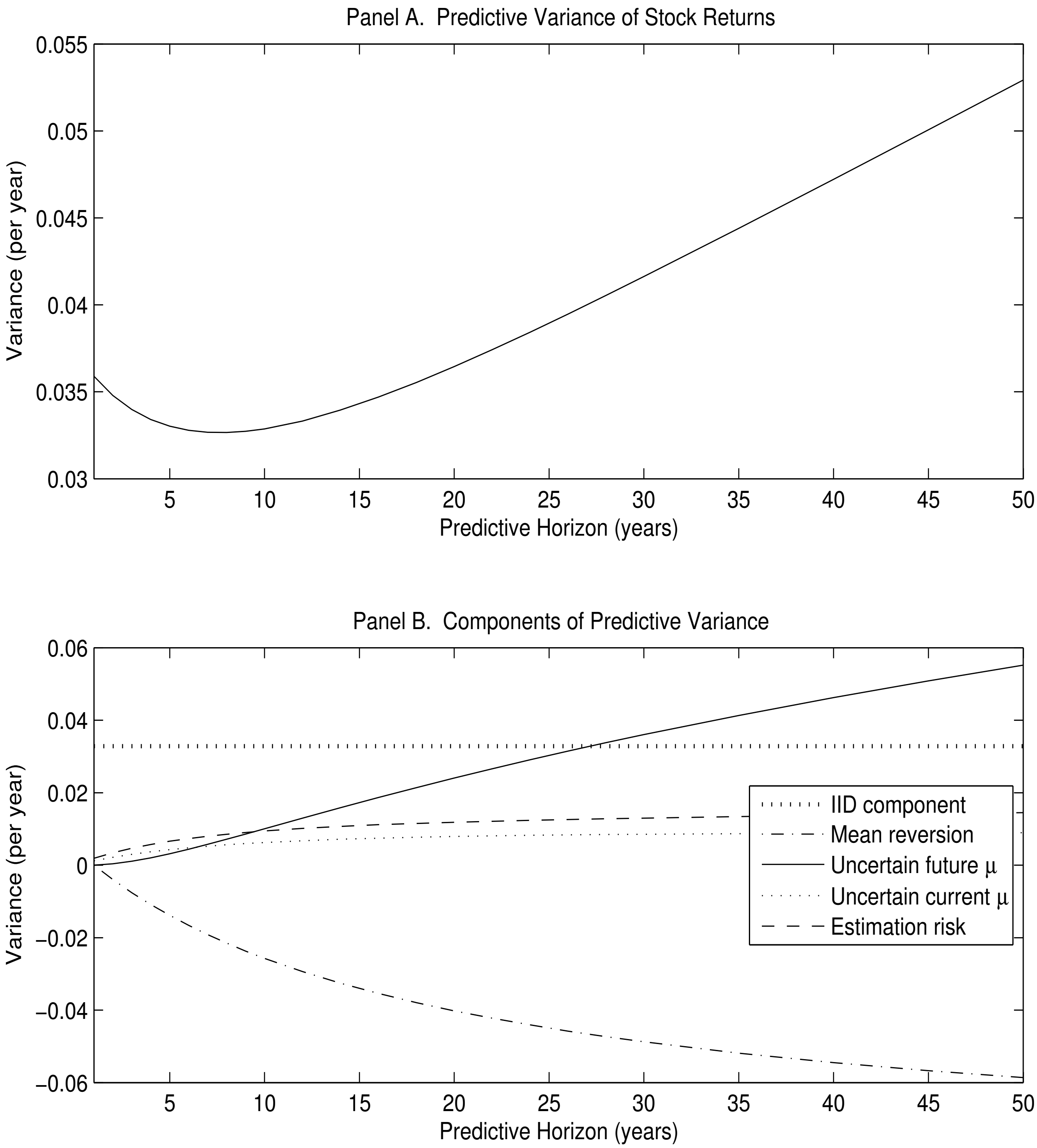


Figure A6. Predictive variance of multiperiod return and its components.
Second subperiod (1905–2007). Counterpart of Figure A2.

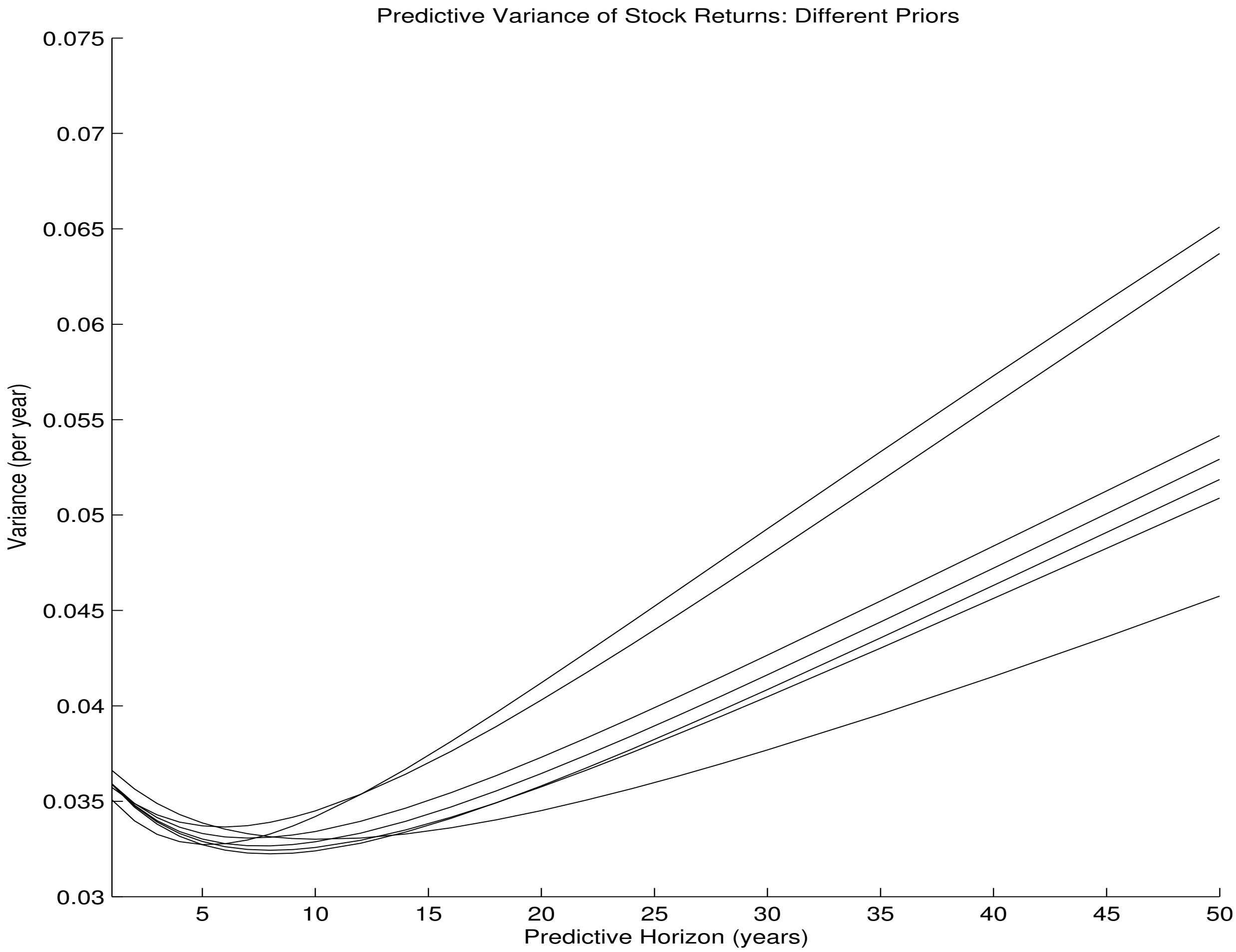


Figure A7. Predictive variance for seven different priors.
Second subperiod (1905–2007). Counterpart of Figure A3.

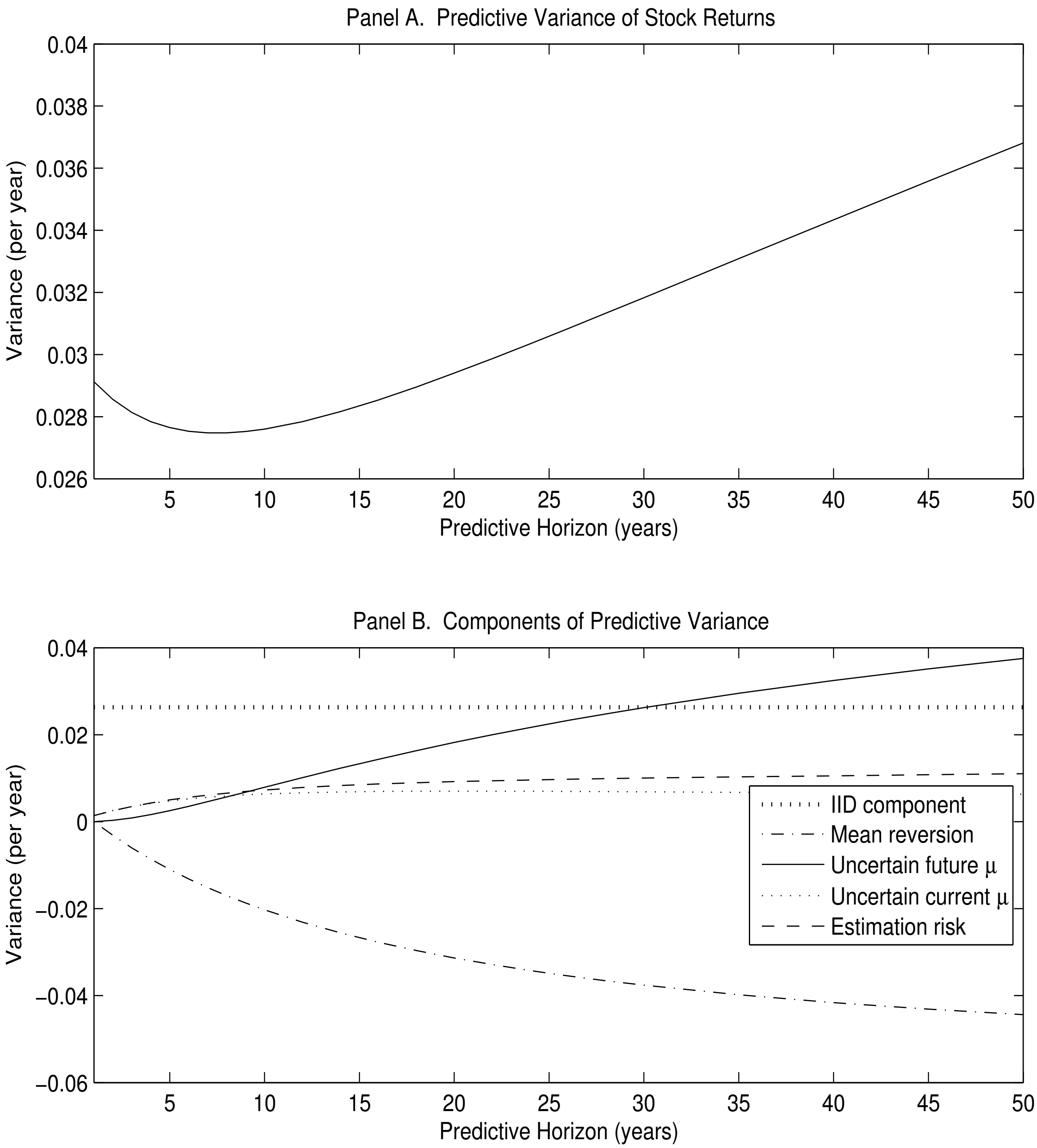


Figure A8. Predictive variance of multiperiod return and its components.
One instead of three predictors (dividend yield). Counterpart of Figure A2.

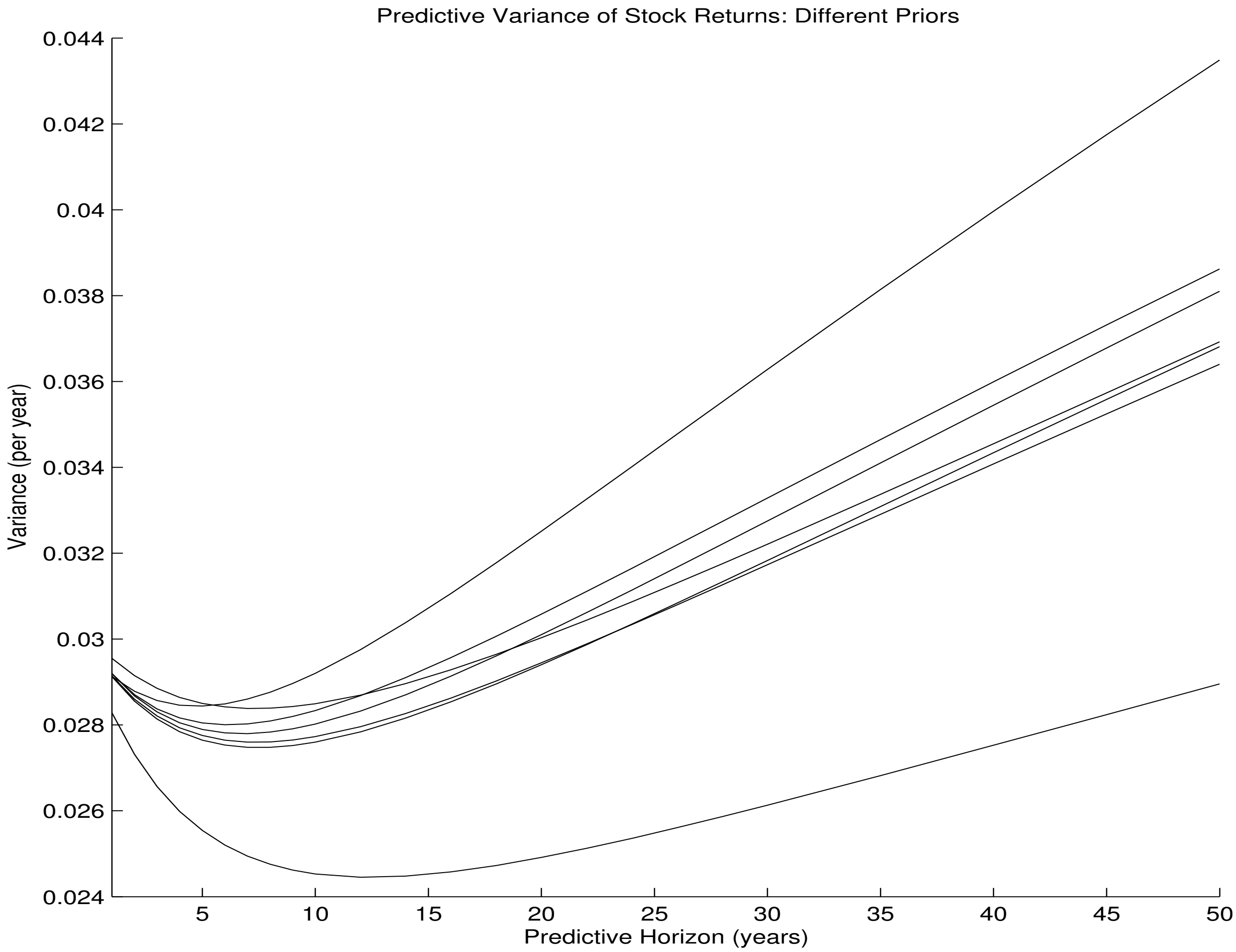
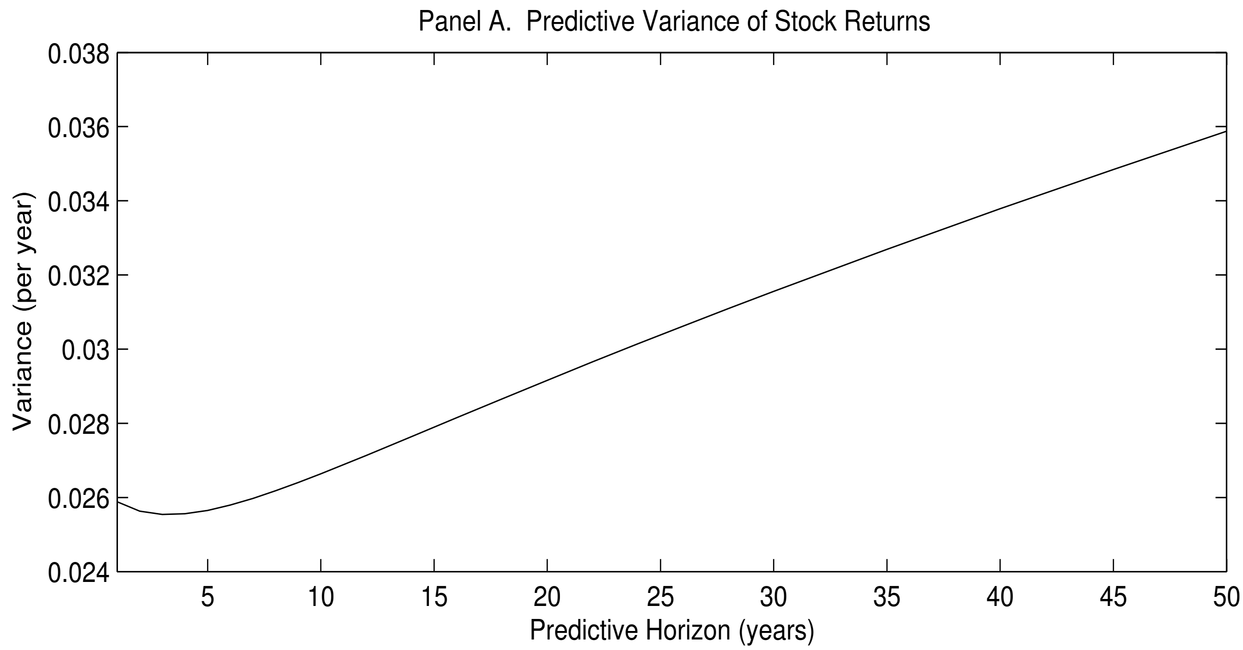


Figure A9. Predictive variance for seven different priors.
One instead of three predictors (dividend yield). Counterpart of Figure A3.



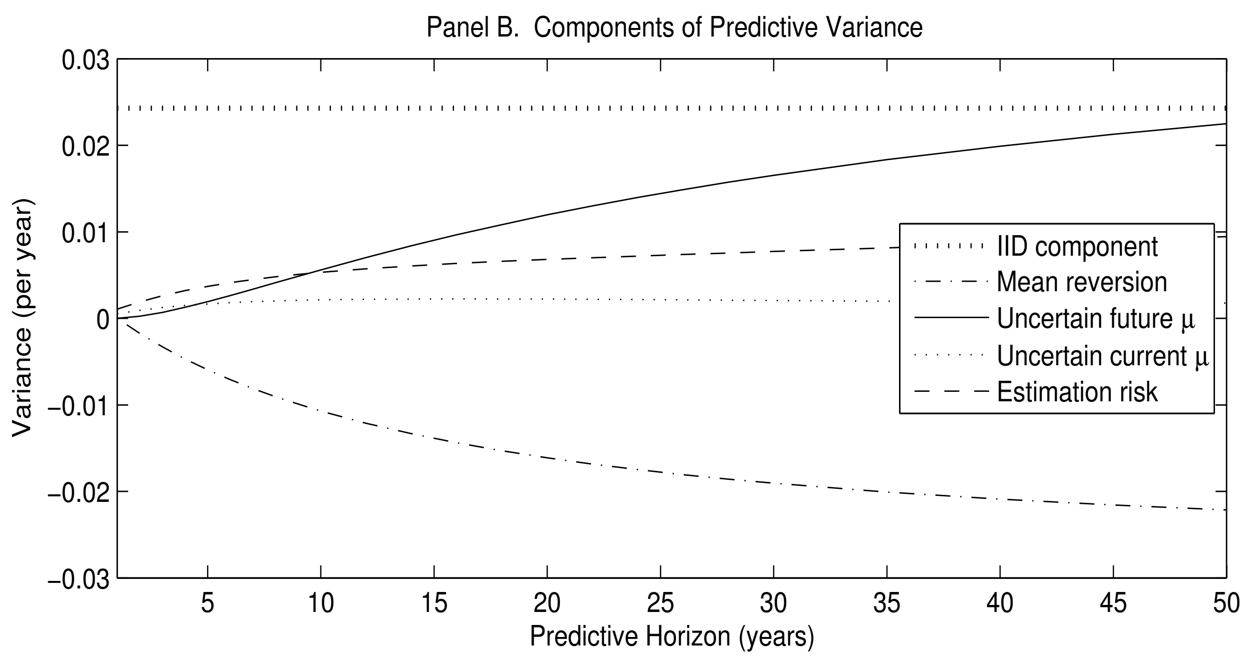


Figure A10. Predictive variance of multiperiod return and its components.
Excess instead of real returns. Counterpart of Figure A2.

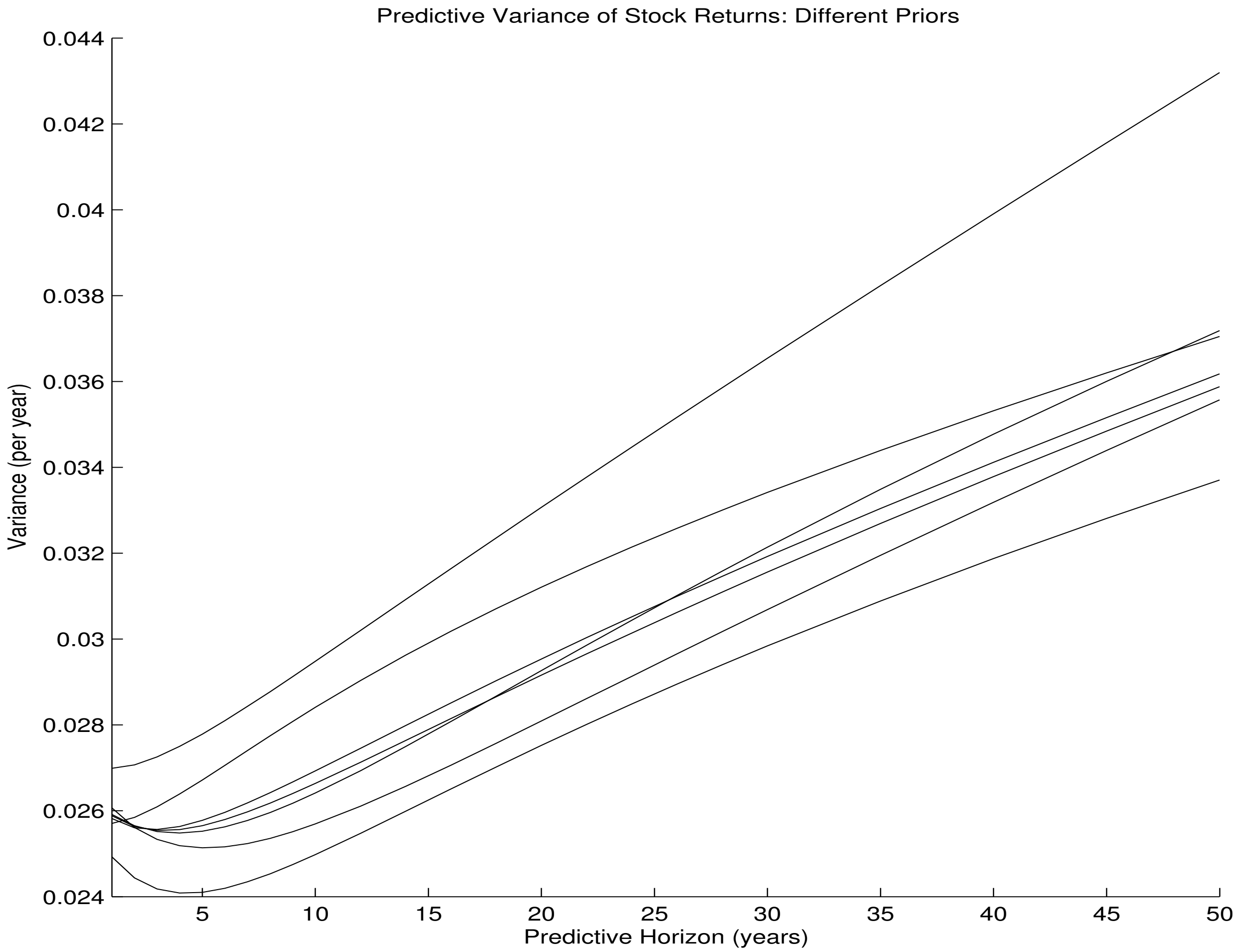


Figure A11. Predictive variance for seven different priors.
Excess instead of real returns. Counterpart of Figure A3.

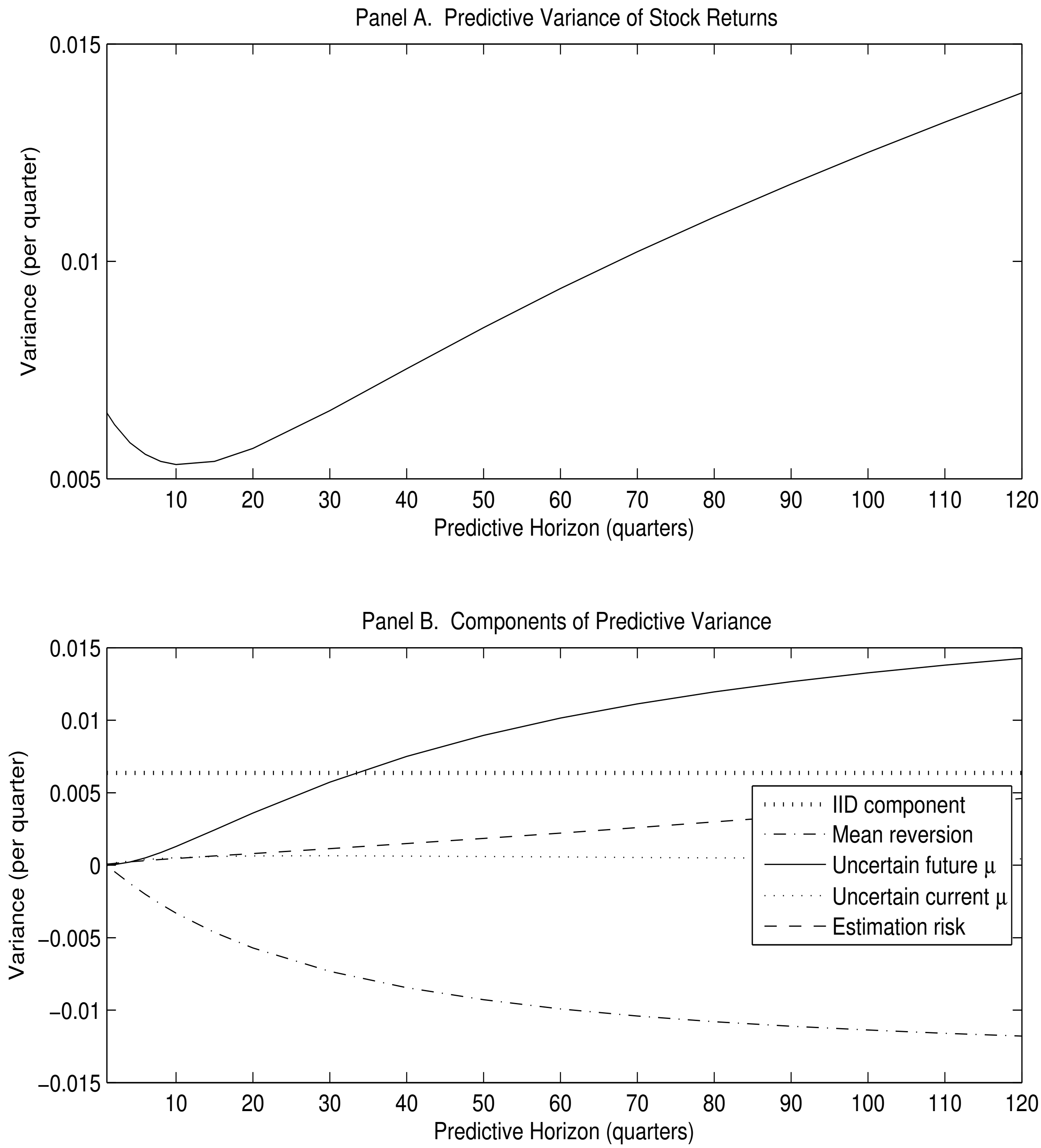


Figure A12. Predictive variance of multiperiod return and its components.
Quarterly data (1952Q1–2006Q4). Counterpart of Figure A2.

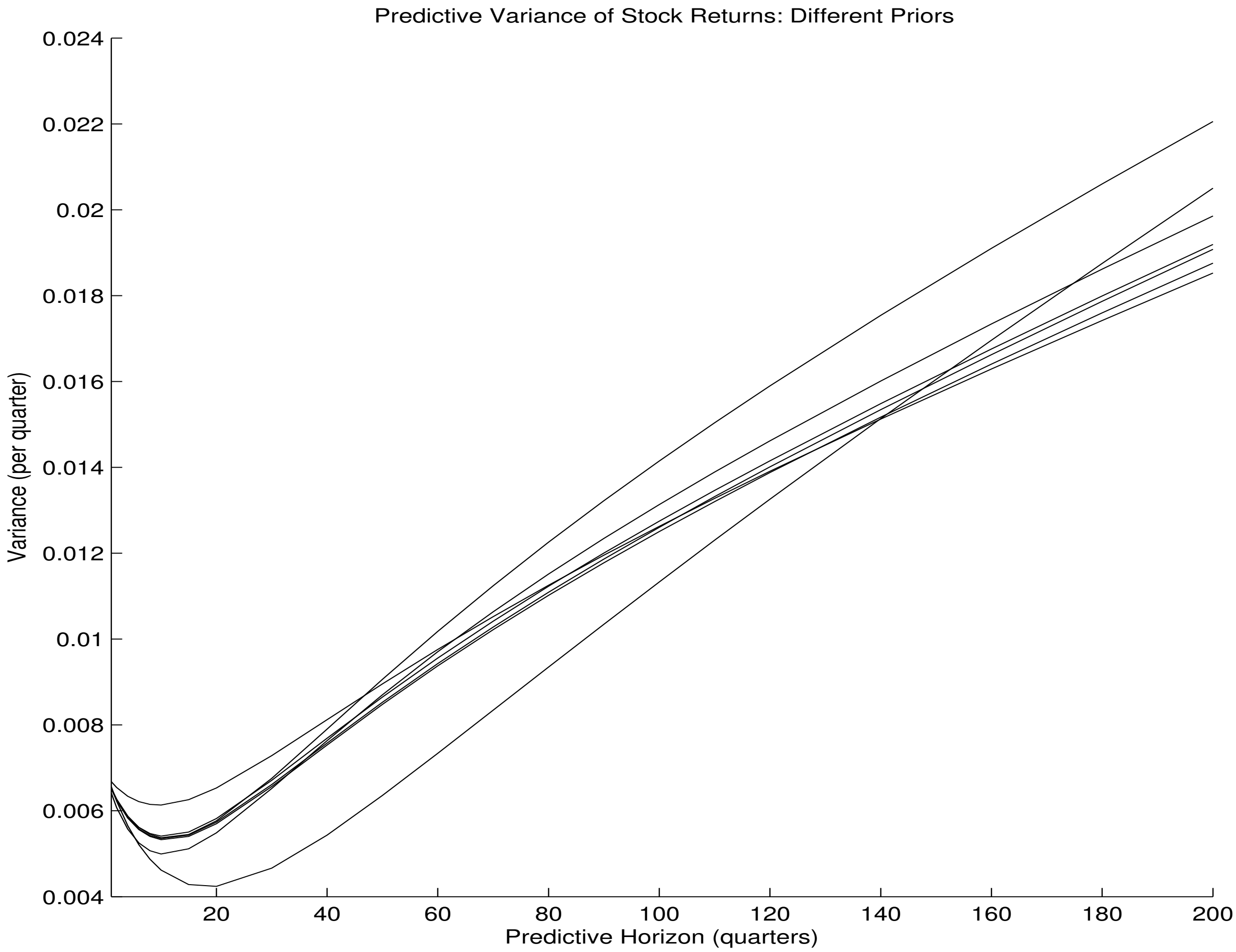


Figure A13. Predictive variance for seven different priors.
Quarterly data (1952Q1–2006Q4). Counterpart of Figure A3.

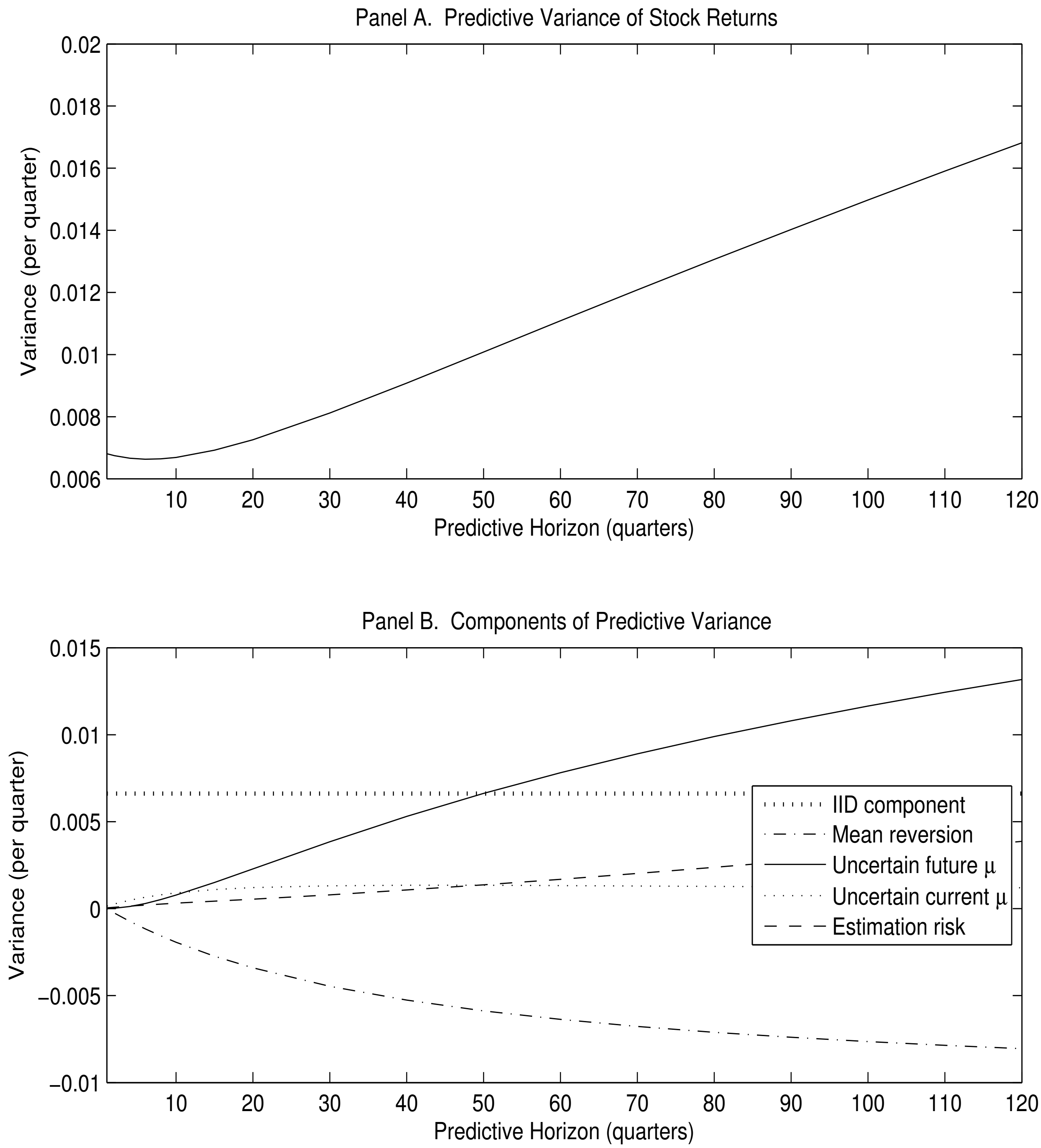


Figure A14. Predictive variance of multiperiod return and its components.
Quarterly data (1952Q1–2006Q4). One instead of three predictors (dividend yield). Counterpart of Figure A2.

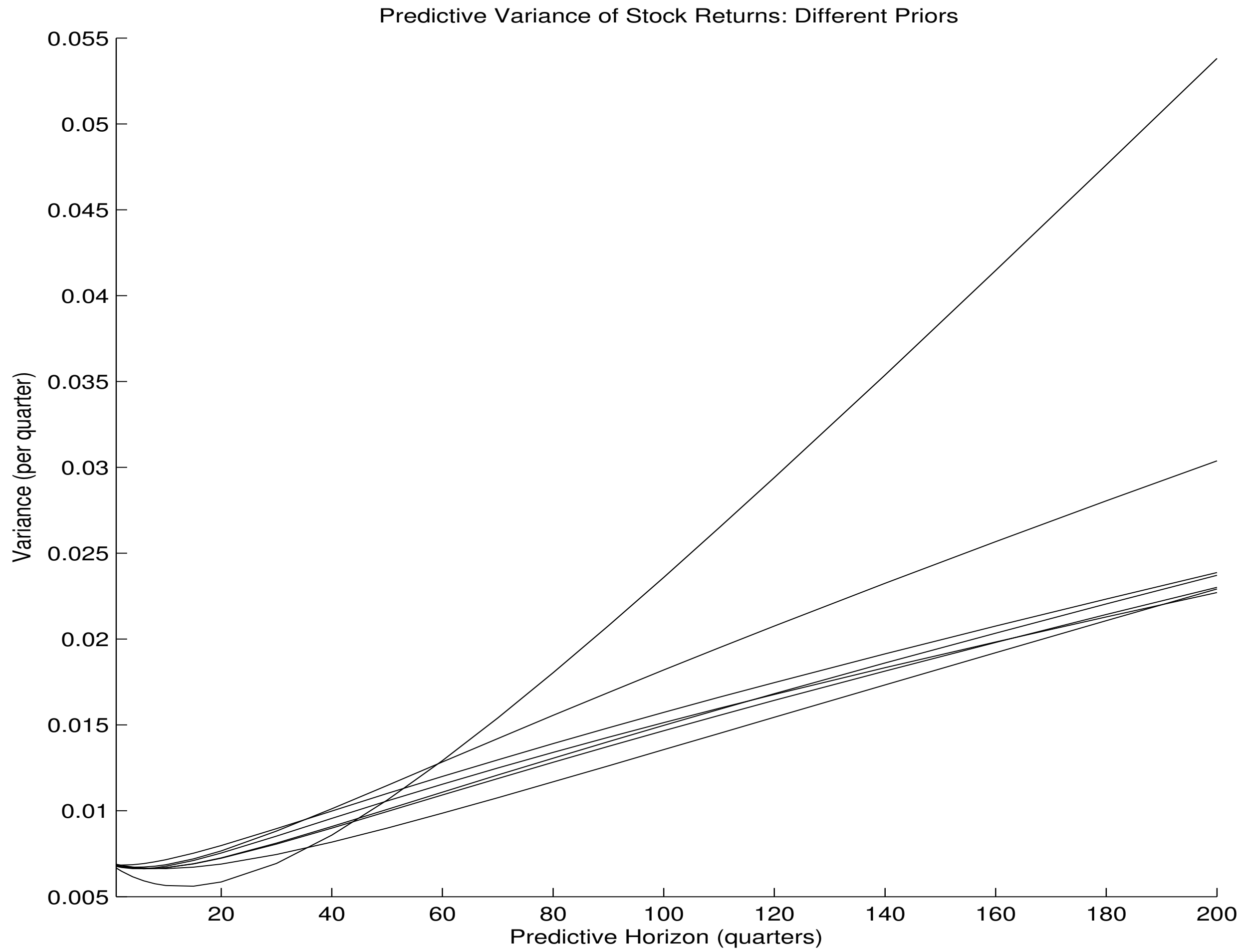


Figure A15. Predictive variance for seven different priors.
Quarterly data (1952Q1–2006Q4). One instead of three predictors (dividend yield). Counterpart of Figure A3.

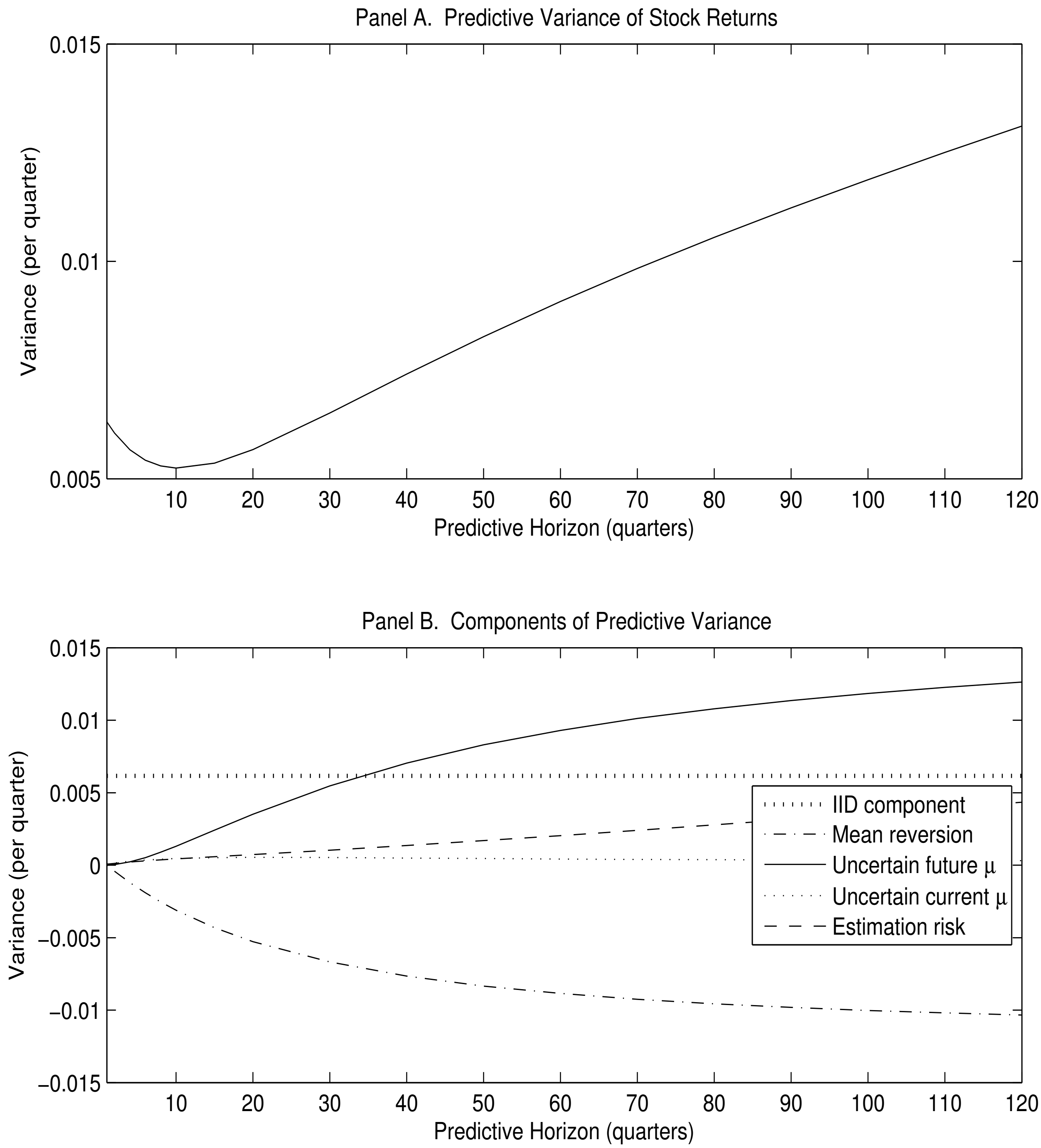


Figure A16. Predictive variance of multiperiod return and its components.
Quarterly data (1952Q1–2006Q4). Excess instead of real returns. Counterpart of Figure A2.

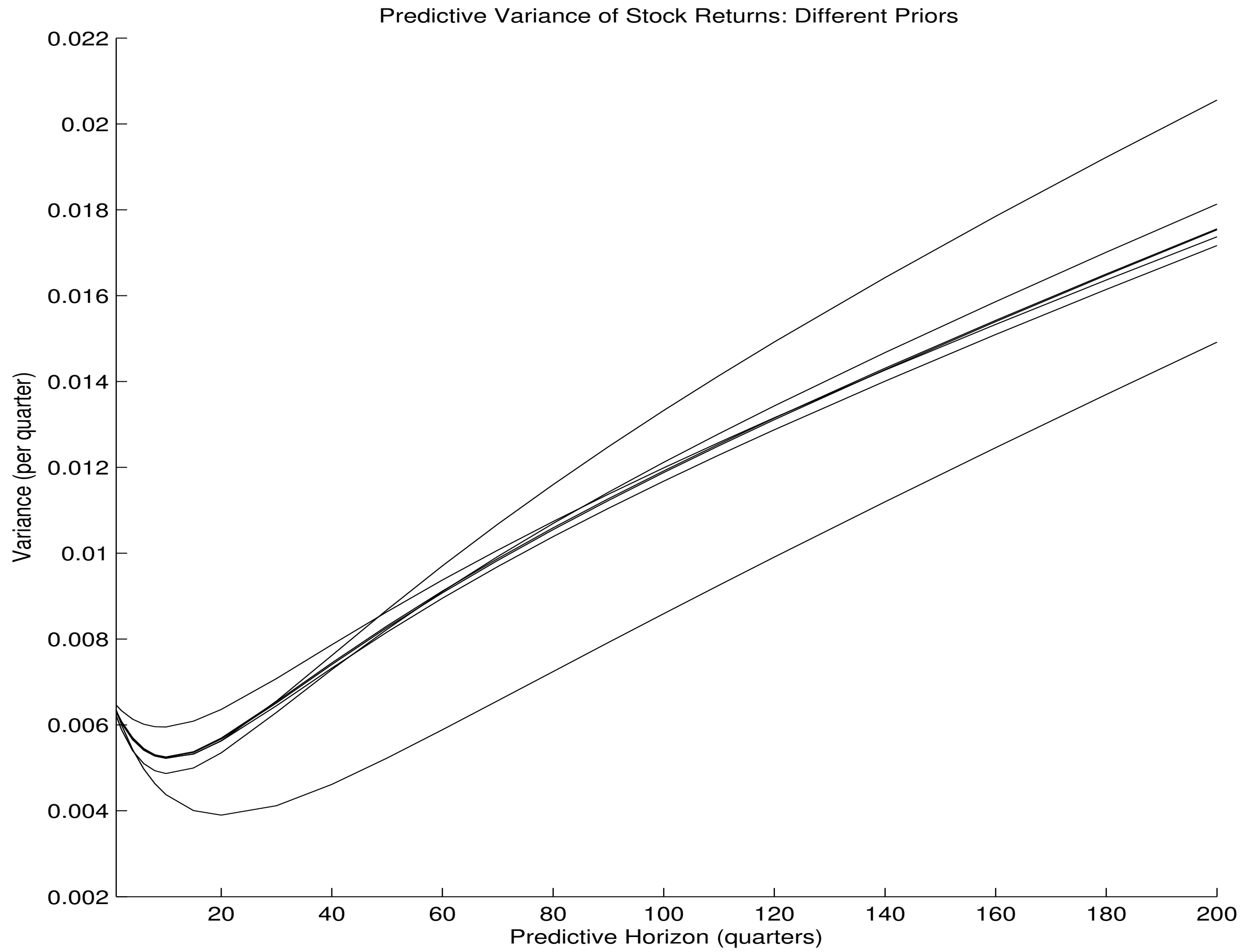


Figure A17. Predictive variance for seven different priors.
Quarterly data (1952Q1–2006Q4). Excess instead of real returns. Counterpart of Figure A3.

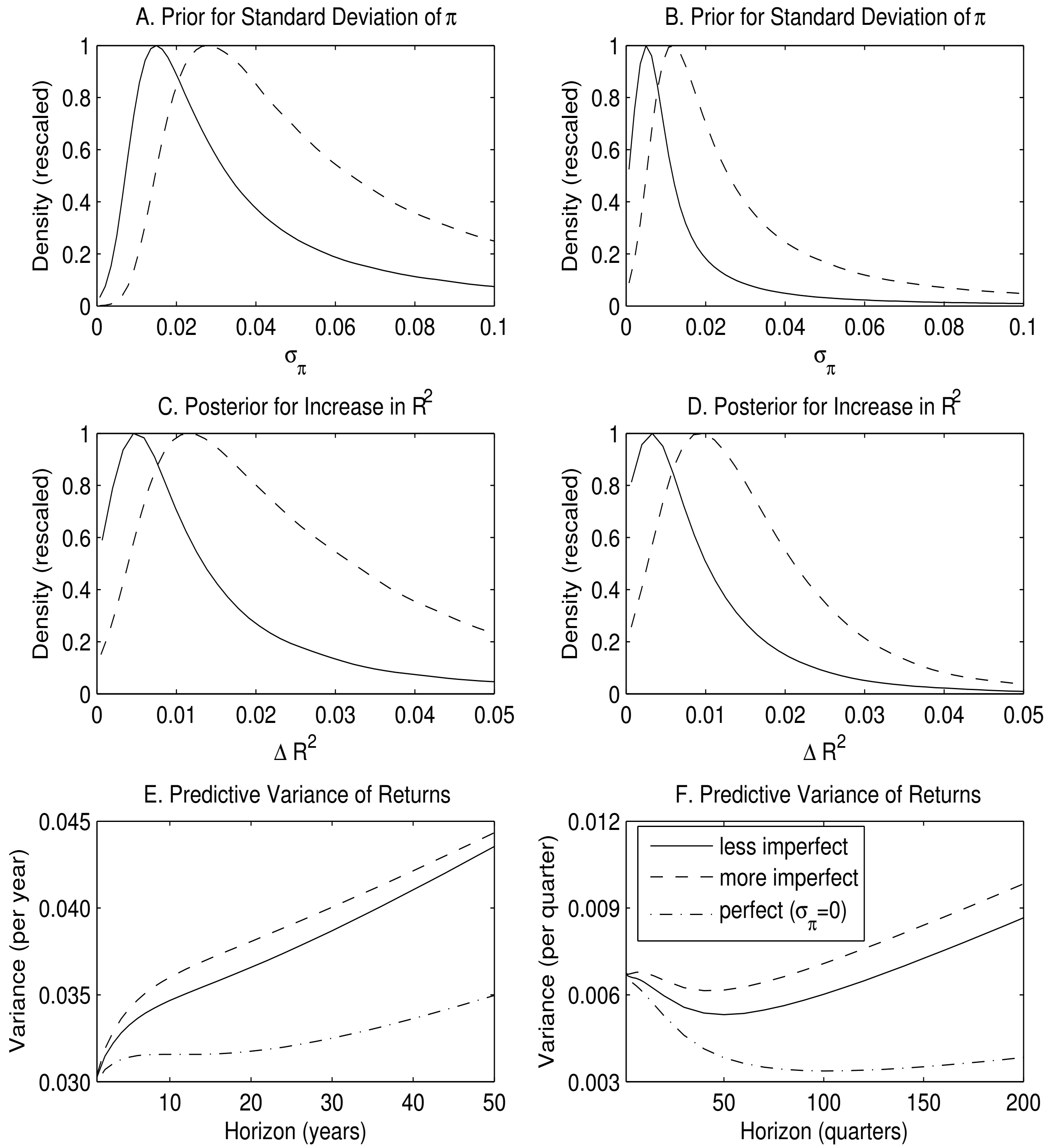


Figure A18 (same as Figure 7 in the paper). Predictive variance and predictor imperfection. The plots display results under the predictive system (System 2) in which expected return depends on a vector of observable predictors, x_t, as well as a missing predictor, π_t, that obeys an AR(1) process. The top panels display prior distributions for σ_π, the standard deviation of π_t, under different degrees of predictor imperfection. The middle panels display the corresponding posteriors of ΔR^2, the "true" R^2 for one-period returns minus the "observed" R^2 when conditioning only on x_t. The bottom panels display the predictive variances for the two imperfect-predictor cases as well for the case of perfect predictors ($\sigma_\pi = \Delta R^2 = 0$). The left-hand panels are based on annual data from 1802–2007 for real U.S. stock returns and three predictors: the dividend yield, the bond yield, and the term spread. The right-hand panels are based on quarterly data from 1952Q1–2006Q4 for real returns and three predictors: the dividend yield, CAY, and the bond yield.

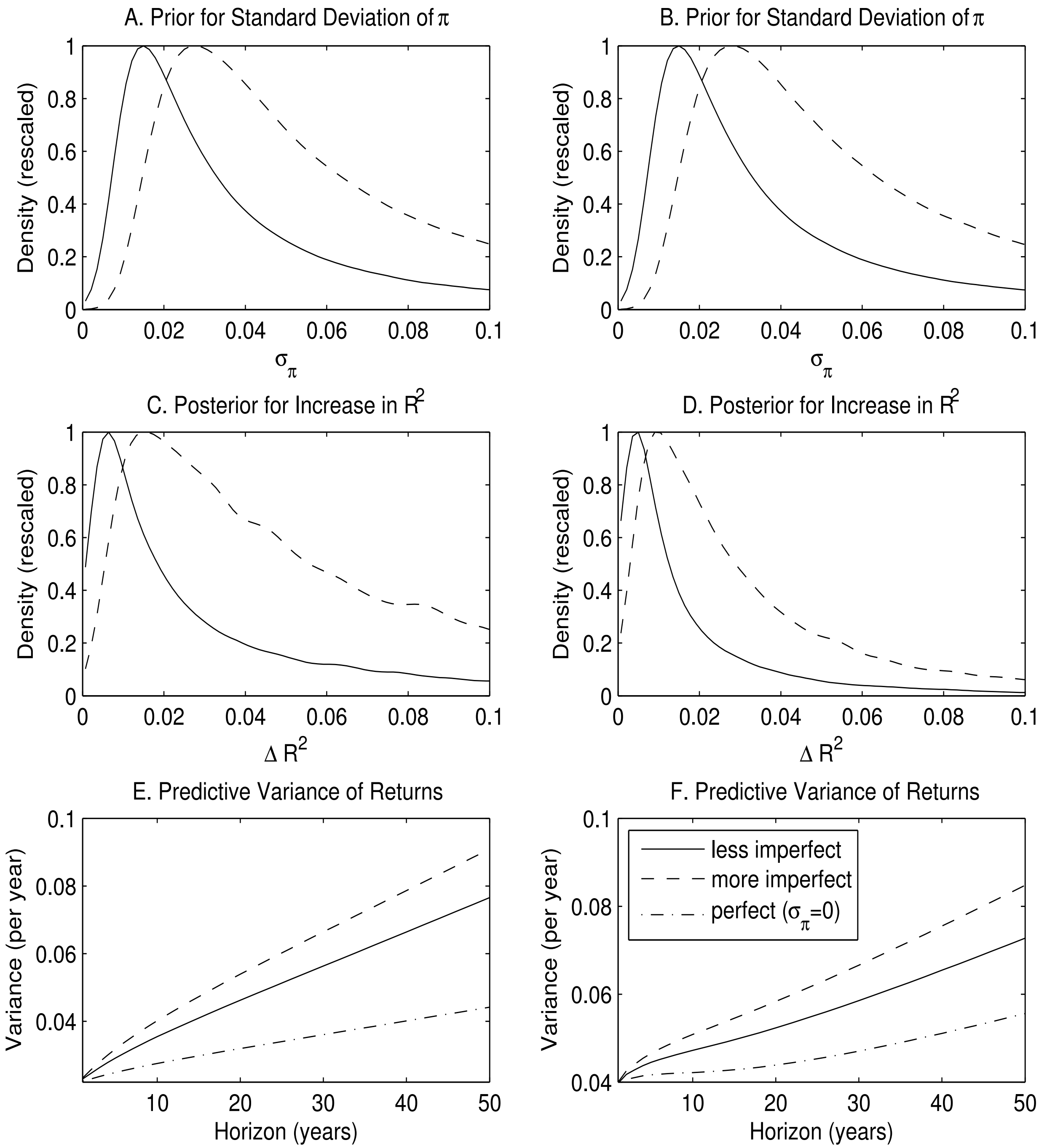


Figure A19. Predictive variance and predictor imperfection. The plots display results under the predictive system (System 2) in which expected return depends on a vector of observable predictors, x_t, as well as a missing predictor, π_t, that obeys an AR(1) process. The top panels display prior distributions for σ_π, the standard deviation of π_t, under different degrees of predictor imperfection. The middle panels display the corresponding posteriors of ΔR^2, the "true" R^2 for one-period returns minus the "observed" R^2 when conditioning only on x_t. The bottom panels display the predictive variances for the two imperfect-predictor cases as well for the case of perfect predictors ($\sigma_\pi = \Delta R^2 = 0$). The left-hand panels are based on annual data from 1802–1904 for real U.S. stock returns and three predictors: the dividend yield, the bond yield, and the term spread. The right-hand panels display the same results for the 1905–2007 period.

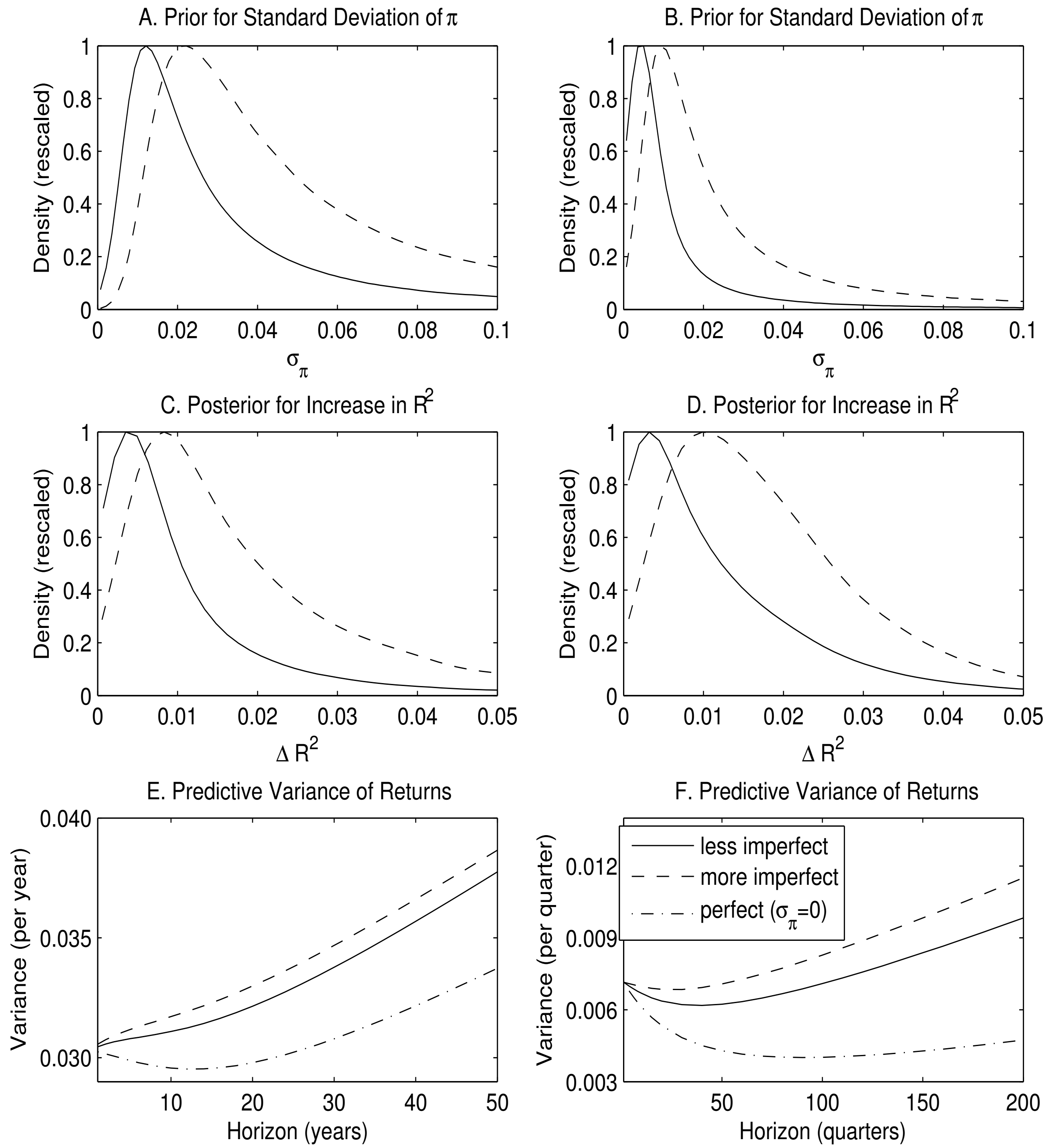


Figure A20. Predictive variance and predictor imperfection. The plots display results under the predictive system (System 2) in which expected return depends on a vector of observable predictors, x_t, as well as a missing predictor, π_t, that obeys an AR(1) process. The top panels display prior distributions for σ_π, the standard deviation of π_t, under different degrees of predictor imperfection. The middle panels display the corresponding posteriors of ΔR^2, the "true" R^2 for one-period returns minus the "observed" R^2 when conditioning only on x_t. The bottom panels display the predictive variances for the two imperfect-predictor cases as well for the case of perfect predictors ($\sigma_\pi = \Delta R^2 = 0$). The left-hand panels are based on annual data from 1802–2007 for real U.S. stock returns and one predictor: the dividend yield. The right-hand panels are based on quarterly data from 1952Q1–2006Q4 for real returns and one predictor: the dividend yield.

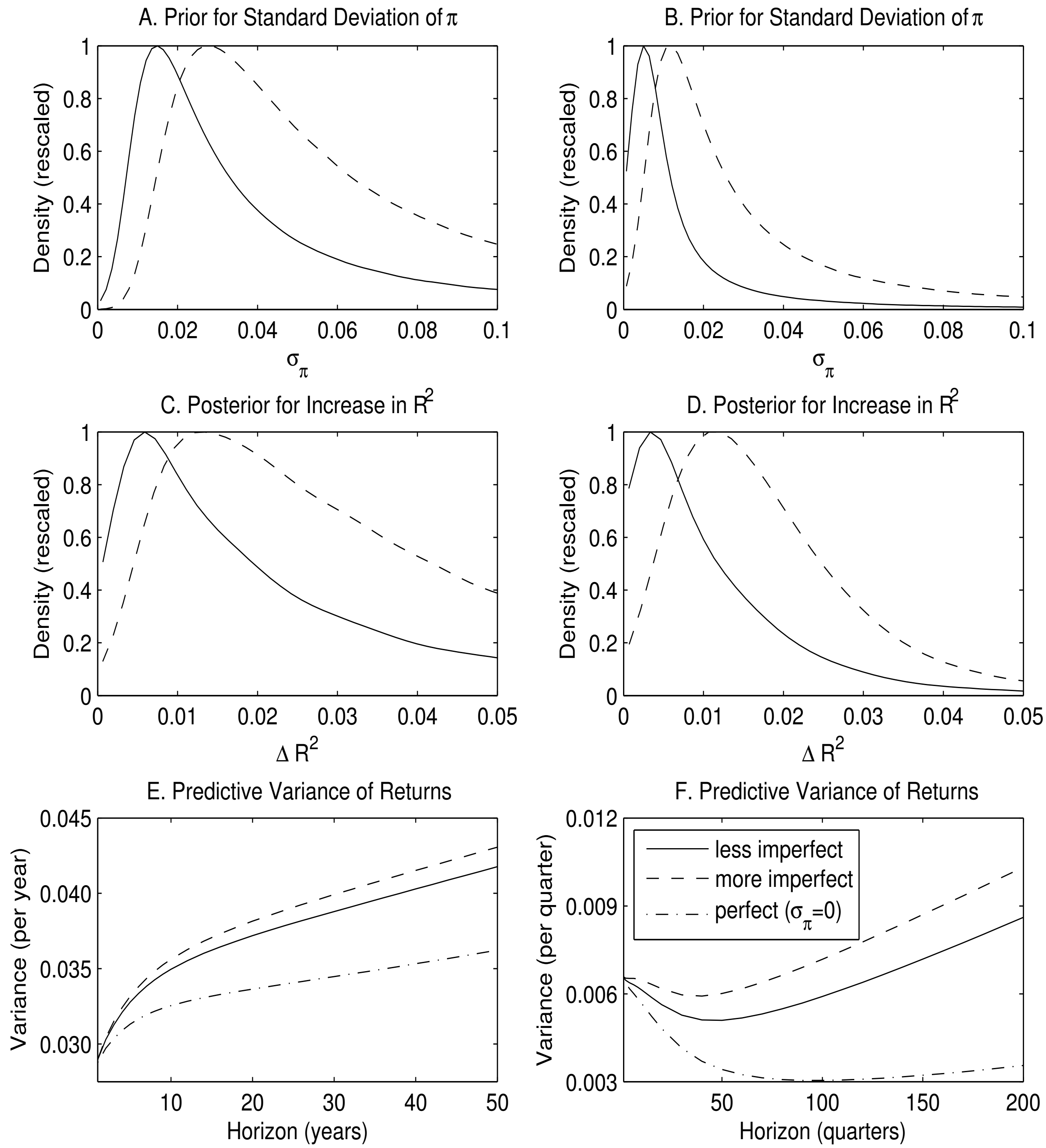


Figure A21. Predictive variance and predictor imperfection. The plots display results under the predictive system (System 2) in which expected return depends on a vector of observable predictors, x_t, as well as a missing predictor, π_t, that obeys an AR(1) process. The top panels display prior distributions for σ_π, the standard deviation of π_t, under different degrees of predictor imperfection. The middle panels display the corresponding posteriors of ΔR^2, the "true" R^2 for one-period returns minus the "observed" R^2 when conditioning only on x_t. The bottom panels display the predictive variances for the two imperfect-predictor cases as well for the case of perfect predictors ($\sigma_\pi = \Delta R^2 = 0$). The left-hand panels are based on annual data from 1802–2007 for excess U.S. stock returns and three predictors: the dividend yield, the bond yield, and the term spread. The right-hand panels are based on quarterly data from 1952Q1–2006Q4 for excess returns and three predictors: the dividend yield, CAY, and the bond yield.

References

Carter, Chris K., and Robert Kohn, 1994, On Gibbs sampling for state space models, *Biometrika* 81, 541–553.

Frühwirth-Schnatter, Sylvia, 1994, Data augmentation and dynamic linear models, *Journal of Time Series Analysis* 15, 183–202.

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

West, Mike, and Jeff Harrison, 1997, *Bayesian Forecasting and Dynamic Models* (Springer-Verlag, New York, NY).

Zellner, Arnold, 1971, *An Introduction to Bayesian Inference in Econometrics* (John Wiley & Sons, New York, NY).