Netspar Network for Studies on Pensions, Aging and Retirement

Bart Diris

Franz Palm

Peter Schotman

Long-Term Strategic Asset Allocation: An Out-of-Sample Evaluation

Discussion Paper 03/2008 - 053

March 18, 2008

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

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

March 18, 2008

Abstract

We investigate the out-of-sample performance of realistic strategic asset allocation models by analyzing plug-in and decision-theoretic approaches with restricted and unrestricted portfolio weights. This paper makes three major contributions. Firstly, this paper refines existing numerical techniques for the calculation of dynamic strategies. Secondly, these refined techniques are used to test myopic, constant proportion and dynamic strategies empirically. Finally, we propose a shrinkage prior with superior performance. The empirical results show that myopic, constant proportion and dynamic strategies could lead to very unstable results unless shrinkage estimators are applied. Dynamic strategies outperform myopic strategies only when using shrinkage. Certainty equivalence returns increase to more than 10% annually since shrinkage-based strategies avoid extreme events and result in less variable portfolio weights. Shrinkage turns bad models into good models and good models into great models. Finally, incorporating parameter uncertainty improves performance when restricting portfolio weights.

Affiliations: Maastricht University[a], Netspar[b], CEPR[c]

*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@finance.unimaas.nl)

1 Introduction

Merton (1969, 1971) showed that under changing investment opportunities, optimal portfolios of short-term and long-term investors differ. The latter contains a term hedging against the changes in these investment opportunities. This fact together with the recent empirical evidence of predictability in asset returns has lead to a revived interest in strategic asset allocation problems[1] in recent years. These problems ask how one should divide his/her funds between asset classes over time.

Papers, such as Campbell, Chan, and Viceira (2003), Jurek and Viceira (2006) and Barberis (2000) to name a few, estimate a model on data and show ex post how an investor should have chosen an optimal portfolio. The literature takes realistic aspects such as predictability, parameter uncertainty and learning effects into account and shows the effect on the optimal allocation for investors. Campbell, Chan, and Viceira (2003) show that the predictability of stock returns implies mean reversion and that this should result in an increased allocation to stocks at longer horizons. Barberis (2000) and Brandt, Goyal, Santa-Clara, and Stroud (2005) argue that parameter uncertainty mitigates this effect and that hedging terms can even be negative when taking learning into account. Investors that want to use such models in reality might therefore be wondering which aspects are important and which ones are not.

Recently, a series of papers question the evidence on return predictability. Goyal and Welch (2007) examine whether one can predict stock returns out-of-sample and their results indicate that the historical average does a better job than all predictor variables considered. Their results would mean that it is not possible to accurately model a time-varying set of asset returns. For strategic asset allocation models, this would mean that investment opportunities might be constant and that optimal long-term and short-term portfolios might be more similar than previously thought. Another series of papers defends the empirical evidence on stock return predictability. Campbell and Thompson (2007) provide evidence on stock return predictability by using economic theory next to data. This would suggest that changing investment opportunities can be modeled, but that prior information needs to be used next to data.

The open question remains how investors can benefit from all these findings in the strategic asset allocation and return predictability literature in real-time. Is it really beneficial for investors to calculate dynamic strategies that allow for market timing and hedging against changes in investment opportunity sets or might it be better to base investment decisions on unconditional moments? In other words, what works best out-of-sample?

This paper tries to answer these questions. We perform an extensive analysis on the out-of-sample performance of realistic strategic asset allocation models using Bayesian methods. We contribute to the literature in three important ways. Firstly, we refine existing numerical procedures to calculate (restricted) portfolio weights in a more efficient way. Secondly, we perform an out-of-sample analysis that covers the post-war period. In the out-of-sample analysis, we compare the relative performance of myopic, dynamic and constant proportion strategies using

[1] Campbell and Viceira (2002) and Brandt (2004) provide extensive surveys of the existing literature

either a plug-in or decision-theoretic approach. Without the refined numerical method, such an extensive analysis would not be computationally feasible in our setting. Finally, we improve out-of-sample results by introducing a Bayesian shrinkage estimator in the portfolio literature based on Ni and Sun (2003). The latter estimator works especially well when combining unrestricted portfolio weights with the plug-in method.

We find that our numerical method is up to 100 times faster than existing numerical methods in the literature with a negligible impact on accuracy. Our empirical results show that the out-of-sample performance of unrestricted strategic asset allocation models can be disastrous without shrinkage. An investor with low risk aversion would have lost all his money if she would have applied such a model in the past 30 years. Furthermore, dynamic strategies are inferior to myopic strategies. Shrinkage estimators change this picture. Shrinkage turns bad models into good models and good models into great models. These results are robust to model misspecification. Dynamic strategies are superior to myopic strategies if shrinkage is used. Our results indicate that another way to avoid extremely bad results consists in imposing constraints on the portfolio weights. This finding is in line with Jagannathan and Ma (2003). However, using shrinkage is substantially better than restricting portfolio weights in order to avoid extreme events. The optimal restricted strategy turns out to be one where parameter uncertainty is taken into account. Additional tests show that the latter is not robust to a (more) misspecified model.

The closest paper to ours is Wachter and Warusawitharana (2007). These authors also consider a shrinkage prior and show that this prior might improve out-of-sample results. However, these authors only consider myopic investors, while we consider investors with a long horizon and implement dynamic strategies. Furthermore, we investigate an extended version of their strategic asset allocation model by considering multiple predictor variables. Finally, our prior leaves less room for subjectivity.

Our paper also contributes to the return predictability literature by investigating the economic significance of return predictability in an out-of-sample context. Only using data leads to unstable results as in Goyal and Welch (2007), but adding restrictions and economic theory makes return predictability economically important in out-of-sample experiments as in Campbell and Thompson (2007).

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

2 Data

Our empirical analysis is based on monthly data for the US stock and bond market. We use data on three assets and several predictor variables.

The monthly data set starts in 1954 and ends in 2006. The first three variables are log returns on different types of assets. 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[2], and log inflation, obtained from Center for Research in Security Prices (CRSP). The second variable is excess log stock return, which is defined as the difference between the value weighted log return on the NYSE, NASDAQ and AMEX market and the log return on the 3-month bill. The third variable, excess log bond return, is defined in a similar way, where we are using the five-year bond return from CRSP.

In total, we use four different predictor variables but only include three of them in our models simultaneously. The first predictor variable is the log nominal yield on the 90-day T-Bill. 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, NASDAY and AMEX market. The next predictor variable is the log yield spread which 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. The final predictor variable is the log of the price-earnings ratio and is obtained from the Irrational Exuberance data, available from the website of Professor Shiller[3]. It is defined as the log of the current price over the lagged sum of earnings over the past 10 years. In our results in section 6, we use the three assets, the nominal yield, the price-earnings ratio and the yield spread in our models. As a robustness check, we replace the price-earnings ratio by the dividend-to-price ratio in section 7.

Table 1 provides summary statistics of our data. In all panels, we correct mean log asset returns by one-half their variance to reflect log simple returns. Note that minima of more than -100% are possible since we annualize monthly values.

[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 gives some intuition about the relative performance of different strategies. The last subsection explains the difference between the plug-in and decision-theoretic method.

3.1 General set-up

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 T-bills, stocks and bonds. We choose power utility as utility function. Portfolios with weights restricted to be non-negative and portfolios with weights unrestricted and thus allowing for going short will be considered.

[2]http://research.stlouisfed.org/fred2
[3]http://www.econ.yale.edu/ shiller/data.htm

More formally, the investor chooses portfolio weights $w_t, ...w_{t+K-1}$ such that the value function at time point t is maximized[4]

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

subject to the budget constraint

$$W_{s+1} = W_s\left(1 + w_s' R_{s+1}\right), s = t, ...t + K - 1, \tag{2}$$

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 returns on the assets in period $s + 1$. Portfolio weights add up to 1 and are restricted or unrestricted.

We consider three types of strategies: a dynamic strategy, a constant proportion strategy and a myopic strategy. Since initial wealth is 1, the following equality holds

$$W_{t+K} = \prod_{r=t}^{t+K-1} \left(1 + w_r' R_{r+1}\right). \tag{3}$$

The dynamic strategy is the optimal solution to the long-horizon problem in equation (1). The constant proportion strategy restricts portfolio weights to be equal in all periods $w_t = ... = w_{t+K-1}$. Finally, the myopic strategy ignores the long horizon, sets portfolio weights as if the remaining horizon is only one period and hence ignores hedging effects. More formally, the dynamic $w_{t,D}$, constant proportion $w_{t,C}$ and myopic strategies $w_{t,M}$ are defined as follows

$$\{w_{t,D}, ..., w_{t+K-1,D}\} = \arg\max E\left\{\frac{\left(\prod_{r=t}^{t+K-1}(1 + w_{r,D}' R_{r+1})\right)^{1-\gamma}}{1-\gamma} \mid Z_t\right\} \tag{4}$$

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

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

If horizon $K = 1$, the three strategies are equal. Section 3.2 gives some intuition about the relative performance of the strategies.

In our empirical 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

[4]Relative risk aversion parameter γ is assumed to be larger than 1, $\gamma = 1$ leads to the log utility function.

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 shortly 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.

The econometric model is needed to evaluate the conditional expectation over conditioning variables and asset returns in equation (1). Following among others Campbell, Chan, and Viceira (2003) and Jurek and Viceira (2006), we model the dynamics of asset returns and state variables (our data) by using a VAR(1) as the econometric model. For this, define the $nx1$ vector y_t as follows

$$y_t = \begin{pmatrix} r_{tbill,t} \\ x_t \\ s_t \end{pmatrix}, \tag{7}$$

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 state variables. Vector s_t either consists of the nominal yield $Y_{nom,t}$, the price-earnings ratio PE_t and the yield spread $Y_{spr,t}$ (refer to section 6) or the nominal yield, the dividend-yield DP_t and the yield spread (refer to section 7).

The VAR(1) model is as follows

$$y_{t+1} = B_0 + B_1 y_t + \epsilon_{t+1}, \tag{8}$$

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 assumption

$$\epsilon_{t+1} \sim N(0, \Sigma). \tag{9}$$

For future reference, it is useful to introduce the following decomposition for Σ, consistent with equation (7)

$$\Sigma = \begin{pmatrix} \sigma^2_{tbill} & \sigma'_{tbill,x} & \sigma'_{tbill,s} \\ \sigma_{tbill,x} & \Sigma_x & \Sigma'_{x,s} \\ \sigma_{tbill,s} & \Sigma_{x,s} & \Sigma_s \end{pmatrix}. \tag{10}$$

In this paper, we take a Bayesian perspective and obtain posterior distributions for the parameters depending on different prior distributions.

In the portfolio choice literature, there are two methods that prescribe how to use these estimation results. The first method treats the parameter estimates as the true parameters.

This is the plug-in method. The second method acknowledges that there might be parameter uncertainty which might be taken into account by a parameter distribution. 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 lead to different portfolio weights. We investigate what leads to the best out-of-sample performance for investors by performing an extensive analysis. We consider the following choices for investors with risk aversion γ is 2, 5 or 10:

- Uniform, shrinkage or no-predictability prior (3 choices)
- Dynamic, myopic or constant proportion strategy (3 choices)
- Plug-in or decision-theoretic method (2 choices)
- Restricted or unrestricted portfolio weights (2 choices).

If we would combine all choices, we would get too many sets of results. Therefore, we do not consider all sets of results. We do not calculate the unrestricted portfolio weights for the decision-theoretic method. Furthermore, we only use the no-predictability prior to calculate benchmark results and only combine it with the plug-in method. In this case, the dynamic, myopic and constant proportion strategy are similar. Next to the above possibilities, we also implement a (non data-based) $1/N$ strategy and repeat the above analysis for a different set of predictor variables.

The solution method we use depends on whether weights are (un)restricted, what kind of strategy we use (myopic, dynamic or constant proportion) and how we use the econometric estimation results (plug-in method or decision-theoretic method).

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 choices made in Wachter and Warusawitharana (2007). 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. In order to be able to compare the three strategies, we give the constant proportion strategy the opportunity to update allocations in every period. In the first period, such an investor uses a constant proportion strategy with horizon K to calculate portfolio weights; in the next period, she uses a strategy with horizon K-1 etcetera. Otherwise, we would give investors that follow the dynamic or myopic strategy an information advantage. We also considered (not in the paper) true constant proportion investors, i.e. investors that keep the same weights for the full horizon, but their performance was inferior to the constant proportion investors we consider in this paper. Finally, we use the certainty equivalence return (CER) as performance criterium. It is the risk-less return that would make investors indifferent between following a strategy or accepting this risk-less return. The CER is a monotone transformation of average realized utility values $\overline{U}$ and is given as follows

$$CER = \left(\bar{U}(1-\gamma)\right)^{\frac{1}{1-\gamma}} - 1. \tag{11}$$

3.2 Comparison of strategies

The aim of this paper is to investigate what strategy works best out-of-sample in terms of expected utility. In case we would know the process that generates asset returns and state variables perfectly, this would be a trivial question to answer. It would be the strategy that takes myopic and hedging demands into account, i.e. the dynamic strategy, and for which the portfolio weights are unrestricted since this strategy encompasses all other portfolio strategies. Or in other words, maximization on a subset can never lead to better results than maximization on a bigger set that encompasses all subsets.

In reality, we do not know the true data generating process (DGP) and have to select and estimate a model. This model is however by definition misspecified and estimates suffer from sampling errors. Optimizing on a bigger set therefore can also lead to a bigger error and it is therefore far from trivial which strategy works best out-of-sample.

We organize our discussion around the value function (1). The multiple period problem above can be written as a single period problem in a relatively straightforward way:

$$V_{t+s}(K-s, W_{t+s}, Z_{t+s}) \quad = \quad \max_{w_{t+s}} E\left\{ \frac{\left(w_{t+s}' R_{t+s+1}\right)^{1-\gamma}}{1-\gamma} \psi\left(K-s-1, Z_{t+s+1}\right) \mid Z_{t+s} \right\}, \tag{12}$$

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

$$\frac{1}{1-\gamma}\psi\left(K-s-1, Z_{t+s+1}\right) = \max_{w_{t+s+1},\ldots,w_{t+K-1}} E\left\{ \frac{\left(\prod_{r=t+s+1}^{t+K-1}\left(w_r' R_{r+1}\right)\right)^{1-\gamma}}{1-\gamma} \mid Z_{t+s+1} \right\}. \tag{13}$$

The conditioning set at time t is summarized by conditioning variables Z_t. The latter consist of asset returns $r_{tbill,t}$ and x_t and state variables s_t at time t. This equation is the Bellman equation for the power utility case. We can solve for the optimal portfolio strategy by solving the sequence of one-period problems by backward induction.

The conditional expectation in equation (12) is taken over asset returns R_{t+s+1} and conditioning variables Z_{t+s+1} simultaneously with $R_{t+s+1} \subset Z_{t+s+1}$. The hedging term in the dynamic strategy depends on this dependence. If the contemporaneous dependence between the other conditioning variables and the asset returns 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.

The constant proportion strategy sets equal portfolio weights in all future periods. This means that such an investor will not speculate on short-term changes in the investment environment. This strategy performs best if the investor is better able to predict asset returns on longer horizons than on shorter horizons. It is especially important to accurately model unconditional means and variances since the predictions of the long-term mean and long-term variance of the portfolio returns converge to these unconditional moments at longer horizons.

3.3 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 and 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,

$$P\left(y_{t+1}|\hat{B},\hat{\Sigma},y_t\right), \tag{14}$$

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 (8) and (9), 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 degree of uncertainty: parameter uncertainty. Returns are not only uncertain because of the error terms but also because parameters might not be estimated correctly. References to studies that use this approach are Campbell and Viceira (2002) and Jurek and Viceira (2006).

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

$$P\left(y_{t+1}|y_t,y_{t-1}...\right)=\int P\left(y_{t+1}|B,\Sigma,y_t\right)P\left(B,\Sigma|y_t,y_{t-1}...\right)d\Sigma dB. \tag{15}$$

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 obtain a posterior distribution that accurately describes what we really know about the parameters. Another disadvantage is that the returns are not lognormally distributed anymore if we do not condition on parameters which implies that we have to rely on numerical methods. 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 (2007), 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 6 variables y_t. If we use the decision-theoretic method, this is not necessarily true. Over time an investor learns more about the true unknown values of the parameters. Therefore, the investment opportunity set changes over time since the posterior parameter distribution changes 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. 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 Z_t[5] to describe the investment opportunity set. This is unfeasible. Note that currently only problems up to 11 conditioning variables are solved in the portfolio literature, please refer to Brandt, Goyal, Santa-Clara, and Stroud (2005). Therefore, 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. 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 investor still learns about the true parameter values through time if new observations are available. The simplification we make is that the investor does not hedge against this learning. Brandt, Goyal, Santa-Clara, and Stroud (2005) show by means of simulations that taking parameter uncertainty into account while ignoring learning leads to improved performance relative to the case without parameter uncertainty.

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 (8) and (9) 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 parameters. Adding only one extra lag means $n^2 = 36$ extra parameters in our setting. Since efficiency is an important issue in our models, 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 homoskedastic, i.e. that risk is constant over time. Modeling heteroskedastic errors would again mean an increase in parameters and reduce efficiency. Therefore, we choose to assume homokcedastic errors. This choice is supported by results of Chacko and Viceira (2005) who find that stock return volatility does not generate large hedging demands.

[5] All distinct parameters plus the current variable values

In order to estimate the VAR(1) model in equations (8) and (9), provide inferences and make forecasts, we use a conditional likelihood function which treats the first observation as fixed. A popular alternative would be to use an exact/unconditional likelihood function as in Schotman and van Dijk (1991), Wachter and Warusawitharana (2007) or Stambaugh (1999) which treats the first observation as stochastic. We do not pursue this alternative in this paper. Under the model given in (8) and (9), the conditional likelihood function is as follows

$$P(Y|B,\Sigma) \propto |\Sigma|^{-T/2} \exp\left\{-\frac{1}{2}tr\left[(Y-XB')'(Y-XB')\Sigma^{-1}\right]\right\}, \tag{16}$$

where T is the number of observations, Y is the Txn matrix of observations on y_t, Y_{-1} is the Txn matrix of lagged observations on y_t, X is the Tx$(n+1)$ matrix $X = [\iota, Y_{-1}]$ and B is the nx$(n+1)$ matrix $B = [B_0, B_1]$.

We are both interested in point estimates for the parameters B_0, B_1 and Σ, please refer to section 6.2 about the plug-in method, and in the posterior distribution of these parameters, please refer to section 6.3 about the decision-theoretic method. We use the posterior means of B and Σ as point estimates.

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

$$p(B,\Sigma) \propto |\Sigma|^{-(n+1)/2}. \tag{17}$$

We refer to this prior as the uniform prior. If we combine this prior with the conditional likelihood function, we get the following posterior distribution

$$P(B,\Sigma|Y) \propto |\Sigma|^{-(T+n+1)/2} \exp\left\{-\frac{1}{2}tr\left[(Y-XB')'(Y-XB')\Sigma^{-1}\right]\right\}. \tag{18}$$

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 $S/(T-n-1)$, 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. This is explained in appendix A.

We consider a second Bayesian estimator which is used among others in Ni and Sun (2003) in the context of a similar VAR model. This estimator shrinks the coefficients towards zero. It is the general prior for a skeptic investor. We refer to this prior as the shrinkage prior, it is as follows

$$p(B,\Sigma) \propto \parallel B \parallel^{-(n(n+1)-2)} |\Sigma|^{-(n+1)/2}. \tag{19}$$

It is the product of a shrinkage prior on 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 when the ML estimator exists, which holds in our setting. Note that the prior on B has a negative exponent. This means that prior draws with a high norm are relatively improbable. Hence, if we combine this prior with the conditional likelihood, the posterior is shrunk towards B values with a low norm, i.e. a matrix with all 0s.

This causes a problem in our setting. Our state variables are highly autocorrelated and shrinking the autocorrelation towards 0 might result in a misspecified VAR system. We solve this issue by transforming the state variables on the left hand side in the VAR system. More specifically, introduce vector $y_{trans,t}$

$$y_{trans,t} = \begin{pmatrix} r_{tbill,t} \\ x_t \\ \Delta s_t \end{pmatrix}. \tag{20}$$

Instead of estimating model (8), we estimate the following auxiliary model

$$y_{trans,t+1} = B_0 + B_1^* y_t + \epsilon_{t+1}, \tag{21}$$

where ϵ_{t+1} is distributed as indicated in equation (9). Note that we can obtain the matrix of slope coefficients B_1 in the original model by adding 1 to the diagonal elements corresponding to the state variables of matrix B_1^*

Shrinking the auxiliary model (21) towards a zero matrix implies that we are shrinking all coefficients in the original model towards zero except for the autocorrelation coefficients of the state variables which we shrink towards 1. Therefore, this prior is in line with the prior beliefs of an investor that is skeptical but not dogmatic about the predictability in the data.

Let Y_{trans} be the Txn matrix of stacked $y_{trans,t}$ vectors. If we combine the prior with the conditional likelihood of the auxiliary model, we get an unknown posterior distribution, i.e. the exact properties are unknown

$$P(B,\Sigma) \propto \| B^* \|^{-(n(n+1)-2)} |\Sigma|^{-(T+n+1)/2} \exp\left\{ -\frac{1}{2} tr \left[(Y_{trans} - XB^{*'})'(Y_{trans} - XB^{*'})\Sigma^{-1} \right] \right\}, \tag{22}$$

where $B^* = [B_0, B_1^*]$. In order to obtain point estimates for the plug-in approach and draw parameter values and future variable values for the decision-theoretic approach, we have to be able to simulate from the posterior and predictive distribution. The simulation algorithm is explained in appendix A.

If the lagged asset returns and state variables are not able to predict asset returns, the first three rows of B_1 in model (8) will be equal to zero rows. Our third prior combines prior (17) with an indicator function that is only 1 when these rows are zero, i.e. it dogmatically sets these rows equal to zero. In the paper, we refer to this specification as the no-predictability prior.

The state variables 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 therefore impose the assumption of stationarity. Numerical results are more stable, since this excludes extreme non-stationary draws. Since the mode of the likelihood function is within the stationary region, we do not impose this assumption explicitly when using the plug-in approach. This only slightly changes the point estimates, has a negligible 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 (8), i.e. we give point estimates for the parameter values for the full data-set. We report the posterior means for the models based on the uniform and shrinkage prior.

Table 2 reports the results for the model where the price-earnings ratio is one of the state variables. From the table it is clear that the state variables are highly autocorrelated. 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. This holds for both priors. There is also a large positive correlation between shocks to the price-earnings ratio and excess stock returns, which means that unexpected positive shocks to stock returns lead to negative future investment opportunities, summarized in the PE variable. A remarkable finding is that the VAR(1) model implies an unconditional mean of 0.0034 for excess stock returns, far below the sample mean 0.0048.

If we compare the posterior means for both priors, it is clear that the posterior means for most coefficients are shrunk towards zero by the shrinkage estimator which means that we allow for less predictability. Hence, unless the data provides sufficient evidence that a variable predicts asset returns positively or negatively, the posterior distribution is (tightly) centered around zero. Furthermore, the autocorrelation of the price-earnings ratio is closer to 1.

[Table 2 about here.]

We constantly re-estimate our models on bigger data-sets that include the newest observations. Therefore, we also provide time-series of estimates of the two most important parameters. 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 February 1974 until December 2006.

We present time series plots of the slope coefficients of (x_s,s_{DP}) and (x_b,s_{SPR}) in figures 1 and 2. It turns out that for both coefficients, the posterior mean for the shrinkage prior is closer to 0 than for the uniform prior. It appears from the figures that there is a lot of uncertainty about the estimated value, since the parameters are extremely variable. Another important point is that 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.]

[Figure 2 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 solving for 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 observation made by Koijen, Nijman, and Werker (2007).

5.1 Analytical method

Given the VAR(1) model in equation (8), returns are lognormally distributed conditional on the parameter values. Jurek and Viceira (2006)[6] use this fact and derive approximate-analytical solutions for the unrestricted plug-in model, for the myopic, constant-proportion strategy and the dynamic strategy. These solutions are all based on the Campbell and Viceira (2002) approximation to log-portfolio returns

$$r_{p,t+1} = r_{tbill,t+1} + w_t' x_{t+1} + \frac{1}{2}\left(w_t'\sigma_x^2 - w_t'\Sigma_{xx} w_t\right), \tag{23}$$

where w_t is the weights vector on the risky assets[7] and σ_x^2 is the vector of diagonal elements of Σ_x. 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.

The (semi-)analytical solution to the dynamic problem is involved. Therefore we refer to Jurek and Viceira (2006) who show that portfolio weights on risky assets are an affine function of the conditioning variables

$$w_t^{K,DYN} = A_0^{(K)} + A_1^{(K)} z_t, \tag{24}$$

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 optimal constant proportion strategy is as follows

$$w_t^{K,CP} = \left((\gamma-1)\Sigma_x^{(K)} + K\Sigma_x\right)^{-1}\left(E_t[x_{t\to t+K}] + \frac{K}{2}\sigma_x^2 + (1-\gamma)\sigma_{1x}^{(K)}\right), \tag{25}$$

where $\Sigma_x^{(K)}$ is the covariance matrix of K-period log excess returns on the risky assets, $E_t[x_{t\to t+K}]$ is the expected value of K-period log excess returns on the risky assets and finally $\sigma_{1x}^{(K)}$ is the vector of covariances of K-period log excess returns with the benchmark asset. Note that the one-period variances and covariances occur in the formula because of the approximation to the 1-period log portfolio returns.

[6]Note that Jurek and Viceira (2006) use the ML estimate as plug-in estimate. We use the posterior mean of the uniform prior, which is very comparable to the ML estimate, and the posterior mean of the shrinkage prior as plug-in estimates instead.

[7]Note that the weight on the benchmark asset is $1 - w_t'\iota$.

If we set $K = 1$ in the above formula, we get the weights for the myopic strategy

$$w_t^{1,M} = (\gamma\Sigma_{xx})^{-1}\left(E_t[x_{t\to t+1}] + \frac{1}{2}\sigma_x^2 + (1-\gamma)\sigma_{1x}\right). \tag{26}$$

5.2 Numerical method

There is no analytical solution available for the restricted plug-in model. Furthermore, returns are not lognormally distributed if parameters are integrated out and therefore there is no analytical solution available for the (un)restricted decision-theoretic model. In these cases we have to use numerical methods instead. Our method is based on Monte Carlo simulations, introduced by Brandt, Goyal, Santa-Clara, and Stroud (2005). Our method consists of a refinement (step 5 of the algorithm below) of van Binsbergen and Brandt (2007). We impose no-shortselling constraints to guarantee the accuracy of the solutions as in the above paper.

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 calculate dynamic strategies more than 300 times, computation time is an important issue. Therefore, we use a refinement in Koijen, Nijman, and Werker (2007)[8] in our setting and parameterize the regression coefficients in regressions that approximate conditional utility by a quadratic function of portfolio weights. This allows us to find the optimal weights along 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. The distribution can or cannot be conditioned on parameter estimates.

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.

[8] Note that Koijen, Nijman, and Werker (2007) 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.

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.

Calculating myopic and constant proportion strategies is easier[9]. We use an adapted version of the algorithm above. We replace step 3, 4 and 5 by steps 6, 7 and 8 given below. Note that we do not have to repeat the steps for multiple periods.

6. Pick one set of portfolio weights from the grid and calculate the realized utility values for all simulated paths. Hence: use the chosen constant proportion weight in all periods.
7. Take the average of the N realized utility values to obtain the conditional utility values.
8. Parameterize the average in the portfolio weights with a quadratic function and choose the portfolio weights that maximize this approximate function for every path. Refer to step 5 for details.

Appendix B gives more details on the parameterization of regression coefficients and the accuracy of our method.

6 Out-of-sample results

In our empirical exercise, we investigate the out-of-sample performance of strategic asset allocation models. These models differ in their method (plug-in or decision-theory), the general strategy (myopic, dynamic or constant proportion), how these translate data into a model (based on the uniform prior or shrinkage prior) and whether the weights are (un)restricted (shortselling allowed or not). We report results 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 Benchmark results

In order to be able to compare results, we give results for three benchmark strategies in table 3. We compare results by reporting their certainty equivalence return (CER), their average terminal wealth and the standard deviation of terminal wealth.

[9]Note that the myopic strategy is identical to the constant proportion strategy (or dynamic strategy) with $K = 1$

The first strategy is the $1/N$ strategy. It invests 33% in stocks, bonds and T-bills irrespective of the data. The next strategy is the unrestricted no-predictability strategy. It is data-based but we use a dogmatic prior that imposes that asset returns cannot be predicted by conditioning on information from the past, i.e. asset allocations are based on unconditional moments only. This is done by setting the slope coefficients in the first three rows of the VAR model equal to zero and the intercepts equal to the sample mean. The final strategy is the restricted no-predictability strategy. It is similar to the previous strategy but without allowing for shortselling. The data-based strategies use the plug-in method. Note that the dynamic, myopic and constant proportion strategies are equivalent when there is no predictability.

[Table 3 about here.]

Table 3 shows that these simple strategies all have positive certainty equivalence returns. Hence, investors are willing to follow these strategies unless they are paid a positive riskless return. Furthermore, the table shows that the certainty equivalence returns decrease when the risk aversion γ increases: more risk averse investors value a riskless return more than less risk averse investors. Finally, we see from the table that investors with low and moderate risk aversion (γ is 2 or 5) prefer an unrestricted no-predictability strategy while a highly risk averse investor (γ is 10) prefers the $1/N$ strategy. For such an investor, the unrestricted no-predictability strategy is even inferior to the restricted no-predictability strategy. Apparently, data-based methods do not necessarily outperform non-data based methods and imposing restrictions might improve out-of-sample performance. The latter is consistent with results in Jagannathan and Ma (2003), who show that imposing weight constraints is a form of shrinkage.

Figure 3 plots a histogram of realized utility values for the unrestricted no-predictability strategy. We set $\gamma = 5$. The picture shows that the utility value distribution is very left skewed. Most values are near zero but there are some large negative outliers. Note that these correspond to extreme events, i.e. low realized terminal wealth values, and that these are important since the utility function gives extra weight to extreme events.

[Figure 3 about here.]

6.2 Plug-in results

This section gives results using the plug-in method. We report results for the myopic, dynamic and constant proportion strategies, either using the uniform or shrinkage prior and with restricted or unrestricted portfolio weights.

Results are given in table 4. The table shows some remarkable results. Firstly, the performance of an investor with low risk aversion (γ is 2) is disastrous under the standard uniform prior when weights are unrestricted. Such an investor is willing to pay a return up to 100% in order to avoid adopting this strategy. This corresponds to expected utility equal to minus infinity. 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 is equal to zero, i.e. an investor loses all her money, which results in average utility of $-\infty$.

[Table 4 about here.]

For higher γ values, the performance is better. CER's are positive and substantially higher than results for benchmark strategies. Note that myopic strategies give a better performance than the theoretical optimal dynamic strategies. Even though the average terminal wealth is higher for these dynamic strategies, the risk increases more than proportionally. The performance of constant proportion strategies is bad. CERs are negative in all cases. Hence, when using the uniform prior, it seems sufficient to just focus on short-term changes in investment opportunities and ignore long-term changes.

If we use the shrinkage prior instead of the uniform prior, we get a completely different picture when using unrestricted weights. Firstly, the performance for all strategies and all risk aversion levels increases substantially. Although the usage of shrinkage estimators reduces average terminal wealth, its standard deviation is more than proportionally reduced. For example, compare the dynamic strategy 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. Most importantly, CERs for investors with $\gamma = 2$ are not equal to -100% anymore. Furthermore, we see that dynamic strategies outperform myopic and constant proportion strategies. The risk for dynamic strategies is still higher but this time the extra average terminal wealth more than offsets this. The performance of constant proportion strategies is close to zero but does not beat the benchmarks of the previous section.

In order to understand how the shrinkage model works, we plot the realized utility values for both the shrinkage and the uniform prior against time in figure 4. The figure shows that both series are heavily autocorrelated due to overlapping intervals and that there is positive correlation between the series. Both strategies perform similarly except for a couple of extreme events, i.e. extremely low realized utility values. While the shrinkage prior manages to keep these losses in check, the losses for the uniform prior are very big. Hence, the shrinkage prior improves results by avoiding extreme losses.

[Figure 4 about here.]

How does the shrinkage estimator reduce losses? In order to answer this question, we plot the stock weights against time in figure 5 for the same strategies considered above. 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 less extreme. The strategy based on the shrinkage estimator allows for market timing but reduces this substantially.

[Figure 5 about here.]

Under the shrinkage prior, the dynamic strategy outperforms the myopic strategy. In order to illustrate this, consider figure 6 which plots the histogram of differences in realized values between a dynamic investor and a myopic investor with $\gamma = 5$. Note that 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.

[Figure 6 about here.]

The table shows that another way to avoid CERs of -100% is to restrict portfolio weights. Jagannathan and Ma (2003) showed that restricting portfolio weights is a form of shrinkage as well. If we do not allow for shortselling, risk and average terminal wealth decrease substantially. For investors with low risk aversion, this leads to a substantial increase in CERs. Note that this performance is still inferior to combining a shrinkage prior with unrestricted portfolio weights. For higher risk aversion levels, the table shows that performance for myopic and dynamic strategies deteriorates when restricting portfolio weights. Hence, it pays off to be able to go long and short in assets. The performances with and without shrinkage are similar with a slight edge to using the shrinkage prior. In all cases, a dynamic strategy (either with or without shrinkage) seems optimal when portfolio weights are restricted.

A remarkable observation is that restricting portfolio weights leads to substantially improved constant proportion strategies. This performance is even further improved when using the shrinkage prior. Apparently, constant proportion strategies only perform satisfactorily when imposing double shrinkage by using the shrinkage estimator and by restricting the portfolio weights.

We can conclude that combining a dynamic strategy with the shrinkage prior leads to superior results in all cases. Using shrinkage is substantially better than restricting portfolio weights in order to avoid extreme events. Differences can be as big as 21% in annualized certainty equivalence returns. Dynamic strategies only work satisfactorily when using the shrinkage estimator. Especially investors with low risk aversion benefit greatly from using the shrinkage estimator compared to alternatives such as no shrinkage or restricting portfolio weights.

6.3 Decision-theoretic results

This section gives results for the decision-theoretic approach. This approach takes parameter uncertainty into account but ignores learning. We only consider the case with restricted portfolio weights to guarantee the accuracy of the solutions.

Table 5 shows results. If we compare table 4 with this table, we see that taking parameter uncertainty into account increases CERs in all cases for the dynamic strategies. The effect is not large however. We use a lot of observations to estimate our models and therefore the level of parameter uncertainty might be small. Parameter uncertainty does not have a clear effect on average terminal wealth or its standard deviation: it only finds a slightly better trade-off between them. Results for the myopic and constant proportion strategy are more mixed. On average, the performance for the myopic strategy slightly increases when using decision-theoretic methods and the performance of the constant proportion strategy slightly decreases.

We illustrate the performance of the decision-theoretic approach in figure 7. 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. Note that positive values indicate outperformance for the decision-theory model. The figure shows that the decision-theoretic model performs better in most cases, i.e. the median is slightly positive. Furthermore, if the decision-theoretic model outperforms the plug-in model, the difference is relatively big, illustrated by larger positive values. Hence, the mean difference is positive and the decision-theoretic model outperforms the plug-in model.

[Figure 7 about here.]

[Table 5 about here.]

As above, shrinkage reduces average terminal wealth and its standard deviation. Results on combining the shrinkage prior with the decision-theoretic method are mixed.

We can conclude that in all cases, taking parameter uncertainty into account when restricting portfolio weights pays off. For the restricted case, a dynamic strategy using this decision-theoretic approach either combined with a uniform or shrinkage prior leads to the best results. Theoretical optimal results turn out to be optimal in an empirical setting as well. Note that combining unrestricted weights with the shrinkage estimator (refer to section 6.2 on the plug-in method) outperforms all models considered in this section.

7 Additional tests

In this section, we perform some tests to investigate the robustness of our results. In the first subsection, we performs some classical tests on the performance differences between the strategies. The last subsection considers a model with the dividend-to-price ratio as one of the predictor variables.

7.1 Robustness 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 benchmark strategies in section 6.1 in a repeated samples context. More precisely, we test whether the difference in average utility between a strategy and its benchmark is statistically different from zero. This is a fundamentally different point of view than previously taken in this paper. A Bayesian would assign probabilities[10] to two competing hypotheses and minimize expected posterior loss, a classical econometrician takes a yes/no decision using hypothesis testing.

As a benchmark, we take the no-predictability strategies of section 6.1, either unrestricted or restricted, depending on the context. Hence, we test whether the extra value of market timing we find in previous sections might be spurious.

Our results depend on whether we are able to forecast asset returns accurately out-of-sample. One could 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 such a test, 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. A problem in our context is the estimation of the variance of the difference series. Our results are based on overlapping observations which means that the series is strongly autocorrelated. Furthermore, there might be other correlation in the series besides the correlation induced by the overlapping observations. This autocorrelation might enforce or mitigate the autocorrelation due to overlapping observations and needs to be taken into account.

One way would be to parameterize the autocorrelation in the difference series by means of an ARMA process. This would be an efficient way to take the correlation structure into account if we would be able to estimate a parsimonious ARMA model. This is not the case however.

Therefore, we estimate the covariance matrix non-parametrically by means of the Newey and West (1987) HAC estimator. In order to do so, we have to specify the lag length. Since the autocorrelation is unknown, we use a data-based selection procedure. We choose to do this non-parameterically and use Newey and West (1994) lag length selection criterium. An alternative would have been to use Andrews' (1991) selection criterium. We do not pursue his alternative, since it would require us to parameterize the series.

Table 6 presents the results. It shows that the constant proportion strategy does not significantly outperform the benchmark strategy in all cases. Furthermore, the performance of unrestricted plug-in strategies based on the uniform prior is only significant 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. If we look at the restricted strategies, the picture looks a bit different. 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.]

[10] A Bayesian might also select a single model if she uses a diagonal loss function

7.2 Using dividend-to-price ratio as a predictor

In section 6, we used the price-earnings ratio as one of the predictor variables. Another variable which has been found in the portfolio literature to improve return predictions 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 7 shows results for the plug-in method. Again, the performance of an investor with low risk aversion ($\gamma = 2$), unrestricted portfolio weights and the uniform prior is very bad with an 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. Furthermore, we see that dynamic strategies suddenly outperform myopic ones. Hence, using our shrinkage estimator turns a bad model into a good model. The table shows that we could have restricted portfolio weights as well, instead of shrinkage, to improve out-of-sample performance. 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.

[Table 7 about here.]

The results in table 7 make clear that double shrinkage does not work for the DP model. Combining the shrinkage estimator with restricted portfolio weights deteriorates results for all cases.

All together we see that the results from section 6.2 are confirmed. Using an unrestricted plug-in strategy without shrinkage potentially leads to bad results. We either have to restrict the weights or use a shrinkage prior to improve them and make dynamic strategies work. In this section, combining shrinkage with restricted portfolio weights slightly deteriorates the results.

Table 8 shows analogous results for the decision-theoretic method. These results are different from the results in section 6.3. For the DP model, taking parameter uncertainty into account deteriorates results. A potential explanation is that the DP model is misspecified and that trying to specify a posterior distribution for the parameters increases this misspecification. In other words, the decision-theoretic method is less robust than the plug-in method.

[Table 8 about here.]

8 Conclusion

In this paper, we investigate the out-of-sample performance of strategic asset allocation models. We perform an extensive analysis where we vary the method (plug-in or decision-theoretic), the estimator (uniform prior, shrinkage prior or no-predictability prior), the strategy (myopic, dynamic or constant proportion) and the portfolio constraints (constrained or unconstrained) for risk aversion levels γ is 2,5 or 10.

We demonstrate that the out-of-sample performance of these models would have been disastrous if a purely data-based view had been used. An investor with low risk aversion might even have lost all her money. We also show that there are obvious ways to improve this performance. Restricting portfolio weights or, even better, using shrinkage estimators leads to significant improvements. Our results suggest that the best choice for an investor who can freely vary his/her portfolio weights is to use a dynamic strategy based on a model that is estimated by our shrinkage estimator. This estimator consistently outperforms all other estimators we consider. It could give annualized certainty equivalence returns of more than 10%. If investors face portfolio constraints, results are less clear. We favor the decision-theoretic model that uses the shrinkage prior, since its results are slightly better than for the other models considered. Note however that this model seems less robust to model misspecification.

Section 5 and appendix B also explain a new numerical method for solving myopic, dynamic and constant proportion strategies for terminal wealth problems. We show that it is not necessary to consider a large grid of portfolio weights since conditional utility is a quadratic function in portfolio weights. It is sufficient to consider a smaller grid and optimize a quadratic function analytically without loss of accuracy. This method can easily be generalized to more assets, more predictor variables, and other utility functions. Our method is very robust due to the grid search and still very fast.

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. However, we investigated the sensitivity of performance with respect to the choice of another predictor variable. An alternative would be to use model selection criteria or Bayesian model averaging. 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 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. However, our paper shows that such claims have to be tested empirically, since 'optimal' models might not be optimal empirically.

There are a number of extensions we consider. Firstly, we want to be able to compare the models in a Bayesian way. This calls for encompassing tests, i.e. compare models by giving them a relative weight/probability. Secondly, the implied mean of our models is substantially below the sample mean in the data. If we are better able to model the unconditional moments, this might improve our results. This is especially important for the performance of the constant proportion strategy.

A Simulating from posterior

This section gives details on how we simulate from the posterior and predictive distribution for both the uniform prior, introduced in equation (17), and the shrinkage prior, introduced in equation (19).

We firstly consider the uniform prior. The posterior distribution is repeated in the following equation

$$P(B,\Sigma|Y) \propto |\Sigma|^{-(T+n+1)/2} \exp\left\{-\frac{1}{2}tr\left[(Y-XB')'(Y-XB')\Sigma^{-1}\right]\right\}. \qquad (27)$$

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

$$P(\Sigma|Y) = iWishart\,(S, T-n-1) \qquad (28)$$

$$P(\beta'|\Sigma,Y) = MVN\left(\hat{\beta}', \Sigma\otimes(X'X)^{-1}\right), \qquad (29)$$

where β' and $\hat{\beta}'$ are equal to vectorized B', $\hat{B}=(X'X)^{-1}X'Y$ 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 β' given Σ from the multivariate normal distribution.

If we impose the assumption of stationarity, we cannot derive the marginal posterior for Σ by integrating over B and can only use conditional posteriors. This implies that we have to use a Gibbs sampler with the conditional posteriors $\beta|\Sigma$, given in equation (29), and $\Sigma|\beta$. The latter distribution is an inverted Wishart distribution where S in equation (28) depends on B instead of $\hat{B}$ and the degrees of freedom are equal to T instead of $T-n-1$. We reject draws for B that are outside the stationary region.

Secondly, consider the shrinkage prior. Again, the posterior distribution is repeated in the following equation

$$P(B,\Sigma) \propto \| B^* \|^{-(n(n+1)-2)} \, |\Sigma|)^{-(T+n+1)/2} \exp\left\{-\frac{1}{2}tr\left[(Y_{trans}-XB^{*'})'(Y_{trans}-XB^{*'})\Sigma^{-1}\right]\right\}. \qquad (30)$$

The above posterior does not belong to a known distribution class and therefore simulation is not trivial. 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

$$P(\Sigma|B^*,\delta,Y) = iWishart\left((Y_{trans}-XB^{*'})'(Y_{trans}-XB^{*'}), T\right) \qquad (31)$$

$$P(\delta|B^*,\Sigma,Y) = iGamma\left(J/2-1, \frac{1}{2}\beta^{*'}\beta^*\right) \qquad (32)$$

$$P(B^*|\delta,\Sigma,Y) = MVN\left(\delta(\Sigma\otimes(X'X)^{-1}+\delta I_J)^{-1}\hat{\beta}^*, (\Sigma^{-1}\otimes X'X+\frac{1}{\delta}I_J)^{-1}\right), \qquad (33)$$

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.

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 parameters 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

$$P(y_{t+1}|y_t, B, \Sigma) = MVN(B_0 + B_1 y_t, \Sigma), \tag{34}$$

where B_0, B_1 and Σ are drawn parameters. 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 discard the first 5,000 draws and draw 10,000 parameter estimates in total. This results in 20,000 asset return and state variable paths. Increasing the burn-in phase or the number of simulations does not influence 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 convergence is reached.

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. 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

$$\max_{w_t} E\left(\frac{W_{t+1}^{1-\gamma}}{1-\gamma} \mid Z_t\right). \tag{35}$$

The standard approach to 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 every path and calculate conditional utility for this grid point, i.e. average realized utility in this simplified example. Repeat this for all grid points and pick the portfolio weights that maximize conditional utility.

Since different portfolio weights lead to different conditional utilities, conditional utility has to be a function of portfolio weights. We illustrate this fact in figure 8 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

$$\max_{w_t} E\left(\frac{W_{t+1}^{1-\gamma}}{1-\gamma} \mid Z_t\right) = \max_{w_t} f(w_t), \tag{36}$$

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

[Figure 8 about here.]

In other words, maximizing conditional utility on an (un)constrained set is equivalent to maximizing a quadratic function on an (un)constrained 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:

$$E\left\{V_{t+1}\left(K-1, W_{t+1}, Z_{t+1}\right) \mid Z_t\right\} = \alpha_{0w_t} + \alpha_{1w_t} Z_t, \tag{37}$$

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

$$\begin{aligned} E(.. \mid Z_t) = {} & (\gamma_{00} + \gamma_{10} w_{t,s} + \gamma_{20} w_{t,b} + \gamma_{30} w_{t,s}^2 + \gamma_{40} w_{t,b}^2 + \gamma_{50} w_{t,s} x_{t,b}) + \\ & (\gamma_{01} + \gamma_{11} w_{t,s} + \gamma_{21} w_{t,b} + \gamma_{31} w_{t,s}^2 + \gamma_{41} w_{t,b}^2 + \gamma_{51} w_{t,s} w_{t,b}) Z_t \end{aligned} \tag{38}$$

$$\begin{aligned} E(.. \mid Z_t) = {} & (\gamma_{00} + \gamma_{01} Z_t) + (\gamma_{10} + \gamma_{11} Z_t) w_{t,s} + (\gamma_{20} + \gamma_{21} Z_t) w_{t,b} + \\ & (\gamma_{30} + \gamma_{31} Z_t) w_{t,s}^2 + (\gamma_{40} + \gamma_{41} Z_t) w_{t,b}^2 + (\gamma_{50} + \gamma_{51} Z_t) w_{t,b} w_{t,s}, \end{aligned} \tag{39}$$

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[11]. The grid size is 100 and the number of paths is equal to 20,000. 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 100 instead of 100x100. Larger grid sizes do not influence the results. 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.

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

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

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 9. 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. The largest inaccuracy is 0.025 for $K = 60$ and $\gamma = 10$. Our method is around 100 times faster since we only have to consider a grid of 100 points instead of 10,000.

[Table 9 about here.]

References

Andrews, Donald, 1991, Heteroskedasticity and autocorrelation consistent covariance matrix estimation, *Econometrica* 59, 817–58.

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).

Brandt, Michael, 2004, Portfolio choice problems, Forthcoming in Handbook of Financial Econometrics.

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

Campbell, John, and Samuel Thompson, 2007, Predicting excess stock returns out of sample: Can anything beat the historical average?, Forthcoming Review of Financial Studies.

Campbell, John, and Luis Viceira, 2002, *Strategic Asset Allocation*.

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.

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

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

Goyal, Amit, and Ivo Welch, 2007, A comprehensive look at the empirical performance of equity premium prediction, Forthcoming Review of Financial Studies.

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

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, 2007, When can life-cycle investors benefit from time-varying bond risk premia, Working paper, Tilburg University.

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.

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.

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.

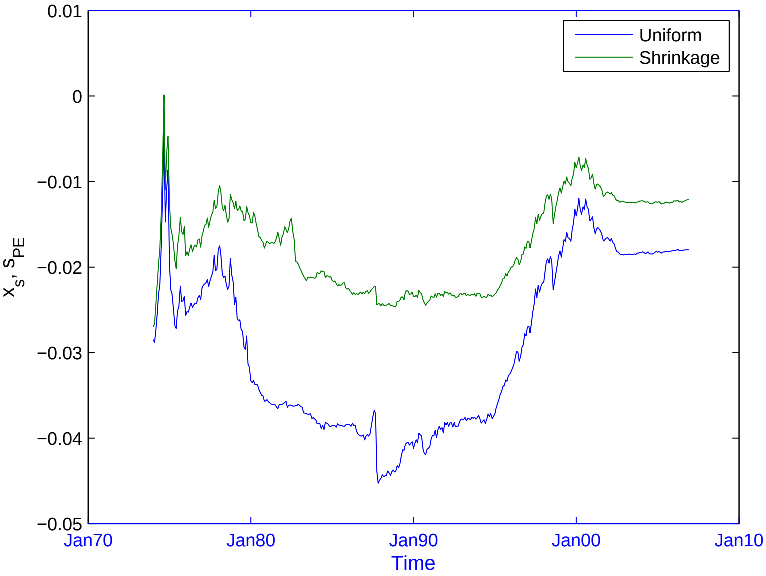
van Binsbergen, Jules, and Michael Brandt, 2007, Solving dynamic portfolio choice problems by recursing on optimized portfolio weights or on the value function, *Computational Economics* 29, 355–367.

Wachter, Jessica, and Missaka Warusawitharana, 2007, Predictable returns and asset allocation: Should a skeptical investor time the market?, Working paper, University of Pennsylvania.

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

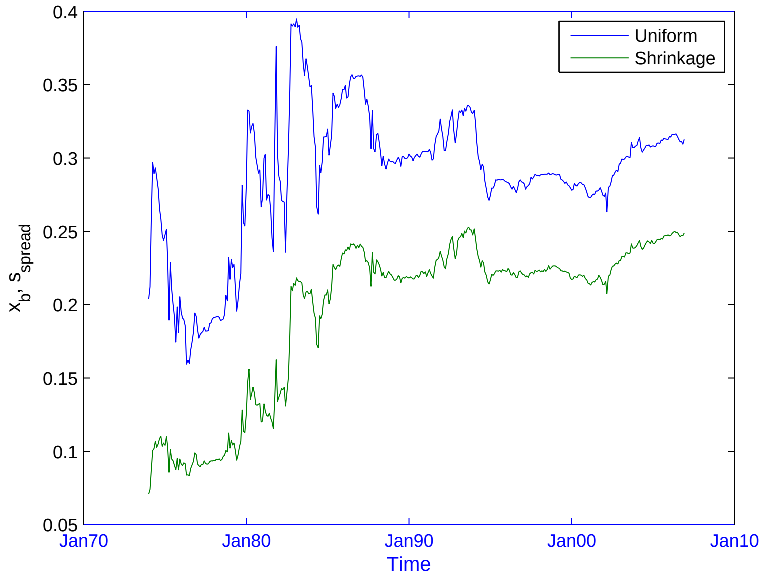
List of Figures

Figure 1: Overview of (x_s, s_{PE}) coefficient over time



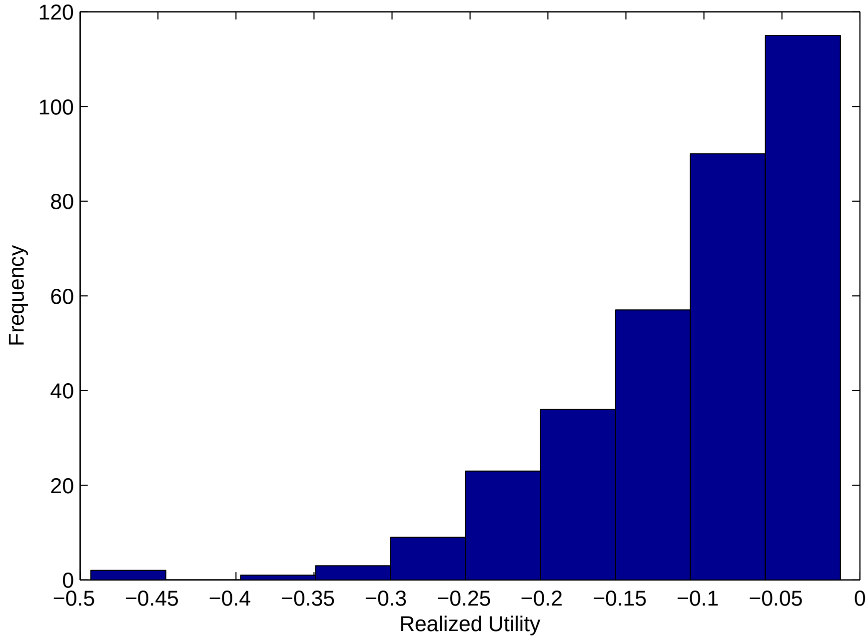
This figure plots the coefficients of x_s,s_{PE} (y-axis) against time (x-axis) for the uniform and shrinkage priors. The model is estimated from February 1954 until the indicated date on the x-axis.

Figure 2: Overview of (x_b, s_{spread}) coefficient over time



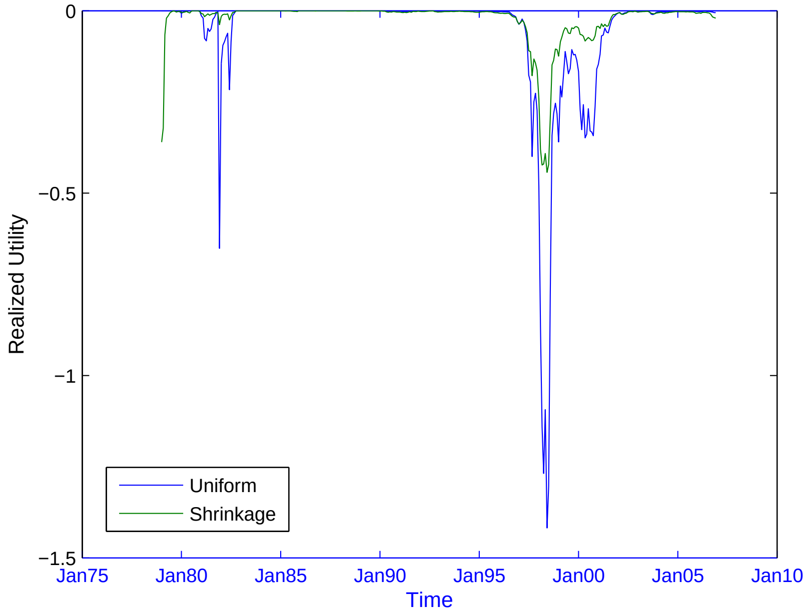
This figure plots the coefficients of x_b,s_{spread} (y-axis) against time (x-axis) for the uniform and shrinkage priors. The model is estimated from February 1954 until the indicated date on the x-axis.

Figure 3: Histogram of realized utility values for no predictability strategy with $\gamma = 5$



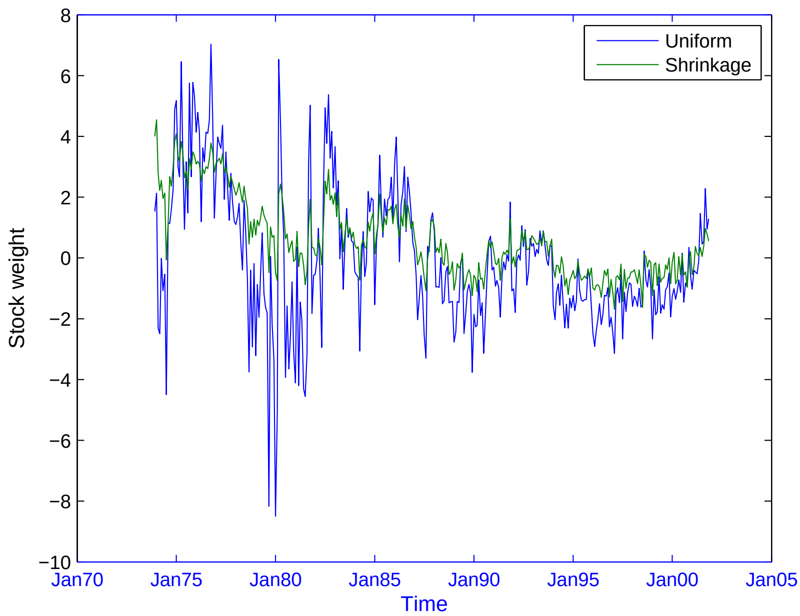
This figure gives a histogram of realized utility for the unrestricted no-predictability strategy with $\gamma = 5$. We use the plug-in method.

Figure 4: Realized utility values against time for different priors



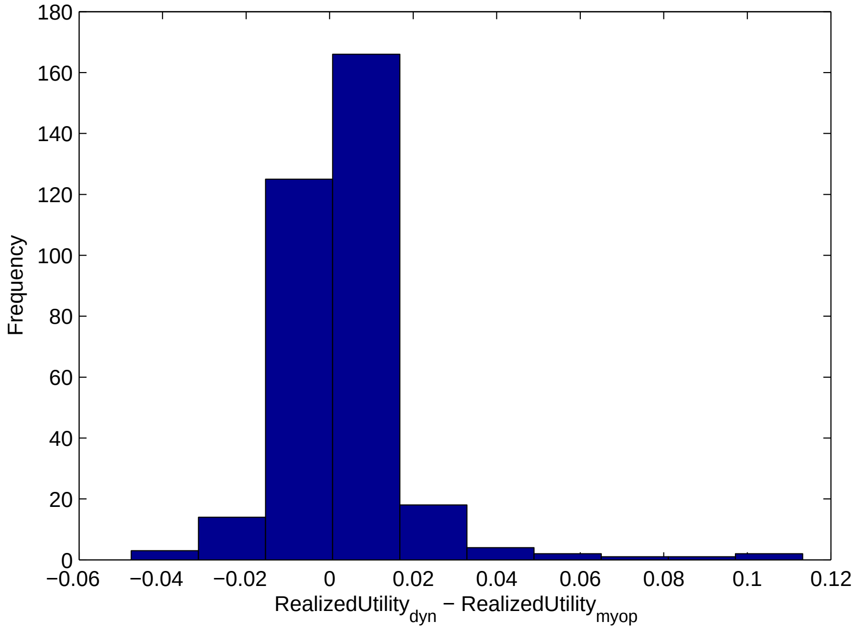
This figure plots realized utility values against time for the uniform (blue) and shrinkage (green) priors using the plug-in method. We consider a dynamic strategy, $\gamma = 5$ and unrestricted portfolio weights.

Figure 5: Portfolio weight in stock against time for different priors



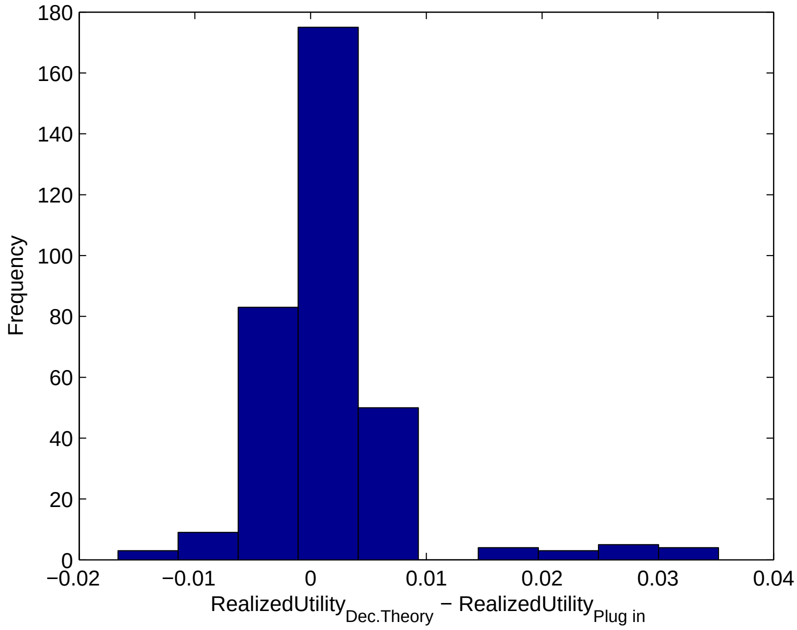
This figure plots the stock weight w_s against time for the uniform (blue) and shrinkage (green) priors using the plug-in method. We consider a dynamic strategy, $\gamma = 5$ and unrestricted portfolio weights.

Figure 6: Histogram of difference in realized utility: dynamic versus myopic



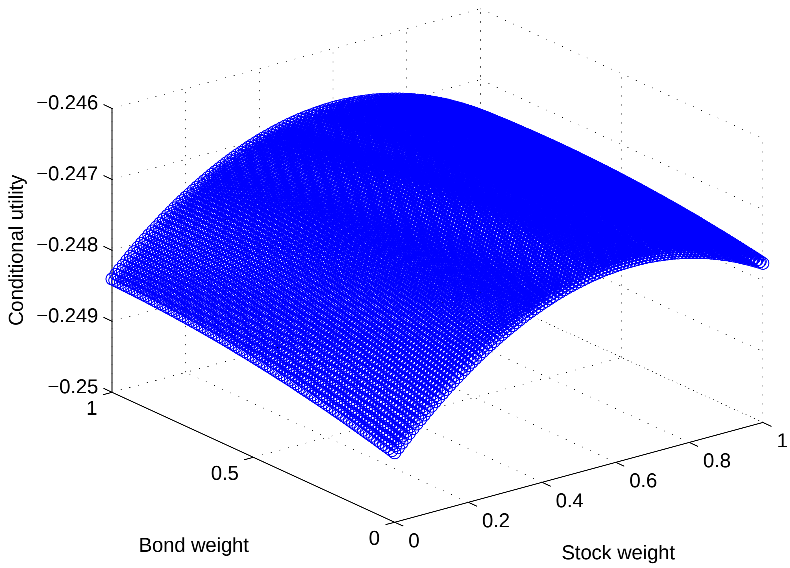
This figure gives a histogram of the difference in realized utility values between a dynamic and myopic strategy. We use the unrestricted shrinkage model with $\gamma = 5$.

Figure 7: Histogram of difference in realized utility: plug-in versus decision theory



This figure gives a histogram of the difference in realized utility values between the decision-theory and the plug-in approach. We consider a dynamic strategy based on the uniform prior and set $\gamma = 5$.

Figure 8: Conditional utility versus portfolio weights



This figure plots conditional utility over terminal wealth (z-axis) against the portfolio weight in stocks (x-axis) and the portfolio weight in bonds (y-axis). 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, AR(1) coefficients and Sharpe ratios 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 data set starts in February 1954 and ends in December 2006. All values are annualized. Percentages are given as fractions.

	R_{tbill}	X_s	X_b	Y_{nom}	DP	PE	Y_{spr}
Mean	0.0120	0.0690	0.0141	0.0501	-3.5339	2.8565	0.0112
Std Dev	0.0102	0.1484	0.0513	0.0261	0.3820	0.4141	0.0091
Min	-0.1349	-3.1279	-0.8306	0.0058	-4.5637	1.8929	-0.0160
Max	0.1349	1.7793	1.0772	0.1443	-2.8452	3.7887	0.0421
AR(1)	0.3831	0.0722	0.1089	0.9837	0.9930	0.9968	0.9193
Sharpe		0.4650	0.2742				

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 the posterior mean for the uniform prior and panel B for the shrinkage prior. In each panel, the slope coefficients, the implied mean and posterior standard deviations are given. Furthermore, 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						
	Parameter estimates						
	r_{tbill}	x_s	x_b	s_y	s_{PE}	s_{spread}	Mean
r_{tbill}	0.3242	0.0027	0.0078	0.0276	0.0011	0.0412	0.0011
	0.0552	0.0028	0.0088	0.0055	0.0003	0.0150	
x_s	1.6434	0.0240	0.3249	-0.3521	-0.0180	-0.0822	0.0034
	0.4910	0.0451	0.1310	0.0836	0.0058	0.2239	
x_b	0.4215	-0.0569	0.0787	0.0410	0.0017	0.3127	0.0011
	0.2007	0.0179	0.0497	0.0321	0.0015	0.0794	
s_y	-0.0813	0.0144	-0.0652	0.9855	-0.0001	0.0206	0.0505
	0.0701	0.0055	0.0214	0.0123	0.0005	0.0254	
s_{PE}	1.3611	0.4168	0.3114	-0.1423	0.9917	0.1227	2.9404
	0.3505	0.0274	0.0820	0.0515	0.0034	0.1378	
S_{spread}	0.0056	-0.0048	-0.0650	0.0070	-0.0002	0.9503	0.0106
	0.0547	0.0038	0.0157	0.0094	0.0004	0.0204	
	Error correlation matrix						
r_{tbill}	0.2686	0.1052	0.0757	-0.0805	0.1727	0.0560	
x_s		4.1814	0.1128	-0.0487	0.7746	-0.0331	
x_b			1.4319	-0.6237	0.0557	0.2208	
s_y				0.4304	-0.0494	-0.8516	
s_{PE}					2.8091	-0.0219	
S_{spread}						0.3483	
	Panel B: Shrinkage prior						
	Parameter estimates						
r_{tbill}	0.2728	0.0029	0.0068	0.0297	0.0012	0.0414	0.0011
	0.0369	0.0026	0.0074	0.0057	0.0003	0.0133	
x_s	0.0097	0.0260	0.1473	-0.2032	-0.0121	-0.0040	0.0034
	0.1157	0.0375	0.0847	0.0666	0.0048	0.0974	
x_b	0.1056	-0.0537	0.0775	0.0409	0.0016	0.2484	0.0011
	0.1038	0.0136	0.0380	0.0284	0.0018	0.0630	
s_y	-0.0173	0.0137	-0.0650	0.9855	-0.0001	0.0315	0.0503
	0.0450	0.0041	0.0117	0.0089	0.0005	0.0202	
s_{PE}	0.1030	0.4143	0.2025	-0.0562	0.9950	0.1357	2.9688
	0.1130	0.0256	0.0610	0.0464	0.0033	0.0786	
S_{spread}	-0.0086	-0.0046	-0.0644	0.0066	-0.0002	0.9480	0.0104
	0.0424	0.0033	0.0096	0.0073	0.0004	0.0171	
	Error correlation matrix						
r_{tbill}	0.2706	0.1109	0.0793	-0.0829	0.1783	0.0566	
x_s		4.2402	0.1196	-0.0536	0.7789	-0.0315	
x_b			1.4432	-0.6247	0.0648	0.2209	
s_y				0.4331	-0.0552	-0.8511	
s_{PE}					2.8543	-0.0202	
S_{spread}						0.3502	

Table 3: **Benchmark Results**

This table gives benchmark results for the $1/N$, unrestricted no predictability (NP) and restricted no predictability strategies ($NP - rest$). 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. Dynamic, myopic and constant proportion strategies are identical when there is no predictability.

	CER	$\overline{TW}$	$\sigma(TW)$
	Panel A: $\gamma = 2$		
$1/N$	0.0465	1.2876	0.2064
NP	0.0851	1.7816	0.8220
$NP - rest$	0.0704	1.5423	0.4718
	Panel B: $\gamma = 5$		
$1/N$	0.0385	1.2876	0.2064
NP	0.0439	1.3445	0.2564
$NP - rest$	0.0435	1.3564	0.2691
	Panel C: $\gamma = 10$		
$1/N$	0.0272	1.2876	0.2064
NP	0.0243	1.2124	0.1460
$NP - rest$	0.0253	1.2258	0.1547

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

This table gives the results for the dynamic (Dyn), myopic (Myop) and constant proportion (CP) 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 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 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	24.5372	71.2909	0.0817	1.5764	0.4508
	Myop	-1.0000	21.4363	59.2318	0.0813	1.5675	0.4337
	CP	-1.0000	4.5241	5.8324	0.0702	1.5022	0.4295
Shrinkage	Dyn	0.2952	10.1565	8.9631	0.0815	1.5613	0.3923
	Myop	0.2750	8.7744	8.0095	0.0806	1.5520	0.3870
	CP	-0.2131	2.7746	2.7292	0.0763	1.5217	0.3727
	Panel B: $\gamma = 5$						
Uniform	Dyn	0.0755	6.5258	7.0541	0.0640	1.5204	0.3996
	Myop	0.0762	4.6320	4.3931	0.0643	1.5128	0.3890
	CP	-0.0188	1.9490	1.1568	0.0520	1.4286	0.3418
Shrinkage	Dyn	0.1229	3.3674	1.6791	0.0672	1.5254	0.3505
	Myop	0.1162	2.8363	1.2084	0.0660	1.4966	0.3304
	CP	-0.0014	1.5965	0.6734	0.0591	1.4553	0.3192
	Panel C: $\gamma = 10$						
Uniform	Dyn	0.0425	3.0156	2.0637	0.0500	1.4872	0.3566
	Myop	0.0550	2.3874	1.2644	0.0448	1.4413	0.3249
	CP	-0.0028	1.4530	0.4975	0.0299	1.3155	0.2274
Shrinkage	Dyn	0.0660	2.0003	0.6170	0.0496	1.4470	0.2818
	Myop	0.0621	1.7855	0.4554	0.0490	1.4000	0.2414
	CP	0.0043	1.3255	0.3199	0.0352	1.3357	0.2312

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

This table gives the results for the dynamic (Dyn), myopic (Myop) and constant proportion (CP) 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 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 different risk aversion levels γ.

	Restricted Weights			
	Panel A: $\gamma = 2$			
		CER	$\overline{TW}$	$\sigma(TW)$
Uniform	Dyn	0.0829	1.5776	0.4342
	Myop	0.0829	1.5743	0.4239
	CP	0.0704	1.5028	0.4268
Shrinkage	Dyn	0.0831	1.5650	0.3747
	Myop	0.0827	1.5613	0.3730
	CP	0.0784	1.5325	0.3671
	Panel B: $\gamma = 5$			
Uniform	Dyn	0.0648	1.5335	0.4180
	Myop	0.0653	1.5135	0.3806
	CP	0.0448	1.5135	0.3136
Shrinkage	Dyn	0.0676	1.5205	0.3480
	Myop	0.0657	1.4921	0.3257
	CP	0.0567	1.4243	0.2944
	Panel C: $\gamma = 10$			
Uniform	Dyn	0.0519	1.4775	0.3415
	Myop	0.0485	1.4448	0.3174
	CP	0.0270	1.2711	0.1925
Shrinkage	Dyn	0.0516	1.4268	0.2575
	Myop	0.0485	1.4039	0.2559
	CP	0.0288	1.2797	0.1995

Table 6: **Classical significance tests**

This table presents classical t-tests which test the hypothesis whether the portfolio strategy and its benchmark performance are statistically similar. We give results for the plug-in approach, the decision-theoretic approach, different risk aversion levels, different type of strategies and for different restrictions. Note that we did not calculate results for the unrestricted decision-theoretic approaches.

		$\gamma = 2$		$\gamma = 5$		$\gamma = 10$	
		Unr	Restr	Unres	Restr	Unr	Restr
	Panel A: plug-in approach						
Uniform	Dyn	$-\infty$	0.9192	1.2582	2.8998	1.3854	4.0347
	Myopic	$-\infty$	0.8757	1.6926	2.6416	2.5229	3.4605
	CP	$-\infty$	-0.0171	-1.7395	1.0520	-1.6546	0.6774
Shrinkage	Dyn	4.4107	0.7892	3.8943	2.5255	3.2826	3.7253
	Myopic	4.2549	0.7134	3.9273	2.2724	3.4567	3.5524
	CP	-1.4846	0.4596	-1.3486	1.9360	-1.2361	3.0636
	Panel B: decision theoretic approach						
Uniform	Dyn	-	0.9964	-	2.8695	-	3.9470
	Myopic	-	0.9878	-	2.6932	-	3.7395
	CP	-	0.0001	-	0.1020	-	0.2172
Shrinkage	Dyn	-	0.8595	-	2.5221	-	3.6503
	Myopic	-	0.8291	-	2.2403	-	3.5423
	CP	-	0.6119	-	1.7624	-	0.9303

Table 7: **Plug-in approach - DP model**

This table gives the results for the dynamic (Dyn), myopic (Myop) and constant proportion (CP) strategies using the plug-in method. 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 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	23.5422	61.6013	0.0756	1.4924	0.3052
	Myop	-1.0000	18.3547	44.4534	0.0756	1.4924	0.3043
	CP	-1.0000	3.8380	4.5221	0.0787	1.5296	0.3470
Shrinkage	Dyn	0.2186	4.5198	3.1416	0.0682	1.4353	0.2660
	Myop	0.2035	4.0139	2.7565	0.0684	1.4359	0.2638
	CP	0.0381	2.0789	1.4753	0.0708	1.4963	0.4044
	Panel B: $\gamma = 5$						
Uniform	Dyn	-0.0888	5.6438	5.6729	0.0661	1.5091	0.3312
	Myop	-0.0020	4.0443	3.6314	0.0646	1.5010	0.3520
	CP	-0.0853	1.8069	1.0965	0.0579	1.4522	0.3141
Shrinkage	Dyn	0.0947	2.1152	0.6618	0.0587	1.4437	0.2862
	Myop	0.0883	1.9057	0.5503	0.0550	1.3945	0.2448
	CP	0.0176	1.4288	0.4874	0.0482	1.3805	0.2901
	Panel C: $\gamma = 10$						
Uniform	Dyn	-0.0517	2.8221	1.8944	0.0495	1.4779	0.3327
	Myop	0.0132	2.2438	1.1920	0.0474	1.4362	0.3194
	CP	-0.0400	1.3919	0.5179	0.0326	1.3294	0.2381
Shrinkage	Dyn	0.0522	1.5530	0.2937	0.0414	1.3627	0.2289
	Myop	0.0482	1.4495	0.2397	0.0378	1.3085	0.1914
	CP	0.0113	1.2521	0.2560	0.0281	1.2861	0.2131

Table 8: **Decision-theoretic approach - DP model**

This table gives the results for the dynamic (Dyn), myopic (Myop) and constant proportion (CP) 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 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.

	Restricted Weights			
	Panel A: $\gamma = 2$			
		CER	$\overline{TW}$	$\sigma(TW)$
Uniform	Dyn	0.0760	1.4981	0.3122
	Myop	0.0765	1.4995	0.3075
	CP	0.0780	1.5306	0.3607
Shrinkage	Dyn	0.0728	1.4667	0.2696
	Myop	0.0731	1.4693	0.2740
	CP	0.0728	1.5059	0.3965
	Panel B: $\gamma = 5$			
Uniform	Dyn	0.0649	1.5119	0.3470
	Myop	0.0640	1.4958	0.3468
	CP	0.0479	1.3966	0.3097
Shrinkage	Dyn	0.0603	1.4490	0.2850
	Myop	0.0570	1.4090	0.2520
	CP	0.0477	1.3752	0.2901
	Panel C: $\gamma = 10$			
Uniform	Dyn	0.0477	1.4794	0.3472
	Myop	0.0472	1.4324	0.3127
	CP	0.0251	1.2804	0.2100
Shrinkage	Dyn	0.0415	1.3550	0.2258
	Myop	0.0383	1.3173	0.2026
	CP	0.0266	1.2713	0.2005

Table 9: **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 (DPS2008). 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.

		BB2007		DPS2008	
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	0.5612	0.4388
	10	0.3000	0.4500	0.2958	0.4469
4	2	1.0000	0.0000	1.0000	0.0000
	5	0.5900	0.4100	0.5881	0.4119
	10	0.3100	0.2900	0.3083	0.2933
8	2	1.0000	0.0000	1.0000	0.0000
	5	0.6100	0.3900	0.6053	0.3947
	10	0.3200	0.3100	0.3172	0.3114
15	2	1.0000	0.0000	1.0000	0.0000
	5	0.6100	0.3900	0.6132	0.3868
	10	0.3100	0.3300	0.3075	0.3382
30	2	1.0000	0.0000	1.0000	0.0000
	5	0.6200	0.3800	0.6201	0.3799
	10	0.2900	0.4200	0.2935	0.4250
60	2	1.0000	0.0000	1.0000	0.0000
	5	0.6900	0.3100	0.6852	0.3148
	10	0.3500	0.4200	0.3243	0.4221