Long-term strategic asset allocation: an out-of-sample evaluation *

Bart Diris[a,b†], Franz Palm[a] and Peter Schotman[a,b,c]

February 15, 2010

Abstract

The objective of this paper is to find out whether the expected potential gains from strategic asset allocation can be realized in an out-of-sample test. Firstly, we find that long-term investors should time the market if they use our proposed shrinkage prior. This prior downplays the predictability of asset returns and leads to superior out-of-sample results compared to a standard uniform prior. Important is the use of a utility metric to evaluate prediction models. Shrinkage limits the losses in extreme negative events and this is what risk-averse investors value the most. Secondly, including the hedge component of strategic portfolios only leads to a modest performance improvement out-of-sample. Repeated myopic strategies perform almost as well as a dynamic asset allocation strategy. Monte Carlo simulations relate this finding to estimation error, i.e. the estimated repeated myopic and dynamic portfolios approximate the true unknown optimal dynamic portfolio equally well. Next, our paper shows that incorporating parameter uncertainty leads to a small performance improvement. Finally, portfolio weight restrictions improve performance for bad models and hurt the good models.

Affiliations: Maastricht University[a], Netspar[b], CEPR[c]
Keywords: Strategic asset allocation, out-of-sample analysis
JEL: G11, C53, C22

---

*We are grateful to Herman van Dijk, Roy Hoevenaars, Frank de Jong, Ronald Mahieu, Jeffrey Pontiff, Marno Verbeek, Peter Vlaar and participants at the 2nd International workshop on Computational and Financial Econometrics - Neuchâtel, 3M in Finance Workshop - Rotterdam, 14th International conference on Computing in Economics and Finance - Paris, Aarhus School of Business, APG Investments, Econometric Society European Meeting 2008 - Milan, Maastricht University, NBER time-series conference - Aarhus, Netspar Pension Workshop, and the Stockholm School of Economics. Part of this paper was written while Diris was a visiting fellow at Harvard University and Schotman a visiting fellow at SIFR in Stockholm.

†Corresponding author: Bart Diris, Limburg Institute of Financial Economics (LIFE) and Quantitative Economics (QE), Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands (email: b.diris@maastrichtuniversity.nl)

1 Introduction

Individuals and institutions (e.g. pension funds) invest financial wealth in different asset classes to meet their long-term goal. Individuals save money for retirement. Pension funds invest on behalf of their participants to provide them with retirement income. Merton (1969, 1971) showed that under changing investment opportunities, the optimal portfolios of these long-term investors (their strategic asset allocations) differ from the ones of short-term investors. Long-term investors hold hedge portfolios that anticipate future changes in the investment opportunities. Empirically, the main driving force in these hedge portfolios is the mean reversion of stock returns, which implies that equity is less risky for long-term investors than other types of assets. A second element of the strategic portfolios is inflation and interest rate risk. Long-term real returns from nominal bonds are subject to inflation risk, making them unattractive for long-term investors. Similarly short-term T-bills are not risk-free in the long-run because they must be rolled over repeatedly. Long-term investors have to take these risks into account in their hedge portfolios. If investment opportunities are changing, optimal long-term portfolio allocation requires that investors dynamically adjust the portfolio weights every period.[1]

By now, there exists a rich literature (e.g. Campbell, Chan and Viceira, 2003 and Brandt, Goyal, Santa-Clara and Stroud, 2005) that shows how to calculate the hedge portfolio and investigates the utility gains from these long-term strategic asset allocations in-sample. However, there are reasons to doubt the utility gains from strategic portfolio choice in practice, since the models of asset returns might be subject to substantial estimation error. First, Goyal and Welch (2008) document the poor out-of-sample predictability of equity returns, thus casting doubt on the mean reversion of stock returns. If returns are indeed nearly unpredictable, the optimal portfolio composition should not exhibit much time variation. Secondly, strategic asset allocation is even more demanding than myopic portfolio choice. The strategic portfolio consists of a speculative component that depends on the predictions of single period returns and a hedge component that is sensitive to the long-run predictions of returns and their covariance with current returns. The strategic portfolio is affected by estimation error in both components, whereas the myopic portfolio is only affected by errors in the speculative component. Therefore, the strategic portfolio is more susceptible to estimation error and might not perform very well in an out-of-sample test. Thirdly, unrestricted optimized portfolios for long-term investors based on estimates of the underlying dynamics show wildly fluctuating portfolio weights. The portfolio composition is even more extreme than the portfolio for short-term investors. This phenomenon is acknowledged by Campbell, Chan and Viceira (2003) among others. These extreme weights are subject to what is called "error maximization" and magnify any small misspecification in the return prediction model.

The performance measurement of strategic portfolios is still an open question in the academic literature, despite the relevance for (institutional) investors and the issues raised above. Therefore, our main objective in this paper is to find out whether the potential gains from

[1] See Campbell and Viceira (2002) for a broad overview of strategic asset allocation.

strategic portfolios can be realized in an out-of-sample test. Such an out-of-sample test has not been performed for long investment horizons. Because the gains from hedge demands apply to long investment horizons, performance evaluation of strategic portfolio choice requires long-term returns. Existing studies, however, use a single period return metric and thus cannot evaluate the out-of-sample utility gains from hedge demands.[2]

Our long-term investor optimizes the expected utility of wealth at a five year horizon using power utility. She is allowed to invest in a real T-bill, a stock index and a 5-year government bond. The predictive state variables are the price-earnings ratio, yield spread, and three-months T-Bill rate.[3] We measure the portfolio performance using the certainty equivalent returns based on the average realized utility over repeated five year horizons. In our analysis, we look at both the certainty equivalent return and the hedge component.

We use Bayesian time-series methods to estimate a model of investment opportunities.[4] We use a general Bayesian shrinkage prior advocated by Berger and Strawdermann (1996) adapted to vector autoregressions by Ni and Sun (2003). Such a Bayesian prior provides more plausible parameter estimates than a uniform prior such that optimal portfolio strategies become less aggressive and therefore avoid implausible extreme positions. More specifically, the prior shrinks slope coefficients in the predictive regressions for excess returns on stocks and bonds to zero, and shrinks the coefficients of the state variables to a random walk. It downplays the predictability in the data and therefore corresponds to the prior information of an investor who is skeptical with respect to the predictability of returns. Its generality allows for applications in larger systems than the setting in Wachter and Warusawitharana (2009).

We analyze the performance of this shrinkage prior, in particular whether it outperforms a standard uniform prior and whether these differences are robust to changes in the set-up. Much of the portfolio choice literature (e.g. Barberis, 2000) advocates the use of Bayesian decision-theory to account for parameter uncertainty. Supposedly, it leads to more robust portfolios and is another way to avoid the extreme "wacky" weights (Cochrane, 2007). The second method we use, called plug-in method, ignores parameter uncertainty and conditions on a given set of estimated parameters (using the posterior mean). A third way to stabilize portfolio weights are short-sell constraints as argued in Jagannathan and Ma (2003). We consider specifications with and without constraints on the portfolio weights.

For the set-up that conditions on parameter estimates (with unrestricted weights), Jurek and Viceira (2006) derived almost closed form solutions for the optimal strategic portfolios conditional on a given set of parameters. For the version of the model that accounts for parameter uncertainty as well as the plug-in version that uses restricted portfolio weights we need numerical optimization. Our performance analysis requires a fast and stable numerical algorithm. We succeed in accelerating the method of Brandt, Goyal, Santa-Clara and Stroud (2005) by intro-

[2]Some recent examples containing short-term out-of-sample results are Campbell and Thompson (2008), Goyal and Welch (2008) and Wachter and Warusawitharana (2009).

[3]As a robustness test, we also consider the dividend-yield as a predictor instead of the price-earnings ratio

[4]Some example from the growing Bayesian literature include Merton (1980), Cremers (2002), Wachter and Warusawitharana (2008), Jorion (1986), Black and Litterman (1992), Avramov (2002) and Pastor and Stambaugh (2000).

ducing a quadratic interpolation step that dramatically reduces the grid size of portfolios that must be evaluated. This makes our extensive out-of-sample analysis feasible.

Not surprisingly we find that a naive implementation of strategic asset allocations based on a uniform prior can lead to disastrous performance in terms of certainty equivalence returns. Weights are wildly fluctuating and this leads to periods with badly performing portfolios. More interestingly, we find that using Bayesian shrinkage priors leads to superior out-of-sample performance for long-term investors. Both the strategic as well as repeated myopic portfolios substantially and significantly outperform an unconditional strategy that ignores predictability and hedging. Changing portfolio allocations over time pays off for a long-term investor. Results are robust to small changes in the setup (such as different predictor variables) and the optimization as long as we use the shrinkage prior.

It turns out that it is very important to use a utility metric for assessing the performance of a prediction model. Risk averse investors evaluate big gains and big losses differently, since they want to avoid big losses at all costs. Due to this asymmetry in the utility function the best return prediction model for a risk averse investor is not necessarily the one that is on average closest to the actual return (for example in terms of mean squared error). It is the model that helps the investor avoid the big extreme (negative) events. It turns out that prediction models based on the shrinkage priors are best at avoiding these extreme events.

In terms of expected utility, the strategic portfolio performs only marginally better than the repeated myopic portfolio, even though both portfolios differ most of the time in terms of their asset mix. We conduct a Monte Carlo study to analyze the performance of the myopic and strategic portfolios rules. In simulated data, containing some predictability, the estimated myopic rule is more aggressive than the true myopic portfolio rule. By being more aggressive, the estimated myopic rule moves towards the optimal strategic rule. The estimated strategic rule is also too aggressive, thereby overshooting the true optimal rule. Compared to the truly optimal strategic portfolio, the estimated myopic rule is not aggressive enough, whereas the estimated strategic rule is too aggressive. In the end the estimated myopic and strategic rules produce almost the same average realized utility. Both rules suffer from estimation error, but the strategic rule is hurt more by estimation error than the (repeated) myopic rule. The hedge component of the strategic portfolio only marginally improves performance compared to a repeated myopic strategy that ignores this hedge component.

Parameter uncertainty improves performance slightly. Brandt, Goyal, Santa-Clara, and Stroud (2005) show that parameter uncertainty mainly has an impact on the weights of the hedge portfolio. As this hedge component does not have a big impact on performance (positively or negatively) in general, it is not surprising that parameter uncertainty does not have a large impact on performance. Portfolio weight restrictions have a larger impact on results. If portfolio weights are restricted, the best models perform worse and the bad models perform better.

The remainder of this article is organized as follows. Section 2 presents the data we use. Sections 3-5 describe respectively the general methodology, the modeling framework and the solution method. Section 6 consists of the out-of-sample results. Section 7 provides some additional tests and finally section 8 concludes. The appendix contains technical details on the estimation techniques and the numerical optimization algorithm.

2 Data

Our empirical analysis is based on monthly data for the US stock and bond market. We use data on three assets and two sets of three predictor variables; i.e. the nominal yield, the yield spread and either the price-earnings ratio or the dividend yield.

The monthly data set starts in February 1954 and ends in December 2006. The first three variables are log returns on different types of assets.[5] The first variable is the ex post real T-bill rate which is the difference between the log return (or lagged yield) on the 3-month T-bill, obtained from the FRED website[6], and log inflation, obtained from the Center for Research in Security Prices (CRSP). The second variable is the excess log stock return, which is defined as the difference between the value weighted log return on the NYSE, NASDAQ and AMEX market (including dividends) and the log return on the 3-month T-bill. The third variable, the excess log bond return, is defined in a similar way, where we use the five-year bond return from CRSP.

The sets of predictor variables have been shown to predict stock and/or bond returns in-sample. However, their out-of-sample predictive power is doubtful as argued in Goyal and Welch (2008) for stock return predictability. Fama and Schwert (1977) and Campbell (1987) among others show that the log nominal yield on the 90-day T-Bill predicts both stock and bond returns. Next, the log dividend-to-price ratio is defined as the log of the ratio of the sum of dividend payments over the past year divided by the current stock price. Dividend payouts are extracted from stock data by combining the value-weighted return including dividends and the index level excluding dividends of the NYSE, NASDAQ and AMEX market. Campbell and Shiller (1998) show that this ratio predicts stock returns. The log yield spread is defined as the difference between the log yield on a 5-year bond obtained from the FRED site and the log yield on the 90-day T-Bill. This spread forecasts stock returns and bond returns according to Campbell (1995) and Fama and French (1989). The log of the price-earnings ratio and is obtained from the Irrational Exuberance data, available from the website of Professor Shiller.[7] It is defined as the log of the current price over the lagged sum of earnings over the past 10 years. Campbell and Shiller (1998) show that this yield is a predictor of stock returns. In section 6, we use the the nominal yield, the price-earnings ratio and the yield spread. As a robustness check, we replace the price-earnings ratio by the dividend-to-price ratio in section 7.

These asset return and predictor variables are commonly used in the strategic asset allocation literature, see e.g. Campbell, Chan, and Viceira (2003) and Jurek and Viceira (2006). Table 1

[5] We use log asset returns when estimating our econometric model. However, we transform the log asset returns into simple returns when evaluating portfolio performance.

[6] http://research.stlouisfed.org/fred2

[7] http://www.econ.yale.edu/ shiller/data.htm

provides summary statistics of our monthly data.

[Table 1 about here.]

3 Methodology

This section describes the methodology we use in this paper. The first subsection explains the general set-up of our out-of-sample analysis. The second subsection explains the difference between the plug-in and decision-theoretic method. For the plug-in method, estimates are substituted for the unknown parameters in the predictive distribution function. The last subsection gives some intuition about the relative performance of different strategies.

3.1 General set-up

Define the $n \times 1$ vector y_t as follows

$$y_t = \begin{pmatrix} r_{tbill,t} \\ x_t \\ s_t \end{pmatrix}, \tag{1}$$

where $r_{tbill,t}$ is the real return on the T-bill, x_t is a vector of excess returns on stocks and bonds, and s_t is a vector of predictor variables. Vector s_t either consists of the nominal yield $Y_{nom,t}$, the price-earnings ratio PE_t and the yield spread $Y_{spr,t}$ or the nominal yield, the dividend-yield DP_t and the yield spread. Hence, $n = 6$.

We consider investors who start with initial wealth normalized to 1 and maximize expected utility over terminal wealth K periods in the future by investing in the real T-bill, a stock index and a government bond. We choose power utility for preferences. We consider both restricted and unrestricted portfolio weights. Restricted weights impose short-sell constraints.

More formally, the investor has power utility with $\gamma > 1$ and chooses portfolio weights $w_t, ...w_{t+K-1}$ such that the value function at time point t is maximized

$$V_t(K, Z_t, W_t) = \max_{w_t,\ldots,w_{t+K-1}} E\left(\frac{W_{t+K}^{1-\gamma}}{1-\gamma} \mid Z_t\right) \tag{2}$$

subject to the budget constraint

$$W_{s+1} = W_s \left(1 + w_s' R_{s+1}\right), s = t, ...t + K - 1, \tag{3}$$

where Z_t are conditioning variables that summarize all information available at time t, W_t is the wealth at time t, γ is a constant relative risk aversion parameter and R_{s+1} is the vector of simple returns on the assets in period $s+1$.[8] Portfolio weights add up to 1. Section 3.2 explains

[8] We obtain R_{s+1} by transforming the log benchmark return $r_{tbill,s+1}$ and the excess log returns x_{s+1} into real simple returns.

that the conditioning variables Z_t are equal to vector y_t under our assumptions and therefore we replace Z_t by y_t in the following.

Since initial wealth is 1, the following equality holds

$$W_{t+K} = \prod_{s=t}^{t+K-1} \left(1 + w_s' R_{s+1}\right). \tag{4}$$

We consider two types of strategies: a dynamic strategy and a myopic strategy. The dynamic strategy is the optimal solution to the long-horizon problem in equation (2) and contains both a myopic as well as a hedging component, defined as the difference between the dynamic and the myopic strategies. The myopic strategy ignores the long horizon, sets portfolio weights as if the remaining horizon is only one period and hence ignores the hedging part. More formally, the dynamic $w_{t,D}$ and myopic strategies $w_{t,M}$ are defined as follows

$$\{w_{t,D}, ..., w_{t+K-1,D}\} = \arg\max E\left\{\frac{\left(\prod_{s=t}^{t+K-1}(1 + w_{s,D}' R_{s+1})\right)^{1-\gamma}}{1-\gamma} \mid y_t\right\} \tag{5}$$

$$\{w_{s,M}\} = \arg\max E\left\{\frac{\left(1 + w_{s,M}' R_{s+1}\right)^{1-\gamma}}{1-\gamma} \mid y_s\right\}, s = t, ...t + K - 1. \tag{6}$$

If horizon $K = 1$, the two strategies are obviously identical.

An econometric model is needed to evaluate the conditional expectation over conditioning variables and asset returns in equation (2). Following among others Campbell, Chan, and Viceira (2003) and Jurek and Viceira (2006), we model the dynamics of log asset returns and state variables (our data) by using a VAR(1) as the econometric model. The VAR(1) model is as follows

$$y_{t+1} = B_0 + B_1 y_t + \epsilon_{t+1}, \tag{7}$$

where B_0 is a vector of intercepts, B_1 is a matrix of slope coefficients and ϵ_{t+1} is a vector of disturbances for which we make the following common assumption

$$\epsilon_{t+1} \sim N(0, \Sigma). \tag{8}$$

For future reference, it is useful to introduce the following decomposition for Σ, consistent with equation (1)

$$\Sigma = \begin{pmatrix} \sigma_{tbill}^2 & \sigma_{tbill,x}' & \sigma_{tbill,s}' \\ \sigma_{tbill,x} & \Sigma_x & \Sigma_{x,s}' \\ \sigma_{tbill,s} & \Sigma_{x,s} & \Sigma_s \end{pmatrix}. \tag{9}$$

We take a Bayesian perspective and obtain posterior distributions for the parameters for various prior distributions. We either use a uniform prior or a shrinkage prior, details are explained below.

In the portfolio choice literature, there are two methods that prescribe how to use these estimation results. The plug-in method substitutes parameter estimates for the true parameters. A second method acknowledges that there might be parameter uncertainty which can be taken into account by the posterior distribution of the parameters. This is the decision-theoretic method.

When making decisions, investors need to translate data into an econometric model and the econometric model into portfolio allocation rules. Different choices in this process lead to different portfolio weights. We mainly focus on whether investors should actively time the stock and bond market, whether they should incorporate the hedge portfolio and whether the shrinkage prior leads to improved results over the uniform prior. In order to tackle these issues, we consider the following choices for investors with risk aversion levels γ ranging from 2 to 5 to 10:

- Uniform or shrinkage prior (2 choices)
- Dynamic or myopic strategy (2 choices)
- Plug-in or decision-theoretic method (2 choices)
- Restricted or unrestricted portfolio weights (2 choices).

We have to carefully consider specifications based on the decision-theoretic method. In this case, the tails of the posterior predictive distribution of asset returns are fatter than the tails of the normal distribution, since we integrate out the parameters. Barberis (2000) shows that optimal portfolios using the decision-theoretic method are not defined in such a setting unless we make slight modifications. Our setting is further complicated, since none of the assets is completely risk-free (due to inflation risk). For the decision-theoretic method combined with restricted portfolio weights, we solve this problem by imposing that the return on the real T-bill rate is always larger than a lowerbound of -20% and by requiring that an investor invests at least 1% of the wealth in this asset. This guarantees that at least some portfolios have finite expected utility.

Portfolios that are based on short-selling do not have finite expected utility under the above assumptions. This implies that the optimal portfolio based on unrestricted portfolio weights exactly coincides with the optimal portfolio based on restricted portfolio weights. Therefore, we do not separately report results for the decision-theoretic method combined with unrestricted weights.

Hence, for all three risk aversion levels we consider 12 different specifications. Furthermore, we also calculate five benchmark specifications. Firstly, the $1/N$ rule that invests one third of the wealth in each asset. This fixed rule does not depend on data. Next, we consider rules that dogmatically impose that excess stock and bond returns are unpredictable, either combined with restricted or unrestricted weights and a myopic or dynamic strategy.[9] Investors that follow

[9] The dynamic and myopic specifications are not equal in this setting, since the expected real T-Bill rate is assumed to vary over time.

these rules do not actively time the stock and bond market. We combine the latter rules only with the plug-in method. Hence, in total we calculate 17 different specifications for each risk aversion level. The solution method we use depends on whether weights are (un)restricted, what kind of strategy we use (myopic or dynamic) and how we use the econometric estimation results (plug-in method or decision-theoretic method).

In the out-of-sample analysis, our first investor has an investment horizon of K months and uses all data available until period t_{start} to choose her first portfolio weights $w_{t_{start}}$. In the next period $t_{start}+1$, her investment horizon is $K-1$ and she updates her information set to choose portfolio weights $w_{t_{start}+1}$ etcetera. In period $t_{start}+K-1$, her investment horizon is 1 period and she uses all data until that period to choose her last portfolio weights $w_{t_{start}+K-1}$. This sequence of K portfolio weights results in exactly one terminal wealth value at time $t_{start}+K$, the end of the horizon. The next investor follows a similar strategy but she starts in period $t_{start}+1$ and ends in period $t_{start}+K+1$ with again exactly one terminal wealth value. We repeat this analysis for many investors, all with horizon K, who start their strategies one month after each other. The last investor starts in period $T-K$ and ends in period T, the end of our sample. In this way, we obtain a time series of terminal wealth values and a time series of realized utility values. This sample of realized utility values is used to measure performance. It provides a measure of out-of-sample performance of investors, since we only use information that is available to investors in real time.

In setting up the out-of-sample experiment, we need to make several choices. Firstly, we choose our starting date t_{start} to be equal to February 1974 in order to have enough initial observations (20 years) to estimate a model and to have a representative out-of-sample period. This choice is identical to the choice made in Wachter and Warusawitharana (2009). Secondly, we choose the investment horizon $K = 60$ months. This is a medium to long-term horizon and gives us almost 7 non-overlapping out-of-sample investment periods. Next, every month we allow investors to use all available information up to this month to update their portfolio holdings. This means that we re-estimate our models every month to include the newest observations using an expanding data window. Finally, we use the certainty equivalence return (CER) as performance criterium. It is the riskfree return that would make investors indifferent between following a strategy or accepting this riskfree return. The CER is a monotone transformation of average realized utility values $\overline{U}$ and is given as follows

$$CER = \left(\bar{U}(1-\gamma)\right)^{\frac{1}{1-\gamma}} - 1. \tag{10}$$

In the tables, we report the annualized certainty equivalence returns.

3.2 Plug-in method versus decision-theoretic method

In this section, we explain how to use the results from the econometric model. The first method is the plug-in method. This method treats the parameter estimates as the true values, ignoring any form of parameter uncertainty. This gives the following result for the conditional distribution of future values y_{t+1} for asset returns and state variables given their current values,

$$P\left(y_{t+1}|\hat{B},\hat{\Sigma},y_t\right), \tag{11}$$

where $\hat{B}$, $\hat{\Sigma}$ are estimates for B and Σ. In other words, the pdf of returns and state variables 1 period in the future is conditioned on estimated values. From the VAR(1) model defined in equations (7) and (8), returns are conditionally lognormally distributed. The current values of asset returns and state variables summarize the conditioning space (next to the parameter estimates).

The advantage of this approach is that it leads to attractive analytical properties and that we do not need to specify a distribution for the parameters. The disadvantage however is that this method ignores an important source of uncertainty: parameter uncertainty. Returns are not only uncertain because of the error terms but also because parameters might not be estimated correctly. This approach is adopted by Campbell and Viceira (2002) and Jurek and Viceira (2006).

The second method is the decision-theoretic method. It uses the following conditional predictive probability density function for asset returns and state variables

$$P\left(y_{t+1}|y_t,y_{t-1}...\right)=\int P\left(y_{t+1}|B,\Sigma,y_t\right)P\left(B,\Sigma|y_t,y_{t-1}...\right)d\Sigma dB. \tag{12}$$

Hence, a (posterior) distribution for parameters (B,Σ) is used to integrate over the parameters, i.e. parameter uncertainty is taken into account.

The advantage of this method is that it takes parameter uncertainty and uncertainty due to the stochastic nature of the variables into account. The disadvantage is that it is difficult to specify a posterior distribution that accurately describes what we really know about the parameters. Another disadvantage is that the posterior predictive distribution of returns in (12) is not a lognormal pdf anymore. This implies that we have to rely on numerical simulation methods for portfolio construction. Analytical properties of returns $L>1$ periods in the future are not known anymore, but we can simulate them. References for this method are Wachter and Warusawitharana (2009), Barberis (2000) and Brandt, Goyal, Santa-Clara, and Stroud (2005).

The dynamic strategy is equal to the myopic strategy plus a term that hedges against changes in the investment opportunity set. In case of the plug-in method, the investment opportunity set is completely determined by the current value of the vector y_t (given the parameter estimates which are treated as the true parameters). However, if we use the decision-theoretic method, this is not necessarily true. An investor learns more about the true unknown values of the parameters over time. This implies that her investment opportunity set also changes over time since the posterior parameter distribution is updated over time. In other words, hedging against a changing investment set means that we have to hedge against the changing posterior distribution due to learning as well when we consider the decision-theoretic approach. In this paper, we ignore this learning aspect however, because it is unfeasible given the size of our VAR(1) system. Since our VAR(1) system is 6 by 7, introducing this aspect would mean that we need 69 conditioning variables in vector Z_t to describe the investment opportunity set.[10] This is infeasible given the current numerical methods: currently only problems up to 11 conditioning variables are solved in the portfolio literature (see Brandt, Goyal, Santa-Clara, and Stroud (2005)).

We follow Barberis (2000) and assume that investors take parameter uncertainty into account, but ignore the impact of changing beliefs on today's asset allocation. They invest as if they only learn about the parameters at the end of their investment horizon. Under this assumption, the current values of y_t summarize the conditioning space at time t (next to the current posterior distribution). Note that our investors still learn about the true parameter values through time if new observations become available. The simplification we make is that they do not hedge against this learning. Brandt, Goyal, Santa-Clara, and Stroud (2005) show by means of simulations that incorporating parameter uncertainty while ignoring learning leads to improved performance relative to the case without parameter uncertainty.

3.3 Comparison of strategies

One of the aims of this paper is to investigate whether investors should take the hedge component of strategic portfolios into account in an out-of-sample test. In order to answer this question we analyze whether a dynamic strategy outperforms repeated myopic strategies. In case we would know the process that generates asset returns and state variables perfectly, this would be a trivial question to answer. A dynamic strategy would be superior to repeated myopic strategies, since the former strategy encompasses the latter (for the same investment horizon).

As we do not know the true data generating process (DGP), we have to select and estimate a model. This model is however by definition misspecified and estimates suffer from sampling errors. For the myopic portfolio, the errors are only related to estimation error in the single period expected returns. The hedge component however is also sensitive to the long-run predictions of returns and their covariance with current returns. Out-of-sample, it is therefore far from trivial which strategy works best.

We organize our discussion around the value function (2). The multiple period problem above can be written as a single period problem in a relatively straightforward way:

$$V_{t+s}(K-s, W_{t+s}, Z_{t+s}) = \max_{w_{t+s}} E\left\{ \frac{\left(w'_{t+s}R_{t+s+1}\right)^{1-\gamma}}{1-\gamma} \psi\left(K-s-1, Z_{t+s+1}\right) \mid y_{t+s} \right\}, \quad (13)$$

where $\psi\left(K-s-1, Z_{t+s+1}\right)$ is given as

$$\frac{1}{1-\gamma}\psi\left(K-s-1, Z_{t+s+1}\right) = \max_{w_{t+s+1},\ldots,w_{t+K-1}} E\left\{ \frac{\left(\prod_{r=t+s+1}^{t+K-1}\left(w'_r R_{r+1}\right)\right)^{1-\gamma}}{1-\gamma} \mid y_{t+s+1} \right\}. \quad (14)$$

The conditioning set at time t is summarized by conditioning variables y_t. This equation is the Bellman equation for the power utility case. We can solve for the optimal portfolio strategy by

[10] All distinct parameters plus the current variable values.

solving the sequence of one-period problems by backward induction.

The hedging component of the dynamic strategy depends on the contemporaneous dependence between the conditioning variables in y_t. If this dependence is misspecified, the myopic strategy might perform better out-of-sample. If there is a strong contemporaneous dependence and if we are able to estimate this dependence accurately, the dynamic strategy is superior.

4 Modeling framework

This section describes how we model the time-varying investment opportunity set and gives estimation results for these models.

4.1 Econometric model and estimation

The VAR(1) model introduced in equations (7) and (8) is restrictive in two ways. Firstly, it is unlikely that all dynamics in the data are modeled by using only one lag, i.e. the error terms are probably still autocorrelated. Note however that adding extra lags leads to an enormous increase in the number of parameters. One extra lag already means $n^2 = 36$ extra parameters in our setting. Since estimation efficiency is an important issue, we choose not to add extra lags. The usual trade-off between misspecification and efficiency applies.

Secondly, it is unlikely that the covariance matrix of the error terms is homoscedastic, i.e. that risk is constant over time. However, modeling heteroscedastic errors would mean a loss of precision of the estimates of the parameters of interest due to a substantial increase in the number of parameters to be estimated. Therefore, we choose to assume homoscedastic errors. This choice is supported by results of Chacko and Viceira (2005) who find that time-varying stock return volatility does not generate large hedging demands.

In order to facilitate the prior choice, we firstly re-parameterize the VAR(1) model by transforming the state variables

$$y_{trans,t} = \begin{pmatrix} r_{tbill,t} \\ x_t \\ \Delta s_t \end{pmatrix} \tag{15}$$

and use the following transformed auxiliary model in the estimation stage

$$y_{trans,t+1} = B_0 + B_1^* y_t + \epsilon_{t+1}. \tag{16}$$

It is possible to obtain the matrix of slope coefficients B_1 in the original model by adding 1 to the diagonal elements that correspond to the predictor variables in matrix B_1^*. In this paper, we are mainly interested in the posterior distributions for B_0 and B_1. Therefore, we generally first obtain the posterior distribution for coefficients B_0 and B_1^* in the auxiliary model and subsequently use the above transformation to obtain the posterior distribution for B_1. We only report and use the latter.

In order to estimate the VAR(1) model in equation (16), provide inference and make forecasts, we use, in line with most of the literature, a conditional likelihood function that conditions on the first observation. The conditional likelihood function is

$$P(Y^*|B,\Sigma) \propto |\Sigma|^{-T/2} \exp\left\{-\frac{1}{2}tr\left[(Y^* - XB^{*'})'(Y^* - XB^{*'})\Sigma^{-1}\right]\right\}, \tag{17}$$

where T is the number of observations, Y^* is the $T \times n$ matrix of observations on $y_{trans,t}$, Y_{-1} is the $T \times n$ matrix of lagged observations on y_t, X is the $T \times (n+1)$ matrix $X = [\iota, Y_{-1}]$ and B^* is the $n \times (n+1)$ matrix $B^* = [B_0, B_1^*]$. A popular alternative would be to use an unconditional likelihood function as in Schotman and van Dijk (1991), Wachter and Warusawitharana (2009) or Stambaugh (1999) that treats the first observation as stochastic. We do not pursue this alternative in this paper. We are both interested in point estimates for the parameters and in the posterior distribution of these parameters. For point estimates we use the posterior means.

Our first prior is a uniform prior on B^* and a Jeffrey's prior on Σ,

$$p(B^*, \Sigma) \propto |\Sigma|^{-(n+1)/2}. \tag{18}$$

We refer to this prior as the uniform prior. It is the most commonly used prior for VAR models. The corresponding posterior is given in equation (23) in the appendix. The posterior mean of B^* is equal to the OLS/ML estimator $\hat{B}^{*'} = (X'X)^{-1}X'Y^*$ and the posterior mean of Σ is equal to $S/(T-2n-2)$, where $S = (Y^* - X\hat{B}^{*'})'(Y^* - X\hat{B}^{*'})$. For the decision-theoretic approach, we need to be able to simulate from the full posterior distribution and its predictive distribution. We explain this in the appendix.[11]

We consider a second Bayesian estimator which is used among others in Ni and Sun (2003) in the context of a similar VAR model. We refer to this prior as the shrinkage prior. This estimator shrinks the coefficients towards zero. The prior is given as

$$p(B^*, \Sigma) \propto \left(b^{*'}b^*\right)^{\frac{-(n(n+1)-2)}{2}} |\Sigma|^{-(n+1)/2}, \tag{19}$$

where $b^* = vec(B^*)$. The exponent is exactly equal to the exponent that Ni and Sun (2003) propose. It is the product of a shrinkage prior for B^* and the Jeffrey's prior on Σ. The prior itself is not proper, but Ni and Sun (2003) show that the posterior is proper in a VAR model when the ML estimator exists, which holds in our setting. Note that the prior has a negative exponent. This means that prior draws with large coefficients are relatively improbable. Shrinking the coefficients in the auxiliary model (16) towards a zero matrix implies that we are shrinking the coefficients in the original model towards zero except for the predictor variables which we shrink towards a random walk.[12]

The kernel of the posterior density is given in equation (26) of the appendix. The shrinkage prior is not conjugate, and hence does not lead to a known posterior density for the parameters.

[11] Results using the uniform prior are equivalent for the original and the auxiliary model.

[12] Note that if we would have combined the shrinkage prior with the original model, we would have shrunk the autocorrelation coefficients of the highly persistent state variables to 0 instead. This would have resulted in a misspecified model.

However, as Ni and Sun (2003) show, a straightforward MCMC sampler exists to draw from the posterior. The simulation algorithm is explained in appendix A.

Our shrinkage prior has a clear economic interpretation. It corresponds to an investor that is very skeptical about predictability. As a result such an investor downplays all the predictability that is found in the data. However, the investor does not dogmatically ignore predictability. If there is sufficient evidence in the data that asset returns are predictable, this investor will take (some) asset return predictability into account.

This particular shrinkage prior has several advantages. Firstly, since the prior is improper, it is relatively uninformative. The likelihood dominates the prior quickly once there is sufficient data. In other words if the data shows a lot of predictability, the posterior will reflect this. Secondly, the prior does not depend on any tuning constants. This avoids all kind of calibration issues that could arise. Finally, the prior leads to a posterior that is relatively easy to calculate using Gibbs sampling. The sampling algorithm is fast and stable, even for large VAR models.

If the lagged asset returns and predictor variables are not able to predict risky asset returns, the second and third row of B_1^* in model (16) are both equal to zero rows. As a benchmark, we consider specifications that dogmatically set these coefficients equal to zero and leave the coefficients in other equations equal to the posterior mean under the uniform prior.[13] We refer to this specification as the no-predictability prior.

The state variables in the model are highly autocorrelated and close to a unit root. It is common in the literature to impose the assumption of stationarity (e.g. Campbell and Viceira (2002) and Stambaugh (1999)). For the decision-theoretic approach, we indeed impose that the original model is stationary. Numerical results are more stable, since this excludes extreme non-stationary draws. Since the mode of the likelihood function is generally within the stationary region, we do not impose this assumption when using the plug-in approach. This only slightly changes the point estimates, has a minor impact on the out-of-sample results but saves on computation time.

4.2 Estimation results

In this section, we give estimation results for the VAR(1) model introduced in equation (7) and (8). We report the posterior mean for the model estimated on the full data-set using either the uniform or shrinkage prior. Firstly, we estimate the posterior moments in the auxiliary model and subsequently transform these estimates into posterior moments for B and Σ in the original model. Table 2 reports posterior moments for B and Σ for the model where the price-earnings ratio is one of the state variables.

The table shows that the state variables are highly autocorrelated under both priors. Furthermore, we see that the nominal yield and the price-earnings ratio predict stock returns negatively, and that the yield spread predicts bond returns positively. There is also a large positive correlation between shocks to the price-earnings ratio and excess stock returns, which means that

[13] Results are similar if we additionally assume that the real T-Bill rate is unpredictable.

unexpected positive shocks to stock returns lead to negative future investment opportunities. This result implies that there is mean-reversion in stock returns.

Comparing the posterior means for both priors, we clearly see that the posterior mean for the return prediction coefficients are shrunk towards zero by the shrinkage estimator except for the autocorrelation coefficients which are shrunk to one. The shrinkage estimator downplays the predictability of asset returns. One way to see this is to look at the lower R^2 values under the shrinkage prior, especially for excess stock returns. The lower R^2 values lead to less aggressive investment strategies.[14]

[Table 2 about here.]

We re-estimate our models on bigger and bigger data-sets that include the newest observations. Since our data set starts in February 1954 and our empirical analysis in February 1974, we estimate models for which the last observation ranges from January 1974 until November 2006. The table shows that the price-earnings ratio and the yield spread are among the most important predictors for respectively excess stock and bond returns. Therefore, we present time series plots of the slope coefficients of (x_s,s_{PE}) and (x_b,s_{SPR}) in figure 1.

From the figure it is clear that the posterior means for the shrinkage prior are closer to 0 than for the uniform prior. There seems to be a lot of uncertainty about the estimated values, since the parameters are extremely variable over time. However, the estimated values for the shrinkage estimator are less variable. Finally, note that the values for the two estimators slowly converge to each other once more observations are available, since the likelihood dominates when the sample size grows.

[Figure 1 about here.]

5 Solution method

This section explains the solution methods we use in this paper. This choice depends on whether we condition on parameter estimates (plug-in approach) or use the posterior distribution of the parameters in a decision-theoretic approach and whether we restrict portfolio weights or not. We use the semi-analytical method in Jurek and Viceira (2006) for calculating the unrestricted plug-in strategies. We have to use numerical methods for all other strategies. We propose a refinement of the method of Brandt, Goyal, Santa-Clara, and Stroud (2005) and van Binsbergen and Brandt (2007) by relying on an important observation made by Koijen, Nijman, and Werker (2009).

5.1 Analytical method

Given the VAR(1) model in equation (7), returns are lognormally distributed conditional on the parameter values. Jurek and Viceira (2006) use this fact to derive approximate-analytical

[14]The R^2 values we provide are implied by the posterior mean of the parameters. The mean of the posterior distribution of R^2 values does not exist when allowing for non-stationary draws.

solutions for the unrestricted plug-in model for the myopic and the dynamic strategy.[15] These solutions are all based on the Campbell and Viceira (2002) approximation to log-portfolio returns

$$r_{p,t+1} = r_{tbill,t+1} + w_t' x_{t+1} + \frac{1}{2}\left(w_t' \sigma_x^2 - w_t' \Sigma_x w_t\right), \tag{20}$$

where w_t is the weights vector on the risky assets and σ_x^2 is the vector of diagonal elements of Σ_x.[16] This approximation, and therefore Jurek and Viceira (2006)'s method, is exact in continuous time and accurate on short time intervals. It is very accurate in our setting since we are using monthly data.

Jurek and Viceira (2006) show that portfolio weights on risky assets are an affine function of the conditioning variables y_t

$$w_t^{K,DYN} = A_0^{(K)} + A_1^{(K)} y_t, \tag{21}$$

where $A_0^{(K)}$ is a coefficient vector and $A_1^{(K)}$ is a coefficient matrix, depending on the (remaining) investment horizon and the parameters. Please refer to their equation (22) for details.

The weights for the myopic strategy are as follows

$$w_t^M = (\gamma \Sigma_x)^{-1}\left(E_t[x_{t+1}] + \frac{1}{2}\sigma_x^2 + (1-\gamma)\sigma_{tbill,x}\right). \tag{22}$$

5.2 Numerical method

There is no analytical solution available for the plug-in model combined with restricted portfolio weights. Furthermore, the predictive distribution of returns is not lognormal if parameters are integrated out and therefore there is no analytical solution available for the restricted decision-theoretic model. In these cases we have to use numerical methods.

Firstly, we consider the dynamic strategy. We solve the sequence of one-period problems by backward induction, i.e. start in period $K-1$ and iterate to period 0. We follow Brandt, Goyal, Santa-Clara, and Stroud (2005) and simulate many trajectories of asset returns and state variables and approximate the conditional expectations we encounter by regressions of the value function at time $t+1$ on conditioning variables that summarize the information set at time-point t. Furthermore, we follow van Binsbergen and Brandt (2007) and set-up a fine grid of portfolio weights, evaluate the conditional expectation for all grid points and pick the maximum. Since we have to re-calculate dynamic strategies almost 400 times, computation time is an important issue. Therefore, we use a refinement in Koijen, Nijman, and Werker (2009) in our setting and parameterize the regression coefficients in regressions that approximate conditional utility by a quadratic function of portfolio weights.[17] This allows us to find the optimal weights along

[15] Note that Jurek and Viceira (2006) use the ML estimate as plug-in estimate. We use the posterior mean as plug-in estimate instead.

[16] Note that the weight on the benchmark asset is $1 - w_t' \iota$ and that portfolio weights add up to 1.

[17] Note that Koijen, Nijman, and Werker (2009) solve a life-cycle model with intermediate consumption and parameterize the first order conditions by an affine function in the portfolio weights. We parameterize the value function instead.

each path analytically by optimizing a quadratic function on a restricted set which can be done analytically. It means that we do not have to use a very fine grid since the parameterization regressions are very accurate.

This gives the following algorithm:

1. Generate N sample paths of length K of asset returns and state variables from the conditional prediction model ("plug-in") or from the predictive distribution ("decision-theory").

2. Set-up a grid of portfolio weights.

For period $K-1$ until period 0 repeat steps 3, 4 and 5.

3. Pick one set of portfolio weights from the grid and calculate the realized utility values for all simulated paths. Hence: use the chosen portfolio weights together with the optimal portfolio weights chosen in previous steps to calculate the realized terminal wealth values for every path. Take the utility over these values to calculate the realized utility values for all paths.

4. Regress the N realized utility values on a constant and functions of the conditioning variables in order to calculate regression coefficients and conditional utility values.

Repeat step 3 and 4 for all portfolio weights on the grid.

5. Parameterize the regression coefficients in a quadratic function of the portfolio weights. This allows us to express the conditional utility as a function of constants, conditioning variables and portfolio weights. Along each path, constants and conditioning variables are known and hence along each path conditional utility is only a function of the unknown portfolio weights. For every path, choose the portfolio weights that maximize this approximate quadratic function. This can be done analytically.

The calculation of the myopic strategy is similar with $K = 1$. Appendix B gives more details on the parameterization of regression coefficients and the accuracy of the algorithm.

The decision-theoretic method combined with restricted portfolio weights gives some problems as indicated above. We guarantee that at least some portfolios have an expected utility value greater than minus infinity by imposing that the return on the real T-Bill rate is always larger than -20% and requiring that investors invest at least 1% of their wealth in the real T-Bill rate.[18] This solution is proposed by Hoevenaars, Molenaar, Schotman, and Steenkamp (2007).

6 Out-of-sample performance

In our empirical analysis, we investigate the out-of-sample performance of strategic asset allocations. These specifications differ in their method (plug-in or decision-theory), the general

[18] In our numerical algorithm, we simply re-sample draws that would violate this boundary.

strategy (dynamic or myopic), how these translate data into a model (uniform prior or shrinkage prior) and whether or not the weights are restricted. We also show benchmark results based on a no-predictability prior. These benchmark results are also based on the plug-in method. Results are reported for investors with risk aversion parameter γ equal to 2, 5 and 10. The following subsections cover respectively benchmark results, results using the plug-in method and results using the decision-theoretic method.

6.1 Results for the benchmark specifications

Firstly, we report results for five benchmark specifications in table 3. We show their certainty equivalence return (CER), their average terminal wealth and the standard deviation of terminal wealth. The first specification is the $1/N$ strategy. It invests 33% in stocks, bonds and T-bills irrespective of the data. The next four specifications are based on the no-predictability prior either combined with a dynamic or myopic strategy and unrestricted or restricted portfolio weights. The no-predictability prior imposes that excess stock and bond returns are not predictable.

[Table 3 about here.]

Firstly, the certainty equivalence returns (CER) in the table are all positive. This means that investors are willing to follow these simple strategies unless they are paid a positive risk-free return. Interestingly, the dynamic strategy outperforms the repeated myopic strategy for all specifications. This implies that hedging (real) interest rate risk boosts performance for long-term investors as argued in Campbell and Viceira (2001) in an in-sample setting. This is also reflected in the slightly higher average terminal wealth values and the slightly lower standard deviation of terminal wealth values. The performance difference is small however. The performance improvement is relatively more important for investors with higher risk aversion levels. Campbell and Viceira (2001) show that the hedge component for such investors is larger than for investors with lower risk aversion.

Remarkably, for a rather conservative investor with $\gamma = 10$, both the $1/N$ strategy and the restricted no-predictability strategies outperform the unrestricted no-predictability strategy. This suggests that imposing restrictions might improve out-of-sample performance and that non-data based methods are not necessarily inferior to data-based strategies. The former is consistent with results in Jagannathan and Ma (2003) who show that imposing weights restrictions is a form of shrinkage that boosts performance. The latter is consistent with DeMiguel, Garlappi, and Uppal (2007), who show that a $1/N$ strategy is tough to beat out-of-sample.

Figure 2 plots a histogram of realized utility values for the unrestricted no-predictability strategy. We set $\gamma = 5$. The figure shows that the utility value distribution is very left skewed. Most values are near zero but there are some large negative outliers (corresponding to low terminal wealth values). However, these negative outliers are the most important values for risk-averse investors. Risk averse investors want to avoid extreme negative events at all costs and will heavily weight every extreme event in their utility function. This suggests that specifications that limit the number and/or size of extreme events are the ones with the highest certainty equivalence returns (and equivalently highest average realized utility). These outliers are further emphasized when $\gamma = 10$, but will be less severe when $\gamma = 2$.

[Figure 2 about here.]

6.2 Results for the plug-in method

Secondly, we show results based upon the plug-in method. We report results for dynamic and myopic strategies either using the uniform or shrinkage prior and either using restricted or unrestricted portfolio weights. Investors that use these specifications time the stock and bond market actively, since the specifications allow for predictable stock and bond returns. Results are given in table 4.

Firstly, we consider specifications using the uniform prior and unrestricted weights. Remarkably, the performance of an investor with low risk aversion ($\gamma = 2$) is disastrous under the standard uniform prior when weights are unrestricted. She is willing to pay a risk-free return of up to -100% to avoid adopting this strategy. The average terminal wealth and its standard deviation show why. The strategy leads to a very high average terminal wealth but with extremely high risk. Due to this risk, at least one of the terminal wealth values in our sample turns out to be zero which means that at least one investor loses all her money during her 5-year investment period. Such an investor obtains a realized utility value of $-\infty$, since this is the outcome that such a risk-averse investor desperately wants to avoid.

[Table 4 about here.]

The performance is better for higher γ values. These investors are less aggressive and avoid the strategies that lead to disaster for the $\gamma = 2$ investor. CERs are positive and higher than the ones for the benchmark strategies. Differences turn out to be economically important. For very risk averse investors, it pays off to time the bond and stock market.

Another important finding is that repeated myopic strategies outperform the theoretically optimal dynamic strategies. Although the average terminal wealth is higher for dynamic strategies, the risk more than proportionally increases. This result implies that the hedging components of dynamic strategies are misspecified and only deteriorate performance. Dynamic strategies are more sensitive to misspecification of any form, since they do not only require us to model the evolution of asset returns correctly, but also of state variables. It is apparently sufficient to only focus on short-term changes in investment opportunities and ignore long-term changes when using the uniform prior.

Secondly, let us consider the shrinkage prior combined with unrestricted weights. It shows a completely different picture. Firstly, the performance for all strategies and all risk aversion levels increases substantially and is much better than for the benchmark strategies. For all risk aversion levels, it pays off to time the bond and stock market. The shrinkage prior makes sure that investors do not take excessive risk. Although the use of the shrinkage estimator reduces average terminal wealth compared with the uniform prior, its standard deviation is more than proportionally reduced. For example, compare the dynamic strategies for an investor with $\gamma = 2$. Although average terminal wealth is reduced with a factor 2.5, its standard deviation is reduced with a factor 8. The result is that the CER for an investors with $\gamma = 2$ is not equal to -100% anymore.

It also turns out that dynamic strategies outperform myopic strategies. Apparently, we are better able to model the hedge component of strategic asset allocations when using the shrinkage prior. The risk for dynamic strategies is still higher but in this case the extra average terminal wealth more than offsets this. In terms of economic performance, the differences between a dynamic and myopic strategy are relatively modest. Different estimation techniques lead to larger performance differences than different strategies.

In order to understand how the shrinkage model works, we plot the realized utility values for risk aversion $\gamma = 5$ using both the shrinkage and the uniform prior against time in panel A of figure 3. The figure shows that both series are heavily autocorrelated due to overlapping intervals and that there is a positive correlation between the series. In general, both strategies perform similarly except for a couple of extremely low realized utility values. The shrinkage prior manages to reduce these losses, while the losses for the uniform prior are very large. The shrinkage prior improves performance by avoiding extreme losses. This is exactly why risk averse investors value this model the most.

[Figure 3 about here.]

How does the shrinkage estimator reduce losses? In order to answer this question, we plot the corresponding stock weights of investors against time in panel B of figure 3. We plot the weights for investors with a remaining horizon of 60 months. The picture shows that the average weights for both strategies are more or less equal. The weights for the shrinkage prior are, however, much less variable and the portfolio holdings much less extreme. An investor that uses the shrinkage prior is still able to time the market. She can still go long in stocks or bonds if market conditions are good and short in stocks or bonds if market conditions are bad. However, the weights are not as extreme anymore and make more sense intuitively. By avoiding overly aggressive market timing, the investor with the shrinkage prior avoids the important extreme events.

The dynamic strategy outperforms the myopic strategy using the shrinkage prior. In order to illustrate this, consider panel A in figure 4 which plots the histogram of differences in realized utility values between a dynamic investor and a myopic investor with $\gamma = 5$. Positive values indicate outperformance by the dynamic model. The figure shows that both strategies lie close to each other in general. The mass to the right of 0 indicates that most observations give a slight edge to the dynamic strategies. The figure also shows that there are more outliers on the right than on the left. However, differences are not very large.

[Figure 4 about here.]

Finally, let us consider what happens if portfolio weights are restricted between 0 and 1 for all three assets. Restricting portfolio weights leads to a substantial reduction in risk and terminal wealth values for investors using either the uniform or shrinkage prior. It helps to avoid CERs of -100%. This is consistent with results in the previous section and with Jagannathan and Ma (2003). The latter show that restrictions are a form of shrinkage. Hence, in this light it is not surprising that portfolio weight restrictions can improve performance. Note that using shrinkage is substantially better than restricting portfolio weights in order to avoid extreme events.

However, restrictions hurt performance for better performing specifications. Restrictions limit the possibilities of investors and lead to much lower CERs. Apparently, going short and very long in assets pays off for long-term investors, especially for those that use the shrinkage prior. Portfolio weight restrictions hurt specifications using the shrinkage prior more than specifications using the uniform prior. On average, the shrinkage prior still outperforms the uniform prior slightly when imposing portfolio restrictions. In all cases, dynamic strategies outperform myopic strategies when restricting portfolio weights. However, economically, the differences are smaller than before. The hedge component of strategic asset allocations improves performance only slightly in this setting.

We conclude that it might not be optimal to time the stock and bond market unless investors use the shrinkage prior. Empirically, differences turn out to be economically important. Using shrinkage avoids extreme portfolio weights and therefore extreme events. Such a specification is heavily favored by risk-averse investors. Dynamic strategies only work satisfactorily in all cases when using the shrinkage estimator, since shrinkage leads to a better modeling of the hedge component. However, differences are economically relatively modest. The effect of the shrinkage prior is the largest when portfolio weights are unrestricted. Such investors can still go short and very long in assets without taking excessive risk. When portfolio weights are restricted to be non-negative, the effect of the shrinkage prior is modest but still positive. Finally, portfolio weight restrictions help the worse performing specifications, but hurt the best performing specifications.

6.3 Results for the decision-theoretic method

This subsection gives results for the decision-theoretic approach. We consider dynamic and myopic strategies either combined with the uniform or shrinkage prior. We only report results for restricted portfolio weights as explained in sections 3.1 and 5.2. Results are given in table 5.

Firstly, we compare results with previous sections. If we consider dynamic strategies, the CERs increase slightly when taking parameter uncertainty into account. Results for myopic strategies are more mixed, but on average results improve when considering parameter uncertainty. Differences however are again very small. Brandt, Goyal, Santa-Clara, and Stroud (2005) show by means of simulation that parameter uncertainty mainly has an impact on the hedging component of a dynamic strategy. Since this hedging component does not have an important impact on performance according to results in the previous subsection, it is not surprising that there is only a small performance difference between the plug-in method and the decision-theoretic method. The specifications that we consider in this section outperform the benchmark strategies that ignore predictability by economically important margins, especially for higher risk aversion levels. This implies that actively timing the bond and stock market also pays off when taking parameter uncertainty into account.

We illustrate the performance of the decision-theoretic approach in panel B of figure 4. This figure plots a histogram of differences in realized utility values between dynamic strategies using the decision-theoretic and plug-in approach. We set $\gamma = 5$ and use the uniform prior. Positive values indicate outperformance by the decision-theoretic method. The figure shows that the plug-in model performs better in most cases, i.e. the median is slightly negative. However, if the decision-theoretic model outperforms the plug-in model, the difference is relatively big, illustrated by larger positive values. On average, the decision-theoretic method slightly outperforms the plug-in method. Overall differences are negligible.

[Table 5 about here.]

Secondly, the table shows that in terms of performance dynamic and myopic strategies are again close to each other with a slight edge for the dynamic strategy. This is not surprising. Brandt, Goyal, Santa-Clara, and Stroud (2005) show that the hedge component is relatively small when taking parameter uncertainty into account. Portfolio weights for dynamic and myopic strategies are therefore close to each other.

Finally, the certainty equivalence returns for specifications involving the shrinkage prior are very close now to specifications using the uniform prior. Apparently, restricting portfolio weights, incorporating parameter uncertainty and using the shrinkage prior leads to portfolios that are a bit too conservative. However, on average the shrinkage prior still outperforms the uniform prior slightly. The difference is negligible however.

We conclude that most results from the previous section still stand. Timing the stock and bond market still pays off for risk-averse investors. Differences are also economically important. Furthermore, the performance of dynamic and myopic strategies are still close to each other with a slight advantage for the dynamic strategies. The performance difference between specifications involving the shrinkage and uniform prior becomes smaller when incorporating parameter uncertainty. Shrinkage is economically less important in such a setting where we also restrict portfolio weights. Note that it still improves performance on average. Finally, incorporating parameter uncertainty leads to specifications with slightly higher certainty equivalence returns.

7 Additional tests

In this section, we perform some additional tests. In the first subsection, we perform classical tests on the performance differences between different specifications. The second subsection investigates the performance differences between myopic and dynamic strategies. Finally, the last subsection considers a model with the dividend-to-price ratio as one of the predictor variables.

7.1 Classical significance tests

So far, we performed a Bayesian analysis. We calculated strategies and compared the distribution of expected utility of different strategies (by means of a histogram of realized utilities) with each other. In this section, we perform an additional robustness check. We investigate the classical statistical significance of the results by comparing the results of the strategies of sections 6.2 and 6.3 with the benchmark strategies in section 6.1 in a repeated samples context. We test whether the difference in average realized utility between a strategy and its benchmark is statistically different from zero. As a benchmark, we take the no-predictability strategies of section 6.1 either unrestricted or restricted and either dynamic or myopic, depending on the context. In other words, we test whether the extra value of market timing we find in previous sections might be spurious.

We view utility as the loss function of forecasts (after implementing strategies). In the forecasting literature, tests of equal forecasting performance are standard and we use the Diebold and Mariano (1995) test on the utility series. Diebold and Mariano (1995) generate the difference series of two forecasts and test whether this difference is equal to zero by means of a standard t-ratio. They show that this test statistic has a standard normal distribution. We estimate the covariance matrix of average realized utility non-parametrically by means of the Newey and West (1987) HAC estimator. In order to choose the lag length, we use the Newey and West (1994) lag length selection criterium.

Table 6 presents results. The performance of unrestricted plug-in strategies based on the uniform prior is only significantly different from its benchmark in one case. If we use the shrinkage prior instead, we see that these strategies become significant. Hence, the impressive performance for the unrestricted plug-in methods based on the shrinkage prior is not spurious and is statistically different from its benchmark. Results are different for specifications that restrict portfolio weights. Results for the dynamic and myopic strategies are significant except for low risk aversion levels. Apparently, a low risk averse investor is especially hurt when weights are restricted.

[Table 6 about here.]

In order to take issues such as autocorrelation and skewness in the realized utility series into account as well as correct for the fact that the benchmark strategies are based on nested models, we also perform an additional Monte Carlo simulation. We generate 100 time-series of asset returns and predictor variables under the null of no predictability. The DGP is based on the parameter estimates obtained using the no-predictability prior on the full data-set. In every Monte Carlo simulation, we generate a time-series of 52 years of asset returns and state variables and perform the same out-of-sample analysis as on the real data-set. In order to make the Monte Carlo analysis feasible, we only consider specifications with unrestricted portfolio weights.[19]

[19] Specifications involving restricted portfolio weights take approximately half a day to calculate. Repeating this 100 times is not feasible.

Table 7 shows the results. Panel A reports the difference in CERs between the uniform prior and the no-predictability prior in the simulations. A positive difference implies outperformance by the uniform prior. We find that the average difference is negative which means that the no-predictability strategies perform better on average. This is not surprising, since the data is generated using the no-predictability prior on the full data-set. If we compare the differences found in the data with the different percentiles in the simulations, we conclude that the results in the data are not significantly different from the results in the simulations that are generated under the null of no predictability.

[Table 7 about here.]

Panel B shows the difference in CERs between the shrinkage prior and the no-predictability prior. The average difference is negative, but less negative than the average difference for the uniform prior. The table shows that the positive differences in the data are in all cases larger than the maxima in the 100 Monte Carlo simulations. Hence, the largest performance difference in the DGP without predictability is less than we find in the data. Therefore, the data has to contain predictability.

7.2 Difference between dynamic and myopic strategies

The results in previous sections show that there is hardly a difference in CERs between dynamic and (repeated) myopic strategies. In order to understand this surprisingly small difference, we perform a Monte Carlo simulation under the null of predictability. The DGP is based on the parameter estimates obtained using the uniform prior on the full data-set. In every simulation, we generate 52 years of data and perform the same out-of-sample analysis as on the real data-set. We base the portfolio weights on the true parameters that we use to simulate the data (this is obviously infeasible in reality) or on estimated parameters based on either the uniform prior or on the shrinkage prior. To make the analysis feasible, we only consider specifications with unrestricted portfolio weights. We use 100 simulations.

The differences in CERs between a dynamic strategy and a myopic strategy are given in table 8. A positive difference implies that the dynamic strategy outperforms the repeated myopic strategy. In panel A, we give results for strategies that are based on the true parameters. The average difference is positive, which indicates outperformance by the dynamic strategy. This is not surprising, since the dynamic strategy should be the optimal strategy when the investor knows the true DGP exactly. The simulations show that the CER gains are economically important for investors with risk aversion levels of γ is 5 or 10. Remarkable, in some simulations the difference is negative even if an investor knows the true parameters.

Panel B shows results for an investor that has to estimate the parameters using the (simulated) data and the uniform prior. The performance differences that we find in the data are in line with the differences in the Monte Carlo simulations. For investors with low risk aversion, the dynamic strategy is on average inferior to the myopic strategy. Apparently, even if the true DGP contains predictability, a dynamic strategy is not necessarily better than a myopic strategy when the parameters need to be estimated. Panel C gives similar results using the shrinkage prior. These results are in line with the results in panel B.

[Table 8 about here.]

Why do the estimated myopic strategies perform as well as the estimated dynamic strategies? In every simulation and in every period, we can calculate the difference between the optimal dynamic portfolio weights (based on the true parameters) and the estimated portfolio weights (either myopic or dynamic). This allows us to calculate the root mean squared portfolio weight error (RMSPE) for both the stock weights as well as the bond weights for every simulation. We calculate the average RMSPE over all simulations. This is a measure of how far the estimated portfolio weights are from the true optimal dynamic portfolio weights. We compare both the estimated myopic and estimated dynamic weights with the true optimal dynamic weights. We do this for investors with a remaining investment horizon of 60 months.

Lines 10 and 11 of panel B and C show the differences in average RMSPE between the estimated dynamic and myopic weights. A positive number implies that the estimated myopic weight is closer to the optimal dynamic portfolio weight. For the uniform prior, the estimated myopic portfolio weights are indeed closer to the optimal dynamic portfolio weights. This holds for both stocks and bonds. The estimated myopic weights approximate the optimal portfolio better than the estimated dynamic weights due to the large estimation error in the long-run predictions of returns and in the covariances with current returns. This explains why the repeated myopic weights outperform dynamic strategies in the data when using the uniform prior.

A closer look at terminal wealth values shows that the portfolio weights using the uniform prior are too aggressive. For example considering $\gamma = 5$, the average (over the simulations) of the average terminal wealth (standard deviations of terminal wealth) is 7.81 (5.90) for the dynamic and 7.02 (5.08) for the myopic strategy using the true parameters. Using the estimated parameters under the uniform prior, we respectively get 8.76 (8.39) and 7.01 (5.76). These results show that both the estimated myopic as well as the estimated dynamic strategy are (way) too risky and aggressive. However, by being too aggressive, the estimated myopic portfolio weights move towards the optimal dynamic weights and approximates the true optimal dynamic strategy better.

Panel C shows that results are more mixed for the shrinkage prior. The estimated myopic strategy approximates the optimal dynamic strategy slightly better for stock weights, but much worse for bond weights. A closer look at terminal wealth values shows that portfolio weights based on the shrinkage prior are much more conservative than weights based on the true parameters or on the uniform prior, since investors that use the shrinkage prior are more skeptical about predictability.

7.3 Using dividend-to-price ratio as a predictor

In section 6, we use the price-earnings ratio as one of the predictor variables. Another commonly used predictor variable is the dividend-to-price ratio. In this section, we give results for the plug-in and the decision-theoretic approach for a model in which the dividend-to-price ratio replaces the price-earnings ratio.

Table 9 shows results for the plug-in method. Firstly, we consider the plug-in method combined with unrestricted weights. Again, the performance for an investor with low risk aversion ($\gamma = 2$), unrestricted portfolio weights and the uniform prior is very bad with a CER of -100%. This time, however, the performance for higher risk aversion levels is very bad as well, i.e. CERs are often negative and are substantially lower than the ones for the benchmark models. The DP model is apparently misspecified.

Under the shrinkage prior, results substantially improve. Negative CERs become positive and benchmark models are outperformed. Investors should again time the stock and bond market when using the shrinkage prior. Despite the misspecified DP model, differences are still economically important. Next, dynamic strategies still outperform myopic strategies when using the shrinkage prior. However, the performance differences is quite small.

The table shows that we could have restricted portfolio weights as well, instead of shrinkage, to improve out-of-sample performance for all risk aversion levels. Apparently, the misspecified DP model only gives acceptable out-of-sample results when using some form of shrinkage: either by using a shrinkage prior or by restricting portfolio weights. The results in table 9 indicate that double shrinkage does not work for the DP model. Combining the shrinkage estimator with restricted portfolio weights deteriorates results for all cases. Again, portfolio weight restrictions help the bad models but hurt the good models.

[Table 9 about here.]

Table 10 shows analogous results for the decision-theoretic method. Incorporating parameter uncertainty improves performance for the least risk-averse investors using the uniform prior and for all investors using the shrinkage prior. Performance however deteriorates for more risk-averse investors that use the uniform prior. Again, dynamic strategies marginally outperform myopic strategies in this setting.

[Table 10 about here.]

We conclude that results from previous sections are confirmed. Investors should actively time the stock and bond market, including hedging components marginally improves performance and the shrinkage prior leads to superior results. However, using double shrinkage deteriorates results in this section.

8 Conclusion

We investigate the out-of-sample performance of strategic asset allocations. Our aim is to evaluate if the potential gains from strategic portfolios can be realized out-of-sample. Optimal strategic portfolios are time-varying and include a hedge component. We analyze the importance of both aspects. Furthermore, we introduce a shrinkage prior that downplays the predictability of asset returns and shrinks the model for the predictor variables to a random walk. We investigate whether the shrinkage prior leads to better results for long-term investors. In our analysis, we consider several specifications. We vary the method (plug-in or decision-theoretic), the estimator (uniform prior or shrinkage prior), the strategy (myopic or dynamic) and the portfolio constraints (constrained or unconstrained) for risk aversion levels γ is 2, 5 or 10.

The first important characteristic of optimal strategic portfolios is that they are time-varying. We find that this potential gain can be realized out-of-sample. Long-term investors should let their asset allocations depend on market conditions when they use our proposed shrinkage prior. Their allocations outperform strategies that ignore asset return predictability by margins that are economically (and statistically) significant. The shrinkage prior makes sure that weights are not wildly fluctuating and not too extreme. The standard uniform prior on the other hand does not give satisfactory results. An investor with low risk aversion would have lost all her money if she would have relied on a VAR model estimated with a uniform prior.

Our analysis shows that it is very important for investors to evaluate a prediction model by an asymmetric utility metric. Risk-averse investors value models by their capability of avoiding a disaster (the extreme negative events). It turns out that the shrinkage prior does exactly this. Investors that use the shrinkage prior can still time the market and benefit from good market conditions. However, what distinguishes the shrinkage prior from the standard uniform prior is that it is capable of limiting the losses in extreme negative events.

The second important characteristic of optimal strategic portfolios is the hedge component. We argue that this component is sensitive to estimation error in both long-run predictions of returns and in their covariance with current returns. Our analysis shows that its potential gain translates into only a modest extra performance out-of-sample. In some cases dynamic portfolios outperform repeated myopic portfolios by economically relevant margins (especially if the shrinkage prior is used), but in general differences are not very large. Monte Carlo simulations show that this result is indeed caused by estimation error. Estimated portfolios are more aggressive as their population counterparts. By being more aggressive, the estimated myopic portfolio moves towards the true (unknown) optimal dynamic portfolio. The estimated dynamic portfolio on the other hand moves away from the optimal portfolio. In the data, both rules approximate the true optimal portfolio almost equally well.

The specifications we consider in the paper also differ in the method and in the restrictions imposed. Some additional results are the following. Taking parameter uncertainty into account leads to very modest improvements over methods that only condition on parameter estimates. Brandt, Goyal, Santa-Clara, and Stroud (2005) among others show that incorporating parameter uncertainty does not significantly alter the weights for myopic portfolios. It has a much bigger impact on weights of the hedge component. Our analysis shows that this hedge component only leads to a modest improvement over myopic portfolios in general. In this light, it it not surprising that the incorporation of parameter uncertainty does not lead to a much better performance in this particular case. The effect of weight restrictions on performance is ambiguous. Badly performing specifications perform better if weights are restricted. However, the best performing specifications are hurt if weights are restricted. Hence, restrictions help bad models and hurt the good models as is commonly the case.

A risk-averse investor should combine the shrinkage prior with the plug-in method and unrestricted weights to maximize her expected utility. She should time the stock and bond market. Such an investor increases performance slightly by combining the shrinkage prior with a dynamic strategy in order to take the hedge component into account.

Our paper has a couple of limitations. Firstly, in our analysis we do not take model uncertainty into account. We assume that investors only use one set of predictor variables. An alternative would be to use model selection criteria or Bayesian model averaging (see Cremers (2002) and Avramov (2002)). We do however investigate the sensitivity of performance with respect to the choice of another predictor variable. Secondly, the data generating process (DGP) of asset return and state variable dynamics is assumed not to change over time. We do not consider time-varying parameters or regime-switching models. Next, we focus on asset only investors that maximize the expected utility over terminal wealth. We ignore realistic aspects such as labor income, liabilities or transaction costs. Finally, we ignore hedging against learning due to infeasibility. Brandt, Goyal, Santa-Clara, and Stroud (2005) show that incorporating learning might improve certainty equivalence returns even further. A challenging task for future research will be to develop a solution method that is capable of incorporating learning in a large VAR model such as ours.

References

Avramov, Doran, 2002, Stock return predictability and model uncertainty, *Journal of Financial Economics* 64, 423–458.

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

Bauwens, Luc, Michel Lubrano, and Jean-François Richard, 1999, *Bayesian Inference in Dynamic Econometric Models* (Oxford University Press).

Berger, James, and William Strawdermann, 1996, Choice of hierarchical priors: admissibility in estimation of normal means, *Annals of Statistics* 24, 931–951.

Black, Fischer, and Robert Littermann, 1992, Global portfolio optimization, *Financial Analyst Journal* 48, 28–43.

Brandt, Michael, Amit Goyal, Pedro Santa-Clara, and Jonathan Stroud, 2005, A simulation approach to dynamic portfolio choice with an application to learning about return predictability, *The Review of Financial Studies* 18, 831–873.

Campbell, John, 1987, Stock return and the term structure, *Journal of Financial Economics* 18, 373–399.

——— , 1995, Some lessons from the yield curve, *Journal of Economic Perspectives* 9, 129–152.

——— , Yeung Chan, and Luis Viceira, 2003, A multivariate model of strategic asset allocation, *Journal of Financial Economics* 67, 41–80.

Campbell, John, and Robert Shiller, 1998, Valuation ratios and the long-run stock market outlook, *Journal of Portfolio Management* 24, 11–26.

Campbell, John, and Samuel Thompson, 2008, Predicting excess stock returns out of sample: Can anything beat the historical average?, *The Review of Financial Studies* 21, 1509–1531.

Campbell, John, and Luis Viceira, 2001, Who should buy long-term bonds, *American Economic Review* 91, 99–127.

——— , 2002, *Strategic Asset Allocation* (Oxford University Press).

Chacko, George, and Luis Viceira, 2005, Dynamic consumption and portfolio choice with stochastic volatility in incomplete markets, *The Review of Financial Studies* 18, 1369–1402.

Cochrane, John, 2007, Portfolio theory, Working paper, University of Chicago.

Cremers, Martijn, 2002, Stock return predictability: A bayesian model selection perspective, *Review of Financial Studies* 15, 1223–1249.

DeMiguel, Victor, Lorenzo Garlappi, and Raman Uppal, 2007, Optimal versus naive diversification: How efficient is the $1/n$ portfolio strategy, Forthcoming Review of Financial Studies.

Diebold, Francis, and Roberto Mariano, 1995, Comparing predictive accuracy, *Journal of Business & Economic Statistics* 13, 253–63.

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

Fama, Eugene, and G.William Schwert, 1977, Asset returns and inflation, *Journal of Financial Economics* 5, 115–146.

Geweke, John, 2005, *Contemporary Bayesian Econometrics and Statistics* (Wiley).

Goyal, Amit, and Ivo Welch, 2008, A comprehensive look at the empirical performance of equity premium prediction, *The Review of Financial Studies* 21, 1455–1508.

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

Jagannathan, Ravi, and Tongshu Ma, 2003, Risk reduction in large portfolios: Why imposing the wrong constraints helps, *Journal of Finance* 58, 1651–1684.

Jorion, Philippe, 1986, Bayes-stein estimation for portfolio analysis, *Journal of Financial and Quantitative Analysis* 21.

Jurek, Jakub, and Luis Viceira, 2006, Optimal value and growth tilts in long-horizon portfolios, Working paper, Harvard University.

Koijen, Ralph, Theo Nijman, and Bas Werker, 2009, When can life-cycle investors benefit from time-varying bond risk premia, Forthcoming Review of Financial Studies.

Merton, Robert, 1969, Lifetime portfolio selection under uncertainty: the continuous time case, *Review of Economics and Statistics* 51, 247–257.

——— , 1971, Optimal consumption and portfolio rules in a continous-time model, *Journal of Economic Theory* 3, 373–413.

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

Newey, Whitney, and Kenneth West, 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703–08.

——— , 1994, Automatic lag selection in covariance matrix estimation, *Review of Economic Studies* 61, 631–53.

Ni, Shawn, and Dongchu Sun, 2003, Noninformative priors and frequentist risks of bayesian estimators of vector-autoregressive models, *Journal of Econometrics* 115, 159–197.

Pastor, Lubos, and Robert Stambaugh, 2000, Comparing asset pricing models: an investment perspective, *Journal of Financial Economics* 56, 335–381.

Schotman, Peter, and Herman van Dijk, 1991, A bayesian analysis of the unit root hypothesis in real exchange rates, *Journal of Econometrics* 49, 195–238.

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

van Binsbergen, Jules, and Michael Brandt, 2007, Solving dynamic portfolio choice problems by recursing on optimized portfolio weights or on the value function, *Computational Economics* 29, 355–367.

Wachter, Jessica, and Missaka Warusawitharana, 2009, Predictable returns and asset allocation: Should a skeptical investor time the market?, Forthcoming Journal of Econometrics.

Zellner, Arnold, 1971, *An Introduction to Bayesian Inference in Econometrics* (Wiley).

A Posterior distributions

This section gives details on how we simulate from the posterior and predictive distribution for both the uniform prior, introduced in equation (18), and the shrinkage prior, introduced in equation (19). The posterior mean for the no-predictability prior is derived from the results for the uniform prior.

We first consider the uniform prior given in equation (18). The posterior distribution is as follows

$$P(B^*, \Sigma|Y) \propto |\Sigma|^{-(T+n+1)/2} \exp\left\{-\frac{1}{2} tr\left[(Y^* - XB^{*'})'(Y^* - XB^{*'})\Sigma^{-1}\right]\right\}. \tag{23}$$

It is well-known in the literature (e.g. Zellner (1971)) that the above posterior is the product of the marginal posterior distribution for Σ and the conditional posterior distribution for B^*. These distribution functions look as follows

$$P(\Sigma|Y) = iWishart\,(S, T - n - 1) \tag{24}$$
$$P(\beta^{*'}|\Sigma, Y) = MVN\left(\hat{\beta}^{*'}, \Sigma \otimes (X'X)^{-1}\right), \tag{25}$$

where $\beta^{*'}$ and $\hat{\beta}^{*'}$ are equal to vectorized $B^{*'}$ and $\hat{B}^{*'} = (X'X)^{-1} X'Y^*$ respectively, and $S = (Y^* - X\hat{B}')'(Y^* - X\hat{B}')$. We can simulate from the above posterior by first drawing Σ from the inverse Wishart distribution and then drawing $\beta^{*'}$ given Σ from the multivariate normal distribution.

If we impose the assumption of stationarity, it is not possible to derive an analytical expression for the marginal posterior for Σ by integrating with respect to B^* over its stationarity region. This implies that we have to rely on a Gibbs sampler with the conditional posteriors $\beta^{*'}|\Sigma$, given in equation (25), and $\Sigma|\beta^{*'}$. The latter distribution is an inverted Wishart distribution where S in equation (24) depends on B^* instead of $\hat{B}^*$ and the degrees of freedom are equal to T instead of $T - n - 1$. We use rejection sampling in order to impose stationarity, i.e. we reject draws for B^* that would result in a non-stationary model.

Secondly, consider the shrinkage prior given in equation (19). The posterior distribution is given in the following equation

$$P(B^*, \Sigma) \propto \left(b^{*\prime}b^*\right)^{\frac{-(n(n+1)-2)}{2}} |\Sigma|^{-(T+n+1)/2} \exp\left\{-\frac{1}{2} tr\left[(Y^* - XB^{*'})'(Y^* - XB^{*'})\Sigma^{-1}\right]\right\}, \tag{26}$$

The above posterior does not belong to a known distribution class. Ni and Sun (2003) developed an algorithm that allows us to simulate from the posterior distribution. In order to do so, they introduced a latent variable δ which is needed to simulate B^*. We use a Gibbs sampler, where the following conditional distributions are important

$$P(\Sigma|B^*, Y) = iWishart\left((Y^* - XB^{*'})'(Y^* - XB^{*'}), T\right) \quad (27)$$

$$P(\delta|B^*, Y) = iGamma\left(J/2 - 1, \frac{1}{2}\beta^{*'}\beta^*\right) \quad (28)$$

$$P(B^*|\delta, \Sigma, Y) = MVN\left(\delta(\Sigma \otimes (X'X)^{-1} + \delta I_J)^{-1}\hat{\beta}^*, (\Sigma^{-1} \otimes X'X + \frac{1}{\delta}I_J)^{-1}\right), \quad (29)$$

with $J = n(n+1)$ and I_J the identity matrix of dimension J. We can simply impose the assumption of stationarity by rejecting non-stationary draws as explained above. In order to increase the accuracy of point estimates, we use Rao-Blackwellization techniques if possible. This means that we average conditional means of the parameter draws in order to obtain the (un)conditional posterior means instead of averaging drawn parameter values.

No matter whether we use the uniform or shrinkage prior, we can simulate from the predictive distribution once we have a sample of simulated parameter values. This conditional distribution is given as follows

$$P(y_{t+1}|y_t, B^{(i)}, \Sigma^{(i)}) = MVN(B_0^{(i)} + B_1^{(i)}y_t, \Sigma^{(i)}), \quad (30)$$

where $B_0^{(i)}$, $B_1^{(i)}$ and $\Sigma^{(i)}$ are drawn parameter values. Note that we first have to transform $B_1^{*(i)}$ into $B_1^{(i)}$ before we are able to simulate future values of y_t. We use antithetic sampling. This means that we simulate two antithetic scenarios of future returns and state variables for each parameter draw. It is a more efficient and accurate way to simulate from the predictive distribution.

We use the ML estimates for the initialization of the Gibbs samplers. We draw 25,000 parameter estimates in total, but discard the first 5,000 draws. This results in 40,000 asset return and state variable paths. Increasing the burn-in phase or the number of simulations does not significantly impact the results. Visual inspection of the posterior draws, CUMSUM statistics proposed in Bauwens, Lubrano, and Richard (1999) and the equality of means test proposed in Geweke (2005) suggest that estimates converge.

B Numerical method

This section elaborates on the numerical methods used in this paper. We show how the parameterization of regression coefficients works and give an indication of the accuracy of our methods. Our method is based on the observation made in Koijen, Nijman, and Werker (2009) in a different setting that the regression coefficients in step 4 of section 5.2 have to be a function of portfolio weights and can be parameterized. This works extremely well in our setting. When using empirical illustrations, we estimate the PE model on the full data-set and assume that the estimates are the true values. Allowing for parameter uncertainty does not change conclusions in this section.

For simplicity, assume that we want to maximize power utility over terminal wealth one month in the future

$$\max_{w_t} E\left(\frac{W_{t+1}^{1-\gamma}}{1-\gamma} \mid Z_t\right). \tag{31}$$

In the main paper, the conditioning variables in Z_t are equal to the asset returns and predictor variables in y_t. For illustration purposes, we set the conditioning variables equal to their historical average in this section. The standard approach for solving this problem is to set up a portfolio weight grid and simulate N asset return paths. Then, take a grid point, calculate realized utility for all paths and calculate conditional expected utility for this grid point by averaging the realized utility values. Finally, repeat this for all grid points and pick the portfolio weight that maximizes conditional expected utility.

Since different portfolio weights lead to different conditional utilities, conditional utility obviously has to be a function of portfolio weights. We illustrate this fact in figure 5 where we plot conditional utility versus the portfolio weights. The picture clearly shows a quadratic relation. In fact, if we regress conditional utility on a quadratic function of portfolio weights we get an R^2 near 1. Hence, the following holds

$$\max_{w_t} E\left(\frac{W_{t+1}^{1-\gamma}}{1-\gamma} \mid Z_t\right) = \max_{w_t} f(w_t), \tag{32}$$

where $f(w_t)$ is a quadratic function in the portfolio weight w_t.

[Figure 5 about here.]

In other words, maximizing conditional expected utility on a constrained set is equivalent to maximizing a quadratic function on this same set. This can be done analytically. Since the R^2 in the parameterization regression is almost 1, we do not have to estimate this parameterization regression on a very fine grid: knowing a couple of points is enough.

We can easily generalize the above to a dynamic setting where the conditional utility depends on conditioning variables. As an illustration, assume that the conditional expectation of the value function at time t depends on one conditioning variable Z_t:

$$E\left\{V_{t+1}\left(K-1, W_{t+1}, Z_{t+1}\right) \mid Z_t\right\} = \alpha_{0w_t} + \alpha_{1w_t} Z_t, \tag{33}$$

where α_{0w_t} and α_{1w_t} are coefficients depending on portfolio weights w_t. If we parameterize both coefficients in a quadratic function of portfolio weights $w_{t,s}$ for stocks and $w_{t,b}$ for bonds depending on coefficient vectors γ_0 and γ_1, we get

$$\begin{aligned} E(.. \mid Z_t) = &(\gamma_{00} + \gamma_{10} w_{t,s} + \gamma_{20} w_{t,b} + \gamma_{30} w_{t,s}^2 + \gamma_{40} w_{t,b}^2 + \gamma_{50} w_{t,s} w_{t,b}) + \\ &(\gamma_{01} + \gamma_{11} w_{t,s} + \gamma_{21} w_{t,b} + \gamma_{31} w_{t,s}^2 + \gamma_{41} w_{t,b}^2 + \gamma_{51} w_{t,s} w_{t,b}) Z_t \end{aligned} \tag{34}$$

$$\begin{aligned} E(.. \mid Z_t) = &(\gamma_{00} + \gamma_{01} Z_t) + (\gamma_{10} + \gamma_{11} Z_t) w_{t,s} + (\gamma_{20} + \gamma_{21} Z_t) w_{t,b} + \\ &(\gamma_{30} + \gamma_{31} Z_t) w_{t,s}^2 + (\gamma_{40} + \gamma_{41} Z_t) w_{t,b}^2 + (\gamma_{50} + \gamma_{51} Z_t) w_{t,b} w_{t,s}, \end{aligned} \tag{35}$$

where the second equality follows after collecting terms. Along each path, the conditioning variables are known. Therefore, maximizing the above conditional expectations boils down to maximizing a quadratic function in portfolio weights where conditioning variables can be treated as constants.

In the empirical section in the paper we use 6 conditioning variables.[20] The grid size is only 10 and the number of paths is equal to 40,000. We use a first order polynomial of the conditioning variables, refer to step 4 in section 5.2, and a second order approximation in the parameterization regressions, refer to step 5. Note that this numerical method is very fast, since we only have to consider a grid size of 10 instead of more than 5,000.[21] Our second-order approximation of the regression parameters on the portfolio weights is very accurate, i.e. the R^2 of these parameterization regressions are all larger than 0.999. Larger grid sizes do not influence the results because of this high R^2.

Van Binsbergen and Brandt (2007) show that their method is accurate by comparing their method with the method of Barberis (2000). Their results are similar and therefore these authors conclude that their method is accurate. We provide evidence that our method is accurate by comparing our numerical method with the one used in van Binsbergen and Brandt (2007). We report results in table 11.

From the table it is clear that the two methods are equally accurate, i.e. the impact on accuracy of using our method is negligible. However, our method is around 500 times faster since we only have to consider a grid of 10 points instead of more than 5,000!

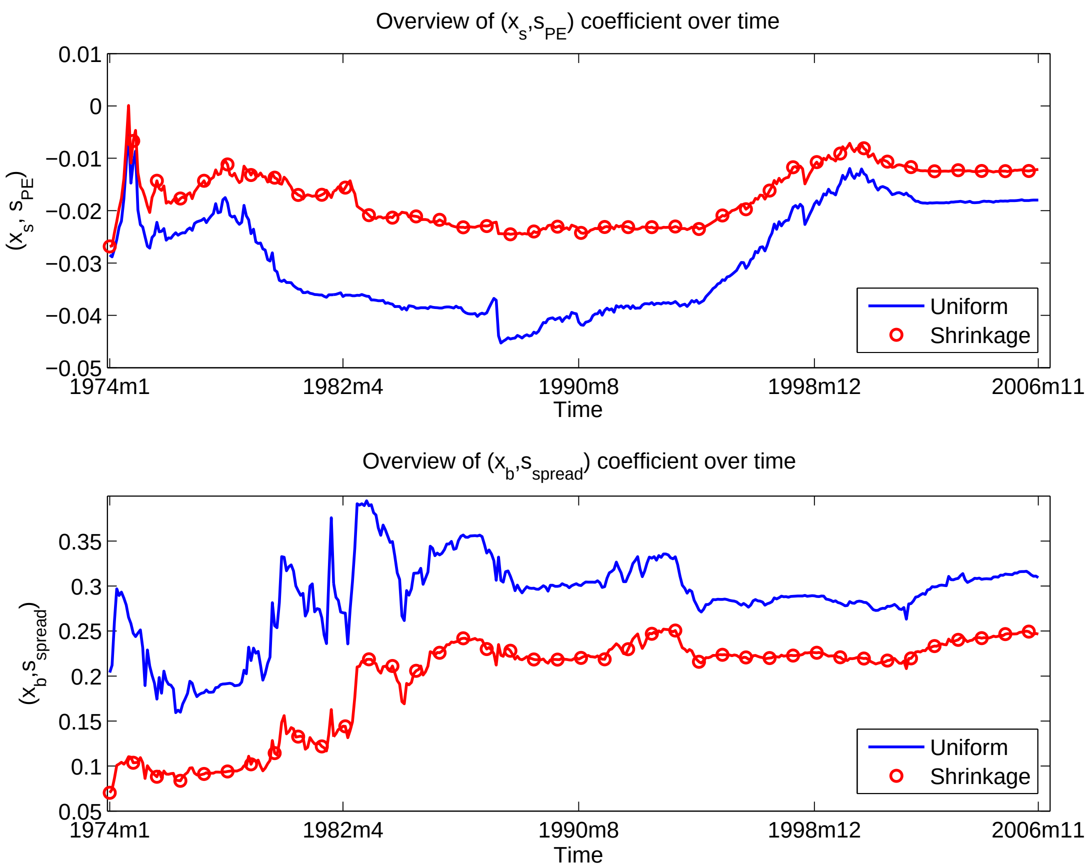
[Table 11 about here.]

[20] The current values of asset returns and state variables

[21] Portfolio weights for the stock index, government bond and real T-bill rate should all be non-negative and add up to 1.

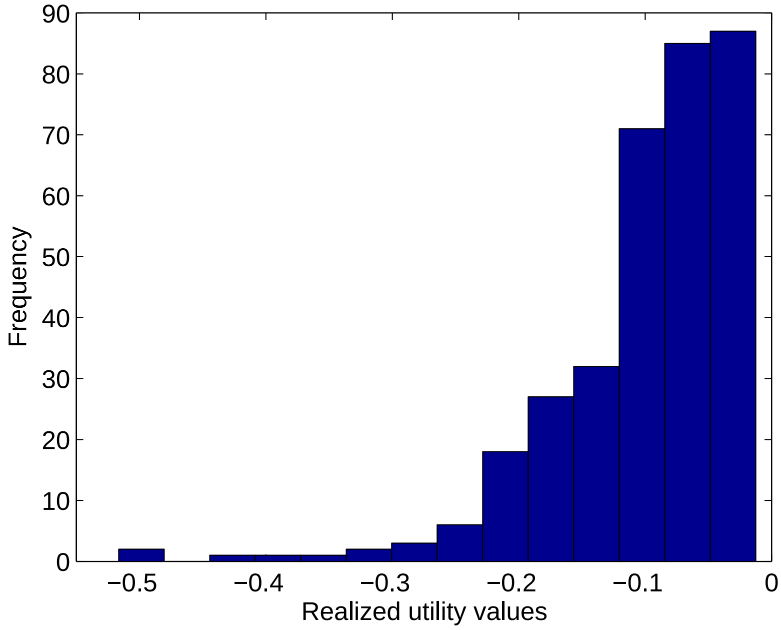
List of Figures

Figure 1: Overview of (x_s,s_{PE}) and (x_b,s_{spread}) coefficients over time



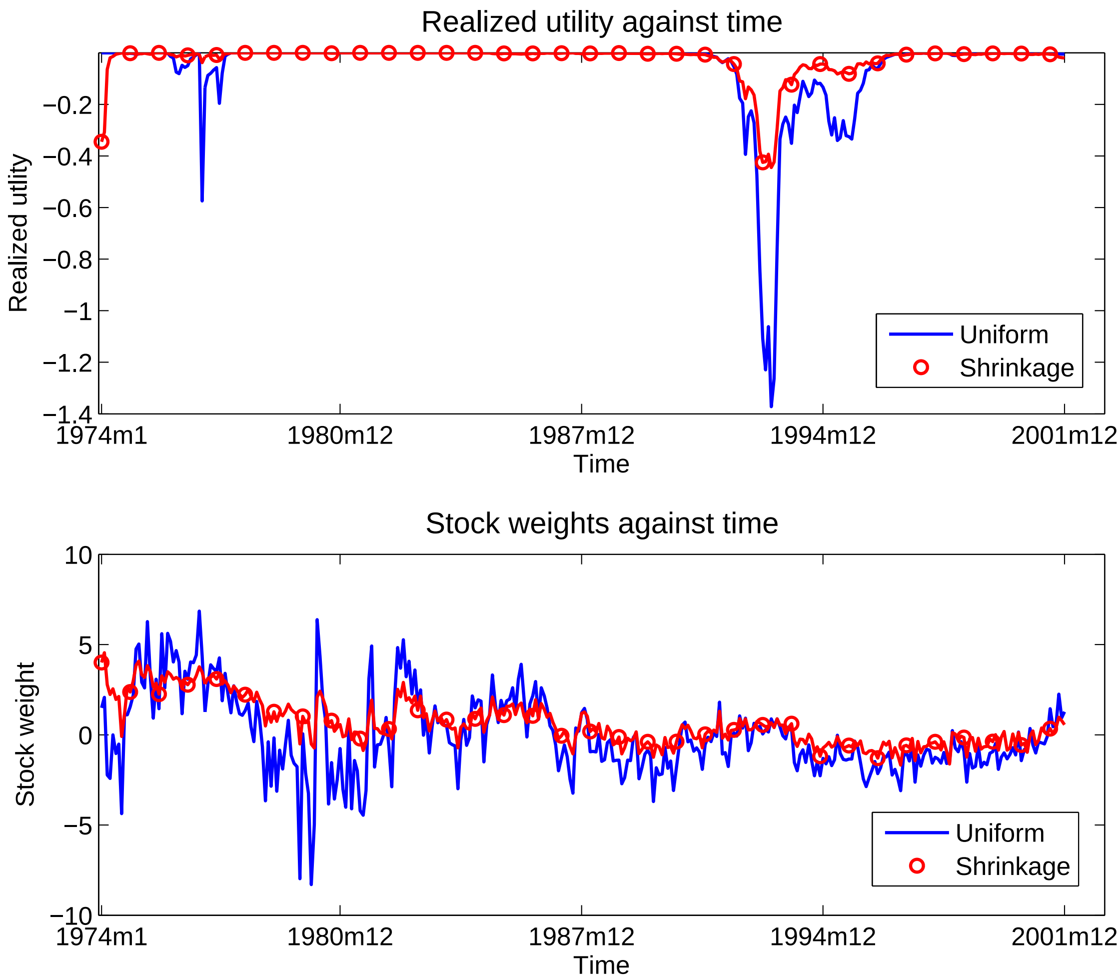
This figure plots the posterior mean of the coefficients (x_s,s_{PE}) and (x_b,s_{spread}) (y-axis) against time (x-axis) for the uniform and shrinkage prior. The model is estimated from February 1954 until the date on the x-axis.

Figure 2: Histogram of realized utility values for benchmark strategy with $\gamma = 5$



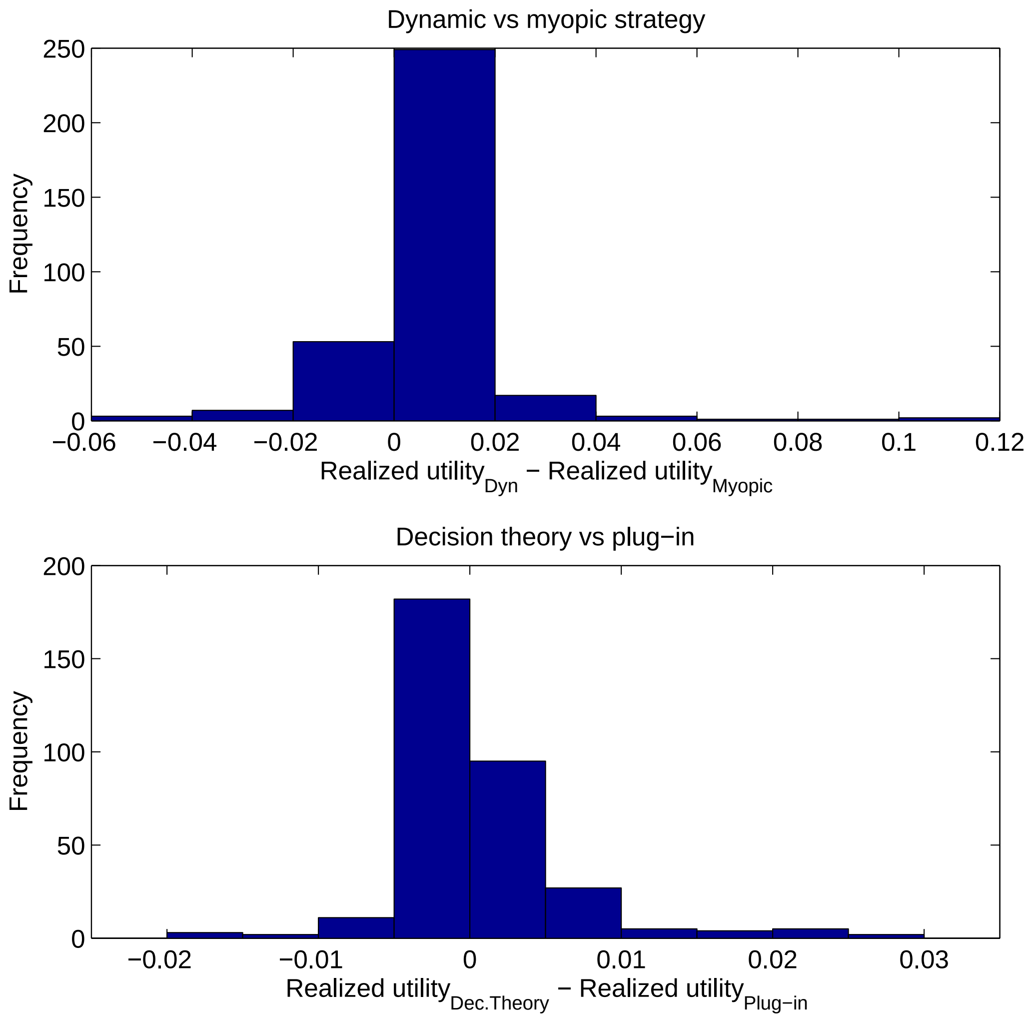
This figure gives a histogram of realized utility values for the unrestricted dynamic no-predictability strategy with $\gamma = 5$. We use the plug-in method.

Figure 3: Realized utility values and stock weights against time for different priors



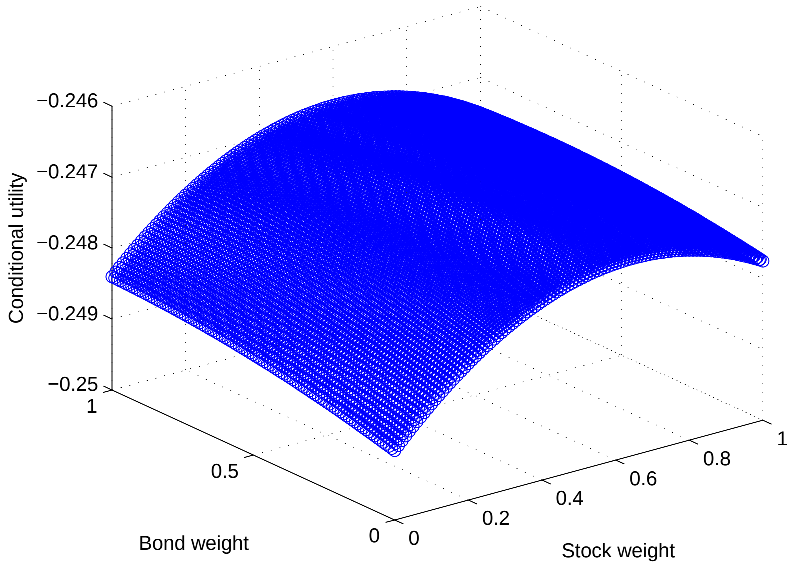
This figure plots realized utility values and stock weights against time for the uniform and shrinkage prior using the plug-in method. We consider a dynamic strategy, $\gamma = 5$ and unrestricted portfolio weights. The x-axis is indexed by the time at which the investors start investing. The second plot shows the stock weights the investors use at the beginning of their investment period.

Figure 4: Histogram of difference in realized utility: dynamic versus myopic and decision-theory vs plug-in



The first figure is a histogram of the difference in realized utility values between a dynamic and myopic strategy, using the unrestricted shrinkage model combined with the plug-in method for an investor with $\gamma = 5$. The second figure gives a histogram of the difference in realized utility values between the decision-theory method and the plug-in method using a restricted dynamic strategy combined with the uniform prior for an investor with $\gamma = 5$.

Figure 5: Conditional utility versus portfolio weights



This figure plots conditional utility over terminal wealth against the portfolio weight in stocks and the portfolio weight in bonds. We impose short-selling constraints which implies that only the subregion for which weights add up to 1 is feasible.

List of Tables

Table 1: **Summary Statistics**

This table reports the means, standard deviations, minima, maxima and AR(1) coefficients for the ex post T-bill rate (R_{tbill}), the excess stock return (X_s), the excess bond return (X_b), the nominal yield (Y_{nom}), the dividend-to-price ratio (DP), the price-earnings ratio (PE) and the yieldspread (Y_{spr}). The monthly data set starts in February 1954 and ends in December 2006. Percentages are given as fractions.

	R_{tbill}	X_s	X_b	Y_{nom}	DP	PE	Y_{spr}
Mean	0.0010	0.0048	0.0011	0.0501	-3.5339	2.8565	0.0112
Std dev.	0.0030	0.0428	0.0148	0.0261	0.3820	0.4141	0.0091
Min	-0.0112	-0.2607	-0.0692	0.0058	-4.5637	1.8929	-0.0160
Max	0.0112	0.1483	0.0898	0.1443	-2.8452	3.7887	0.0421
AR(1)	0.3831	0.0722	0.1089	0.9837	0.9930	0.9968	0.9193

Table 2: **Estimation results PE model**

This table reports estimates for the VAR(1) model based on the full data-set where we use PE among the state variables. Panel A gives results for the uniform prior and panel B for the shrinkage prior. In each panel, columns 2-7 show the posterior mean of the slope coefficients and their posterior standard deviations. The last column shows the implied R^2 (implied by the the posterior mean). Finally, the correlation matrix of the error terms is given. The elements on the diagonal are the standard deviations(x100) of the error terms, the off-diagonal elements are the correlations.

	Panel A: Uniform prior						
	r_{tbill}	x_s	x_b	s_y	s_{PE}	s_{spread}	R^2
	Parameter estimates						
r_{tbill}	0.3242	0.0027	0.0078	0.0276	0.0011	0.0412	0.1808
	0.0383	*0.0026*	*0.0076*	*0.0058*	*0.0003*	*0.0136*	
x_s	1.6434	0.0240	0.3249	-0.3521	-0.0180	-0.0822	0.0579
	0.5972	*0.0400*	*0.1181*	*0.0897*	*0.0054*	*0.2117*	
x_b	0.4215	-0.0569	0.0787	0.0410	0.0017	0.3127	0.0764
	0.2040	*0.0137*	*0.0401*	*0.0308*	*0.0019*	*0.0723*	
s_y	-0.0813	0.0144	-0.0652	0.9855	-0.0001	0.0206	0.9730
	0.0615	*0.0041*	*0.0120*	*0.0093*	*0.0006*	*0.0217*	
s_{PE}	1.3611	0.4168	0.3114	-0.1423	0.9917	0.1227	0.9954
	0.4025	*0.0269*	*0.0790*	*0.0602*	*0.0036*	*0.1421*	
S_{spread}	0.0056	-0.0048	-0.0650	0.0070	-0.0002	0.9503	0.8541
	0.0497	*0.0033*	*0.0098*	*0.0075*	*0.0005*	*0.0176*	
	Error correlation matrix						
r_{tbill}	0.2702	0.1052	0.0757	-0.0805	0.1727	0.0560	
x_s		4.2049	0.1128	-0.0487	0.7746	-0.0331	
x_b			1.4400	-0.6237	0.0557	0.2208	
s_y				0.4328	-0.0494	-0.8516	
s_{PE}					2.8249	-0.0219	
S_{spread}						0.3503	
	Panel B: Shrinkage prior						
	Parameter estimates						
r_{tbill}	0.2730	0.0029	0.0068	0.0297	0.0012	0.0414	0.1782
	0.0371	*0.0026*	*0.0075*	*0.0057*	*0.0003*	*0.0134*	
x_s	0.0099	0.0261	0.1481	-0.2037	-0.0121	-0.0039	0.0381
	0.1173	*0.0374*	*0.0840*	*0.0673*	*0.0048*	*0.0978*	
x_b	0.1061	-0.0537	0.0775	0.0410	0.0016	0.2487	0.0705
	0.1054	*0.0135*	*0.0377*	*0.0285*	*0.0018*	*0.0631*	
s_y	-0.0173	0.0137	-0.0650	0.9854	-0.0001	0.0314	0.9729
	0.0456	*0.0041*	*0.0117*	*0.0089*	*0.0005*	*0.0202*	
s_{PE}	0.1038	0.4144	0.2031	-0.0564	0.9950	0.1360	0.9953
	0.1132	*0.0257*	*0.0610*	*0.0469*	*0.0033*	*0.0792*	
S_{spread}	-0.0086	-0.0046	-0.0644	0.0066	-0.0002	0.9480	0.8541
	0.0428	*0.0033*	*0.0097*	*0.0074*	*0.0004*	*0.0170*	
	Error correlation matrix						
r_{tbill}	0.2706	0.1110	0.0792	-0.0828	0.1783	0.0566	
x_s		4.2400	0.1195	-0.0535	0.7789	-0.0316	
x_b			1.4432	-0.6247	0.0647	0.2209	
s_y				0.4331	-0.0552	-0.8511	
s_{PE}					2.8541	-0.0202	
S_{spread}						0.3503	

Table 3: **Benchmark Results**

This table gives benchmark results for the $1/N$, the unrestricted no predictability and the restricted no predictability strategies (NP). We either calculate a dynamic or a myopic strategy for the latter two. Specifications are based on the plug-in method. We show annualized certainty equivalence returns (CER), average terminal wealth ($\overline{TW}$) and the standard deviation of terminal wealth ($\sigma(TW)$) for three different risk aversion levels γ.

		Unrestricted Weights			**Restricted Weights**		
		CER	$\overline{TW}$	$\sigma(TW)$	CER	$\overline{TW}$	$\sigma(TW)$
	Panel A: $\gamma = 2$						
$1/N$					0.0466	1.2885	0.2069
NP	Dyn	0.0886	1.8030	0.8159	0.0706	1.5444	0.4731
	Myop	0.0855	1.7859	0.8251	0.0706	1.5444	0.4731
	Panel B: $\gamma = 5$						
$1/N$					0.0386	1.2885	0.2069
NP	Dyn	0.0478	1.3678	0.2503	0.0450	1.3560	0.2540
	Myop	0.0443	1.3475	0.2571	0.0439	1.3586	0.2686
	Panel C: $\gamma = 10$						
$1/N$					0.0273	1.2885	0.2069
NP	Dyn	0.0265	1.2355	0.1457	0.0279	1.2350	0.1452
	Myop	0.0247	1.2151	0.1463	0.0258	1.2253	0.1519

Table 4: **Plug-in approach - PE model**

This table gives the results for the dynamic (Dyn) and myopic (Myop) strategies using the plug-in method. The results are based on a VAR(1) model with PE as one of the predictors. We report results under the uniform and shrinkage prior, either using restricted or unrestricted portfolio weights. We report annualized certainty equivalence returns (CER), average terminal wealth ($\overline{TW}$) and the standard deviation of terminal wealth ($\sigma(TW)$) for three different risk aversion levels γ.

		Unrestricted Weights			**Restricted Weights**		
	Panel A: $\gamma = 2$						
		CER	$\overline{TW}$	$\sigma(TW)$	CER	$\overline{TW}$	$\sigma(TW)$
Uniform	Dyn	-1.0000	23.7044	68.0114	0.0821	1.5796	0.4544
	Myop	-1.0000	20.7321	56.5669	0.0814	1.5683	0.4351
Shrinkage	Dyn	0.2961	10.1712	8.9929	0.0815	1.5617	0.3935
	Myop	0.2752	8.7933	8.0463	0.0808	1.5539	0.3881
	Panel B: $\gamma = 5$						
Uniform	Dyn	0.0769	6.3876	6.8324	0.0645	1.5260	0.4046
	Myop	0.0785	4.5576	4.2663	0.0643	1.5137	0.3901
Shrinkage	Dyn	0.1231	3.3727	1.6823	0.0670	1.5222	0.3468
	Myop	0.1164	2.8407	1.2112	0.0659	1.4967	0.3320
	Panel C: $\gamma = 10$						
Uniform	Dyn	0.0430	2.9673	2.0021	0.0492	1.4826	0.3568
	Myop	0.0552	2.3598	1.2337	0.0452	1.4421	0.3235
Shrinkage	Dyn	0.0661	2.0022	0.6181	0.0500	1.4405	0.2747
	Myop	0.0622	1.7870	0.4561	0.0489	1.4000	0.2421

Table 5: **Decision-theoretic approach - PE model**

This table gives the results for the dynamic (Dyn) and myopic (Myop) strategies using the decision-theoretic approach. The results are based on a VAR(1) model with PE as one of the predictors. We report results under the uniform and shrinkage priors and use restricted portfolio weights. We report annualized certainty equivalence returns (CER), average terminal wealth ($\overline{TW}$) and the standard deviation of terminal wealth ($\sigma(TW)$) for three different risk aversion levels γ.

	Restricted Weights			
	Panel A: $\gamma = 2$			
		CER	$\overline{TW}$	$\sigma(TW)$
Uniform	Dyn	0.0830	1.5770	0.4312
	Myop	0.0826	1.5722	0.4227
Shrinkage	Dyn	0.0831	1.5647	0.3729
	Myop	0.0825	1.5596	0.3731
	Panel B: $\gamma = 5$			
Uniform	Dyn	0.0650	1.5313	0.4119
	Myop	0.0652	1.5127	0.3814
Shrinkage	Dyn	0.0682	1.5220	0.3483
	Myop	0.0656	1.4911	0.3255
	Panel C: $\gamma = 10$			
Uniform	Dyn	0.0516	1.4723	0.3372
	Myopic	0.0483	1.4441	0.3182
Shrinkage	Dyn	0.0509	1.4238	0.2571
	Myopic	0.0486	1.4038	0.2565

Table 6: **Classical significance tests**

This table presents classical t-statistics to test whether the performance of the portfolio strategy and its benchmark are statistically significant from each other. We give results for the plug-in approach, the decision-theoretic approach, different risk aversion levels, different types of strategies and for different weight restrictions.

		$\gamma = 2$		$\gamma = 5$		$\gamma = 10$	
		Unr	Restr	Unres	Restr	Unr	Restr
	Panel A: plug-in approach						
Uniform	Dyn	NaN	0.9208	1.1372	2.7858	1.1683	3.7736
	Myopic	NaN	0.8615	1.7996	2.6184	2.5115	3.4862
Shrinkage	Dyn	4.3796	0.7716	3.6705	2.4392	2.6949	3.5770
	Myopic	4.2474	0.7136	3.9038	2.2459	3.4290	3.5485
	Panel B: decision-theoretic approach						
Uniform	Dyn		0.9757		2.7631		3.7395
	Myopic		0.9476		2.6611		3.7302
Shrinkage	Dyn		0.8541		2.4885		3.4696
	Myopic		0.8039		2.2149		3.5315

Table 7: **Monte Carlo simulation without predictability**

This table gives results for 100 Monte Carlo simulations assuming no predictability. The specifications differ in the strategy (dynamic or myopic) and in the risk aversion level. The entries in panel A and B are respectively defined as $CER_{Uni} - CER_{Nopred}$ and $CER_{Shr} - CER_{Nopred}$. Data indicates the result found in the actual data-set (based on table 4). Mean, median, min, 1st, 5th, 95th, 99th and max respectively indicate the average difference in CERs, median difference, minimum difference, 1st percentile of differences, 5th percentile of differences, 95th percentile of difference, 99th percentile of difference and maximum difference in the Monte Carlo simulations.

	Dynamic			Myopic		
	$\gamma = 2$	$\gamma = 5$	$\gamma = 10$	$\gamma = 2$	$\gamma = 5$	$\gamma = 10$
	Panel A: Uniform prior					
Data	-1.0886	0.0291	0.0165	-1.0855	0.0342	0.0305
Mean	-0.1965	-0.0567	-0.0264	-0.1325	-0.0387	-0.0176
Median	-0.1397	-0.0524	-0.0237	-0.1090	-0.0339	-0.0154
Min	-1.2249	-0.2571	-0.1394	-1.0669	-0.1481	-0.0745
1st	-1.1952	-0.2559	-0.1337	-0.7526	-0.1477	-0.0732
5th	-0.8486	-0.1782	-0.0840	-0.3667	-0.1171	-0.0556
95th	0.0062	0.0242	0.0162	0.0079	0.0108	0.0069
99th	0.0828	0.0438	0.0243	0.0956	0.0535	0.0298
Max	0.1101	0.0456	0.0283	0.1455	0.0649	0.0314
	Panel B: Shrinkage prior					
Data	0.2075	0.0753	0.0396	0.1897	0.0721	0.0375
Mean	-0.0720	-0.0261	-0.0121	-0.0576	-0.0186	-0.0084
Median	-0.0629	-0.0227	-0.0109	-0.0589	-0.0181	-0.0061
Min	-0.3270	-0.1586	-0.0820	-0.2227	-0.0896	-0.0458
1st	-0.3070	-0.1453	-0.0737	-0.2178	-0.0855	-0.0434
5th	-0.2081	-0.1035	-0.0543	-0.1614	-0.0647	-0.0317
95th	0.0392	0.0284	0.0159	0.0413	0.0213	0.0116
99th	0.1045	0.0419	0.0248	0.1151	0.0539	0.0294
Max	0.1265	0.0422	0.0261	0.1540	0.0620	0.0297

Table 8: **Monte Carlo simulation with predictability**

This table gives results for 100 Monte Carlo simulations assuming predictability. We compare $(CER_{dyn} - CER_{myop})$ for specifications using either the true parameters (panel A), the estimated parameters under a uniform prior (Panel B) or the estimated parameters under a shrinkage prior (Panel C). Secondly, we give the average differences (over the simulations) in root mean squared portfolio weight error between the dynamic and myopic strategies: $\overline{\Delta RMSPE_{w_s}} = RMSPE_{w_s,dyn} - RMSPE_{w_s,myopic}$ and $\overline{\Delta RMSPE_{w_b}} = RMSPE_{w_b,dyn} - RMSPE_{w_b,myopic}$. Specifications differ in their risk aversion level. Data indicates the result found in the actual data-set.

	$\gamma = 2$	$\gamma = 5$	$\gamma = 10$
	Panel A: True		
Mean	0.0146	0.0179	0.0111
Median	0.0200	0.0231	0.0124
Min	-0.2904	-0.1237	-0.0551
1st	-0.2654	-0.0717	-0.0335
5th	-0.0275	-0.0072	-0.0013
95th	0.0541	0.0386	0.0225
99th	0.0696	0.0475	0.0290
Max	0.0785	0.0508	0.0311
	Panel B: Uniform		
Data	0.0000	-0.0016	-0.0122
Mean	-0.1649	-0.0037	0.0023
Median	-0.0171	0.0100	0.0098
Min	-2.0126	-0.1805	-0.1062
1st	-1.9797	-0.1794	-0.1057
5th	-1.4106	-0.1057	-0.0609
95th	0.0723	0.0547	0.0375
99th	0.1548	0.0790	0.0479
Max	0.1878	0.0842	0.0498
$\overline{\Delta RMSPE_{w_s}}$	0.2113	0.2232	0.1628
$\overline{\Delta RMSPE_{w_b}}$	0.3210	0.3321	0.2442
	Panel C: Shrinkage		
Data	0.0209	0.0067	0.0039
Mean	-0.0041	-0.0006	0.0002
Median	0.0084	0.0075	0.0043
Min	-0.2320	-0.1430	-0.0857
1st	-0.2265	-0.1362	-0.0773
5th	-0.0940	-0.0740	-0.0413
95th	0.0443	0.0352	0.0228
99th	0.0656	0.0499	0.0307
Max	0.0730	0.0546	0.0339
$\overline{\Delta RMSPE_{w_s}}$	0.0029	0.0040	0.0422
$\overline{\Delta RMSPE_{w_b}}$	-0.2874	-0.2454	-0.1641

Table 9: **Plug-in approach - DP model**

This table gives the results for the dynamic (Dyn) and myopic (Myop) strategies using the plug-in method. We use a VAR(1) model with DP as one of the predictors for a robustness check. We report results under the uniform and shrinkage priors and either use restricted or unrestricted portfolio weights. We report annualized certainty equivalence returns (CER), average terminal wealth ($\overline{TW}$) and the standard deviation of terminal wealth ($\sigma(TW)$) for three risk aversion levels γ.

		Unrestricted Weights			**Restricted Weights**		
	Panel A: $\gamma = 2$						
		CER	$\overline{TW}$	$\sigma(TW)$	CER	$\overline{TW}$	$\sigma(TW)$
Uniform	Dyn	-1.0000	22.7307	58.7851	0.0755	1.4929	0.3081
	Myop	-1.0000	17.7485	42.5112	0.0758	1.4941	0.3061
Shrinkage	Dyn	0.2192	4.5321	3.1417	0.0682	1.4351	0.2661
	Myop	0.2041	4.0242	2.7603	0.0683	1.4357	0.2647
	Panel B: $\gamma = 5$						
Uniform	Dyn	-0.0857	5.5379	5.4991	0.0661	1.5109	0.3324
	Myop	0.0010	3.9889	3.5361	0.0647	1.5024	0.3532
Shrinkage	Dyn	0.0946	2.1192	0.6630	0.0588	1.4420	0.2827
	Myop	0.0883	1.9085	0.5513	0.0552	1.3956	0.2449
	Panel C: $\gamma = 10$						
Uniform	Dyn	-0.0500	2.7785	1.8400	0.0499	1.4814	0.3335
	Myop	0.0139	2.2197	1.1653	0.0474	1.4365	0.3177
Shrinkage	Dyn	0.0521	1.5547	0.2946	0.0414	1.3587	0.2241
	Myop	0.0482	1.4506	0.2403	0.0376	1.3080	0.1914

Table 10: **Decision-theoretic approach - DP model**

This table gives the results for the dynamic (Dyn) and myopic (Myop) strategies using the decision-theoretic approach. We use a VAR(1) model with DP as one of the predictors as a robustness check. We report results under the uniform and shrinkage priors and use restricted portfolio weights. We report annualized certainty equivalence returns (CER), average terminal wealth ($\overline{TW}$) and the standard deviation of terminal wealth ($\sigma(TW)$) for three risk aversion levels γ.

	Restricted Weights			
	Panel A: $\gamma = 2$			
		CER	$\overline{TW}$	$\sigma(TW)$
Uniform	Dyn	0.0756	1.4939	0.3084
	Myop	0.0759	1.4947	0.3041
Shrinkage	Dyn	0.0723	1.4620	0.2655
	Myop	0.0730	1.4680	0.2726
	Panel B: $\gamma = 5$			
Uniform	Dyn	0.0647	1.5085	0.3448
	Myop	0.0636	1.4923	0.3443
Shrinkage	Dyn	0.0599	1.4446	0.2778
	Myop	0.0570	1.4102	0.2530
	Panel C: $\gamma = 10$			
Uniform	Dyn	0.0476	1.4732	0.3434
	Myop	0.0471	1.4320	0.3142
Shrinkage	Dyn	0.0413	1.3515	0.2191
	Myop	0.0382	1.3181	0.2039

Table 11: **Comparison accuracy numerical methods**

This table compares the portfolio weights obtained by the simulation method in van Binsbergen and Brandt (2007)(BB2007) with the portfolio weights obtained by using the refined method of this paper (DPS2010). We give the portfolio weights for a dynamic strategy with K periods remaining for stocks, w_s, and bonds, w_b. Results are based on the plug-in method. We vary parameter K and risk aversion γ. State variables are set to their historical average.

		BB2007		DPS2010	
K	γ	w_s	w_b	w_s	w_b
1	2	1.0000	0.0000	1.0000	0.0000
	5	0.5600	0.4400	0.5644	0.4356
	10	0.3000	0.4400	0.2974	0.4428
4	2	1.0000	0.0000	1.0000	0.0000
	5	0.6000	0.4000	0.5959	0.4041
	10	0.3100	0.2900	0.3110	0.2853
8	2	1.0000	0.0000	1.0000	0.0000
	5	0.6200	0.3800	0.6185	0.3815
	10	0.3200	0.2700	0.3198	0.2736
15	2	1.0000	0.0000	1.0000	0.0000
	5	0.6500	0.3500	0.6530	0.3470
	10	0.3400	0.3000	0.3420	0.2963
30	2	1.0000	0.0000	1.0000	0.0000
	5	0.6900	0.3100	0.6923	0.3077
	10	0.3700	0.3100	0.3651	0.3150
60	2	1.0000	0.0000	1.0000	0.0000
	5	0.7500	0.2500	0.7469	0.2531
	10	0.4100	0.2000	0.4097	0.2013