On the Long Run Volatility of Stocks

March 4, 2015

Abstract

In this paper we investigate whether or not the conventional wisdom that stocks are more attractive for long horizon investors hold. Taking the perspective of an investor, we evaluate the predictive variance of k-period returns for different models and prior specifications and conclude, that stocks are indeed less volatile in the long run. Part of the developments include an extension of the modeling framework to incorporate time varying volatilities and covariances in a constrained prior information set up.

1 Introduction

The view that stocks represent safer and more attractive investment for long horizon portfolios is widespread. This conventional wisdom is key in justifying the large allocations in stocks suggested by financial advisors with regards to retirement portfolios, and the now very popular, target-date mutual funds. This notion has been explored and validated through the years in various empirical studies. Siegel [2008] provides a comprehensive assessment of the behavior of stock portfolios with an extensive dataset of equity returns in the U.S. dating back to 1802 and concludes, rather emphatically, that stocks are indeed very attractive to long horizon portfolios.

A number of the studies that attempt to characterize the variance of stocks over different horizons can be viewed as incomplete for they ignore the effects of estimation risk. They also fail to consider the perspective of a forward looking investor that relies on all available current information to evaluate the predictive variance of stocks. Note that dealing with this problem from the investor perspective requires the integration of all sources of uncertainty with regards to the current information set, i.e., the posterior distribution of all unknowns. Therefore, taking

a Bayesian perspective is not a choice but rather a necessity in answering this question. In addition, due to the low signal-to-noise ratio in modeling returns, sensible priors that really capture the available information are needed.

To our knowledge only 3 papers explore this question from the investor's perspective while simultaneously accounting for the effects of parameter uncertainty (estimation risk). First, Barberis [2000] works with the predictive regression framework [Stambaugh, 1999] and his results are in line with the conventional view. Pettenuzzo and Timmermann [2011] work in the same predictive regression framework and incorporate model instability through structural breaks; their results contradict conventional wisdom and suggest that the potential for future structural breaks significantly increases the predictive variance in the long run. Finally, Pástor and Stambaugh [2012] using a very flexible model (the predictive systems framework of Pástor and Stambaugh [2009]) conclude that once all sources of uncertainty are taken into account stocks are actually more volatile over longer horizons. The last two papers are based on more realistic models for stocks returns, use the same data as Siegel [2008], and their results cast doubts on a number of investment strategies that rely on the conventional wisdom being correct.

Our paper attempts to settle this disagreement by carefully exploring the effects of priors and model specifications in the estimation of the predictive variance. We work with a similar model framework as in Pástor and Stambaugh [2012] and try to isolate and understand the effects of priors on different parameters in the final results. Taking a conservative approach we start our analysis by ignoring the potential predictability, via covariates, of expected returns and focus solely on the temporal properties of the time series of returns. This strategy allow us to more easily isolate the effects of a small set of parameters and assumptions. We them extend our analysis to incorporate predictors and a very important element ignored in previous papers: time varying variances and covariances. This modification requires a significant innovation in modeling multivariate stochastic volatiles in the presence of prior constraints and associated computation strategy for posterior sampling.

Our conclusion is that conventional wisdom is correct after all. Unless investors possess very unusual beliefs about key quantities in the model the predictive variance per period of stocks decreases as a function of the investment horizon. This result is very robust and can be achieved in a variety of settings including the very flexible time varying volatilities and covariances. The paper starts by describing the basic problem and main model framework in Section 2 followed by an extensive empirical analysis in Section 3. Section 4 explore some extensions to the basic

framework and describes the time-varying modeling strategy for the covariance matrix of returns, expected returns and predictors.

2 The Basic Model

A state-space model provides a simple yet realistic representation of the data generating process for stock returns. The model takes the form:

$$\begin{aligned} r_{t+1} &= \mu_t + u_{t+1} \\ \mu_{t+1} &= \alpha + \beta\mu_t + w_{t+1} \end{aligned} \tag{1}$$

with

$$\begin{pmatrix} u_{t+1} \\ w_{t+1} \end{pmatrix} \sim N(0, \mathbf{\Sigma}) \quad \text{and} \quad \mathbf{\Sigma} = \begin{pmatrix} \sigma_u^2 & \sigma_{uw} \\ \sigma_{wu} & \sigma_w^2 \end{pmatrix}, \tag{2}$$

where r_{t+1} denotes the continuously compounded excess return from time t to $t+1$, and μ_t is the expected return (equity premium) conditional on all information at time t. Notice that unlike traditional state-space models the observation equation connects r_{t+1} to μ_t in order to emphasize the fact that μ_t represents the expectation for returns at time $t+1$ given the information set at time t. In addition, it is common to assume a stationary process for expected returns with $\beta \in (0, 1)$ and the correlation between shocks

$$\rho_{uw} = \frac{\sigma_{uw}}{\sigma_u \sigma_w} \in (-1, 0).$$

Although very simple, this model encompasses what is now seen as facts in the literature (see Cochrane [2005] for a comprehensive review): (i) expected returns are time varying and mean reverting and (ii) contemporaneous shocks to expected returns are negatively correlated with shocks to returns, i.e., when expected returns go up, returns tend to go down. This is due to the fact that asset prices tend to fall as discount rates (expected returns) rise (see Pástor and Stambaugh [2009]).

Taking the perspective of an investor that is focused on the $k-$period return and has information up to time T (denote the information set D_T), the question of whether or not stocks are more attractive in the long run is answered by calculating the predictive variance per period of

the random variable

$$r_{T,T+k} = r_{T+1} + r_{T+2} + \cdots + r_{T+k},$$

denoted by $\text{var}(r_{T,T+k}|D_T)$. If

$$\frac{\text{var}(r_{T,T+k}|D_T)}{k} < \text{var}(r_{T,T+1}|D_T), \tag{3}$$

for k corresponding to relevant investment horizons, and given that $\text{E}(r_{T,T+k}|D_T)$ grows approximately linearly in k, investors would be more attracted to stocks in the long run as the Sharpe-ratio (a measure of risk adjusted return)

$$\frac{\text{E}(r_{T,T+k}|D_T)}{\sqrt{\text{var}(r_{T,T+k}|D_T)}} \tag{4}$$

would grow with the horizon k.

This paper's main goal, from this point forward, is to evaluate $\text{var}(r_{T,T+k}|D_T)$ under alternative model specifications and different prior assumptions in order to check whether the inequality in (3) holds. This evaluation is done with respect to the investor's joint posterior distribution of all states and parameters of the model so to incorporate all sources of uncertainty. Note that, unlike the three papers referenced in the introduction, our initial model does not rely on any predictor for μ_t and looks to model the univariate series of returns. It is our view that understanding the time series patterns of returns and expected returns is a necessary first step in studying the main quantities and assumptions driving the resulting predictive variance. If one can show that the inequality in (3) holds without the use of predictors it follows that conditioning on any additional information capable of predicting expected returns can only reduce the resulting predictive variance.

To help facilitate the understanding of the main quantity of interest it is useful to look at a simpler version of the model in (1) where $\mu_t = \mu$ (for all t) so that returns are i.i.d. normal with mean μ and variance σ_u^2. This simple assumption is associated with stock prices following a random walk. By denoting $\theta = (\mu, \sigma_u^2)$ the predictive variance is computed via

$$\begin{aligned}
\text{var}(r_{T,T+k}|D_T) &= \text{E}_\theta\left[\text{var}(r_{T,T+k}|D_T,\theta)\right] + \text{var}_\theta\left[\text{E}(r_{T,T+k}|D_T,\theta)\right] \\
&= \text{E}_\theta(k\sigma_u^2) + \text{var}_\theta(k\mu) \\
&= k\text{E}_\theta(\sigma_u^2) + k^2\text{var}_\theta(\mu)
\end{aligned} \tag{5}$$

so that $(1/k)\text{var}(r_{T,T+k}|D_T)$ increases with the horizon k and, in turn, falsifies the inequality in (3).

Applying the same logic to the time-varying model in (1), Pástor and Stambaugh [2012] decompose the predictive variance into 5 components and observed that $\rho_{uw} < 0$ is a necessary condition for the inequality in (3) to hold (see Appendix 1). This decomposition is very useful in understanding the sources of uncertainty faced by the investor and we explore this notion in the empirical examples that follow. In the meantime, we can summarize the main variables that affect the predictive variance: β and ρ_{uw}. Simply put, β close to 1 means "momentum", i.e., abnormally large expected returns today lead to likely abnormally large expected returns tomorrow and this high persistence leads to very high unconditional variance for μ_t which in turn, leads to quick growth in the predictive variance of k-period returns. A strong negative ρ_{uw} offsets the "momentum" effect as abnormally large expected returns today are associated with negative shocks to observed returns tomorrow, mitigating the growth of the predictive variance with the horizon. In essence, the answer to the question of whether or not stocks are good for the long run is a function of the investor's posterior distribution on β and ρ_{uw} and our main objective in the first part of the paper is to understand how much information in the data is available about these two quantities.

2.1 Priors and Data

Our analysis focus on the annual data used in Pástor and Stambaugh [2012] consisting of observations from 1802 to 2007 as compiled by Siegel [2008]. The returns are annual real (excess) log returns from the U.S. equity markets. The evaluation of the predictive variance from the investor's perspective can only be achieved by integration over the posterior distribution of all unknowns. The problem in hand is a notoriously low-signal environment in which carefully chosen priors play a central role in the analysis. Our objective here is to compute our target, the predictive variance $\text{var}(r_{T,T+k}|D_T)$, by working with a few different prior specifications that reasonably represent the investor's beliefs. For robustness of analysis, we attempt to be as "non-informative" as possible even when we believe this may represent an overstatement of the uncertainty faced by the investor. Using the same rationale, we assume the investor does not observe any potential predictive variable for returns. This is a simplification to the model used in Pástor and Stambaugh [2012] that reduces the investor's information set and can only bring more variance for the long run. As mentioned above, this provides a more clean framework in

which to evaluate the role of different parameters in determining the long run predictive variance.

The basis for our analyses is the model defined by (1) and (2). Throughout we define uninformative priors for the initial state μ_0 and for both α and β while respecting the constraint that $\beta \in (0,1)$ (see draws in Figure 1). As for $\mathbf{\Sigma}$, we use a Cholesky prior that results in a very uninformative prior for σ_u^2 but somewhat informative for σ_w^2. This is due to the understanding that we are dealing with a very low signal-to-noise ratio time series and the prior consensus belief that expected returns are significantly less volatile than actual returns.

To be specific, we follow the Cholesky representation in Daniels and Pourahmadi [2002] and rewrite the joint distribution of $(u_{t+1}, w_{t+1})'$, as

$$\begin{aligned} u_{t+1} &= \sigma_u z_1 \\ w_{t+1} &= \phi_{wu} u_{t+1} + \sigma_w z_2 \end{aligned}$$

where z_1 and z_2 are $N(0,1)$. Under this parametrization, $\rho_{uw} = \phi_{wu}\sigma_u\sigma_w^{-1}$ so we can easily enforce $\rho_{uw} \in (-1,0)$ by the use of a truncated normal prior for ϕ_{wu}. The results presented in the next Section are based on two implied priors for ρ_{uw}. In what we call "weak prior" regime, ρ_{uw} is approximate uniformly distributed between -1 and 0 while the "strong prior" regime refers to a prior where ρ_{uw} is concentrated around -0.8 (a strong but yet plausible belief in the literature see Barberis [2000], Pástor and Stambaugh [2012]). In both cases we use the priors

$$\sigma_u^{-2} \sim \text{Ga}(5, 0.15) \quad \text{and} \quad \sigma_w^{-2} \sim \text{Ga}(5, 0.002)$$

with $\phi_{wu} \sim N(-0.016, 0.07^2)1_{\phi_{wu}<0}$ and $\phi_{wu} \sim N(-0.097, 0.07^2)1_{\phi_{wu}<0}$ in the "weak" and "strong" prior specifications respectively. Draws from the priors of β and Σ in both specifications are presented in Figures 1 and 2.

3 Results

Figure 1 presents draws from priors and posteriors for the mean reversion coefficient β and the standard deviations of the innovations (σ_u,σ_w) under three model/prior specifications. The first row of plots refers to a model where $\rho_{uw} = 0$ whereas the second and third rows are the results for the "weak prior" and "strong prior" specifications, respectively. Notice that in all cases the information in the data about σ_u and σ_w are very robust validating our initial beliefs about the

signal-to-noise ratio in the problem. As for β, the story is quite different as very little information is available in the data so that posterior learning is very weak. However, a few things can be noted: first, by forcing $\rho_{uw} = 0$, β is inferred to be much smaller than when we allow $\rho_{uw} < 0$. Second, in all cases, there is little posterior evidence for $\beta > 0.9$. As we show below, this has a great impact in the resulting computation of the predictive variance.

Figure 2 shows prior and posterior draws for ρ_{uw} in both "weak prior" (left panel) and "strong prior" (right panel) setup. In both cases the data is suggestive of a somewhat strong negative correlation between the error terms. As noted before, whether or not stocks are less volatile and more appealing to long horizon investors depends, in one direction, on how large β is and, how negative ρ_{uw} is the other. The results from these analysis show that β is unlikely to be very high and that ρ_{uw} is probably smaller than -0.5.

Turning to our main quantity of interest, Figure 3 presents the resulting predictive standard deviation per period, i.e.,

$$\sqrt{\frac{\text{var}(r_{T,T+k}|D_T)}{k}}$$

in all model specifications presented so far plus the addition of the a few benchmarks and a time varying Σ_t specification to be described later (see Section 4). Two important benchmarks are the i.i.d. model (black line) and the standard state-space model where $\rho_{uw} = 0$; The former can be seen as an upper bound for the predictive variance as it is associated with the a random walk for stock prices. The latter is the simplest, standard state-space model that does not take advantage of the economic motivated prior that $\rho_{uw} < 0$. Not surprisingly, the results for these models are quite similar to each other. By ignoring the important economic fact that innovations to expected returns and innovations in returns are negatively correlated, the noise level of returns makes it very hard to filter out the path of expected returns and therefore β is inferred to be quite small. In other words, without the important prior information that $\rho_{uw} < 0$ the inferences in the state-space model indicate the 207 observations available look essentially like independent normal draws and therefore the predictive variance (red line) grows with the horizon as in the derivation in equation (5). Now turning to the results where $\rho_{uw} < 0$, both the "weak prior" and "strong prior" specification lead to the conclusion that stocks are attractive for the long run investor as their predictive variance per period decays (up to a point) as a function of the horizon. We have seen that in posterior for β is very similar in both cases so the difference in the results is solely due to the prior on ρ_{uw} – the stronger the investor's belief about the negative correlation the more attractive stocks are for his/hers long run portfolio. It is also important

to notice that both the blue and green curves eventually start to climb up (for horizons of 30 years or more) as future uncertainty about expected returns eventually outpaces the effects of ρ_{uw}. However, even looking at 50 year horizons, the inequality in (3) still holds.

A final point to notice out of the right panel in Figure 3 is the grey line labeled "unconditional"; this is a model-free assessment of the volatility per period at different horizons showing the in-sample variance of 1-year, 2-years, 3-years returns and so on. This is computed based on a rolling sum of k-period returns. This line is often used as the justification for the conventional wisdom (see Siegel [2008]) that long run stock portfolios are relatively less risky than short horizon ones. This line is obviously problematic as it ignores many sources of uncertainty, in particular as the horizon grows (the sample size for its computation decreases with horizon). However, it is also very reassuring to see that at least looking at 10 or 20 years horizons this model-free assessment is in qualitative agreement with the results from the blue and green line. Once all the uncertainty about parameters are taking into account we see a shift up from the simplistic unconditional line however, in all settings where the economic information about $\rho_{uw} < 0$ is used, the resulting predictive variance line still decreases with horizon and the inequality in (3) holds.

To check that these results are not an artifact of the model specification that forces ρ_{uw} to be negative, the left panel in Figure 3 shows the prior predictive variance per period plotted along with the posterior predictive in the "strong prior" case. The result is clear: under this model/prior specification the investor's initial belief is that stocks are very bad for long run portfolios but after observing 207 years of data, he/she concludes the opposite – so we conclude that no, these results are not an artifact of the model/prior specification but rather a function of the information in the data. Additionally, we randomly shuffled the 207 observations 1000 times. This is an attempt to to break its dynamics and turn the data into approximately i.i.d. draws. The left panel of Figure 4 shows a histogram of draws from the posterior of ρ_{uw} using data from one of the shuffled data sets. As expected, the posterior evidence is that ρ_{uw} is likely to be close to 0 (contrast with Figure 2). The right panel shows the different predictive variances (grey area) obtained by this exercise under the "weak" prior setting. It is clear that the results obtained from the "weak" and "strong" prior settings are indeed a function of the dynamic patterns in the actual, observed data.

The analyses above give us enough evidence that the conventional wisdom behind larger allocations in stocks for long horizon investors is not so wrong after all and puts the results from

Pástor and Stambaugh [2012] into question. Using what is a very simple but representative model for returns we see that the inequality in (3) holds even for investors that have very uninformative priors about the crucial parameters in the model. Our view is that the results in Figure 3 are enough to validate the notion that stocks are safer for the long run but a deeper understanding of the trade-offs associated with the assumptions is necessary and explored further below.

Figure 5 uses the decomposition of the predictive variance presented in Pástor and Stambaugh [2012] (see Appendix 1) to help us understand the relevance and magnitude of the different sources of uncertainty in determining the predictive variance. Again, the only source of reduction of the predictive variance as a function of the horizon is the component labeled "mean reversion", i.e., the effect associated with ρ_{uw} whereas the uncertainty about "future μ" is the main driver bringing the predictive variance up. Recall that μ is modeled as a mean-reverting auto-regressive quantity and therefore its future uncertainty in primarily determined by the mean reverting coefficient β (see Appendix 1). The closer β is to one, the larger the uncertainty about future μ's will be[1]. By comparing the two panels we can see that the difference between the results in the "weak prior" (left) versus the "strong prior" (right) set up is the "mean reversion" component – the stronger negative beliefs about ρ_{uw} result in a lower mean reversion component leading to a reduction in the long run predictive variance.

Second, we investigate the effects of β. One particular interesting aspect of the results so far is the lack of information in the data about β and therefore the investor's prior about this quantity will be particularly influential in the computation of the predictive variance. To further our understanding about the role of β, Figure 6 shows the predictive volatility per period in the "weak prior" set up for a variety of fixed values of β. The plot on the left emphasizes the notion mentioned before that higher β's lead to higher predictive variance. In particular, we see that if an investor believes in very persistent expected returns (say $\beta > 0.94$) stocks are not attractive for the long run as the inequality in (3) doesn't hold. On the other hand, the panel on the right shows that fast mean-reversion also makes stocks less attractive for long horizons. Increasing β from 0.1 to 0.85 makes stocks more and more attractive but at some point (see the red line where $\beta = 0.875$) the high persistency starts increasing the long run volatility. From this analysis, one can conclude that if $0.1 < \beta < 0.93$ conventional wisdom is right. The analysis so far has shown that even with very little information available in the data, the posteriors for β in Figure 1 are concentrated in this range. In simple terms, it's hard to learn about β but we learn that β is not

[1]In the limit of $\beta = 1$, μ turns into a random-walk which would lead to a diverting uncertainty about future μ.

too high nor too small which in turn provides additional evidence that stocks are indeed good for long run portfolios.

To explore this point further Figure 7 displays (left panel) the predictive distribution of μ_{T+30} when $\beta = 0.945$ (dashed line) and when β is free to vary under the "weak prior" set up. It also shows the resulting predictive variance decomposition when $\beta = 0.945$. It is clear that the reason for the growth of the predictive variance as a function of the horizon is the resulting uncertainty about future μ. Using the non-informative setting of the "weak prior" regime, a investor faces a 95% predictive interval for expected returns in $T+30$, between -3% and 15%whereas if $\beta = 0.945$ the interval is significantly wider: -6.5% to 19%. One could argue that either case results in too wide of an interval. As mentioned before, we believe that the "weak prior" set up is too vague and most likely overstates the uncertainty faced by the investor, but even in that case, the resulting predictive variance per period decreases with the horizon.

It is important to note that our results are in direct contradiction with the strong claims in Pástor and Stambaugh [2012]. In this basic model, without predictors, the only different between their approach and ours is the prior distribution over Σ. We believe the Cholesky prior used here is a natural, interpretable and flexible choice. Nonetheless, in an attempt to better understand the nature of the differences in our results if compare to Pástor and Stambaugh [2012] we re-ran the analysis using the "non", "less" and "more" informative inverse-Wishart priors with hyper parameters described in a series of their papers. Figure 8 and 9 present the results obtained from both theirs and our prior specifications. The "non" informative specification does not impose the information that $\rho_{uw} < 0$ and therefore the noise overwhelms the signal leading to results similar to the standard dynamic model in our analysis where $\rho_{uw} = 0$. The other specifications, "less" and "more" informative, impose the negative correlation and the results are in qualitative agreement with our results presented so far. Once again, unless the investor possesses very unusual views a priori, or ignores the fundamental economic fact that $\rho_{uw} < 0$, the data consistently points in the direction that stocks are more attractive to long run portfolios.

In summary, the past 207 years of data provides enough evidence that the inequality in (3) holds and therefore the conventional wisdom for larger allocation in stocks for long run portfolios is justified. In order to violate (3) an investor has to hold very strong beliefs about the mean reversion coefficient β – beliefs that (i) are not supported by the data and (ii) imply an artificial amount of uncertainty regarding future expected returns as seen in Figure 7.

3.1 Portfolio Implications

To illustrate the implications of the different results obtained so far we present a stylized portfolio problem. Assume a buy-and-hold investor that can only choose between the stock index or a risk free bond. The investor's initial wealth is $W_T = 1$ and ω is the allocation to stocks so that the terminal wealth at horizon H is defined by:

$$W_{T+H} = (1-\omega)\exp(r_f H) + \omega \exp(r_f H + r_{T,T+H})$$

where r_f is the risk-free rate. The investor's preferences over terminal wealth is defined as in Barberis [2000] and Pástor and Stambaugh [2012] by a constant relative risk-aversion power utility function of the form

$$u(W) = \frac{W^{1-\delta}}{1-\delta}$$

so that when looking to determine its allocation to stocks the investor solves the following problem:

$$\max_{\omega} E\left\{\frac{\left[(1-\omega)\exp(r_f H) + \omega\exp(r_f H + r_{T,T+H})\right]^{1-\delta}}{1-\delta}\right\} \tag{6}$$

where the expectation is taken with respect to the predictive distribution of $r_{T,T+H}$.

In our example we set $r_f = 0.02$ (2% a year) and $\delta = 5$ (without loss of generality and following the examples in both Barberis [2000] and Pástor and Stambaugh [2012]). Using simulated draws from the posterior of all unknowns of each model we are able to evaluate the expectation in (6) via Monte Carlo and consequently determine the investor's optimal allocation. Figure 10 shows the allocations as a function of the horizon under 3 different model specifications. The results are clear: a 30-year horizon investor that believes stock returns are i.i.d. will hold roughly 50% of their portfolio in stocks while the investors that uses the model in (1) with beliefs that $\beta = 0.945$ will hold a little less than 50% in stocks. Meanwhile, the 30-year investor with beliefs described by our "strong prior" setting will hold more than 70% is stocks, a result very much in line with the current conventional wisdom with benefits well cataloged in Siegel [2008].

4 Model Extensions

In this section we extend the analysis conducted so far in a two directions. First we add predictive variables to the analysis using the predictive systems framework of Pástor and Stambaugh [2012].

The addition of conditional information can only help reduce the predictive variance and therefore we are simply trying to evaluate how much this information helps reduce the variability of stocks in the long run. It is important to notice that the conclusions in Pástor and Stambaugh [2012] that stocks are more volatile into long run are based on a model with predictors, a conclusion we were able to refute without this additional information.

The second extension is allowing the covariance matrix to vary in time. The fact that volatility of stocks returns changes in time has been extensively documented in the literature and we seek to understand how this feature affects the predictive variance of long run portfolios.

4.1 Adding Predictors

Working with a simple yet comprehensive model, we have concluded that the predictive variance per period of a k-period return decreases with the horizon. In other words, the inequality in (3) holds even if an investor has very vague initial beliefs. We now extend the analysis to incorporate predictors of expected returns. There is a vast literature on return predictability (see for example Welch and Goyal [2008], Johannes et al. [2014], Pettenuzzo and Ravazzolo [2014]) but to maintain the coherence of the analysis we follow the predictive systems approach as in Pástor and Stambaugh [2012]. This is done through the following state-space model:

$$\begin{aligned} r_{t+1} &= \mu_t + u_{t+1} \\ \mu_{t+1} &= \alpha + \beta\mu_t + w_{t+1} \\ \mathbf{x}_{t+1} &= \mathbf{A} + \mathbf{B}\mathbf{x}_t + \mathbf{v}_{t+1} \end{aligned} \tag{7}$$

with

$$\begin{pmatrix} u_{t+1} \\ w_{t+1} \\ \mathbf{v}_{t+1} \end{pmatrix} \sim N(0, \mathbf{\Sigma}) \quad \text{and} \quad \mathbf{\Sigma} = \begin{pmatrix} \sigma_u^2 & \sigma_{uw} & \mathbf{\Sigma}_{ux} \\ \sigma_{wu} & \sigma_w^2 & \mathbf{\Sigma}_{wx} \\ \mathbf{\Sigma}_{xu} & \mathbf{\Sigma}_{xw} & \mathbf{\Sigma}_x \end{pmatrix},$$

where $\mathbf{x}_{t+1}$ is a vector of variables (predictors) related to expected and realized returns. This follows the specification of Pástor and Stambaugh [2012] and provides a framework where investors use the information in the variables $\mathbf{x}$ to learn about expected returns (and returns) via the model in (7). Pástor and Stambaugh [2012] call this a "imperfect predictors" framework as, unlike most of the predictability literature, expected returns are not solely a linear function of the $\mathbf{x}$'s but a latent variable (μ) that is informed by the predictors via Σ. As in the models

in Section 2, we assume that $\beta \in (0,1)$, $\rho_{uw} < 0$ and that $\mathbf{B}$ is a diagonal matrix with entries $b_{ii} \in (-1,1)$ for all i so to guarantee stationarity.

In our empirical analysis, we use the exact same predictors as Pástor and Stambaugh [2012]. They are: dividend yield on U.S. equity, a measure of bond yields and a measure of term spread as defined in Pástor and Stambaugh [2012]. We also extend the Cholesky prior to the now (5×5) covariance matrix $\mathbf{\Sigma}$ with the same implied priors for σ_u^2, σ_w^2 and ρ_{uw} as defined in the "weak prior" and "strong prior" specification in our previous analysis. In addition, a very uninformative prior is implied for the remaining elements of $\mathbf{\Sigma}$.

Figure 11 shows the results for the predictive variance in both "weak" and "strong" prior specifications and compares the results with and without the inclusion of the predictors $\mathbf{x}$. The results are very similar; in the "weak" prior setting the predictors provide some information and are able to slightly reduce the the predictive variance of portfolios with horizon longer than 20 years. However, when the "strong" priors are used the predictors appear to be irrelevant as the resulting long run predictive variances are essentially the same. This fact should come with no surprise; these predictors, while extensive explored and discussed in the literate are able to capture a very small fraction of the variation in stock returns and many authors (see Welch and Goyal [2008]) have pointed to the fact that taking advantage of these signals out-of-sample is a very hard task.

Once again, these results add evidence to our original conclusions that under mild assumptions about β and ρ_{uw}, conventional wisdom is correct and the inequality in (3) is satisfied, making stocks very attractive to the long run investor.

4.2 Time-Varying Σ_t

We now modify the model in (7) with

$$\begin{pmatrix} u_t \\ w_t \\ \mathbf{v}_t \end{pmatrix} \sim N(0, \mathbf{\Sigma}_t) \tag{8}$$

and follow the specification of Lopes et al. [2014] where the joint distribution of $(u_t, w_t, \mathbf{v}_t)'$, is written as a series of regressions

$$\begin{aligned}
u_t &= \sigma_{tu} Z_{t1} \qquad (9)\\
w_t &= \phi_{twu} u_t + \sigma_{tw} Z_{t2}\\
v_{t1} &= \phi_{t11} u_t + \phi_{t12} w_t + \sigma_{t1} Z_{t3}\\
v_{t2} &= \phi_{t21} u_t + \phi_{t22} w_t + \phi_{t23} v_{t1} + \sigma_{t2} Z_{t4}\\
v_{t3} &= \phi_{t31} u_t + \phi_{t32} w_t + \phi_{t33} v_{t1} + \phi_{t34} v_{t2} + \sigma_{t3} Z_{t5}\\
&\dots\\
v_{tp} &= \phi_{tp1} u_t + \phi_{tp2} w_t + \phi_{tp3} v_{t1} + \ldots + \phi_{tp(p+1)} v_{t(p-1)} + \sigma_{tp} Z_{t(p+2)}
\end{aligned}$$

where all of the Z_{ti} are iid $N(0,1)$. This model provides a one-to-one correspondence between the parameters $(\sigma_{tu}, \phi_{twu}, \sigma_{tw}, \{\phi_{tij}\}, \{\sigma_{ti}\})$ and $\mathbf{\Sigma}_t$ and is a straightforward vehicle to include time variation by modeling each row of the system in (9) independently, while guaranteeing a proper variance covariance matrix as a result. We use the standard stochastic volatility approach and let

$$\sigma_{t,i} = \exp\left(\frac{s_{t,i}}{2}\right), \text{ with } s_{t,i} = d_i + g_i s_{t-1,i} + h_i \epsilon_{t,i}.$$

where the ϵ are independent standard normal. Similarly, for each (ij) we model

$$\phi_{t,ij} = d_{ij} + g_{ij} \phi_{t-1,ij} + h_{ij} \epsilon_{t,ij}.$$

Assuming the use of standard priors, posterior sampling for the modification in (9) relies on well studied dynamic linear model [West and Harrison, 1997, Prado and West, 2010] and stochastic volatility techniques [Jacquier et al., 2002, Chib et al., 2002]. The complication arises from the fact that in our model we have prior information about the sign of $\rho_{t,uw}$, the correlation between u_t and w_t. In order to respect this constraint, we develop a new approach to prior specification and posterior computation that allow for easy of expression beliefs about $\rho_{t,uw}$. First, we let

$$\sigma_{t,u} = \exp\left(\frac{\theta_{t,1}}{2}\right), \ \sigma_{t,w} = \exp\left(\frac{\theta_{t,2}}{2}\right), \ \phi_{twu} = \theta_{t,3},$$

so that at time t, the three key time-varying parameters of interest are $\theta_t = (\theta_{t,1}, \theta_{t,2}, \theta_{t,3})$.

Note that with our system of regressions setup (9), we can analyze the first two equations

separately from the others so that we can focus on the two equation system:

$$(u_t, w_t)' \sim N(0, \Sigma(\theta_t)), \quad \theta_t = (\theta_{t,1}, \theta_{t,2}, \theta_{t,3})$$

$$\begin{aligned} u_t &= \exp\left(\frac{\theta_{t,1}}{2}\right) Z_{t1} \\ w_t &= \theta_{t,3}\, u_t + \exp\left(\frac{\theta_{t,2}}{2}\right) Z_{t2} \end{aligned}$$

Viewing the above as our observation equation (in a conditional sense, within the Gibbs sampler), we now need to complete our model with a state equation. We need to do this so that we can control the resulting prior on $\rho_{t,uw}$. Note that

$$\rho_{t,uw} = \rho(\theta_t) = \rho(\theta_{t,1}, \theta_{t,2}, \theta_{t,3}) = \frac{\theta_{t,3}}{\left\{\theta_{t,3}^2 + \exp(\theta_{t,2} - \theta_{t,1})\right\}^{1/2}}.$$

We could proceed by specifying dynamics for each $\theta_{t,i}$ so that the resulting $\rho_{t,uw}$ is in accordance with the our prior beliefs. However, this is complicated due to the non-linear relationship between ρ_t and the θ_t's. Instead, we modified the standard AR(1) specification to directly inject beliefs about the correlation at each t. The standard state equation would be given by:

$$q(\theta_t \mid \theta_{t-1}) = \Pi_{i=1}^3\, q(\theta_{t,i} \mid \theta_{t-1,i}),$$

with each $q(\theta_{ti} \mid \theta_{t-1,i})$ an AR(1)

$$\theta_{t,i} = a_i + b_i \theta_{(t-1),i} + c_i W_{i,t}.$$

where each W is $N(0,1)$. We modify q to construct a transition distribution $p(\theta_t \,|\theta_{t-1})$ which includes direct beliefs about each θ_t via a function $f(\cdot)$:

$$p(\theta_t \,|\theta_{t-1}) \propto q(\theta_t \mid \theta_{t-1})\; f(\theta_t), \tag{10}$$

with,

$$f(\theta_t) = \exp\left\{\frac{-(\rho(\theta_t) - \bar{\rho})^2}{\kappa}\right\}.$$

The smaller κ is, the more θ_t is pushed towards values such that $\rho(\theta_t) \approx \bar{\rho}$. We can choose priors on (a_i, b_i, c_i) in the usual way by considering the desired smoothness of the state paths.

In addition, we can choose κ and $\bar{\rho}$ to push the θ_t path towards ones have a desired correlation level. If $q(\theta_0)$ was prior choice for the initial state, we can use the same function f to modify it:

$$p(\theta_0) \propto q(\theta_0) f(\theta_0). \tag{11}$$

While this approach makes the prior specification relatively straightforward, it complicates posterior draws since evaluation of the transition distribution involves the normalizing constant in (10). The computation details are presented in Appendix 2.

Returning to our empirical application where $\bar{\rho}$ and κ were chosen to mimic the strong prior setting used in Section 3 we confirm, once again, that the inequality in (3) is satisfied and the predictive variance decays with horizon. The predictive variance results are displayed in Figure 3. We also present in Figure 12 the implied prior for two elements of Σ_t over time, along with their updated posterior where we see a clear indication of time variation in the variance of stock returns but a rather stable behavior in ρ_{uw}. Again, even when we add an extra layer of flexibility in the model and let variances and covariances to vary in time our results confirm that conventional wisdom is right in stocks become more attractive for the typical long run investor.

5 Conclusion

The conclusion from all on the analyses presented are very clear: under mild assumptions about prior beliefs for a representative investor, the predictive variance per period of equity returns decrease with horizon. The results validate the conventional wisdom that stocks are more and more attractive as the investment horizon grows and provide substantial evidence to justify a number of popular retirement investment strategies currently available. Our claim is based primarily on flexible, state-space models that look to learn the time series behavior of returns without the use of conditioning information from predictors. However, our conclusions are robust to the inclusion of predictors and the added complexity of time-varying variances and covariances as seen in Section 4.

Our view is that by working with simple but flexible models we were able to identify in what direction the predictive variance per period moves as a function of the horizon. A more precise estimation of the predictive variance level, however, can be achieved by the use of a number of predictors (beyond the 3 considered here) that have been proposed in the literature and more complex predictive models [Welch and Goyal, 2008, Johannes et al., 2014, Pettenuzzo and

Ravazzolo, 2014]. The ability to use better conditioning information in the estimation of expected returns can only reduce the uncertainty faced by the investor in the long run and strengthen the results presented here. However, as the usual case when dealing with the bias-variance trade-off in larger models one has to be careful in assessing the impact of prior specification. These extensions, particularly the use of non-linear models for expected returns, are the focus of our future work.

Appendix 1: Predictive Variance Decomposition

Let $\Theta = (\alpha, \beta, \Sigma)$ represent all the fixed parameters in the model defined in (1). Using basic properties of expectations and variances, the predictive variance in (3), can be written as:

$$\begin{aligned}
\mathrm{Var}\left(r_{T,T+k}|D_T\right) &= \mathrm{E}\left\{\mathrm{Var}(r_{T,T+k}|\mu_T,\Theta,D_T)\right\} + \mathrm{Var}\left\{\mathrm{E}(r_{T,T+k}|\mu_T,\Theta,D_T)\right\} \quad (12)\\
&= \mathrm{E}\left\{\mathrm{Var}(r_{T,T+k}|\mu_T,\Theta,D_T)\right\} \\
&+ \mathrm{E}\left\{\mathrm{Var}_{\mu_T}(\mathrm{E}(r_{T,T+k}|\mu_T,\Theta,D_T)\right\} + \mathrm{Var}\left\{\mathrm{E}_{\mu_T}\left[\mathrm{E}(r_{T,T+k}|\mu_T,\Theta,D_T)\right]\right\}. \quad (13)
\end{aligned}$$

The first term of the right hand side can new expanded by using the properties of a MA(∞) process as describe in the Appendix section of Pástor and Stambaugh [2012] and it takes the following form:

$$\mathrm{Var}\left(r_{T,T+k}|\mu_T,\Theta,D_T\right) = k\sigma_u^2\left[1 + 2\bar{d}\rho_{uw}A(k) + \bar{d}^2B(k)\right] \quad (14)$$

where

$$\begin{aligned}
A(k) &= 1 + \frac{1}{k}\left(-1 - \beta\frac{1-\beta^{k-1}}{1-\beta}\right) & (15)\\
B(k) &= 1 + \frac{1}{k}\left(-1 - 2\beta\frac{1-\beta^{k-1}}{1-\beta} + \beta^2\frac{1-\beta^{2(k-1)}}{1-\beta^2}\right) & (16)\\
\bar{d}^2 &= \frac{1+\beta}{1-\beta}\frac{R^2}{1-R^2} & (17)\\
R^2 &= \frac{\sigma_w^2}{\sigma_u^2(1-\beta^2)+\sigma_w^2}. & (18)
\end{aligned}$$

The remaining terms follow directly from the forecast function of a state-space model and depend of the posterior mean and variance of μ_T given D_T, m_T and C_T respectively and can be

written as:

$$\begin{aligned}
\mathrm{Var}_{\mu_T}\left\{\mathrm{E}(r_{T,T+k}|\mu_T,\Theta,D_T)\right\} &= \left(\frac{1-\beta^k}{1-\beta}\right)^2 C_T && (19)\\
\mathrm{E}_{\mu_T}\left[\mathrm{E}(r_{T,T+k}|\mu_T,\Theta,D_T)\right] &= k\frac{\alpha}{1-\beta}+\frac{1-\beta^k}{1-\beta}\left(m_T-\frac{\alpha}{1-\beta}\right) && (20)
\end{aligned}$$

Therefore we can decompose the predictive variance into 5 interpretable components following the nomenclature of Pástor and Stambaugh [2012]:

$$\begin{aligned}
\mathrm{Var}\left(r_{T,T+k}|D_T\right) &= \mathrm{E}\left\{k\sigma_u^2|D_T\right\} \quad \text{(i.i.d uncertainty)} && (21)\\
&+\mathrm{E}\left\{k\sigma_u^2 2\bar{d}\rho_{uw}A(k)|D_T\right\} \quad \text{(mean reversion)} && (22)\\
&+\mathrm{E}\left\{k\sigma_u^2\bar{d}^2B(k)|D_T\right\} \quad \text{(future } \mu \text{ uncertainty)} && (23)\\
&+\mathrm{E}\left\{\left(\frac{1-\beta^k}{1-\beta}\right)^2 C_T|D_T\right\} \quad \text{(current } \mu \text{ uncertainty)} && (24)\\
+\mathrm{Var}\left\{k\frac{\alpha}{1-\beta}+\frac{1-\beta^k}{1-\beta}\left(m_T-\frac{\alpha}{1-\beta}\right)|D_T\right\} &\quad \text{(estimation risk)} && (25)
\end{aligned}$$

Each of the terms are evaluated via Monte Carlo taking as inputs the values of m_T, C_T and Θ in each step of the MCMC used for model fitting.

Appendix 2: MCMC details: Time-Varying Σ_t

From Section 4.2, we have the observation equations:

$$(u_t, w_t)' \sim N(0, \Sigma(\theta_t)), \quad \theta_t = (\theta_{t,1}, \theta_{t,2}, \theta_{t,3})$$

$$\begin{aligned}
u_t &= \exp\left(\frac{\theta_{t,1}}{2}\right) Z_{t1}\\
w_t &= \theta_{t,3}\, u_t + \exp\left(\frac{\theta_{t,2}}{2}\right) Z_{t2},
\end{aligned}$$

state equation

$$p(\theta_t\,|\theta_{t-1}) \propto q(\theta_t\,|\,\theta_{t-1})\; f(\theta_t), \tag{26}$$

and initial state prior

$$p(\theta_0) \propto q(\theta_0) f(\theta_0).$$

As discussed in Section 4.2, $q(\theta_t \,|\, \theta_{t-1})$ and $q(\theta_0)$ represent a "standard" specification and the function $f(\theta_t)$ is used to inject prior beliefs about θ_t.

While this approach makes the prior specification relatively straighforward, it complicates posterior draws since evaluation of the transition distribution involves the normalizing constant in 26. Recall that the AR(1) parameters in $q(\theta_{t,i} \,|\, \theta_{t-1,i})$ are denoted by (a_i, b_i, c_i). Let $(a, b, c) = \{(a_i, b_i, c_i), i = 1, 2, 3\}$. Making (a, b, c) explicit in the above, we have:

$$\begin{aligned} p(\theta_t \,|\theta_{t-1}, a, b, c) &\propto q(\theta_t \,|\, \theta_{t-1}, a, b, c)\; f(\theta_t) \\ &= q(\theta_t \,|\, \theta_{t-1}, a, b, c)\; f(\theta_t)\; K(\theta_{t-1}, a, b, c) \end{aligned}$$

where K denotes the normalizing constant. This normalizing constant will be present in draws of the states $\{\theta_t\}$ given (a, b, c) and draws of (a, b, c) given the states using the usual Gibbs sampler approach conditional on everthing else.

Our approach is to discretize each of the three states so that $\theta_{ti} \in G_i = \{g_{i1}, g_{i2}, \ldots, g_{in_i}\}$, $i = 1, 2, 3$, giving a three dimensional grid of possible θ_t vectors. While three dimensional grids are large, we have found that be carefully keeping track of what has been already computed and using parallel computation of K when needed (using openmp in C++) we can get draws in a reasonable amount of time. For each i, we draw the sequence $\{\theta_{ti}\}$ conditional on the other two state sequences and (a, b, c) using the forward filtering backward sampling algorithm (cite()). Notice this is possible because the discretization of the state space enables us to do the forward filtering and backward sampling exactly. Finally, we use a random walk Metropolis to draw (a, b, c) using joint proposals for (a_i, b_i, c_i).

References

N. Barberis. Investing for the long run when returns are predictable. *The Journal of Finance*, 55(1):225–264, 2000.

S. Chib, F. Nardari, and N. Shephard. Markov chain monte carlo methods for stochastic volatility models. *Journal of Econometrics*, 108(2):281–316, 2002.

J. H. Cochrane. *Asset pricing*, volume 41. Princeton university press Princeton, 2005.

M. J. Daniels and M. Pourahmadi. Bayesian analysis of covariance matrices and dynamic models for longitudinal data. *Biometrika*, 89(3):553–566, 2002.

E. Jacquier, N. G. Polson, and P. E. Rossi. Bayesian analysis of stochastic volatility models. *Journal of Business & Economic Statistics*, 20(1):69–87, 2002.

M. Johannes, A. Korteweg, and N. Polson. Sequential learning, predictability, and optimal portfolio returns. *The Journal of Finance*, 69(2):611–644, 2014.

H. Lopes, R. McCulloch, and R. Tsay. Parsimony inducing priors for large scale state-space models. Technical report, 2014.

L. Pástor and R. F. Stambaugh. Predictive systems: Living with imperfect predictors. *The Journal of Finance*, 64(4):1583–1628, 2009.

L. Pástor and R. F. Stambaugh. Are stocks really less volatile in the long run? *The Journal of Finance*, 67(2):431–478, 2012.

D. Pettenuzzo and F. Ravazzolo. Optimal portfolio choice under decision-based model combinations. Technical report, 2014.

D. Pettenuzzo and A. Timmermann. Predictability of stock returns and asset allocation under structural breaks. *Journal of Econometrics*, 164(1):60–78, 2011.

R. Prado and M. West. *Time series: modeling, computation, and inference*. CRC Press, 2010.

J. J. Siegel. *Stocks for the Long Run by Jeremy Siegel*. 2008.

R. F. Stambaugh. Predictive regressions. *Journal of Financial Economics*, 54(3):375–421, 1999.

I. Welch and A. Goyal. A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, 21(4):1455–1508, 2008.

M. West and J. Harrison. *Bayesian forecasting and dynamic models*, volume 18. Springer New York, 1997.

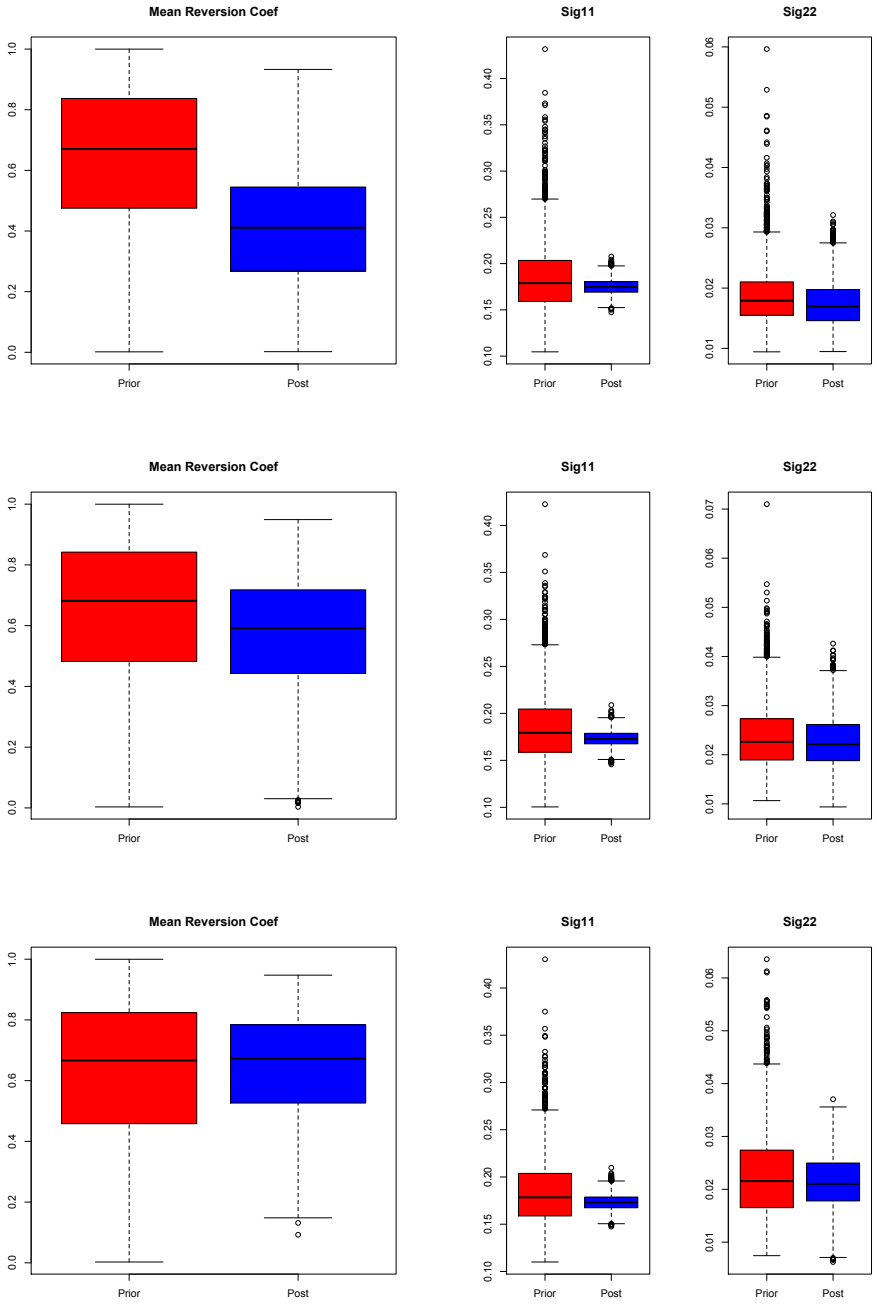

Figure 1: Boxplots of draws from priors (red) and posteriors (blue) for β in the first column and σ_u (sig11) and σ_w (sig22) in the second and third columns. The rows are associated with the $\rho_{uw} = 0$, "weak prior" and "strong prior" specifications respectively. . As a benchmark, the unconditional sample standard deviations of r in this dataset is equal to 0.175, i.e., 17.5% a year.

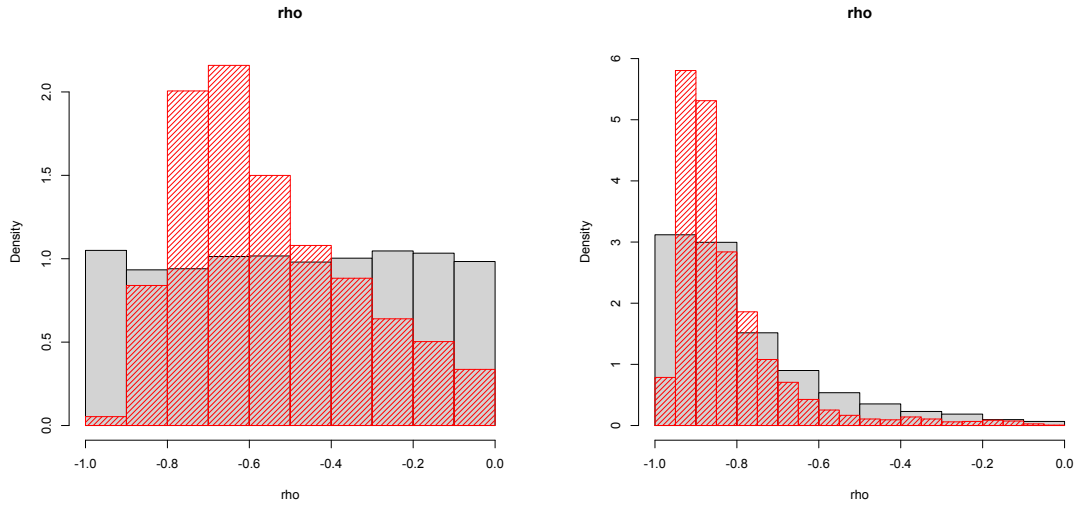

Figure 2: Histograms of draws from priors (gray) and posteriors (red) for ρ_{uw} in the "weak prior" (left) and "strong prior" (right) specifications.

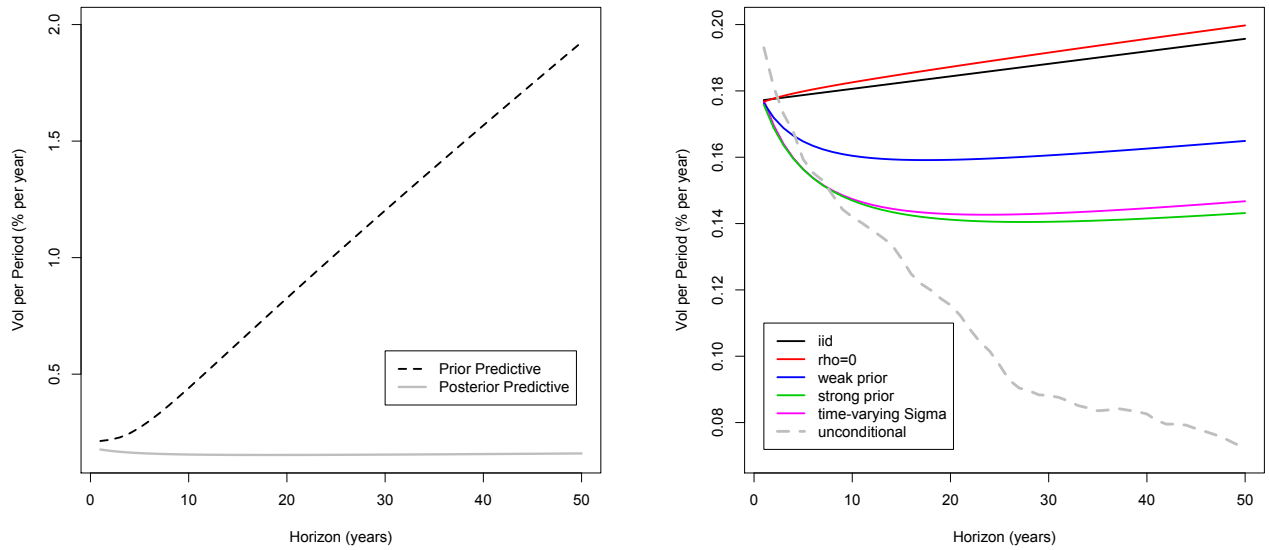

Figure 3: Predictive volatility per period plotted for different horizons, i.e., $\sqrt{\frac{\text{var}(r_{T,T+k}|D_T)}{k}}$. The left panel compares the prior predictive with the posterior predictive in the "strong prior" set up. The right panel compares the different model specifications and benchmarks the results with the i.i.d. case and the model-free unconditional line.

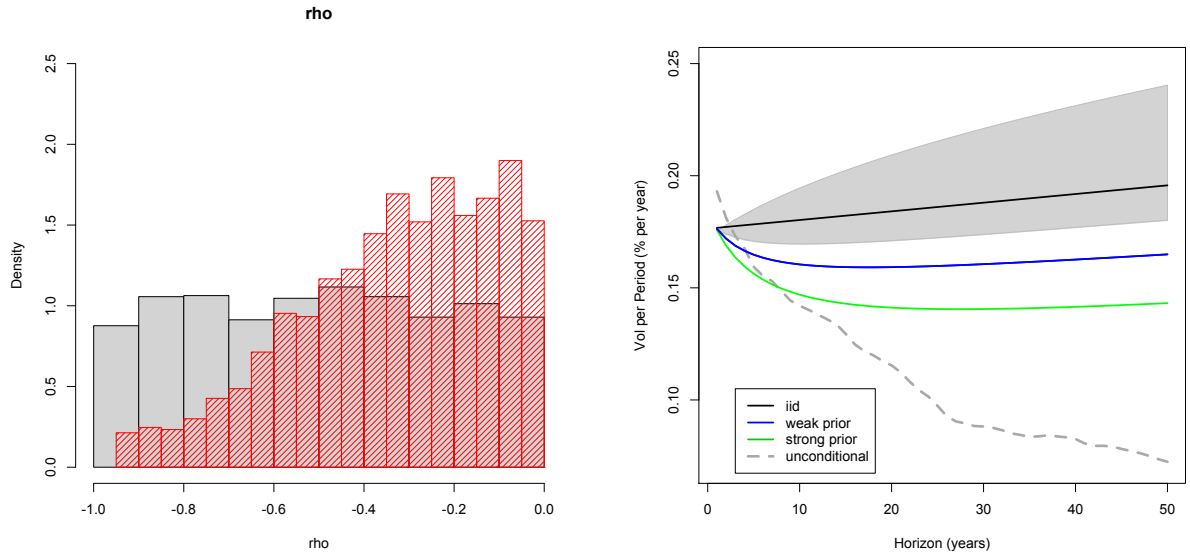

Figure 4: Histograms of draws from priors (gray) and posteriors (red) for ρ_{uw} in the "weak prior" set up where the 207 observations have been reshuffled so to break its dynamics.

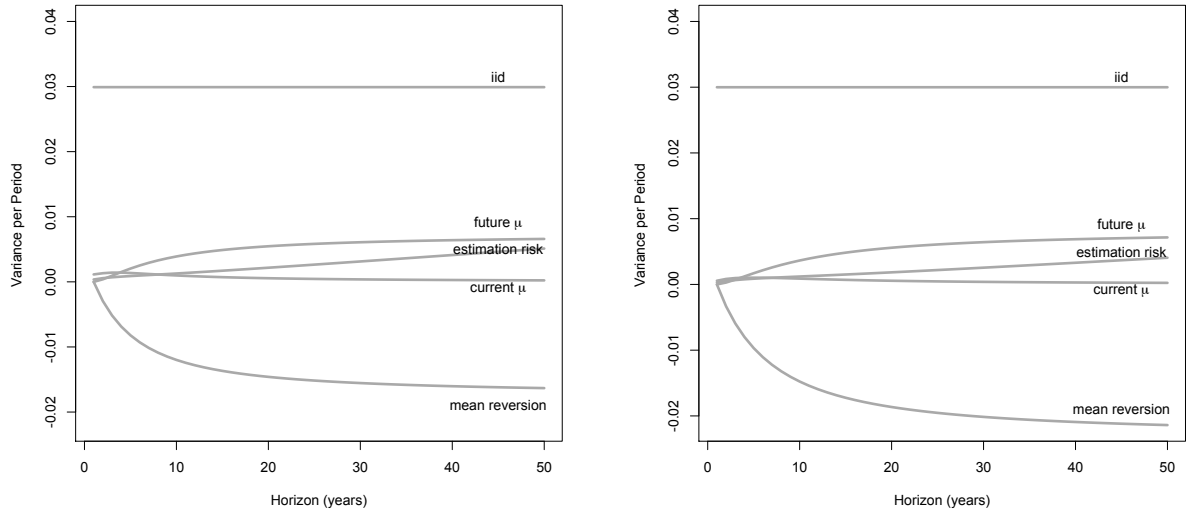

Figure 5: Decomposition of the predictive variance per period. The left panel is the results from the "weak prior" set up while the "strong prior" is in the right panel. The labels (plotted at the same coordinates in both plots to facilitate the comparison) are the same as the decomposition presented in the appendix.

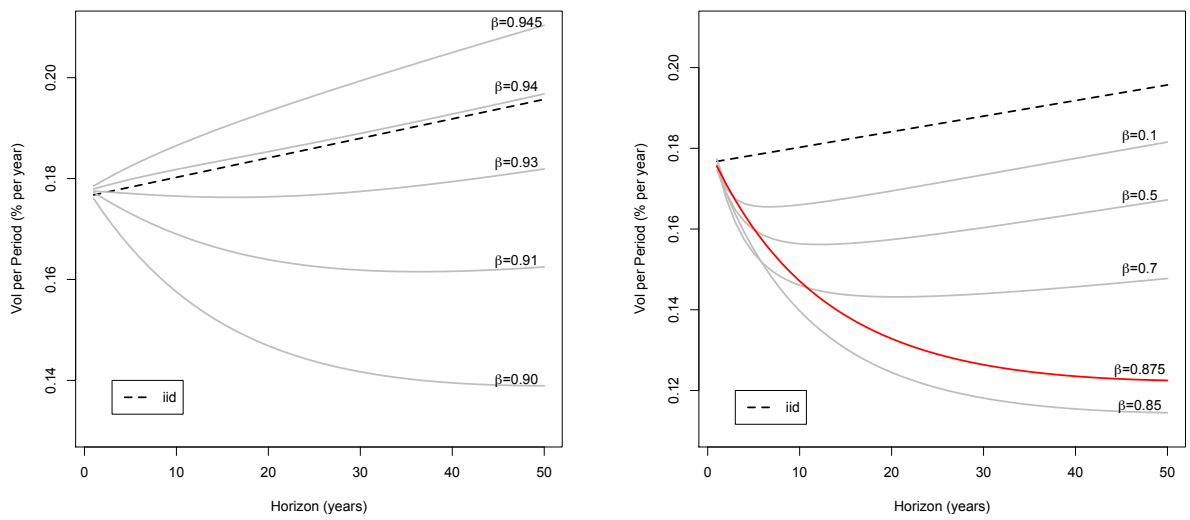

Figure 6: Predictive volatility per period plotted for different horizons (as in Figure 3) for fixed values of β in the "weak prior" set up.

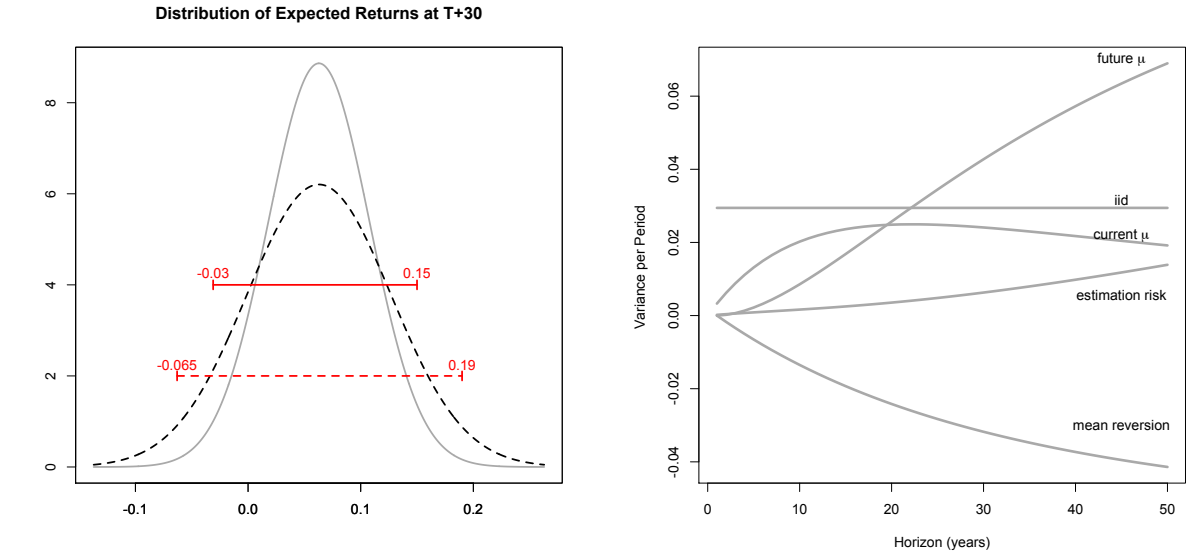

Figure 7: Left panel shows the posterior distribution of μ_{T+30} . The solid line in the right panel results from the "weak" prior set-up while the dashed line fixes $\beta = 0.945$. The right panel shows the decomposition of the predictive variance per period for the case where $\beta = 0.945$.

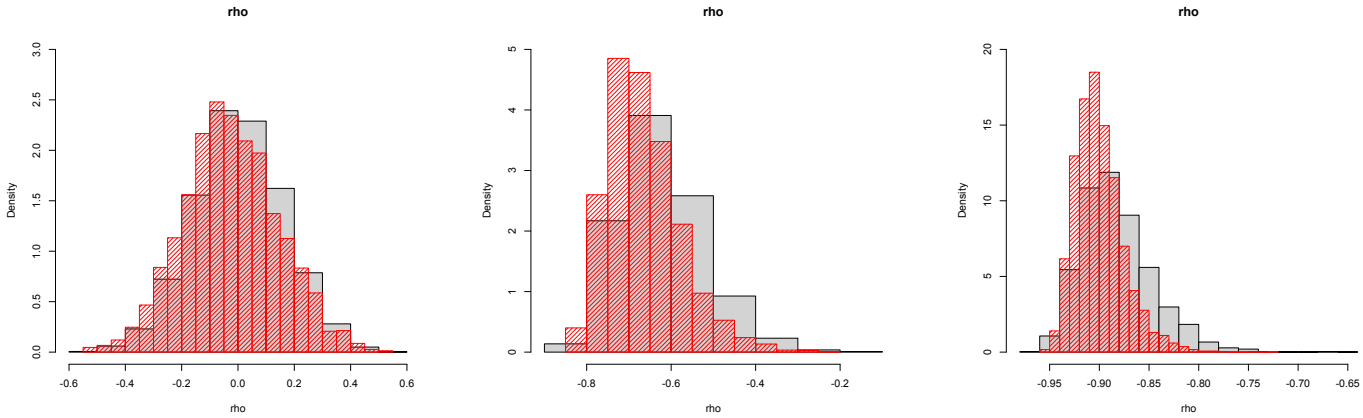

Figure 8: Priors (gray) and posteriors (red) draws of ρ_{uw} using priors from Pastor and Stambaugh 2012. In their terminology, from left to right: "non-informative", "less informative" and "more informative".

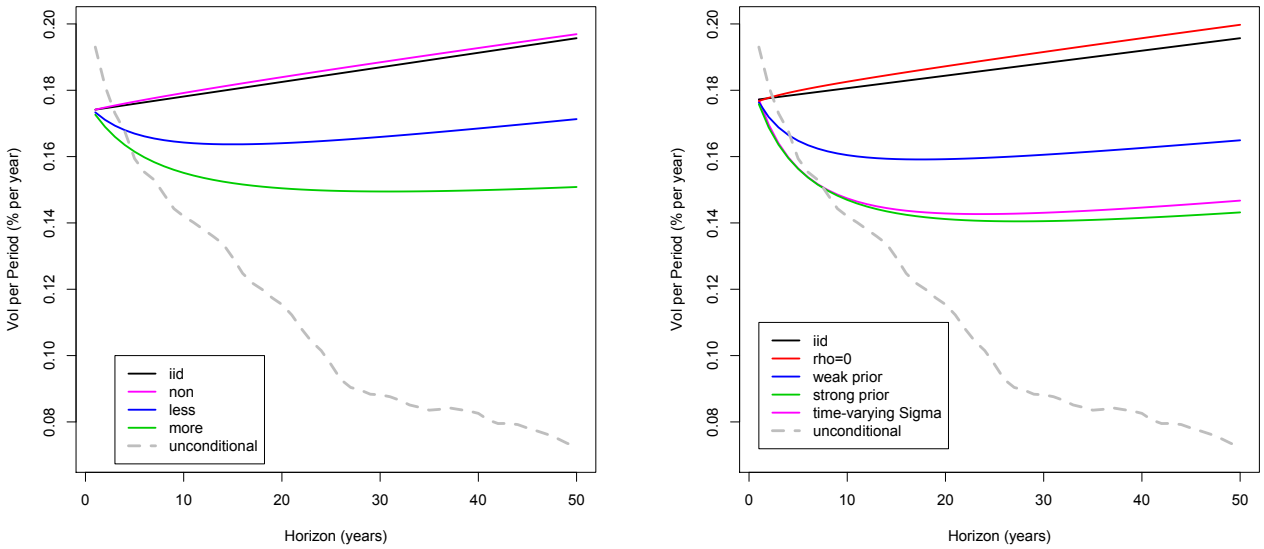

Figure 9: Comparison of our results (right) to the results using the priors in Pastor and Stambaugh 2012 (left).

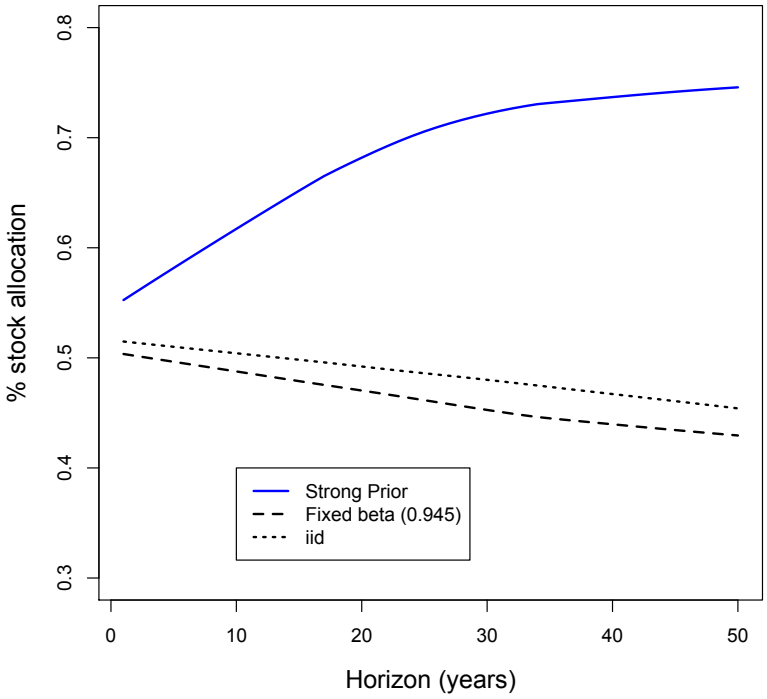

Figure 10: Portfolio Implications: percentage allocated to stocks for different horizon investor under the i.i.d. model and state-space representation under different prior beliefs.

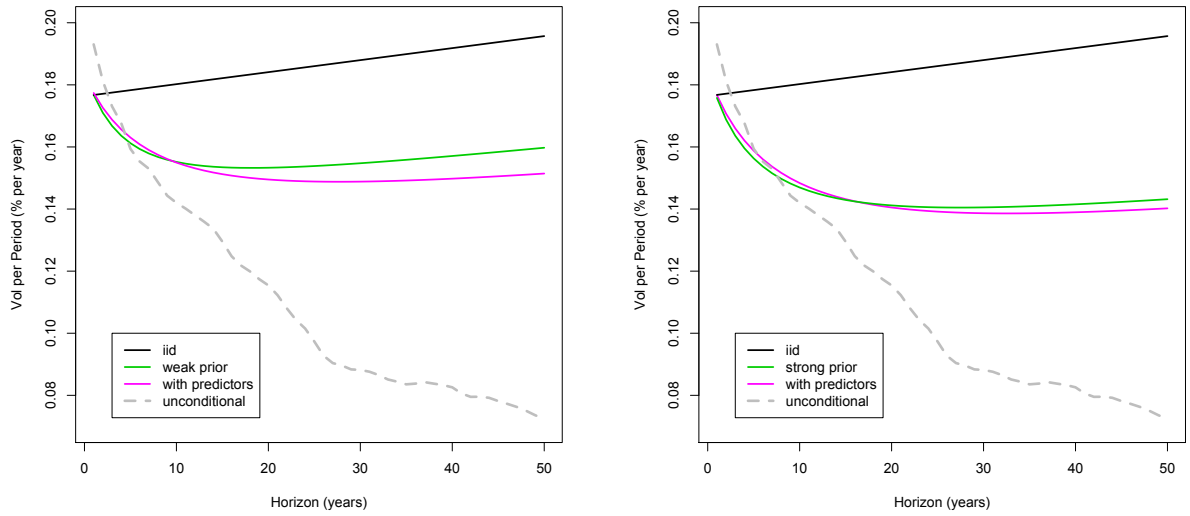

Figure 11: Predictive volatility per period plotted for different horizons when predictors are added. Results are for the "weak prior" (left) and "strong prior" set up.

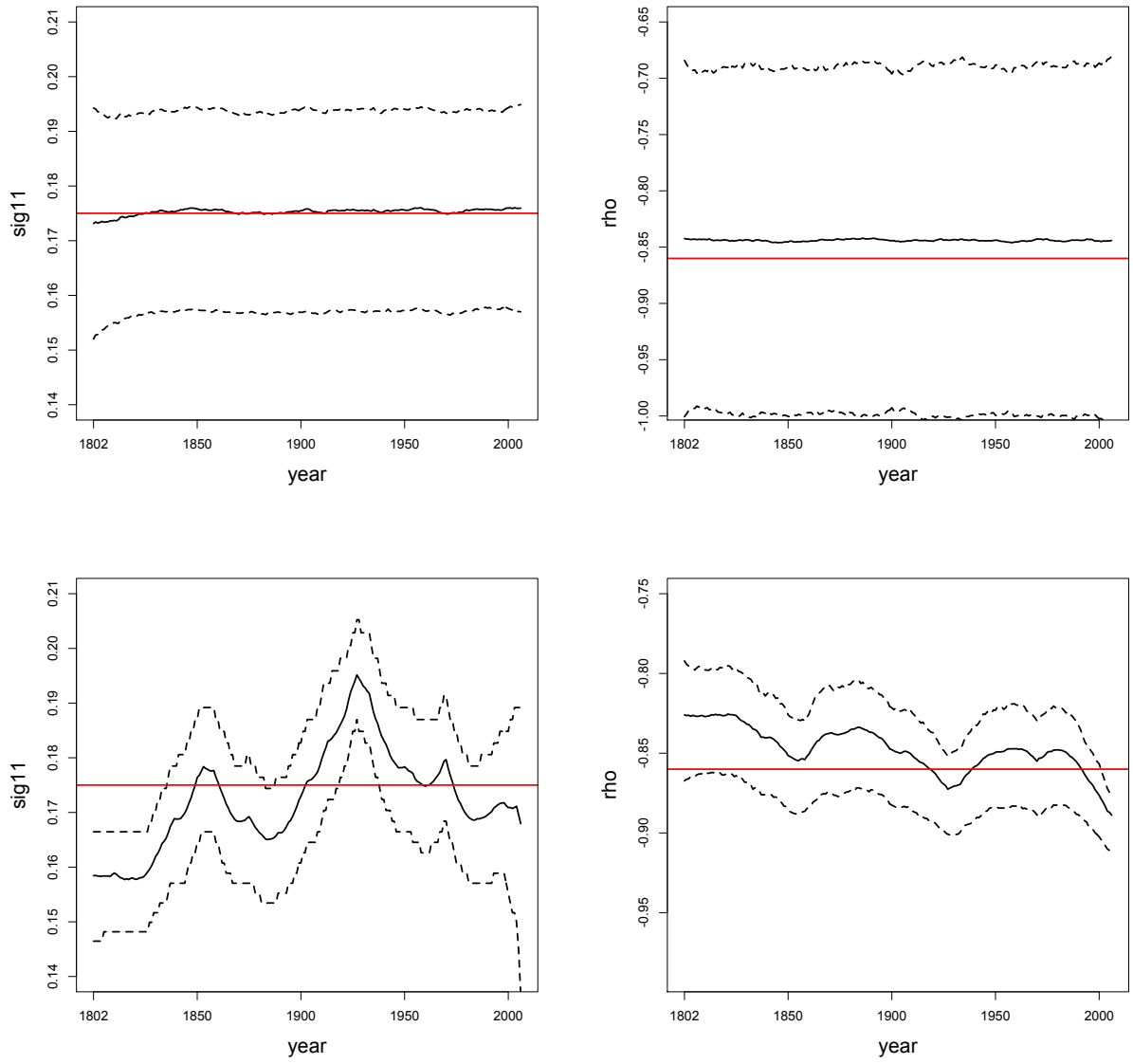

Figure 12: Prior (top row) and posterior summaries of $\sigma_{11,t}$ and ρ_t in the model with time-varying Σ_t. The solid black line in each plot represent the mean of each quantity while the two dashed lines give the 25th and 75th quantiles of the distributions. The red line are the posterior mean of both quantities in the static Σ model.