Model-Based Priors for Predicting the Equity Premium*

Mathias Kruttli†

July 10, 2014

Abstract

This paper attempts to improve the forecast performance of single variable predictive regressions used in the equity premium prediction literature through Bayesian priors derived from consumption-based asset pricing models. To implement these model-based priors, I develop a Bayesian procedure which is rooted in the macroeconometrics literature. The priors are derived from four asset pricing models: Habit Formation, Habit Formation Term Structure, Long Run Risk, and Prospect Theory. The model-based priors can substantially increase the explanatory power of the single variable predictive regressions. Further, they help to assess consumption-based asset pricing models in a novel way.

---

*I thank Tarun Ramadorai and Kevin Sheppard for excellent supervision. I thank Andrew Patton, Narayan Naik, Dimitri Vayanos, Mungo Wilson, and seminar participants at the Saïd Business School and the Oxford-Man Institute of Quantitative Finance for useful comments. Financial support from the David Walton Memorial Fund is thankfully acknowledged.

†Mathias Kruttli is at the Department of Economics, University of Oxford, and the Oxford-Man Institute of Quantitative Finance. Email: mathias.kruttli@economics.ox.ac.uk.
1 Introduction

Predicting aggregate stock returns has been of great interest to finance practitioners and academic finance economists alike. There exists an extensive literature that uses a variety of variables to predict stock market returns. Several papers, such as Campbell and Shiller (1988) and Fama and French (1988), have found that valuation ratios are correlated with subsequent returns. Other variables used to predict the equity premium are yields on corporate bonds and treasuries (see, for example, Campbell (1987) and Fama and French (1989)). More recently, new predictor variables have been suggested, for example, by Baker and Wurgler (2000) and Lettau and Ludvigson (2001). These papers use data on corporate payout and financing activity and the level of consumption relative to wealth, respectively. Another example is Kruttli, Patton, and Ramadorai (2014), who have shown that an aggregate measure of hedge fund illiquidity can predict excess returns across several asset classes.

In this paper, I impose parameter restrictions on typical regression models used to forecast the equity premium. The restrictions are deduced from the economic theory of consumption-based asset pricing models, and they help to substantially improve the out-of-sample (OOS) forecasts. This paper adds to a growing literature of imposing economic constraints to sharpen forecasts of the equity premium (see, for example, Campbell and Thompson (2008), Pastor and Stambaugh (2009), Pastor and Stambaugh (2012), and Pettenuzzo, Timmermann, and Valkanov (2013)).

Goyal and Welch (2008), henceforth Goyal and Welch, provide a comprehensive analysis of the in-sample and out-of-sample (OOS) predictive power of all the major variables used in the literature up to that point. They estimate a single variable predictive regression in-sample and OOS – the equity premium is predicted OOS by estimating the parameters with data up to $t$ and forecasting the equity premium in $t + 1$:

$$ Equity\Premium_{t+1} = \beta_0 + \beta_1 x_t + \epsilon_{t+1}, \text{ where } \epsilon_{t+1} \sim N(0, \sigma^2), $$

(1)
where the equity premium, defined as the rate of return on the stock market minus the prevailing short term interest rate, is regressed on a lagged predictor $x_t$. Goyal and Welch estimate this model for various predictors, for different time periods, and at multiple frequencies. Their findings are rather sobering: while most of the predictors are not even significant in-sample, those that have some predictive power in-sample still suffer from very poor OOS performance. Goyal and Welch compare the predictive regressions with the historical average, i.e. predicting that the equity premium in $t + 1$ will be what it has been on average up to $t$, i.e. $\beta_1 x_t$ is eliminated from the model. The historical average outperforms the predictive regression OOS for the majority of the predictors. Hence, an investor would have been better off using the historical average to forecast the equity premium rather than any of the available variables.

Campbell and Thompson (2008), henceforth Campbell and Thompson, respond to the claims of Goyal and Welch by imposing restrictions on the parameter estimates of the predictive regressions. Through these restrictions, the equity premium can be forecast OOS such that a real time investor could profit. This paper attempts to further investigate the effects of economic constraints on model parameters. I address the claim of poor OOS predictability of the equity premium by implementing restrictions through well specified Bayesian priors. Consumption-based asset pricing models act as a source for these priors – hence, I call them “model-based priors” – which are then imposed on the single variable predictive regression. The procedure can be considered as estimating the parameters in equation (1) with a data set that combines empirical data with data simulated from the asset pricing model. I find that the model-based priors can raise the OOS $R^2$ by several percentage points for all of the predictor variables considered. The results are particularly good for the postwar data sample starting in 1947.

Besides trying to overcome the weak predictability of the equity premium OOS, my approach can also be interpreted as a different perspective on the assessment
of asset pricing models. Generally, consumption-based asset pricing models are evaluated based on their ability to match empirical data moments: in this paper, the effectiveness of their priors is scrutinised.

The asset pricing models used to derive the model-based priors share a common origin. They have been developed to offer a solution to the equity premium puzzle (see Mehra and Prescott (1985)). I use four models to derive model-based priors. These models are the Habit Formation (HF) model (see Campbell and Cochrane (1999)), the Prospect Theory (PT) model (see Barberis, Huang, and Santos (2001)), the Long Run Risk (LRR) model (see Bansal and Yaron (2004)), and the Habit Formation Term Structure (HFTS) model (see Wachter (2006)). I briefly describe these models in Appendix B.

The structure of this paper is as follows. In Section 2, I explain the motivation behind the Bayesian method. Section 3 discusses the Bayesian procedure. In Section 4, I describe the empirical and simulated data used for my analysis. Section 5 presents the results. Section 6 discusses the working of the model-based prior in detail, and Section 7 concludes.

2 Bayesian Motivation

The Bayesian method is related to Del Negro and Schorfheide (2004), who combine a dynamic stochastic general equilibrium (DSGE) model with a vector autoregression (VAR) and create a DSGE-VAR. I use prior distributions derived from consumption-based asset pricing models to estimate the parameters of the single variable predictive regression in equation (1). These model-based priors can improve the OOS forecasts of the equity premium in two ways. First, they reduce the overfitting of the coefficient estimates. The signals of the predictors are generally very weak, and thus, parameters estimated solely on empirical data suffer from the influence of noise. The idea of Bayesian priors improving the forecast performance
of a regression model by reducing the overfitting dates back to the Minnesota prior of Doan, Litterman, and Sims (1984). Second, because of the iterative nature of the OOS forecasting procedure, the OOS estimates of $\beta_0$ and $\beta_1$, denoted $\hat{\beta}_0$ and $\hat{\beta}_1$, respectively, can be very different from their in-sample counterparts at some points in the sample. In these cases, the fundamental relationship between the predictor and the equity premium, which holds over the total sample, fails to hold temporarily. This leads inevitably to inaccurate predictions of the equity premium. Depending on the theoretical estimates implied by the asset pricing model, the model-based prior can bring the parameter estimates closer to their fundamental values during such periods – the model-based prior helps to bring the OOS performance closer to the in-sample performance without introducing an in-sample look-ahead bias.

The data sets used by the authors to calibrate the asset pricing models overlap partially with my OOS forecasting period. Hence, there exists an implicit look-ahead bias. To address this issue, I also use the model-based priors to predict the equity premium over a subsample of the data which has no overlap with the calibration data, and the results are robust to this adjustment of the sample period.

How does my methodology compare to that of Campbell and Thompson? Campbell and Thompson impose three restrictions on the parameter estimates of the predictive regression sequentially – where restrictions 2 and 3 only apply to valuation ratios. First, whenever $\hat{\beta}_1$ takes on a sign which is not in line with the in-sample estimate, it is set to zero. Second, they set the intercept to zero whenever it is negative and the coefficient to one when greater than one and zero when negative. Third, they set $\hat{\beta}_0$ equal to zero and $\hat{\beta}_1$ equal to one regardless of the estimates. For each of the three restrictions, Campbell and Thompson successfully improve the OOS forecasts of the single variable predictive regressions. The third restriction is particularly interesting: Campbell and Thompson set the constant and the coefficient to zero and one, respectively, because these are the theoretical values implied by the steady state theory of Gordon (1962). This can be considered as imposing
a model-based prior very dogmatically, i.e. the prior is very tight and no empirical
data are used to estimate $\beta_0$ and $\beta_1$.\footnote{Campbell and Thompson also improve the OOS forecast performance through changing the predictor variable by adding a growth term derived from the Gordon (1962) growth model or restricting the equity premium forecast to be greater or equal to zero. Since I solely focus on the parameter estimation of the predictive regression rather than on transforming the predictor or imposing sign restrictions on the forecast, I do not discuss this aspect of their paper.}

This paper builds on Campbell and Thompson and develops their parameter
restrictions further. My methodology differs from theirs as the model-based priors
are based on a well specified Bayesian procedure. This has the advantage of allowing
us to adjust the tightness of the prior gradually rather than imposing it dogmatically.
Further, the model-based prior can reduce the overfitting of parameter estimates
which do not violate any of the boundaries, i.e. an estimate of $\beta_0$ which lies above
zero or a coefficient estimate which is between zero and one. The model-based prior
can also push negative estimates above zero and positive estimates below one, where
the restrictions used by Campbell and Thompson only set them as equal to zero or
one, respectively. The two approaches are compared in Table 8 in Appendix C.

The model-based prior is rooted in the economic theory of the consumption
based asset pricing models. An alternative prior is the Minnesota prior, introduced
by Doan et al. (1984), which is a purely statistical way of reducing overfitting in
parameter estimates. A comparison of the performance of the model-based priors
with an adaption of the Minnesota prior is performed in Section 5.7. This comparison
makes it possible to assess whether the economic theory of the asset pricing models
helps to improve the OOS forecasting performance beyond the statistical shrinking
of the parameter estimates by the Minnesota prior.

### 3 Econometric Methodology

This section lays out the econometric methodology. First, I describe the empirical
procedure for forecasting the equity premium and the assessment of the forecast
performance. Second, I illustrate the Bayesian procedure.
3.1 Empirical procedure

The predictive regressions are specified as in Goyal and Welch and Campbell and Thompson. The equity premium is denoted by \( y_t \) and is defined as the rate of return on the stock market minus the prevailing short term interest rate. The equity premium is regressed on a constant and an independent variable lagged by one period, the predictor \( x_{t-1} \):

\[
y_t = \beta_0 + \beta_1 x_{t-1} + \epsilon_t, \quad \text{where} \quad \epsilon_t \overset{i.i.d.}{\sim} N(0, \sigma^2).
\] (2)

For a sample which starts in period 0 and ends in period \( T \), the in-sample regression uses all available data, i.e. \( T \) data points, to run the regression in (2). The OOS predictions of the equity premium are generated through recursive forecasts. I take all available observations up to period \( t \) and estimate the model in (2). Based on the resulting estimate \( \hat{\beta} \) of the \( 2 \times 1 \) vector \( \beta = [\beta_0, \beta_1]' \) and the observation \( x_t \), one can forecast the equity premium in \( t + 1 \). The predicted equity premium is denoted by \( \hat{y}_{t+1} \). Although I have observations up to \( T \), observations after \( t \) are not used to estimate \( \beta \): hence, a real time investor who is interested in predicting the equity premium can implement this procedure. Since at the beginning of the sample there is an insufficient number of observations to estimate \( \beta \) accurately, the recursive OOS forecast starts after period \( z \) where \( z < T \), i.e. observations from period 1 to \( z \) are used as burn-in data.\(^2\)

To measure the OOS forecast performance, I compare the predictive regression forecasts with predictions based on the historical average model. The historical average model forecasts the equity premium in \( t + 1 \), denoted by \( \hat{y}_{t+1}^h \), as its mean up to period \( t \) – we conjecture that the best forecast for the equity premium in \( t + 1 \) is what it has been on average up to \( t \). This is equivalent to estimating the

\(^2\)I also attempted a rolling forecast procedure, which does not use the whole sample to estimate \( \beta \). Instead, a number of periods \( j \) is determined and the regression in (2) is estimated based on observations from \( t-j \) to \( t \). The model-based prior is also successful for rolling windows. However, the recursive forecast performs generally better than the rolling forecast.
regression in (2) with $\beta_1$ set equal to zero.

The two models are compared via an OOS $R^2$, which is computed as

$$R^2_{OS} = 1 - \frac{\sum_{t=z+1}^{T} (y_t - \hat{y}_t^p)^2}{\sum_{t=z+1}^{T} (y_t - \hat{y}_t^h)^2}.$$  \hfill (3)

In the denominator, contrary to the in-sample $R^2$ for which the average equity premium is used to compute the residuals, the forecasting errors are based on the historical average equity premium $\hat{y}_t^h$. A high OOS $R^2$ is a sign that the predictive regression is the superior model. Campbell and Thompson emphasize the importance of the OOS $R^2$ statistic for a real time investor. They show that even a small positive OOS $R^2$ can lead to considerable utility gains and an increase in expected returns for investors with standard mean-variance preferences.

### 3.2 Model-based prior

This section describes the model-based prior. A detailed description of the Bayesian procedure and the derivation of the posterior estimates is provided in Appendix A. The implementation of the prior is based on Del Negro and Schorfheide (2004).

The posterior distribution of the predictive regression parameters is

$$p(\beta, \sigma^2 \mid Y, X) = \frac{p(Y, X \mid \beta, \sigma^2)p(\beta, \sigma^2 \mid \theta)}{p(Y, X \mid \theta)},$$ \hfill (4)

where $Y$ is a $T \times 1$ vector with elements $y_t$, $X$ is a $T \times 2$ matrix with rows $[1 \ x_t]$, $\beta$ and $\sigma^2$ are the parameters of the single variable predictive regression, and $\theta$ stands for the parameters of the consumption-based asset pricing model. The parameters of the asset pricing model are set equal to the values calibrated by the authors of the respective model. The prior for $\beta$ and $\sigma^2$ derived from the asset pricing model is given by $p(\beta, \sigma^2 | \theta)$. I specify the prior in two different ways. The first specification is called the population moments (PM) prior. When imposing the PM prior, the
modes of the posterior distributions of $\beta$ and $\sigma^2$ are given by

$$\tilde{\beta}_{PM}(\theta) = (\lambda T \Gamma^*_{xx}(\theta) + X'X)^{-1}(\lambda T \Gamma^*_{xy}(\theta) + X'Y)$$ and

$$\tilde{\sigma}^2_{PM}(\theta) = \frac{1}{(\lambda + 1)T}[(\lambda T \Gamma^*_{yy}(\theta) + Y'Y)$$

$$- (\lambda T \Gamma^*_{yx}(\theta) + Y'X)(\lambda T \Gamma^*_{xx}(\theta) + X'X)^{-1}(\lambda T \Gamma^*_{xy}(\theta) + X'Y)],$$

where $\lambda T \Gamma^*_{yy}(\theta)$, $\lambda T \Gamma^*_{yx}(\theta)$, and $\lambda T \Gamma^*_{xx}(\theta)$, with $\Gamma^*_{yy}(\theta) = E[y_t^2]$, are the population moments of the data simulated from the asset pricing model scaled by $T$ and the hyperparameter $\lambda$, which determines the weight of the model-based prior – how strongly we believe the model implied parameter estimates relative to the empirical estimates. The population moments are estimated through a Monte Carlo procedure, discussed in Section 4.2.1. The estimators in (5) and (6), respectively, can be interpreted as OLS estimators for a data set that combines empirical and simulated data. The second specification of the prior is the empirical state variables (ESV) prior. The modes of the posterior distributions of $\beta$ and $\sigma^2$ when imposing the ESV prior are

$$\tilde{\beta}_{ESV}(\theta) = (\lambda \bar{X}'\bar{X}(\theta) + X'X)^{-1}(\lambda \bar{X}'\bar{Y}(\theta) + X'Y)$$ and

$$\tilde{\sigma}^2_{ESV}(\theta) = \frac{1}{(\lambda + 1)T}[(\lambda \bar{Y}'\bar{Y} + Y'Y)$$

$$- (\lambda \bar{Y}'\bar{X} + Y'X)(\lambda \bar{X}'\bar{X} + X'X)^{-1}(\lambda \bar{X}'\bar{Y} + X'Y)].$$

The state variables of all the asset pricing models considered in this paper depend on at least one of the three variables: consumption growth, dividend growth, and inflation rate. The asset pricing models have assumed data generating processes for these variables, which are used to simulate the model randomly. For example, the HF model specifies the consumption growth process as $\Delta c_{t+1} = g + v_{t+1}$, where $g$ is a constant and $v_{t+1} \overset{i.i.d.}{\sim} N(0, \sigma^2_v)$. However, instead of drawing random shocks $v_{t+1}$ and simulating a consumption growth series, one can feed the empirical consumption growth data into the model. The resulting time series of the predictor and the equity
premium are then used for the model-based ESV prior. The model data simulated with empirical state variables are denoted by \((\bar{Y}, \bar{X})\). For notational simplicity, the dependency on \(\theta\) is not stated explicitly. As with the population moments, the simulated data can be scaled by \(\lambda\) to give them more weight in the posterior distribution.

The PM and the ESV prior can be interpreted as two different ways of assessing the asset pricing models. The PM prior requires that the theoretical relationship between the equity premium and the predictive variable implied by the model provides useful information for estimating the coefficients. Hence, the model is assessed regarding the empirical validity of its theory. The ESV prior is a more difficult challenge for the asset pricing model. The ESV prior also assesses the theoretical relationship between the equity premium and the predictor implied by the simulated data, but the simulated data are based on empirical consumption growth, dividend growth, or inflation, and not on the model’s assumed data generating processes.

The procedure laid out above is based on the total sample of size \(T\). For a recursive forecasting exercise, I must implement the model-based prior in every period \(t\) with data up to \(t\). The resulting posterior coefficient estimates, given in equations (5) and (7), are then used to forecast the equity premium in \(t + 1\).

For the model-based prior to be of any value for a real time investor, the weight of the prior has to be estimated OOS. Hence, \(\lambda\) is estimated by maximising the data density \(p(Y, X \mid \theta)\) in equation (4) numerically over a grid with data up to period \(z\) and held constant for the whole forecasting exercise.\(^3\) Generally speaking, the \(\lambda\) value which maximises the data density is higher when the simulated data match the empirical data more closely. The exact forms of the data densities for the PM and ESV prior are given in Appendix A.3.

\(^3\)Reestimating the \(\lambda\) before every forecast changes the results only marginally.
4 Data

The availability of variables which can be used to implement the model-based priors is determined by the four asset pricing models. The variables of the equity premium prediction literature for which at least one model provides theoretical predictions are: the equity premium, the dividend-price ratio, the dividend yield, the long term yield, the short term yield, and the term spread. The equity premium is the difference between the log stock returns and the log of the risk-free rate. The dividend-price ratio is computed as the difference between the log dividends and the log price. The dividend yield is defined as the difference between the log dividends and the log price lagged by one period. The long term yield is the annualised rate on long term U.S. government bonds. The short term yield is given by the annualised rate on the short term U.S. government debt. The term spread is the difference between the two.

4.1 Empirical data

The empirical time series of the equity premium and the predictor variables are taken from Amit Goyal’s website. This ensures that I use the same data as Goyal and Welch did, with the only difference being that my sample ends in 2011 instead of 2005. The returns and price of the S&P 500 index are provided by Amit Goyal from 1872 to 2011. The original source of the time series up to 1926 is Robert Shiller’s website. From 1926, returns from The Center for Research in Security Prices are used. The risk-free rate is the 3-month U.S. Treasury bill rate. Dividends on the S&P 500 index are 12-month moving sums, and the time series starts in 1871 and ends in 2011. For the yields, the time series of the annualised rates on the 3-month U.S. Treasury bill and the long term U.S. government bonds start in 1920 and 1919, respectively.
4.2 Simulated data

An overview of how I simulate data from each asset pricing model is given below. I simulate the asset pricing models at the frequency proposed by the respective authors.

4.2.1 Monte Carlo simulation

Three out of the four asset pricing models lack analytical solutions. Hence, I estimate the model-based population moments, which are needed to implement the PM prior, through a Monte Carlo simulation.\(^4\) The set-up of the Monte Carlo simulation is the same for all models. I randomly simulate data \((Y^*, X^*)\), where \(Y^*\) is a vector containing the simulated equity premium and \(X^*\) is a matrix containing a constant and the simulated predictor variable, based on the assumed data generating processes of the respective model for \(S\) periods and estimate the population moments \(\Gamma_{yy}^*(\theta), \Gamma_{yx}^*(\theta),\) and \(\Gamma_{xx}^*(\theta)\), through the sample moments. The estimators are 
\[
\hat{\Gamma}_{yy}^*(\theta) = E[y_t^2], \quad \hat{\Gamma}_{yx}^*(\theta) = E[y_t^*(1, x_{t-1}^*)],\quad \text{and} \quad \hat{\Gamma}_{xx}^*(\theta) = E[(1, x_{t-1}^*)'(1, x_{t-1}^*)].
\]

This procedure is iterated \(P\) times, which results in a total of \(S \times P\) observations, and the estimates of the population moments are averaged across the \(P\) simulations. Since the equity premium and the predictive variables generated by the models are covariance stationary, averaging the \(P\) estimates should result in a precise estimate of the population moments conditional on \(S\) and \(P\) being large enough.\(^5\)

4.2.2 Habit Formation model

Campbell and Cochrane (1999) simulate the HF model at a monthly frequency and aggregate the simulated data to match annual moments of the empirical data. I follow them and use priors from the HF model at an annual frequency. The model

\(^4\)The LRR model has approximate analytical solutions, but these and the Monte Carlo estimates are identical.

\(^5\)I use the number of simulated periods by the authors of the respective asset pricing model as a guideline for my simulation.
generates data for the equity premium, the dividend-price ratio, and the dividend yield.\footnote{The HF model has two specifications. I use their baseline specification which assumes perfect correlation between consumption and dividend growth.}

For the PM prior, I implement the Monte Carlo procedure by simulating data for 10,000 years of data, estimating the population moments, and iterating the procedure 10 times.

To simulate data for the ESV prior, we must feed empirical data for real log consumption growth. Real log consumption growth is defined as log growth in real per capita consumption expenditure on non-durables and services in the U.S. The annual data are taken from John Campbell’s website and extended with data from Federal Reserve Economic Data (FRED). The consumption data series from John Campbell’s website starts in 1890 at an annual frequency. To simulate the model at a monthly frequency when feeding real log consumption growth, the annual consumption growth rate is divided into monthly consumption growth rates.

4.2.3 Prospect Theory model

The PT model by Barberis et al. (2001) is simulated at an annual frequency in their paper. To be consistent with the model calibration of the authors, I use priors from the PT model at an annual frequency. The model generates data for the equity premium, the dividend-price ratio, and the dividend yield.

To estimate the population moments for the PM prior, I simulate 10,000 years of data and average the estimates over 10 iterations.

To simulate data for the ESV prior, only one empirical time series is required. This is the time series for the error terms of the real dividend growth process. Since I have the real dividend growth rate and the parameter values provided by Barberis et al. (2001), I can back out the error terms.\footnote{Barberis et al. (2001) run their model with different parameter values for \(b_0\) and \(k\). I use the specification which is most successful in matching the empirically observed moments of asset returns. This specification sets \(b_0\) equal to 100 and \(k\) equal to 8. More details are provided in Appendix B.}
4.2.4 Long Run Risk model

Similar to Campbell and Cochrane (1999), Bansal and Yaron (2004) simulate the model at a monthly frequency and aggregate the simulated data to match the annual empirical moments. Hence, priors from the LRR model are used at an annual frequency in this paper. The simulated variables are the equity premium, the dividend-price ratio, and the dividend yield.

For the PM prior, the Monte Carlo procedure is implemented by estimating the population moments from simulated data for 100,000 months aggregated to annual levels and averaging them over 100 iterations.

Real log consumption growth and real log dividend growth act as input variables to generate data for the ESV prior. Further, the two state variables, the long run risk component and the time varying volatility, are required. I follow Bansal, Kiku, and Yaron (2007), who provide a simple procedure for extracting the two state variables from the data. The consumption data are the same as for the HF model.

4.2.5 Habit Formation Term Structure model

Wachter (2006) simulates the HFTS model at a quarterly frequency and assesses the model against quarterly empirical moments. Thus, I use priors from the HFTS model at a quarterly frequency. The simulated variables are: the equity premium, the long term yield, the short term yield, and the term spread.\footnote{I could also simulate data on the dividend-price ratio and the dividend yield, but the theoretical relationship between these variables and the equity premium is the same as implied by the HF model.}

For the PM prior, the population moments are estimated from 40,000 quarters of simulated data and averaged over 10 iterations.

To implement the ESV prior, two empirical time series are required. The first one is real log consumption growth. FRED provides quarterly data on consumption expenditure starting in 1947. The second one is inflation. I compute inflation based on CPI values from FRED.
5 Results

5.1 Habit Formation model-based priors

The HF model generates priors for the dividend-price ratio and the dividend yield. The results of imposing these priors on predictive regressions with annual forecast horizons are shown in Table 1. Panel A shows the results for the PM prior, and Panel B presents the results for the ESV prior. For both priors, I forecast the equity premium with annual returns and overlapping annual returns. For annual returns, the forecasting procedure is exactly as described in Section 3.1, with \( t \) being a calendar year. For overlapping annual returns, the forecasts are also at annual horizons but iterated on a monthly basis, with \( t \) comprising 12 months of data. The motivation behind overlapping annual returns is to increase the number of predictions. This is particularly important when only subsamples of the data set are used, as in Section 5.6.

The empirical OOS statistics are computed by running the predictive regression solely on empirical data. I show the results when imposing the model-based prior for a value of \( \lambda \) which is estimated as explained in Section 3.2. To estimate \( \lambda \), I only use data up to and not including the first period of the forecasting exercise. The OOS \( R^2 \) assesses how the predictive regression with or without model-based prior performs relatively to the historical average. In the \( \Delta R^2_{OS} \) column, the difference between the OOS \( R^2 \) of the predictor with and without model-based prior is shown – a positive value indicates that the model-based prior leads to an improvement in the OOS forecast.

I analyse the performance of both priors over two time periods. The first period covers the total sample. For the PM prior, only empirical return data and dividend data are required. Hence, the time period starts in 1872. For the ESV prior, we also need empirical consumption data which is fed to the model. The sample period starts in 1890 for the dividend-price ratio and 1891 for the dividend yield, since
Annual consumption data are unavailable before this. The second sample period used for the analysis is the postwar period, from 1947 to 2011. When choosing the starting period of the OOS forecast, I use at least 20 years of data for the initial regression. However, I do not start the OOS exercise before 1926 since high quality S&P 500 return data from CRSP are available from 1926 onwards – this start date has been motivated by Campbell and Thompson.

The PM prior is successful in improving the OOS performance of the predictive regression for both predictors and sample periods regardless of whether annual returns or overlapping annual returns are used. The OOS $R^2$ are increased up to several percentage points. The ESV prior fares far worse for the total sample than for the postwar sample. This can be explained by the great challenge the HF model faces when trying to replicate the market movements of the great depression and the two world wars. However, the performance of the ESV prior for the postwar sample is remarkable. Improving the OOS $R^2$ from a negative value up to 6.96% through model-based data which is simulated by feeding an empirical consumption growth process is impressive.

For the PM prior, the values for $\lambda$ are higher than for the ESV prior. This means that we put more weight on the simulated data when estimating the constant and coefficient of the predictive regression when imposing the PM prior. The increase in OOS $R^2$ might appear small. However, Campbell and Thompson show that even a small positive OOS $R^2$ can translate into substantially higher expected returns for an investor with mean-variance preferences. They derive the difference in expected returns for an investor who forecasts the equity premium with a predictor and an investor who relies on the historical average return model as

$$\frac{1}{\gamma}(\frac{R^2_{OS}}{1 - R^2_{OS}})(1 + S^2),$$

where $\gamma$ is the risk aversion of the mean-variance preferences, and $S$ is the Sharpe
ratio of the equity premium. The Sharpe ratio of the annual equity premium from 1967 to 2011 was 0.231. An investor who uses the dividend-price ratio as a predictor and imposes the ESV prior of the HF model over the postwar sample achieves and OOS $R^2$ of 6.48%. Hence, the difference in annual expected returns in percentage points compared to an investor who relies on the historical average model is 7.30 for a risk aversion of 1 and 2.43 for a risk aversion of 3.

5.2 Prospect Theory model-based priors

The PT model provides priors for the dividend-price ratio and the dividend yield at an annual frequency. The results for implementing these model-based priors are shown in Table 2.

The PM prior improves the OOS $R^2$ regardless of the predictor, the data sample, and whether we use annual returns or annual overlapping returns. The PM prior fares extremely well for the postwar sample by pushing the OOS $R^2$ up to 5%. In most cases, the $\lambda$ values are smaller than 1, implying that the empirical data are dominant in the posterior, except for the dividend yield in the postwar sample. There, the simulated data are weighted much stronger in the posterior than the empirical data. This indicates that the PT model matches the empirical data well. The ESV prior performs poorly for the total sample and fails to improve the OOS performance for both predictor variables. However, for the postwar sample, the ESV prior improves the OOS predictability substantially, with the OOS $R^2$ being of similar magnitude as when using the PM prior.

5.3 Long Run Risk model-based priors

Through the LRR model, we can obtain priors for the dividend-price ratio and the dividend yield. The priors are implemented at an annual frequency and shown in

---

9Since the PT model is simulated at an annual frequency, I only use the PM prior for the overlapping forecast procedure and not the ESV prior.
Table 1: Results for the Habit Formation Model-Based Priors

Reported are the results for the model-based priors from the HF model. The weight of the prior, $\lambda$, is estimated as explained in Section 3.2. The $\Delta R^2_{OS}$ stands for the difference between the $R^2_{OS}$ of the predictor with and without model-based prior. A positive value indicates an improvement in the OOS forecasts through the prior. The data sample ends in December 2011. For the overlapping annual returns, the forecasting procedure is repeated in monthly steps, but the forecasts are at annual horizons.

Panel A: PM Prior

<table>
<thead>
<tr>
<th>Predictor: Div.-Price Ratio</th>
<th>Total Sample</th>
<th>Postwar Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Empirical OOS Statistics:</td>
<td>$-0.39%$</td>
<td>$-0.87%$</td>
</tr>
<tr>
<td>PM OOS Statistics:</td>
<td>$0.87%$</td>
<td>$1.11%$</td>
</tr>
<tr>
<td>$\Delta R^2_{OS}$</td>
<td>$0.01%$</td>
<td>$2.64%$</td>
</tr>
<tr>
<td>$\lambda R^2_{OS}$</td>
<td>$1.58%$</td>
<td>$2.66%$</td>
</tr>
<tr>
<td>$\Delta R^2_{OS}$</td>
<td>$0.67%$</td>
<td>$0.86%$</td>
</tr>
<tr>
<td>$\lambda R^2_{OS}$</td>
<td>$2.75%$</td>
<td>$0.96%$</td>
</tr>
</tbody>
</table>

Panel B: ESV Prior

<table>
<thead>
<tr>
<th>Predictor: Div.-Price Ratio</th>
<th>Total Sample</th>
<th>Postwar Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Empirical OOS Statistics:</td>
<td>$-0.28%$</td>
<td>$-0.87%$</td>
</tr>
<tr>
<td>ESV OOS Statistics:</td>
<td>$0.08%$</td>
<td>$0.73%$</td>
</tr>
<tr>
<td>$\Delta R^2_{OS}$</td>
<td>$84.18%$</td>
<td>$6.48%$</td>
</tr>
<tr>
<td>$\lambda R^2_{OS}$</td>
<td>$0.13%$</td>
<td>$-0.10%$</td>
</tr>
<tr>
<td>$\Delta R^2_{OS}$</td>
<td>$-66.37%$</td>
<td>$2.50%$</td>
</tr>
<tr>
<td>$\lambda R^2_{OS}$</td>
<td>$-66.50%$</td>
<td>$2.60%$</td>
</tr>
</tbody>
</table>
Table 2: Results for the Prospect Theory Model-Based Priors

Reported are the results for the model-based priors from the PT model. The weight of the prior, $\lambda$, is estimated as explained in Section 3.2. The $\Delta R^2_{OS}$ stands for the difference between the $R^2_{OS}$ of the predictor with and without model-based prior. A positive value indicates an improvement in the OOS forecasts through the prior. The data sample ends in December 2011. For the overlapping annual returns, the forecasting procedure is repeated in monthly steps, but the forecasts are at annual horizons.

<table>
<thead>
<tr>
<th>Panel A: PM Prior</th>
<th>Total Sample</th>
<th>Postwar Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Annual Returns</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda$</td>
<td>$R^2_{OS}$</td>
<td>$\Delta R^2_{OS}$</td>
</tr>
<tr>
<td><strong>Predictor: Div.-Price Ratio</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Empirical OOS Statistics:</td>
<td>-0.39%</td>
<td></td>
</tr>
<tr>
<td>PM OOS Statistics:</td>
<td>0.22</td>
<td>0.07%</td>
</tr>
<tr>
<td><strong>Predictor: Div. Yield</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Empirical OOS Statistics:</td>
<td>-0.35%</td>
<td></td>
</tr>
<tr>
<td>PM OOS Statistics:</td>
<td>0.24</td>
<td>0.63%</td>
</tr>
<tr>
<td><strong>Overlap. Ann. Returns</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda$</td>
<td>$R^2_{OS}$</td>
<td>$\Delta R^2_{OS}$</td>
</tr>
<tr>
<td><strong>Predictor: Div.-Price Ratio</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Empirical OOS Statistics:</td>
<td>-1.41%</td>
<td></td>
</tr>
<tr>
<td>PM OOS Statistics:</td>
<td>0.27</td>
<td>-0.15%</td>
</tr>
<tr>
<td><strong>Predictor: Div. Yield</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Empirical OOS Statistics:</td>
<td>-0.85%</td>
<td></td>
</tr>
<tr>
<td>PM OOS Statistics:</td>
<td>0.33</td>
<td>0.42%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B: ESV Prior</th>
<th>Total Sample</th>
<th>Postwar Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Annual Returns</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda$</td>
<td>$R^2_{OS}$</td>
<td>$\Delta R^2_{OS}$</td>
</tr>
<tr>
<td><strong>Predictor: Div.-Price Ratio</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Empirical OOS Statistics:</td>
<td>-0.28%</td>
<td></td>
</tr>
<tr>
<td>ESV OOS Statistics:</td>
<td>0.56</td>
<td>-0.60%</td>
</tr>
<tr>
<td><strong>Predictor: Div. Yield</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Empirical OOS Statistics:</td>
<td>0.55%</td>
<td></td>
</tr>
<tr>
<td>ESV OOS Statistics:</td>
<td>0.50</td>
<td>-0.40%</td>
</tr>
</tbody>
</table>
Table 3: Results for the Long Run Risk Model-Based Priors

Reported are the results for the model-based priors from the LRR model. The weight of the prior, \( \lambda \), is estimated as explained in Section 3.2. The \( \Delta R^2_{OS} \) stands for the difference between the \( R^2_{OS} \) of the predictor with and without model-based prior. A positive value indicates an improvement in the OOS forecasts through the prior. The data sample ends in December 2011. For the overlapping annual returns, the forecasting procedure is repeated in monthly steps, but the forecasts are at annual horizons.

<table>
<thead>
<tr>
<th>Panel A: PM Prior</th>
<th>Total Sample</th>
<th>Postwar Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annual Returns</td>
<td>Sample Begin: 1872</td>
<td>Sample Begin: 1947</td>
</tr>
<tr>
<td></td>
<td>Forecast Begin: 1926</td>
<td>Forecast Begin: 1967</td>
</tr>
<tr>
<td>Predictor: Div.-Price Ratio</td>
<td>( \lambda )</td>
<td>( R^2_{OS} )</td>
</tr>
<tr>
<td>Empirical OOS Statistics:</td>
<td>-0.39%</td>
<td>1.86%</td>
</tr>
<tr>
<td>PM OOS Statistics:</td>
<td>5.05</td>
<td>5.05</td>
</tr>
<tr>
<td>Predictor: Div. Yield</td>
<td>-0.35%</td>
<td>9.10</td>
</tr>
<tr>
<td>Empirical OOS Statistics:</td>
<td>2.32%</td>
<td>5.05</td>
</tr>
<tr>
<td>PM OOS Statistics:</td>
<td>5.05</td>
<td>1.97%</td>
</tr>
<tr>
<td></td>
<td>Forecast Begin: 12M 1926</td>
<td>Forecast Begin: 12M 1967</td>
</tr>
<tr>
<td>Predictor: Div.-Price Ratio</td>
<td>( \lambda )</td>
<td>( R^2_{OS} )</td>
</tr>
<tr>
<td>Empirical OOS Statistics:</td>
<td>-1.41%</td>
<td>2.86%</td>
</tr>
<tr>
<td>PM OOS Statistics:</td>
<td>4.95</td>
<td>4.95</td>
</tr>
<tr>
<td>Predictor: Div. Yield</td>
<td>-0.85%</td>
<td>15.32</td>
</tr>
<tr>
<td>Empirical OOS Statistics:</td>
<td>2.72%</td>
<td>15.32</td>
</tr>
<tr>
<td>PM OOS Statistics:</td>
<td>4.95</td>
<td>1.87%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B: ESV Prior</th>
<th>Total Sample</th>
<th>Postwar Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annual Returns</td>
<td>Sample Begin: 1892</td>
<td>Sample Begin: 1947</td>
</tr>
<tr>
<td></td>
<td>Forecast Begin: 1926</td>
<td>Forecast Begin: 1967</td>
</tr>
<tr>
<td>Predictor: Div.-Price Ratio</td>
<td>( \lambda )</td>
<td>( R^2_{OS} )</td>
</tr>
<tr>
<td>Empirical OOS Statistics:</td>
<td>-0.55%</td>
<td>-1.23%</td>
</tr>
<tr>
<td>ESV OOS Statistics:</td>
<td>0.38</td>
<td>0.38</td>
</tr>
<tr>
<td>Predictor: Div. Yield</td>
<td>-0.85%</td>
<td>0.37</td>
</tr>
<tr>
<td>Empirical OOS Statistics:</td>
<td>-4.46%</td>
<td>0.08</td>
</tr>
<tr>
<td>ESV OOS Statistics:</td>
<td>0.08</td>
<td>-4.23%</td>
</tr>
<tr>
<td></td>
<td>Forecast Begin: 12M 1934</td>
<td>Forecast Begin: 12M 1967</td>
</tr>
<tr>
<td>Predictor: Div.-Price Ratio</td>
<td>( \lambda )</td>
<td>( R^2_{OS} )</td>
</tr>
<tr>
<td>Empirical OOS Statistics:</td>
<td>-3.12%</td>
<td>-0.79%</td>
</tr>
<tr>
<td>ESV OOS Statistics:</td>
<td>0.38</td>
<td>0.38</td>
</tr>
<tr>
<td>Predictor: Div. Yield</td>
<td>-11.84%</td>
<td>0.37</td>
</tr>
<tr>
<td>Empirical OOS Statistics:</td>
<td>-5.27%</td>
<td>0.29</td>
</tr>
<tr>
<td>ESV OOS Statistics:</td>
<td>0.29</td>
<td>-6.57%</td>
</tr>
</tbody>
</table>
Table 4: Results for the Habit Formation Term Structure Model-Based Priors

Reported are the results for the model-based priors from the HFTS model. The weight of the prior, $\lambda$, is estimated as explained in Section 3.2. The $\Delta R^2_{OS}$ stands for the difference between the $R^2_{OS}$ of the predictor with and without model-based prior. A positive value indicates an improvement in the OOS forecasts through the prior. The data sample ends in December 2011.

<table>
<thead>
<tr>
<th>Panel A: PM Prior</th>
<th>Total Sample</th>
<th>Postwar Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quarterly Returns</td>
<td>Sample Begin: 1Q 1920</td>
<td>Sample Begin: 1Q 1947</td>
</tr>
<tr>
<td>Forecast Begin: 1Q 1940</td>
<td>Forecast Begin: 1Q 1967</td>
<td></td>
</tr>
<tr>
<td>$\lambda$</td>
<td>$R^2_{OS}$</td>
<td>$\Delta R^2_{OS}$</td>
</tr>
<tr>
<td>Predictor: Short Term Yield</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Empirical OOS Statistics:</td>
<td>-0.19%</td>
<td>-3.50%</td>
</tr>
<tr>
<td>PM OOS Statistics:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Predictor: Long Term Yield</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Empirical OOS Statistics:</td>
<td>-1.72%</td>
<td>-3.44%</td>
</tr>
<tr>
<td>PM OOS Statistics:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Predictor: Term Spread</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Empirical OOS Statistics:</td>
<td>0.03%</td>
<td>-3.61%</td>
</tr>
<tr>
<td>PM OOS Statistics:</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B: ESV Prior</th>
<th>Postwar Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quarterly Returns</td>
<td>Sample Begin: 3Q 1947</td>
</tr>
<tr>
<td>Forecast Begin: 3Q 1967</td>
<td></td>
</tr>
<tr>
<td>$\lambda$</td>
<td>$R^2_{OS}$</td>
</tr>
<tr>
<td>Predictor: Short Term Yield</td>
<td></td>
</tr>
<tr>
<td>Empirical OOS Statistics:</td>
<td>-3.08%</td>
</tr>
<tr>
<td>ESV OOS Statistics:</td>
<td></td>
</tr>
<tr>
<td>Predictor: Long Term Yield</td>
<td></td>
</tr>
<tr>
<td>Empirical OOS Statistics:</td>
<td>-3.34%</td>
</tr>
<tr>
<td>ESV OOS Statistics:</td>
<td></td>
</tr>
<tr>
<td>Predictor: Term Spread</td>
<td></td>
</tr>
<tr>
<td>Empirical OOS Statistics:</td>
<td>-3.03%</td>
</tr>
<tr>
<td>ESV OOS Statistics:</td>
<td></td>
</tr>
</tbody>
</table>
Table 3.

The PM prior leads to a substantial increase in OOS $R^2$ for the dividend-price ratio and the dividend yield. This performance over the total sample is not matched by the HF or the PT model. Further, the values for $\lambda$ are very high. For the total sample, $\lambda$ is around 5 for both predictors. For the postwar sample, the performance improvement is even more pronounced. The OOS $R^2$ is pushed from a negative value to over 7%. Hence, a real time investor could have profited significantly by forecasting the equity premium and imposing parameter restrictions derived from the LLR model through Bayesian priors. However, the performance of the ESV prior does not match the success of the PM prior. The ESV prior causes the OOS performance to deteriorate in all but two cases. A possible explanation is that when simulating data from the LRR for the ESV prior, four different time series are required as input variables: the two state variables, consumption growth, and dividend growth. This makes it more difficult for the model to generate data through which the ESV prior can improve the accuracy of the predictions.

5.4 Habit Formation Term Structure model-based priors

Model-based priors for the long term yield, the short term yield, and the term spread are derived from the HFTS model. Table 4 shows the results obtained from implementing these priors on predictive regressions with quarterly data.\(^{10}\) For the PM prior the data sample starts in the 1\(^{st}\) Quarter of 1920 and the OOS forecast begins in the 1\(^{st}\) Quarter of 1940. The postwar sample starts in the 1\(^{st}\) Quarter of 1947 for the PM prior. For the ESV prior, the data sample starts in the 3\(^{rd}\) Quarter of 1947, since quarterly consumption data from FRED, which is needed to compute the log consumption growth process and generate data for the ESV prior, is unavailable before this.

The PM prior is again successful: it improves the OOS $R^2$ for all three predictive

\(^{10}\)No overlapping returns are used as the HFTS model is simulated and matches empirical moments at a quarterly frequency.
variables and over both data samples. The $\lambda$ values are very small, which implies that the majority of the weight is placed on the empirical data when estimating the parameters. The ESV prior can improve the OOS $R^2$ for both the short term and the long term yields by several percentage points despite the small values of $\lambda$. However, the OOS $R^2$ remain negative, as the OOS $R^2$ without imposing a prior are very low.

5.5 Sensitivity of the model-based prior’s OOS performance to $\lambda$

The results above only consider a particular value of $\lambda$, namely the $\lambda$ which maximizes the data density over the initial estimation period. An interesting question is whether the results are sensitive with respect to $\lambda$. There exist three cases. First, imposing a model-based prior helps to raise the OOS $R^2$ for small values of $\lambda$, but the performance deteriorates and the OOS $R^2$ falls below the no-prior level for a large $\lambda$. Second, any value of $\lambda$ improves the forecasting performance. For this case, simply taking the theoretical parameter estimates implied by the asset pricing model will perform better than using empirical data for the estimation of the single variable predictive regression. This finding is similar in spirit to the result in Campbell and Thompson, which shows that using the theoretical parameter values implied by the steady state theory of Gordon (1962) and ignoring the empirical estimates of the parameters in the predictive regression can lead to more accurate forecasts. Third, the model-based prior fails to improve the OOS $R^2$ for any $\lambda$.

Figure 1 shows these three cases. For the PM prior from the HF model imposed on the dividend-price ratio over the postwar sample and forecasting annual returns, a $\lambda$ between 0 and 3 raises the OOS $R^2$. For a $\lambda$ greater than 3, the OOS performance deteriorates. When imposing the PM prior from the LRR model, the OOS $R^2$ converges to a level above 6%. Hence, the prior improves the predictive power for all values of $\lambda$. The opposite case is shown by the third graph, where the ESV prior
from the LLR model imposed on the dividend yield fails to improve the OOS $R^2$ for any value of $\lambda$.

Generally, the performance of the model-based priors is stable with respect to $\lambda$: a prior which improves the predictability of the equity premium does so for a range of $\lambda$ values.

5.6 Is the model-based prior truly OOS?

The results above show the forecast of the equity premium OOS. The equity premium is predicted recursively, and $\lambda$ is estimated over an initial estimation period with no look-ahead bias. However, the parameters of the asset pricing models have been calibrated by the authors based on a data sample ending in the 1990s or 2000s, respectively, depending on the model.

To assess whether the improvement in the predictive power holds up after the calibration data sample ended, I extend the initial estimation period up to the last month or quarter of the calibration data sample of a particular asset pricing model. For the model-based priors imposed at an annual frequency, overlapping annual returns are used to increase the number of OOS predictions.

The results are reported in Table 5. For the HF model, the improvement in OOS predictability holds up, with the exception of the PM prior imposed on the dividend yield. The PM prior from the PT model is successful for both predictor variables (as described above, the ESV prior of the PT model is not used for overlapping annual returns due to the authors’ model calibration). The priors from the LRR model are effective for the dividend yield but do not increase the OOS $R^2$ for the dividend-price ratio. The priors of the HFTS model are robust – the OOS $R^2$ is raised for all predictors and priors, except for the ESV prior used for the term spread. However, the ESV prior is also unsuccessful in improving the performance of the term spread in Table 4.

These results show that the model-based priors are successful when forecasting
Figure 1: Sensitivity of $R^2_{OS}$ to $\lambda$

HF PM Prior: Dividend–Price Ratio 1947 to 2011

LRR PM Prior: Price–Dividend Ratio 1947 to 2011
This figure depicts the sensitivity of $R^2_{OS}$ to $\lambda$ for three cases. The horizontal axis shows the value of $\lambda$ and, in parentheses, the proportion of the simulated data relative to the total data (empirical and simulated data combined) used to estimate the parameters of the single variable predictive regression.
the equity premium over a sample period which does not contain any data used to calibrate the asset pricing model. However, the caveat is that the OOS periods are relatively short and based on a small number of independent predictions.

5.7 The model-based prior versus the Minnesota prior

The model-based priors impose economic restrictions on the parameter estimates of the single variable predictive regressions. An alternative is the Minnesota prior from Doan et al. (1984), which is purely statistical and not based on the economic theory implied by an economic model. The general purpose of the Minnesota prior is to reduce overfitting of VARs used in macroeconomics by shrinking the parameter estimates towards a random walk specification. I adapt the Minnesota prior such that it shrinks the estimate of $\beta_0$ towards the historical average and the estimate of $\beta_1$ towards zero – my version of the Minnesota prior shrinks the parameters towards the historical average return model. Comparing the forecast performance of the model-based prior and the Minnesota prior, helps to answer the question whether the economic theory derived from the asset pricing models adds value when being incorporated in a prior: the model-based priors should outperform the Minnesota prior if this is true.

My implementation of the Minnesota prior assumes the following prior distribution:

$$\beta \sim N(\beta, H), \text{ where } H = \begin{bmatrix} \phi_0 s^2_y & 0 \\ 0 & \phi_1 \end{bmatrix} \text{ and } \beta = \begin{pmatrix} \mu_y \\ 0 \end{pmatrix}. \quad (10)$$

The parameter $\mu_y$ is the mean and $s^2_y$ the variance of the equity premium over the estimation period. The prior tightness parameters for $\beta_0$ and $\beta_1$ are $\phi_0$ and $\phi_1$, respectively. The posterior distribution is then:

$$\bar{\beta} = \overline{H}\beta + (\hat{\sigma}^{-2}X)\gamma, \text{ where } \overline{H} = [H^{-1} + (\hat{\sigma}^{-2}X'X)]^{-1}. \quad (11)$$
Table 5: Results for the Post Calibration Period

Reported are the results for the forecast begin set such that the OOS forecasting period only contains data not used by the authors to calibrate the asset pricing model. The weight of the prior, $\lambda$, is estimated as explained in Section 3.2. The $\Delta R^2_{OS}$ stands for the difference between the $R^2_{OS}$ of the predictor with and without model-based prior. A positive value indicates an improvement in the OOS forecasts through the prior. The postwar sample is used in all cases and the data end in December 2011. For the overlapping annual returns, the forecasting procedure is repeated in monthly steps, but the forecasts are at annual horizons.

Panel A: Habit Formation
Overlapping Annual Returns (Forecast Begin: December 1996)

<table>
<thead>
<tr>
<th>Predictor: Div.-Price Ratio</th>
<th>PM Prior</th>
<th>ESV Prior</th>
</tr>
</thead>
<tbody>
<tr>
<td>PM OOS Statistics:</td>
<td>$0.87$</td>
<td>$-5.58%$</td>
</tr>
</tbody>
</table>

Panel B: Prospect Theory
Overlapping Annual Returns (Forecast Begin: December 1996)

<table>
<thead>
<tr>
<th>Predictor: Div.-Price Ratio</th>
<th>PM Prior</th>
</tr>
</thead>
<tbody>
<tr>
<td>Empirical OOS Statistics:</td>
<td>$-8.62%$</td>
</tr>
<tr>
<td>PM OOS Statistics:</td>
<td>$0.11$</td>
</tr>
</tbody>
</table>

Panel C: Long Run Risk
Overlapping Annual Returns (Forecast Begin: December 1999)

<table>
<thead>
<tr>
<th>Predictor: Div.-Price Ratio</th>
<th>PM Prior</th>
<th>ESV Prior</th>
</tr>
</thead>
<tbody>
<tr>
<td>Empirical OOS Statistics:</td>
<td>$16.36%$</td>
<td>$16.36%$</td>
</tr>
<tr>
<td>PM OOS Statistics:</td>
<td>$5.28$</td>
<td>$10.10%$</td>
</tr>
</tbody>
</table>

Panel D: Habit Formation Term Structure
Quarterly Returns (Forecast Begin: 4Q 2004)

<table>
<thead>
<tr>
<th>Predictor: Short Term Yield</th>
<th>PM Prior</th>
<th>ESV Prior</th>
</tr>
</thead>
<tbody>
<tr>
<td>Empirical OOS Statistics:</td>
<td>$-3.98%$</td>
<td>$-4.21%$</td>
</tr>
<tr>
<td>PM OOS Statistics:</td>
<td>$0.28$</td>
<td>$-2.44%$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Predictor: Long Term Yield</th>
<th>PM Prior</th>
<th>ESV Prior</th>
</tr>
</thead>
<tbody>
<tr>
<td>Empirical OOS Statistics:</td>
<td>$-1.04%$</td>
<td>$-1.12%$</td>
</tr>
<tr>
<td>PM OOS Statistics:</td>
<td>$0.41$</td>
<td>$-0.28%$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Predictor: Term Spread</th>
<th>PM Prior</th>
<th>ESV Prior</th>
</tr>
</thead>
<tbody>
<tr>
<td>Empirical OOS Statistics:</td>
<td>$-3.93%$</td>
<td>$-3.94%$</td>
</tr>
<tr>
<td>PM OOS Statistics:</td>
<td>$0.50$</td>
<td>$-2.10%$</td>
</tr>
</tbody>
</table>
To estimate $\hat{\sigma}^{-2}$, the OLS residuals of the single variable predictive regression, updated for each forecast, are used.

I set $\phi_0 = 1000$ such that the prior on $\beta_0$ is very diffuse – a diffuse prior on the constant of the model is common in applications of the Minnesota prior. For $\phi_1$, I specify a Gamma distribution and average the posterior estimate $\bar{\beta}$ over the possible values of $\phi_1$, i.e. I specify a hierarchical prior. This reflects the uncertainty of a real time investor when choosing $\phi_1$. The Gamma distribution is used since the value of $\phi_1$ has to be positive.

The results of a comparison between the two priors are reported in Table 6. For the model-based priors, the better performing prior, PM or ESV, for the respective period (total or postwar sample) and predictor of each asset pricing model is selected. The three different versions of the Minnesota prior correspond to changes in the scale parameter of the Gamma distribution. Version 1 corresponds to a Gamma distribution for $\phi_1$ with a shape parameter of 2 and a scale parameter of 0.05. Versions 2 and 3 also use a shape parameter of 2, but the scale parameters are 0.01 and 0.001, respectively. These Gamma distributions lead to an average $\phi_1$ which is line with the tightness parameters used in Doan et al. (1984). Adjusting the magnitude of the shape and scale parameter downwards changes the OOS $R^2$ only marginally. When the shape and scale parameters are increased, the performance of the Minnesota prior deteriorates.

For the dividend-price ratio and the dividend yield, the model-based prior raises the OOS $R^2$ to a higher level than the Minnesota prior in the majority of cases. This implies that the economic theory of the asset pricing models is valuable for the purpose of forecasting the equity premium. A purely statistical prior like the Minnesota prior does not achieve the same performance. For the yield related variables, the Minnesota prior generally fares slightly better than the model-based priors from the HFTS model. Hence, the economic insights from the HFTS model do not appear to result in a prior which is more useful than a purely statistical method for addressing
the problem of overfitting.

6 How the Model-Based Prior Improves Predictive Accuracy

To gain further insights into the working of the model-based prior, I analyse the two ways in which the model-based prior improves the OOS forecast performance of the single variable predictive regression in more detail. As mentioned in Section 2, these two ways are: the reduction in the overfitting of the parameter estimates and bringing the estimates of $\beta_0$ and $\beta_1$ closer to their in-sample counterparts during periods when the fundamental relationship between the predictor and the equity premium, which holds over the total sample, fails to hold temporarily.

Figure 2 depicts an example of a prior improving the OOS forecasts of the equity premium by smoothing the estimates of $\beta$ and bringing them closer to their in-sample estimates. The figure shows the in-sample estimates over the total postwar sample, the empirical OOS estimates, and the OOS estimates when imposing the PM prior from the LRR model. The empirical estimates are more volatile since the overfitted parameter estimates pick up noise. Imposing the model-based prior helps to reduce the impact of noise on the parameter estimates: the posterior estimates are smoother when plotted over time. Further, the graph shows that the PM prior pushes the OOS estimates closer to the in-sample estimates for most of the OOS period. The empirical OOS estimates are far higher than their in-sample counterparts up to 1998, but the PM prior pushes the OOS estimates down. This also has the effect that for the end of the sample, from 1998 to 2011, the posterior estimates of the PM prior are lower than the in-sample estimates.

Table 7 reports the parameter estimates and the standard errors over the total sample. The second and third column in the table show the in-sample estimates of the empirical data, of the data simulated from the LRR model through the Monte
Table 6: Comparison of the Model-Based Prior and the Minnesota Prior

Reported is the increase or decrease in $R^2_{OS}$ when either imposing the model-based prior or the Minnesota prior. The baseline is the $R^2_{OS}$ when no prior is imposed. For the model-based priors, the better performing prior, PM or ESV, for the respective period (total or postwar sample) and predictor of each model is selected. The initial estimation period is at least 20 years, but the OOS forecast does not start before 1926 for annual returns or December 1926 for overlapping annual returns. Version 1 of the Minnesota prior corresponds to the implementation of the prior with a Gamma distribution for $\phi_1$ with a shape parameter of 2 and a scale parameter of 0.05. Versions 2 and 3 also use a shape parameter of 2, but the scale parameters are 0.01 and 0.001, respectively.

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Model-Based Prior $\Delta R^2_{OS}$</th>
<th>Minnesota Prior $\Delta R^2_{OS}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HF</td>
<td>PT</td>
</tr>
<tr>
<td><strong>Predictor: Div.-Price Ratio</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Annual Returns 1872-2011</td>
<td>0.01%</td>
<td>0.32%</td>
</tr>
<tr>
<td>Overlapping Annual Returns 12M 1872-12M 2011</td>
<td>1.16%</td>
<td>1.26%</td>
</tr>
<tr>
<td>Annual Returns 1947-2011</td>
<td>7.35%</td>
<td>6.52%</td>
</tr>
<tr>
<td>Overlapping Annual Returns 12M 1947-12M 2011</td>
<td>7.11%</td>
<td>5.94%</td>
</tr>
<tr>
<td><strong>Predictor: Dividend Yield</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Annual Returns 1872-2011</td>
<td>1.02%</td>
<td>0.98%</td>
</tr>
<tr>
<td>Overlapping Annual Returns 12M 1872-12M 2011</td>
<td>1.59%</td>
<td>1.71%</td>
</tr>
<tr>
<td>Annual Returns 1947-2011</td>
<td>2.60%</td>
<td>5.06%</td>
</tr>
<tr>
<td>Overlapping Annual Returns 12M 1947-12M 2011</td>
<td>2.92%</td>
<td>5.82%</td>
</tr>
<tr>
<td><strong>Predictor: Short Term Yield</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quarterly Returns 1Q 1920-4Q 2011</td>
<td>0.31%</td>
<td></td>
</tr>
<tr>
<td>Quarterly Returns 1Q 1947-4Q 2011</td>
<td>2.26%</td>
<td></td>
</tr>
<tr>
<td><strong>Predictor: Long Term Yield</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quarterly Returns 1Q 1920-4Q 2011</td>
<td>1.38%</td>
<td></td>
</tr>
<tr>
<td>Quarterly Returns 1Q 1947-4Q 2011</td>
<td>3.33%</td>
<td></td>
</tr>
<tr>
<td><strong>Predictor: Term Spread</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quarterly Returns 1Q 1920-4Q 2011</td>
<td>0.16%</td>
<td></td>
</tr>
<tr>
<td>Quarterly Returns 1Q 1947-4Q 2011</td>
<td>2.52%</td>
<td></td>
</tr>
</tbody>
</table>
Figure 2: Parameter Estimates over the Postwar Sample

This figure shows the parameter estimates over the postwar sample with the OOS period starting in 1967 for the dividend-price ratio and annual returns. The PM prior is from the LRR model and imposed with $\lambda = 1.9$. 

32
Table 7: Parameter Estimates for the Postwar Sample

Reported are the in-sample parameter estimates of $\beta_0$ and $\beta_1$ using the postwar sample with and without model-based prior for the dividend-price ratio and annual returns. Further, the standard deviation of the parameter estimates over the OOS period with forecast begin in 1967. The estimates of the simulated data refer to the data generated by feeding empirical consumption growth into the HF model. The Monte Carlo simulation for the LLR model is described in Section 4.2.1.

<table>
<thead>
<tr>
<th>Habit Formation Model</th>
<th>In-Sample Estimates</th>
<th>OOS Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\beta_0$</td>
<td>$\beta_1$</td>
</tr>
<tr>
<td>Empirical Data 1947-2011</td>
<td>0.467</td>
<td>0.118</td>
</tr>
<tr>
<td>Simulated Data 1947-2011</td>
<td>0.378</td>
<td>0.116</td>
</tr>
<tr>
<td>ESV Prior 1947-2011 ($\lambda = 0.73$)</td>
<td>0.269</td>
<td>0.066</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Long Run Risk Model</th>
<th>In-Sample Estimates</th>
<th>OOS Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\beta_0$</td>
<td>$\beta_1$</td>
</tr>
<tr>
<td>Empirical Data 1947-2011</td>
<td>0.467</td>
<td>0.118</td>
</tr>
<tr>
<td>Monte Carlo Simulation</td>
<td>0.089</td>
<td>0.016</td>
</tr>
<tr>
<td>PM Prior 1947-2011 ($\lambda = 1.90$)</td>
<td>0.197</td>
<td>0.048</td>
</tr>
</tbody>
</table>

We see that the theoretical estimates implied by the LRR model, $\hat{\beta}_0 = 0.089$ and $\hat{\beta}_1 = 0.016$, are substantially lower than the empirical estimates, $\hat{\beta}_0 = 0.467$ and $\hat{\beta}_1 = 0.118$. Thus, when imposing the PM prior, the resulting posterior estimates, $\hat{\beta}_0 = 0.197$ and $\hat{\beta}_1 = 0.048$, are lower than the empirical estimates.

The fourth and fifth column report the standard deviations of the estimates over the OOS periods. The standard deviations of the empirical estimates are higher than those of the posterior estimates. This is a result of the smoothing effect of the model-based prior as shown in Figure 2.

Figure 3 shows why pushing the OOS estimates closer to the in-sample estimates for most of the sample period helps to increase the OOS $R^2$. The graph presents the difference between the cumulative sum of squared errors (SSE). I take the cumulative SSE of the historical average model and subtract the cumulative in-sample regression residuals, the cumulative SSE of the OOS predictive regression estimated only with empirical data, and the cumulative SSE of the OOS predictive regression with the

Carlo procedure described in Section 4.2.1, and of the combined data, i.e. when imposing the PM prior, with the weight on the simulated data being $\lambda = 1.9$. We see that the theoretical estimates implied by the LRR model, $\hat{\beta}_0 = 0.089$ and $\hat{\beta}_1 = 0.016$, are substantially lower than the empirical estimates, $\hat{\beta}_0 = 0.467$ and $\hat{\beta}_1 = 0.118$. Thus, when imposing the PM prior, the resulting posterior estimates, $\hat{\beta}_0 = 0.197$ and $\hat{\beta}_1 = 0.048$, are lower than the empirical estimates.

The fourth and fifth column report the standard deviations of the estimates over the OOS periods. The standard deviations of the empirical estimates are higher than those of the posterior estimates. This is a result of the smoothing effect of the model-based prior as shown in Figure 2.

Figure 3 shows why pushing the OOS estimates closer to the in-sample estimates for most of the sample period helps to increase the OOS $R^2$. The graph presents the difference between the cumulative sum of squared errors (SSE). I take the cumulative SSE of the historical average model and subtract the cumulative in-sample regression residuals, the cumulative SSE of the OOS predictive regression estimated only with empirical data, and the cumulative SSE of the OOS predictive regression with the
LRR PM prior imposed with \( \lambda = 1.9 \), respectively. A model which outperforms the historical average model, has a positive difference in cumulative SSE. We can see that the in-sample regression underperforms the OOS regression up to around 1997, but performs better at the end of our sample period, because the empirical model loses its previous gains over a window of four years. This is because during the Dot-com bubble the dividend-price ratios were very low and the empirical data up to this point implied a strong relationship between a low dividend-price ratio and a subsequent low equity premium. Hence, during the years leading up to the bursting of the Dot-com bubble, the empirical model predicts a low positive or negative equity premium. However, falling stock market prices did not materialise until the year 2000, which leads to the drop in the difference between the cumulative SSE for the predictive regression estimated only with empirical data. Since the in-sample coefficients are estimated over the total sample, the estimator can trade-off current losses and future gains. The in-sample estimates tolerate larger squared errors around 1975 in order to limit the squared errors in the late 1990s. Since the in-sample estimates include this period of low predictability in the late 1990s, they are smaller than the OOS estimates up to 1998. They have a forward-looking bias. The PM prior helps to avoid the pitfall of the Dot-com bubble and brings the difference in cumulative SSE curve closer to the in-sample curve. The PM prior achieves this as the LRR model implies that the dividend-price ratio has weak predictive power. Hence, the PM prior helps to anticipate periods of weak predictability, which leads to a reduction in forecasting errors.

The ESV prior of the HF model implies lower predictability and has the same effect as the PM prior from the LRR model. Table 7 reports the estimates of \( \beta \) with and without the ESV prior from the HF model. Further, the estimates of the simulated data generated by feeding empirical consumption growth to the HF model are also provided. The ESV prior of the HF model pushes the parameter estimates down, since the model implies less predictive power of the dividend-price ratio than
This figure shows the difference between the cumulative sum of squared errors (SSE) of the historical average model and either the cumulative in-sample regression residuals, the cumulative SSE of the OOS predictive regression estimated only with empirical data, or the cumulative SSE of the OOS predictive regression with the LRR PM prior imposed with $\lambda = 1.9$. A model which outperforms the historical average model has a positive difference in cumulative SSE.
found in the empirical data. Hence, the ESV prior of the HF model also reduces the forecasting errors during periods of low predictability. The posterior estimates are even lower than the theoretical estimates implied by the model. For small values of $\lambda$, the estimates can be pushed down below or pushed up above the model implied estimates. However, as $\lambda \to \infty$, the posterior estimates converge to the estimates implied by the asset pricing model.

7 Conclusion

In this paper, I impose restrictions on the parameter estimates of typical single variable predictive regressions used in the equity premium prediction literature. My parameter restrictions are based on a well specified Bayesian framework which has been inspired by the work of Del Negro and Schorfheide (2004). I adapt their framework to derive priors from consumption-based asset pricing models. These priors are termed model-based priors and are derived from from four asset pricing models: Habit Formation (see Campbell and Cochrane (1999)), Prospect Theory (see Barberis et al. (2001)), Long Run Risk (see Bansal and Yaron (2004)), and Habit Formation Term Structure (see Wachter (2006)).

I find that implementing the priors on the single variable predictive regressions can lead to increases in the OOS $R^2$ of up to several percentage points. The model-based priors perform particularly well for the postwar period when using the dividend-price ratio or the dividend yield as predictors. The prior tightness is determined by the hyperparameter $\lambda$, which is estimated from the data. The estimation of $\lambda$ is only based on data which are available before the beginning of the OOS forecast. Hence, a real time investor could implement the estimation procedure described in this paper and achieve substantially higher expected returns from the higher OOS $R^2$.

I also compare the performance of the model-based priors with a purely statistical
one, namely the Minnesota prior, which does not incorporate any economic theory derived from an asset pricing model. In most cases, the model-based priors raise the OOS $R^2$ by more than the Minnesota prior: this implies that the economic theory of the asset pricing models is valuable for the purpose of forecasting the equity premium. Further, the results are also robust when restricting the OOS period to a subsample which does not include any data used for the calibration of the asset pricing model by the respective authors.
A Model-Based Priors

A.1 The prior and posterior distributions

The model-based prior is based on the DSGE-VAR approach of Del Negro and Schorfheide (2004). The likelihood function for the predictive regression based on the total empirical data sample of size \( T \), denoted by \((Y, X)\), is

\[
p(Y, X | \beta, \sigma^2) \propto (\sigma^2)^{-\frac{T}{2}} \exp \left\{ -\frac{1}{2} \sigma^{-2} (Y'Y - \beta'X'Y - Y'X\beta + \beta'X'X\beta) \right\}.
\]

(12)

The likelihood function for the predictive regression based on data simulated from the asset pricing model \((Y^*, X^*)\) is accordingly

\[
p(Y^*(\theta), X^*(\theta) | \beta, \sigma^2) \propto (\sigma^2)^{-\frac{\lambda T}{2}} \times \exp \left\{ -\frac{1}{2} \sigma^{-2} (Y'^*Y'^* - \beta'X'^*Y'^* - Y'^*X^*\beta + \beta'X'^*X^*\beta) \right\}.
\]

(13)

The asset pricing model parameters are denoted by \(\theta\), and \(\lambda\) is the hyperparameter which determines the weight of the simulated data in the posterior. These individual likelihoods are combined to form the joint likelihood:

\[
p(Y^*(\theta), Y, X^*(\theta), X | \beta, \sigma^2) = p(Y^*(\theta), X^*(\theta) | \beta, \sigma^2)p(Y, X | \beta, \sigma^2),
\]

(14)

where \(p(Y^*(\theta), X^*(\theta) | \beta, \sigma^2)\) can be interpreted as a prior for the parameters of the predictive regression. I specify this prior in two different ways: the PM prior and the ESV prior.

Through the use of scaled population moments, we can rewrite (13) as

\[
p(\beta, \sigma^2 | \theta) = c_{PM}(\theta)^{-1}(\sigma^2)^{-\frac{\lambda T + 2}{2}} \times \exp \left\{ -\frac{1}{2} \lambda T \sigma^{-2} (\Gamma_{yy}^*(\theta) - \beta'\Gamma_{xy}^*(\theta) - \Gamma_{yx}^*(\theta)\beta + \beta'\Gamma_{xx}^*(\theta)\beta) \right\}.
\]

(15)
We incorporate an improper prior \( p(\beta, \sigma^2) \propto \sigma^{-2} \) together with \( c_{PM}(\theta) \), a normalisation factor which ensures that the density integrates to one. The form of \( c_{PM}(\theta) \) is given in Appendix A.2. The PM prior distribution of the predictive regression parameters is of the Inverted-Wishart Normal form:

\[
\sigma_{PM}^2 \mid \theta \sim IW(\lambda T \sigma_{PM}^2, \lambda T - 2, 1) \text{ and }
\]

\[
\beta_{PM} \mid \sigma_{PM}^2, \theta \sim N(\beta_{PM}^*(\theta), \sigma_{PM}^2 \otimes (\lambda T \Gamma_{xx}(\theta))^{-1}),
\]

where

\[
\beta_{PM}^*(\theta) = \Gamma_{xx}^*(\theta)^{-1} \Gamma_{xy}^*(\theta) \quad \text{and} \quad \sigma_{PM}^2(\theta) = \Gamma_{yy}(\theta) - \Gamma_{yx}(\theta) \Gamma_{xx}^*(\theta)^{-1} \Gamma_{xy}^*(\theta).
\]

The ESV prior takes the form

\[
p(\beta, \sigma^2 \mid \theta) = c_{ESV}(\theta)^{-1}(\sigma^2)^{-\frac{\lambda T + 2}{2}} \times \exp\left\{-\frac{1}{2} \lambda \sigma^{-2}(\bar{Y}'\bar{Y} - \beta'\bar{X}'\bar{Y} - \bar{Y}'\bar{X}\beta + \beta'\bar{X}'\bar{X}\beta)\right\}.
\]

Again, we add an improper prior \( p(\beta, \sigma^2) \propto \sigma^{-2} \) together with the normalisation factor \( c_{ESV}(\theta) \). The form of \( c_{ESV}(\theta) \) is given in Appendix A.2. For notational simplicity, the dependency of \( \bar{X} \) and \( \bar{Y} \) on \( \theta \) is not explicitly stated. The ESV prior distribution of the predictive regression parameters is

\[
\sigma_{ESV}^2 \mid \theta \sim IW(\lambda T \sigma_{ESV}^2, \lambda T - 2, 1) \text{ and }
\]

\[
\beta_{ESV} \mid \sigma_{ESV}^2, \theta \sim N(\beta_{ESV}^*(\theta), \sigma_{ESV}^2 \otimes (\lambda \bar{X}'\bar{X})^{-1}),
\]

where

\[
\beta_{ESV}^*(\theta) = (\bar{X}'\bar{X})^{-1}\bar{X}'\bar{Y} \quad \text{and} \quad \sigma_{ESV}^2(\theta) = \frac{1}{T}(\bar{Y}'\bar{Y} - \bar{Y}'\bar{X}(\bar{X}'\bar{X})^{-1}\bar{X}'\bar{Y}).
\]
For the PM prior, the maximum-likelihood estimates of $\beta$ and $\sigma^2$ based on simulated and empirical data take the following form

$$
\tilde{\beta}_{PM}(\theta) = (\lambda T\Gamma_{xx}^*(\theta) + X'X)^{-1}(\lambda T\Gamma_{xy}^*(\theta) + X'Y)
$$

and

$$
\tilde{\sigma}_{PM}^2(\theta) = \frac{1}{(\lambda + 1)T}[(\lambda T\Gamma_{yy}^*(\theta) + Y'Y)

- (\lambda T\Gamma_{yx}^*(\theta) + Y'X)(\lambda T\Gamma_{xx}^*(\theta) + X'X)^{-1}(\lambda T\Gamma_{xy}^*(\theta) + X'Y)].
$$

The posterior distribution of the predictive regression parameters when implementing the PM prior is given by

$$
\sigma_{PM}^2 | Y, \theta \sim IW((\lambda + 1)T\tilde{\sigma}_{PM}^2(\theta), (\lambda + 1)T - 2, 1) \quad \text{and} 
$$

$$
\beta_{PM} | Y, \sigma_{PM}^2, \theta \sim N(\tilde{\beta}_{PM}(\theta), \sigma_{PM}^2 \otimes (\lambda T\Gamma_{xx}^*(\theta) + X'X)^{-1}).
$$

For the ESV prior, the maximum-likelihood estimates of $\beta$ and $\sigma^2$ are

$$
\tilde{\beta}_{ESV}(\theta) = (\lambda \bar{X}'\bar{X} + X'X)^{-1}(\lambda \bar{X}'\bar{Y} + X'Y)
$$

and

$$
\tilde{\sigma}_{ESV}^2(\theta) = \frac{1}{(\lambda + 1)T}[(\lambda \bar{Y}'\bar{Y} + Y'Y)

- (\lambda \bar{Y}'\bar{X} + Y'X)(\lambda \bar{X}'\bar{X} + X'X)^{-1}(\lambda \bar{X}'\bar{Y} + X'Y)].
$$

Hence, the posterior distribution takes the form

$$
\sigma_{ESV}^2 | Y, \theta \sim IW((\lambda + 1)T\tilde{\sigma}_{ESV}^2(\theta), (\lambda + 1)T - 2, 1) \quad \text{and} 
$$

$$
\beta_{ESV} | Y, \sigma_{ESV}^2, \theta \sim N(\tilde{\beta}_{ESV}(\theta), \sigma_{ESV}^2 \otimes (\lambda \bar{X}'\bar{X} + X'X)^{-1}).
$$

Since I take $\theta$ as given by the authors of the asset pricing models – $p(\theta)$ in Del Negro and Schorfheide (2004) is set equal to one – drawing from the posterior distribution is trivial. For $\beta$, the modes are $\tilde{\beta}_{PM}(\theta)$ and $\tilde{\beta}_{ESV}(\theta)$, depending on the prior. For $\sigma^2$, the modes are $\tilde{\sigma}_{PM}^2$ and $\tilde{\sigma}_{ESV}^2$, respectively.
A.2 Normalisation factors

For the PM prior, the normalisation factor \( c_{PM}(\theta) \) in (15) is

\[
c_{PM}(\theta) = 2\pi | \lambda T \Sigma_{xx}(\theta) |^{-\frac{1}{2}} | \lambda T \sigma_{PM}^2(\theta) |^{-\frac{\lambda T - 2}{2}} 2^{\frac{\lambda T - 2}{2}} \Gamma[(\lambda T - 2)/2], \tag{33}
\]

and for the ESV prior, the normalisation factor \( c_{ESV}(\theta) \) in (20) takes the form

\[
c_{ESV}(\theta) = 2\pi | \lambda \bar{X}' \bar{X} |^{-\frac{1}{2}} | \lambda T \sigma_{ESV}^2(\theta) |^{-\frac{\lambda T - 2}{2}} 2^{\frac{\lambda T - 2}{2}} \Gamma[(\lambda T - 2)/2], \tag{34}
\]

where \( \Gamma(\cdot) \) is the gamma function.

A.3 Data densities

For the PM prior, the data density \( p_{PM}^\lambda(Y, X | \theta) \) is

\[
p_{PM}^\lambda(Y, X | \theta) = \frac{| \lambda T \Sigma_{xx}^*(\theta) + X'X |^{-\frac{1}{2}} | (\lambda + 1)T \sigma_{PM}^2(\theta) |^{-\frac{(\lambda+1)T-2}{2}}}{| \lambda T \Sigma_{xx}^*(\theta) |^{-\frac{1}{2}} | \lambda T \sigma_{PM}^2(\theta) |^{-\frac{\lambda T - 2}{2}}} \times \left( \frac{(2\pi)^{-\frac{T}{2}} 2^{\frac{((\lambda + 1)T-2)}{2}} \Gamma[((\lambda + 1)T - 2)/2]}{2^{\frac{\lambda T - 2}{2}} \Gamma[(\lambda T - 2)/2]} \right), \tag{35}
\]

and for the ESV prior, the data density \( p_{ESV}^\lambda(Y, X | \theta) \) takes the form

\[
p_{ESV}^\lambda(Y, X | \theta) = \frac{| \lambda \bar{X}' \bar{X} + X'X |^{-\frac{1}{2}} | (\lambda + 1)T \sigma_{ESV}^2(\theta) |^{-\frac{(\lambda+1)T-2}{2}}}{| \lambda \bar{X}' \bar{X} |^{-\frac{1}{2}} | \lambda T \sigma_{ESV}^2(\theta) |^{-\frac{\lambda T - 2}{2}}} \times \left( \frac{(2\pi)^{-\frac{T}{2}} 2^{\frac{((\lambda + 1)T-2)}{2}} \Gamma[((\lambda + 1)T - 2)/2]}{2^{\frac{\lambda T - 2}{2}} \Gamma[(\lambda T - 2)/2]} \right). \tag{36}
\]

These forms are derived by using the normalization constants for the Inverted-Wishart Normal distribution as shown in Zellner (1971).
B Asset Pricing Models

B.1 By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior

Campbell and Cochrane (1999) use a standard representative-agent consumption-based asset pricing model but add a slow-moving habit to the basic power utility function. This slow-moving habit leads to a slowly time-varying risk aversion and an equity risk premium which is higher at business cycle troughs than at peaks.

The agents are identical and maximise their utility given by

$$E \left[ \sum_{t=0}^{\infty} \delta^t \frac{(C_t - X_t)^{(1-\gamma)} - 1}{1 - \gamma} \right],$$

(37)

where $X_t$ is the level of habit, and $\delta$ is the time discount factor. A surplus consumption ratio $S_t \equiv (C_t - X_t)/C_t$ is defined – a small value of $S_t$ indicates that the economy is in a bad state.

A process is specified for $s_t = \ln(S_t)$ which ensures that $C_t$ is always above $X_t$. This process is given by

$$s_{t+1} = (1 - \phi)\bar{s} + \phi s_t + \Psi(s_t)(c_{t+1} - c_t - g),$$

(38)

with $\phi$ reflecting habit persistence, i.e. how quickly $s_{t+1}$ returns to the steady state value $\bar{s}$. The function $\Psi(s_t)$ is specified as

$$\Psi(s_t) = \begin{cases} \frac{1}{2} \sqrt{1 - 2(s_t - \bar{s})} - 1, & s_t \leq s_{max} \\ 0, & s_t > s_{max} \end{cases},$$

(39)

with the parameter $s_{max}$ set equal to $\bar{s} + \frac{1}{2}(1 - \bar{S}^2)$. The steady state value of $s_t$, i.e. $\bar{s}$, is given by $\ln(\sigma \sqrt{\gamma/(1 - \phi)})$. The evolution of $s_{t+1}$ is based on consumption.
growth being an i.i.d. lognormal process:

$$\Delta c_{t+1} = g + v_{t+1}, \text{ where } v_{t+1} \sim N(0, \sigma_v^2).$$ (40)

Stocks represent a claim to the consumption stream. The price-consumption ratio for a consumption claim satisfies

$$\frac{P_t(s_t)}{C_t} = E_t \left[ M_{t+1} \frac{C_{t+1}}{C_t} \left[ 1 + \frac{P_{t+1}}{C_{t+1}}(s_{t+1}) \right] \right].$$ (41)

The underlying assumption is that dividend growth is perfectly correlated with consumption growth in (40).\(^{11}\)

The price-consumption ratio is correlated with the business cycles as it depends on \(s_t\). The ratio is high at business cycle peaks and low at troughs. Due to the slowly time-varying risk aversion, the equity premium is also correlated with the business cycle, but this correlation is negative. Hence, the model generates an equity risk premium which is predictable by the price-consumption ratio. A high price-consumption ratio predicts a low equity premium.

I apply the Fixed-Point Method to solve for the price-consumption and the price-dividend ratio (Wachter (2005)).

\subsection*{B.2 A Consumption-Based Model of the Term Structure of Interest Rates}

Wachter’s (2006) model is based on the HF model but differs in two ways. First, Wachter departs from the constant risk-free rate assumption by using a different specification for \(\bar{S}\). Second, she adds an inflation process, and thus, a nominal term structure can be computed.

\(^{11}\)Results for a model specification which assumes imperfectly correlated consumption and dividend processes are given in the paper. However, in this chapter, I focus only on the baseline specification.
Wachter specifies $\Psi(s_t)$ in the same way as Campbell and Cochrane (1999). However, $\bar{s}$ which is set equal to $\ln(\sigma \sqrt{\gamma/(1 - \phi - b/\gamma))}$ differs due to the parameter $b$. Hence, the value $s_{max}$, defined as $\bar{s} + \frac{1}{2}(1 - \bar{S}^2)$, is also different. In the HF model, $b$ is set equal to zero, which implies a constant risk-free rate. In the HFTS model, the log risk-free rate is given by

$$r_{f,t+1} = -\ln(\delta) + \gamma g - \frac{\gamma(1 - \phi) - b}{2} + b(\bar{s} - s_t).$$

(42)

The parameter $b$ is chosen to be positive. Thus, an increase in $s_t$ leads to a lower risk-free rate.

Wachter denotes the exogenous price level $\Pi_t$, which implies that inflation is given by $ln(\Pi_{t+1}/\Pi_t) = \Delta \pi_{t+1}$. Thus, the nominal bond pricing equation is

$$P_{n,t}^\delta = E_t \left[ M_{t+1} \frac{\Pi_t}{\Pi_{t+1}} P_{n-1,t+1}^\delta \right] = E_t \left[ \delta \left( \frac{S_{t+1} C_{t+1}}{S_t C_t} \right)^{-\gamma} \frac{\Pi_t}{\Pi_{t+1}} P_{n-1,t+1}^\delta \right],$$

(43)

with $P_{0,t}^\delta = 1$. $M_{t+1}$ is the pricing kernel, and $n$ denotes the time to maturity.

While consumption growth is specified as an i.i.d. lognormal process of the form

$$\Delta c_{t+1} = g + \sigma_c \epsilon_{c,t+1}, \text{ where } \epsilon_{c,t+1} \sim N(0,1),$$

(44)

the inflation process is given by

$$\Delta \pi_{t+1} = \eta_0 + \eta Z_t + \sigma_\pi \epsilon_{\pi,t+1},$$

(45)

where the expected inflation $Z_t$ takes the form

$$Z_t = (\Delta \pi_{t+1} - \eta_0 - \sigma_\pi \epsilon_{\pi,t+1}).$$

(46)
$Z_t$ is univariate and follows an AR(1) process:

$$Z_t = \mu + \Phi Z_{t-1} + \Sigma \epsilon_{\pi,t}. \quad (47)$$

The shocks to consumption growth and to inflation are imperfectly correlated. The correlation is denoted by $\rho$.

Due to the inflation process, the nominal bond price depends on two state variables, namely $Z_t$ and $s_t$. This complicates solving (43) as there are two random shocks, $\epsilon_{c,t+1}$ and $\epsilon_{\pi,t+1}$. However, through conditioning on $\sigma_c \epsilon_{c,t+1}$ and the Law of Iterated Expectations, the nominal bond price equals

$$P_{n,t}^s = F_n^s(s_t) \exp\{A_n + B_n Z_t\}, \quad (48)$$

where

$$F_n^s(s_t) = E_t [M_{t+1} \exp\{\xi_n \sigma_c \epsilon_{c,t+1} \} F_{n-1}^s(s_{t+1})]$$

$$= E_t [\exp\{\ln(\delta) - \gamma g - \gamma (1 - \phi)(\bar{s} - s_t)$$

$$+ [\xi_n - \gamma (\Psi(s_t) + 1)] \sigma_c \epsilon_{c,t+1} \} F_{n-1}^s(s_{t+1})], \quad (49)$$

and does not depend on $\epsilon_{\pi,t+1}$. Furthermore, $A_n$ and $B_n$ are scalars and defined recursively. The exact forms of $A_n$ and $B_n$ as well as the derivation of (48) can be found in Wachter (2006).

The yields are correlated with the business cycles as they depend on $s_t$. The correlation is negative due to the positive parameter $b$ in equation (42). As the equity premium is also negatively correlated with the business cycles, low yields predict a low equity premium in the HFTS model.

I solve the model by using standard numerical integration methods.
B.3 Prospect Theory and Asset Prices

In the model of Barberis et al. (2001), the agent does not only derive utility from consumption but also from fluctuations of her or his financial wealth. There are two important aspects in the way financial wealth fluctuations affect the utility of an economic agent. First, the agent is loss averse. Second, the degree of loss aversion depends on prior investment outcomes. Prior gains lead to less loss aversion, and prior losses lead to more loss aversion. Hence, the risk aversion of the agent varies over time, as in the HF model.

Aggregate consumption growth and dividend growth follow the i.i.d. lognormal processes given by

\[ \Delta c_{t+1} = g_c + \sigma_c \epsilon_{c,t+1}, \quad \text{where } \epsilon_{c,t+1} \sim \text{i.i.d. } N(0,1) \] (50)

and

\[ \Delta d_{t+1} = g_d + \sigma_d \epsilon_{d,t+1}, \quad \text{where } \epsilon_{d,t+1} \sim \text{i.i.d. } N(0,1), \] (51)

with \( \text{corr}(\epsilon_{c,t+1}, \epsilon_{d,t+1}) = \omega. \)\(^{12}\)

The agent’s maximization problem is set up as

\[
E \left[ \sum_{t=0}^{\infty} \left( \delta^t \frac{C_t^{1-\gamma}}{1-\gamma} + b_0 \bar{C}_t^{-\gamma} \delta^{t+1} v(X_{t+1}, S_t, z_t) \right) \right]. \tag{52}
\]

The second term captures that the agent cares about fluctuations in financial wealth. The variable \( X_{t+1} \) denotes the change of the financial wealth between time \( t \) and \( t + 1 \) and is defined as

\[ X_{t+1} \equiv S_t R_{t+1} - S_t R_{f,t}. \] (53)

The variable \( S_t \) measures the value of the agent’s risky assets at time \( t \). The variable

\(^{12}\)Barberis et al. (2001) consider two different specifications: Economy I, in which dividends equal consumption, and Economy II, in which consumption and dividends follow separate processes. The simulated moments of Economy II are much more successful in matching the empirical moments, and hence, I do not consider Economy I.
$z_t$ accounts for prior gains and losses up to time $t$ and is defined as $Z_t/S_t$, where $Z_t$ is a historical benchmark level for the value of the risky asset. If $z_t$ is smaller than one, the agent has prior gains, if $z_t$ is greater than one, the agent faces prior losses. The discount factor is $\delta$, and $b_0\tilde{C}_t^{-\gamma}$ is a scaling term. The form of the utility function $v(.)$ is different conditional on prior gains or prior losses. More details can be found in the appendix.

The dynamics of $z_t$ are given by the process

$$z_{t+1} = \eta \left( z_t \frac{R}{R_{t+1}} \right) + (1 - \eta).$$ (54)

The benchmark level $Z_t$ reacts sluggishly to changes in the stock price. When $S_t$ increases, $Z_t$ should increase by less in order to allow for prior gains. The sluggishness is determined by the parameter $\eta \in [0, 1]$. The closer $\eta$ is to one, the more sluggish the benchmark level becomes. The parameter $\bar{R}$ is chosen such that the median value of $z_t$ is around one.

The price-dividend ratio is assumed to be a function of the state variable $z_t$:

$$f_t \equiv P_t/D_t = f(z_t).$$ (55)

The real stock returns are thus given as

$$R_{t+1} = \frac{1 + f(z_{t+1})}{f(z_t)} e^{\theta_d^t + \sigma_d^e e_{d, t+1}}. \quad (56)$$

Barberis et al. (2001) show that the equilibrium is characterized by a constant real risk-free rate,

$$R_f = \delta^{-1} e^{\gamma \sigma_c - \gamma^2 \sigma_d^2 / 2},$$ (57)
and a price-dividend ratio given by

\[ 1 = \delta e^{g_d - \gamma g_c + \gamma^2 \sigma_d^2 (1 - \omega^2)/2} E_t \left[ 1 + f(z_{t+1}) e^{(\sigma_d - \gamma \omega \sigma_c) c_{d,t+1}} \right] \]
\[ + b_0 \delta E_t \left[ \hat{v} \left( 1 + f(z_{t+1}) e^{g_d + \sigma_d c_{d,t+1} + \sigma d_{t+1}} z_t \right) \right], \tag{58} \]

where the utility function \( \hat{v}(R_{t+1}, z_t) \) is equal to \( v(R_{t+1}, S_t, z_t)/S_t \).

As the HF model, the PT model is able to generate predictability in returns. A decrease in \( z_t \) leads to both, a higher price-dividend ratio and a less risk averse investor. When \( z_t \) falls, subsequent returns will be lower as less compensation for risk is required. Hence, a high price-dividend ratio predicts a low equity risk premium.

I solve the model by following the process laid out by Barberis et al. (2001).

### B.4 Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles

Bansal and Yaron (2004) propose a solution to the equity premium puzzle through a consumption-based asset pricing model with Epstein and Zin (1989) preferences. Their model differs from other consumption-based asset pricing models in two ways. First, they include a small persistent expected growth rate component in the consumption and dividend growth rate processes. This causes consumption and the return on the market portfolio to covary positively, and hence, the economic agents require a higher risk premium. Second, they allow for time-varying volatility, which accounts for fluctuating economic uncertainty, in both processes: this additional source of systematic risk increases the risk premium further.

The asset pricing restriction for the real return on the market portfolio \( R_{m,t+1} \), according to the Epstein and Zin (1989) preferences, is

\[ E_t \left[ \delta^\theta c_{c,t+1}^{\theta - \beta} c_{t+1}^{-\beta} R_{m,t+1} \right] = 1, \tag{59} \]
where \( G_{c,t+1} \) is the aggregate gross growth rate of consumption, \( R_{c,t+1} \) denotes the real return on an asset which pays aggregate consumption as dividends, and \( \delta \) is the time discount factor. The parameter \( \theta \) is defined as \( (1-\gamma)/(1-\frac{1}{\psi}) \), where \( \gamma \) is the risk aversion parameter, and \( \psi \) accounts for the Intertemporal Elasticity of Substitution. To derive the real returns, the authors use the standard approximation of Campbell and Shiller (1988). The real log return for the claim to aggregate consumption is

\[
r_{c,t+1} = \kappa_0 + \kappa_1 z_{t+1} - z_t + g_{c,t+1},
\]

where \( g_{c,t+1} \) is the log consumption growth, and \( z_t \) denotes the log price-consumption ratio. The specification for the real log return on the market portfolio is

\[
r_{m,t+1} = \kappa_{0,m} + \kappa_{1,m} z_{m,t+1} - z_{m,t} + g_{d,t+1},
\]

where \( g_{d,t+1} \) is the log dividend growth rate, and \( z_{m,t} \) denotes the log price-dividend ratio. The \( \kappa \)'s and \( \kappa_m \)'s are constants which depend on the average level of the price-consumption ratio or the price-dividend ratio, respectively.\(^{13}\)

The dynamics of log consumption growth and log dividend growth, which incorporate a small persistent predictable component \( x_t \), the long run risk component, and a time-varying volatility \( \sigma_t \), reflecting fluctuating economic uncertainty, are

\[
\begin{align*}
x_{t+1} &= \rho x_t + \varphi c \sigma_t \epsilon_{t+1} \\
g_{c,t+1} &= \mu_c + x_t + \sigma_t \eta_{t+1} \\
g_{d,t+1} &= \mu_d + \phi x_t + \varphi_d \sigma_t u_{t+1} \\
\sigma_{t+1}^2 &= \sigma^2 + v_1 (\sigma_t^2 - \sigma^2) + \sigma_w w_{t+1},
\end{align*}
\]

with \( \epsilon_{t+1}, \eta_{t+1}, \) and \( w_{t+1} \), having i.i.d. standard Normal distributions. The

\(^{13}\) Bansal, Kiku, and Yaron (2010) define \( \kappa_1 \) as \( \exp(\bar{z})/(1+\exp(\bar{z})) \) and set \( \kappa_0 \) equal to \( \ln(1+\exp(\bar{z})) - \kappa_1 \bar{z} \), where \( \bar{z} \) is the mean log price-consumption ratio. Accordingly, \( \kappa_{1,m} \) is defined by \( \exp(\bar{z}_m)/(1+\exp(\bar{z}_m)) \), and \( \kappa_{0,m} \) is set equal to \( \ln(1+\exp(\bar{z}_m)) - \kappa_{1,m} \bar{z}_m \), with \( \bar{z}_m \) being the mean log price-dividend ratio.
state variables, which determine the price-consumption and price-dividend ratio, are \( x_t \) and \( \sigma_t \). The solutions for \( z_t \) and \( z_{m,t} \) are

\[
\begin{align*}
    z_t &= A_0 + A_1 x_t + A_2 \sigma_t^2 \\
    z_{m,t} &= A_{0,m} + A_{1,m} x_t + A_{2,m} \sigma_t^2.
\end{align*}
\] (63)

The derivation of the \( A \) and \( A_m \) can be found in Bansal and Yaron (2004) and also in Bansal, Kiku, and Yaron (2010).

The model generates excess returns which are predictable by the price-dividend ratio. Equation (61) shows that the lagged price-dividend ratio has a negative impact on future returns. Hence, the relation implied between the price-dividend ratio and future returns is the same as in the HF and PT model.
### C Additional Tables

**Table 8: Comparison of the Model-Based Prior and the Campbell and Thompson Restrictions**

Reported is the increase or decrease in $R_{OS}^2$ when either imposing the model-based priors or the parameter restrictions from Campbell and Thompson (2008). For the model-based priors, the better performing prior, PM or ESV, for the respective period (total or postwar sample) and predictor of each model is selected. Bounded Slope means that the coefficient of the predictor must be of the same sign as the in-sample estimate, otherwise it is set to zero. The restriction Bounded Intercept/Slope ensures that the intercept is positive, otherwise it is set to zero, and the slope is between zero and one, otherwise it is truncated at these values. Fixed Coefficients stands for setting the slope equal to zero and the coefficient equal to 1. For Bounded Intercept/Slope and Fixed Coefficients simple returns are used, and the valuation ratios are not in logs. Further, these two restrictions only apply to valuation ratios. The initial estimation period is at least 20 years, but the OOS forecast does not start before 1926 for annual returns or December 1926 for overlapping annual returns. For the dividend-price ratio and the dividend yield, annual and overlapping annual returns are used. For the yields, quarterly returns are used.

<table>
<thead>
<tr>
<th></th>
<th>Model-Based Prior $\Delta R_{OS}^2$</th>
<th>CT Restrictions $\Delta R_{OS}^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HF</td>
<td>PT</td>
</tr>
<tr>
<td><strong>Div.-Price Ratio</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Returns 1872-2011</td>
<td>0.01%</td>
<td>0.32%</td>
</tr>
<tr>
<td>Overl. Ret. 1872-2011</td>
<td>1.16%</td>
<td>1.26%</td>
</tr>
<tr>
<td>Returns 1947-2011</td>
<td>7.35%</td>
<td>6.52%</td>
</tr>
<tr>
<td>Overl. Ret. 1947-2011</td>
<td>7.11%</td>
<td>5.94%</td>
</tr>
<tr>
<td><strong>Dividend Yield</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Returns 1872-2011</td>
<td>1.02%</td>
<td>0.98%</td>
</tr>
<tr>
<td>Overl. Ret. 1872-2011</td>
<td>1.59%</td>
<td>1.71%</td>
</tr>
<tr>
<td>Returns 1947-2011</td>
<td>2.60%</td>
<td>5.06%</td>
</tr>
<tr>
<td>Overl. Ret. 1947-2011</td>
<td>2.92%</td>
<td>5.82%</td>
</tr>
<tr>
<td><strong>Short Term Yield</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ret. 1Q 1920-4Q 2011</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ret. 1Q 1947-4Q 2011</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Long Term Yield</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ret. 1Q 1920-4Q 2011</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ret. 1Q 1947-4Q 2011</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Term Spread</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ret. 1Q 1920-4Q 2011</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ret. 1Q 1947-4Q 2011</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
References


