Return Predictability under Equilibrium Constraints on the Equity Premium

Davide Pettenuzzo
Department of Economics and International Business School
Brandeis University

Allan G. Timmermann
Rady School of Management
University of California, San Diego (UCSD)
Centre for Economic Policy Research (CEPR)

Rossen I. Valkanov
Rady School of Management
University of California, San Diego (UCSD)

October 2008

Abstract

This paper proposes a new approach for incorporating theoretical constraints on return forecasting models such as non-negativity of the conditional equity premium and sign restrictions on the coefficients linking state variables to the equity premium. Our approach makes use of Bayesian methods that update the estimated parameters at each point in time in a way that optimally exploits information in these constraints. Using a variety of predictor variables from the literature on predictability of stock returns, we find that theoretical constraints have an important effect on the coefficient estimates and can significantly reduce biases and estimation errors in these. In out-of-sample forecasting experiments we find that models that exploit the theoretical restrictions produce better forecasts than unconstrained models.

Keywords: return predictability, constraints, out-of-sample forecasts

JEL Classifications: G12, G14, C22

1   Introduction

Over the last twenty years, the stock return predictability literature has influenced a broad range of areas such as performance evaluation, asset pricing and

asset allocation.[1] This influence is largely due to studies by Campbell (1987), Campbell and Shiller (1988), Fama and French (1988, 1989), Ferson and Harvey (1991), and Keim and Stambaugh (1986) who provided convincing economic arguments and in-sample empirical results that some of the fluctuations in returns are predictable because of persistent time variation in expected returns. The in-sample evidence for predictability is accumulating as various new variables have been suggested as predictors of excess returns (Hodrick (1992), Pontiff and Schall (1998), Lamont (1998), Baker and Wurgler (2000), Lettau and Ludvigson (2001), Polk, Thompson, and Vuolteenaho (2006), among others). The out-of-sample predictability evidence, however, has been much less conclusive. Recent studies by Paye and Timmermann (2006) and Lettau and Van Nieuwerburgh (2007) argued that predictability weakened or disappeared during the 1990s. Bossaerts and Hillion (1999) and Goyal and Welch (2003, 2007) provide an even sharper critique by arguing that predictability was largely an in-sample or *ex-post* phenomenon which disappears once the forecasting models are used to guide forecasts on new, out-of-sample, data.

A shortcoming of the forecasting models used throughout the finance literature is that, while the common state variables are broadly guided by theoretical considerations, finance theory provides little guidance for the choice of functional form of the forecasting model. Largely as a consequence, linear forecasting models are used almost exclusively. As pointed out by Campbell and Thompson (2007, CT henceforth), a problem with these models is that the implied conditional equity premium often turns negative. It is difficult to imagine an equilibrium setting where risk-averse investors would hold stocks if their expected compensations were negative. CT argue that the out-of-sample or ex-ante forecasting performance of return prediction models can be improved by imposing theoretical constraints such as non-negativity of the conditional equity premium or a sign constraint on the coefficient of a given predictor variable. Whenever any of these constraints is violated, CT impose their constraints by truncating the return forecast at the unconditional estimate of the equity premium, i.e. the prevailing mean. While this can be viewed as a first approximation to imposing moment or parameter constraints, the approach fails to make full use of the information in the theoretical constraints.

In this paper we propose a new method that optimally incorporates theoretical constraints including, but not limited to, those proposed by CT. Our approach is based on Bayesian techniques which make imposing an arbitrary number of constraints computationally feasible. We show how to efficiently update the estimates of the restricted forecasting model every time new observations on returns and the predictor variable become available. Theoretical

---

[1] Papers on time-series predictability of stock returns include Campbell (1987), Campbell and Shiller (1988), Fama and French (1988, 1989), Ferson and Harvey (1991), Keim and Stambaugh (1986) and Pesaran and Timmermann (1995). Examples of asset allocation studies under return predictability include Ait-Sahalia and Brandt (2001), Barberis (2000), Brennan, Schwartz and Lagnado (1997), Campbell and Viceira (1999), Kandel and Stambaugh (1996) and Xia (2001). Avramov and Wermers (2006) and Ferson and Schadt (1996) consider mutual fund performance under time-varying investment opportunities.

constraints on the conditional equity premium can indeed have a big impact on parameter estimates of the return forecasting model. To see this, suppose, for example, that a new observation on returns and the predictor variable becomes available that, under the previous parameter estimates, imply a negative conditional equity premium at some point in the sample. Since our approach makes use of this observation to inform the updated parameter estimate, the old parameter estimates would need to be revised so as to ensure that the conditional equity premium is always positive. The theoretical constraint therefore allow investors to more efficiently update their beliefs about the parameters of the forecasting model. We argue that this is a highly attractive feature of our method since the constraints proposed by Campbell and Thompson—such as non-negativity of the conditional equity premium—are overwhelmingly supported by equilibrium arguments and hence should be fully exploited.

When implemented along the lines proposed in our paper, the economically motivated constraints turn out to be highly informative and lead to far more precise estimates of the parameters of the return forecasting model. Intuition for this surprising finding is that every time a new pair of observations on the predictor variable and returns becomes available, the non-negativity constraint on the conditional equity premium is used to rule out values of the parameter that are infeasible given the sign constraint. Since the conditional equity premium must be non-negative at each point in time, in a sample of T observations, we have T constraints rather than just a single constraint.

Through a set of Monte Carlo simulations we show that the better performance of our new forecasting approach can be understood in terms of a reduction in the bias of the slope coefficient known as the Stambaugh bias (Stambaugh (1986, 1999)) as well as smaller parameter estimation errors, i.e. a reduction in estimation uncertainty. The bias shifts estimates of the coefficient on variables such as the dividend yield away from zero. To see how this bias will be reduced (and eventually removed as the sample size increases) in our context, suppose that the true coefficient on the dividend yield is zero but that the bias is such that the coefficient estimate is positive in the absence of any constraints on the equity premium. As new observations of the dividend yield below its sample average emerge, the effect of imposing a sign constraint on the conditional equity premium is to shrink the distribution of the estimated coefficient towards zero - otherwise the predicted value would become negative. The smaller the value of the dividend yield, the stronger this effect is likely to be and so the approach ensures faster learning in the sense that the dispersion of the distribution of the estimated parameter gets reduced more rapidly than in the absence of any constraints. Individual observations—particularly those at odds with the theoretical constraints—can therefore lead to large (and instantaneous) shifts in the entire distribution of the parameter estimates. In contrast, approaches that ignore theoretical constraints when updating the parameter estimates will repeatedly make the same mistakes (i.e. predict negative stock returns).[2]

---

[2] The reduced bias associated with the constrained forecasting models also means that our approach also provides a new way to handle spurious predictability, a phenomenon that could

Predictability from individual forecasting variables in monthly return regressions is likely to be relatively weak and the literature on predictability of asset returns has documented predictability from several regressors. We therefore next consider making use of multivariate forecasting models. Return predictability in the context of multivariate regressions poses complications since it is difficult to impose sign restrictions on the coefficients of the individual predictor variables. Yet, this is often the type of restriction that economic theory implies. To simultaneously deal with sign restrictions on the individual predictor variables and incorporate information from several predictor variables, we propose to use combination methods that combine forecasts from several univariate return forecasting models each of which imposes such sign restrictions and also do not allow the conditional equity premium to be negative. We implement this strategy using Bayesian Model Averaging, a technique that has also been used in the return forecasting literature by Avramov (2004). This approach succeeds in both preserving the individual sign restrictions and in imposing that the conditional equity premium be non-negative. We find that there are considerable gains from forecast combinations that satisfy these sign constraints and use multivariate information.

The plan of the paper is as follows. Section 2 shows how to efficiently incorporate theoretical constraints on the forecasting models and outlines our proposed methodology. Section 3 presents empirical estimation results for a range of predictor variables while Section 4 studies the forecasting performance of both unconstrained and constrained return models. Section 5 considers the effect of the constraints on the bias and estimation error in the model parameters while Section 6 concludes.

2 Methodology

This section describes our new methodology to estimate the return forecasting model subject to a set of constraints motivated by finance theory. These take the form of inequality constraints on the conditional equity premium or constraints on the sign of coefficients relating state variables to the equity premium. Constraints on the signs of state variables are best understood in a univariate context since many of the predictor variables proposed in the literature are strongly correlated with *each other and their* signs can change in multivariate regressions.

2.1 Constraints on the Return Forecasting Model

The literature on predictability of stock returns is extensive. Early studies such as Campbell and Shiller (1988) and Fama and French (1988) found evidence that stock returns could be predicted by means of valuation ratios, while Fama and Schwert (1977), Keim and Stambaugh (1986) and Campbell (1987)

---

well explain the difference between the apparently strong in-sample predictability and weak out-of-sample predictability of returns (see Ferson et al. (2003))

found predictability from the T-bill rate or yields on long-term corporate and government bonds. Subsequent studies have explored information in corporate financing (Baker and Wurgler (2000)), consumption-wealth ratios (Lettau and Ludvigsson (2001)) and the value of high versus low beta stocks (Polk, Thompson and Vuolteenaho (2006)).

Almost invariably, return predictability has been explored in the context of the following simple unconstrained linear forecasting model for the stock return at time t, r_t, measured in excess of a risk-free rate:

$$r_t = \mu + \beta x_{t-1} + \epsilon_t. \tag{1}$$

Here x_{t-1} is the lagged value of the predictor variable and ϵ_t has zero mean and variance σ^2.

This model is attractive since it is simple to interpret and only requires estimating two mean parameters, μ and β. Finance theory generally does not restrict the functional form of the mapping from the state variable, x_{t-1}, to the excess return, r_t, so the use of the linear specification in (1) should be viewed as an approximation. Campbell and Thompson (2007) argue that finance theory can be used to improve on the model. In particular, the conditional equity premium should be non-negative since it is difficult to imagine that markets for stocks can clear while the conditional equity premium is negative. They implement this insight by proposing a truncated forecast which is simply the largest of the unconstrained OLS forecast and zero:

$$\hat{r}_t = \max(0, \hat{\mu} + \hat{\beta} x_{t-1}), \tag{2}$$

where $\hat{\mu}$ and $\hat{\beta}$ are the OLS estimates from (1). While this truncation prevents the predicted equity premium from becoming negative, the theoretical constraint is not used to obtain improved estimates of μ and β. While potentially an improvement over the simple unconstrained model, this approach therefore does not make efficient use of the theoretical constraints.

To efficiently exploit the information embedded in the constraint that the conditional equity premium is non-negative, the parameters μ and β should be estimated subject to the conditional equity premium constraint that $\mu + \beta x_{\tau-1} \geq 0$ for $\tau = 1, ..., t$:

$$r_\tau = \mu + \beta x_{\tau-1} + \epsilon_\tau \quad \mu + \beta x_{\tau-1} \geq 0 \quad (\tau = 1, ..., t) \ . \tag{3}$$

Although the conditional equity premium constraint is not directly a constraint on the model parameters, $\theta = (\mu, \beta)$, it clearly affects these parameters which have to be selected so as to be consistent with $\mu + \beta x_{\tau\text{-}1} \geq 0$ for $\tau = 1, ..., t$. Note that the conditional equity premium constraint has to hold at each point in time, so the number of constraints grows in proportion with the length of the sample size. The seemingly simple equity premium constraint therefore potentially yields a very powerful way to tie down the parameters of the return forecasting model and obtain more precise estimates.

In many situations finance theory is informative about the sign of the slope coefficient, β, relating returns to the state variable, x. To cover such cases, we

also consider a specification that estimates the forecasting model by first imposing a sign constraint on β and then imposing the conditional equity premium constraint:

$$r_\tau = \mu + \beta x_{\tau-1} + \epsilon_\tau \quad \beta I_x \geq 0,\ \mu + \beta x_{\tau-1} \geq 0 \quad (\tau = 1, ..., t)\ , \qquad (4)$$

here I_x is an indicator function that is either +1 or -1, depending on the sign of the constraint.

We next explain how the models are estimated and how the constraints are imposed.

2.2 Accounting For Constraints through Investors' Prior Beliefs

The theoretical constraints incorporated in the models (3) and (4) are naturally interpreted as reflecting the forecaster's prior beliefs on return predictability. Viewed this way, they can best be imposed using Bayesian techniques. Estimation of return forecasting models subject to these constraints therefore requires specifying priors for the regression coefficients and introducing inequality constraints on the model parameters as specified by models (3) and (4) through the priors. To this end we study two sets of priors: a set of Normal-Gamma priors and a set of conditional Normal-Jeffreys' priors. We consider both types of priors to establish the robustness of our results. In what follows we first introduce the priors without constraints and then demonstrate how to incorporate the constraints.

2.2.1 Basic Priors

The first prior we consider for the unconstrained univariate return prediction models is the standard Normal-Gamma prior. Under this prior the parameters of the return model, i.e. the mean parameters $\theta \equiv (\mu, \beta)$ and the precision *parameter* $h = \sigma^{-2}$, follow independent Normal-Gamma distributions:

$$\begin{aligned} \theta &\sim N(\underline{\theta}, \underline{V}) \qquad\qquad\qquad (5) \\ h &\sim G(\underline{v}, \underline{s}^2). \end{aligned}$$

The first two moments of the parameters of interest are: $E(\theta) = \underline{\theta}$, $\text{var}(\theta) = \underline{V}$, $E[h] = \underline{v} \times \underline{s}^{-2}$ and $var(h) = \underline{v} \times \underline{s}^{-4}$ and so the parameters $\underline{\theta}$, $\underline{V}$, $\underline{s}^2$ and $\underline{v}$ fully characterize the priors. Moreover, it follows from the independence assumption that the joint prior distribution is simply the product of each part

$$P(\theta, h) = P(\theta)P(h). \qquad\qquad (6)$$

Following Wachter and Warusawitharana (2007), the second prior is a conditional Normal-Jeffreys prior. We assume the investor holds non-informative (Jeffreys prior) beliefs on the intercept and precision parameter, while the prior

concerning return predictability, measured by the slope coefficient, β, is allowed to depend on the other parameters:

$$\beta | \mu, h \sim N(\underline{\beta}, \underline{\varphi}/h). \tag{7}$$

Letting $P(\mu, h)$ be the prior probability density on (μ, h), the joint prior distribution on the whole set of parameters is

$$P(\theta, h) = P(\beta | \mu, h) P(\mu, h). \tag{8}$$

We start by deriving a limiting Jeffreys prior on the full set of parameters (θ, h). Following Stambaugh (1999),

$$P(\theta, h) \propto (\det I(\theta, h))^{1/2}$$

where $I(\theta, h)$ is the Fisher information matrix. Lastly, combining this part with the conditional Normal density, we get

$$P(\theta, h) \propto (\underline{\varphi})^{-1/2} h^{1/2} \exp(-\frac{h(\beta - \underline{\beta})^2}{2\underline{\varphi}}) h^{-1/2} = \underline{\varphi}^{-1/2} \exp(-\frac{(\beta - \underline{\beta})^2}{2\underline{\varphi}} h). \tag{9}$$

Next we show how to incorporate constraints on these priors.

2.2.2 Prior Beliefs under Constraints

Both the constraint on the sign of the slope coefficient, β, and the conditional equity premium constraint restricts the mean parameters $\theta = (\mu, \beta)$ and hence take the following form:

$$\theta \sim F_0 \times I(\theta \in A), \tag{10}$$

where F_0 is the prior distribution of θ without constraints. In the context of this paper, under the Normal-Gamma prior, $F_0 = N(\underline{\theta}, \underline{V})$, while under the Normal-Jeffreys prior, $F_0 = N(0, \underline{\varphi}/h)$. $I(\theta \in A)$ is an indicator function that equals unity if $\theta \in A$ and is zero otherwise. A is the admissible region for the regression coefficients as reflected in the theoretical constraints on the models (3) and (4). Specifically, under (3) A is the set of parameter values satisfying that the predicted return, $\hat{r}_\tau \geq 0$, $\tau = 1, ..., t$. Hence the conditional equity premium constraint is equivalent to restricting θ to lie in the set A:

$$
\begin{aligned}
A &= \{\mu + x_{\tau-1} \cdot \beta \geqslant 0, \text{for } \tau = 1, 2, ..., t\} \\
&= \{\Lambda \theta \geqslant 0\}, \tag{11} \\
\text{where } \Lambda &= \begin{pmatrix} 1 & \max(X_{\tau-1}) \\ 1 & \min(X_{\tau-1}) \end{pmatrix}, \tau = 1, 2, ..., t
\end{aligned}
$$

Similarly, letting $\underline{sign(\beta)} = \pm 1$ denote the prior belief on the sign of the slope coefficient, for the model that imposes constraints on both the sign of β and on

the conditional equity premium, we have:

$$
\begin{aligned}
A &= \{\mu + x_{\tau-1} \cdot \beta \geqslant 0, \text{for } \tau = 1, 2, ..., t; \text{ and } \underline{sign(\beta)} \cdot \beta \geqslant 0\} \\
&= \{\Psi\theta \geqslant 0\}, \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (12) \\
\text{where } \Psi &= \begin{pmatrix} 1 & f(X) \\ 0 & \underline{sign(\beta)} \end{pmatrix}, \\
\text{and } f(X) &= \begin{cases} \min(X_{\tau-1}) & \text{if } \underline{sign(\beta)} = 1 \\ \max(X_{\tau-1}) & \text{if } \underline{sign(\beta)} = -1 \end{cases}, \tau = 1, 2, ..., t.
\end{aligned}
$$

2.3 Choice of Priors

So far we have explained how we impose the constraints implied by the priors that the conditional equity premium is non-negative *and/or the sign of the slope coefficient* . We next explain our specific choice of prior parameters $(\underline{\theta}, \underline{V}, \underline{v}, \underline{s}^2)$ in the Normal-Gamma case and $(\underline{\beta}, \underline{\varphi})$ in the Normal-Jeffreys case. We assume, first, that investors hold prior beliefs that stock returns are not predictable and, second, that investors hold diffuse priors about the remaining parameters.

Starting with the Normal-Gamma case, our prior reflects the "no predictability" view that the best predictor of the stock return is the historical average. At each point in time t, we therefore center the prior *intercept*, $\underline{\mu}$, on the prevailing mean of historical excess returns, while the prior slope coefficient is centered on zero, $\underline{\beta} = 0$. The prior precision for the mean parameters is $\underline{V} = \psi \times I_k \text{ where } I_k$ denotes the k-dimensional identity matrix (we only consider univariate models so $k = 2$) and ψ is a scaling factor that controls the tightness of the prior. We consider values $\psi = 0.1$ and $\psi = 1$. For the prior *belief on h*, we set $\underline{v} = 1$ and $\underline{s}^2 = 4$ so the Gamma distribution reduces to an exponential distribution which has a significant probability mass near zero *and reflects a diffuse view*.[3]

Turning to the Normal-Jeffreys prior, we continue to assume "no predictability", i.e. $\underline{\beta} = 0$. *Besides,* it is reasonable to let investors' priors on β depend on the variation in the predictor variable (x): A high variance of the predictor variable (captured by σ_x) might lower the spread of the prior on β. Thus, we rewrite $\underline{\varphi}$ in terms of σ_x and a prior scale parameter σ_β, that is common across all predictor variables irrespective of their variance: $\underline{\varphi} \equiv \underline{\sigma_\beta^2 \times \sigma_x^{-2}}$. We consider values $\underline{\sigma_\beta = 0.02}$ and $\underline{\sigma_\beta = 0.2}$ and, at each point in time, set σ_x equal to the historical standard deviation of the explanatory variable, x. This conditional prior on β also reflects a prior on the population R^2, since $R^2 = \beta^2 \sigma_x^2 / (\beta^2 \sigma_x^2 + \sigma_\mu^2)$, see Wachter and Warusawitharana (2007).

---

[3]To see this, consider the probability density function (pdf) of the gamma distribution: $f(x; v, s^2) = x^{v-1} \frac{s^{2v} \exp(-s^2 x)}{\Gamma(v)}$, for $x > 0$. In general, this pdf has a hump shape, but for $v = 1$ it reduces to $f(x; s^2) = s^2 \exp(-s^2 x)$, which is the pdf of an exponential distribution. The value of this function at zero is infinite and it exhibits an exponential decay thereafter.

2.4 Posterior Distributions

We next derive the posterior distributions under the two priors considered so far. Assuming that returns are normally distributed, we can use standard results from Bayesian analysis to obtain formulas for the posterior distribution of the mean and precision parameters.

First we introduce some notations needed for our analysis. Note that the return model can be written in matrix notations

$$Y^t = X^t \theta + \varepsilon^t,$$

where

$$
\begin{aligned}
Y^t &= [r_1, r_2, ..., r_t]' \\
X^t &= \begin{bmatrix} 1, 1, 1, ..., 1 \\ x_0, x_1, x_2, ..., x_{t-1} \end{bmatrix}' \\
Z^t &= \{Y^t, X^t\} \\
\varepsilon^t &= [e_1, e_2, e_3, ..., e_t]' \sim N(0, \sigma^2 I_t)
\end{aligned}
$$

In the following, for simplicity we ignore the subscript t.

Under the Normal-Gamma prior, conditional on h and the data up to time t, the posterior density of θ is given by:

$$\theta | h, Z \sim N(\bar{\theta}, \overline{H}^{-1}) \times I(\theta \in A), \tag{13}$$

where $\bar{\theta}$ and $\overline{H}^{-1}$ is the posterior mean and covariance of the parameters. These moments of the posterior distribution are given by $\overline{H} = \underline{H} + hX'X$ and $\bar{\theta} = \overline{H}^{-1}(\underline{H}\underline{\theta} + hX'Y) = \overline{H}^{-1}hX'Y$.[4]

Conditional on θ and the data, Z, the posterior density of h is given by:

$$h | \theta, Z \sim G(\bar{v}, \overline{S^2}) \tag{14}$$

where $\bar{v} = T + 2$ and $\overline{S^2} = \underline{s}^2 + (Y - X\theta)'(Y - X\theta)$. The unconstrained model (1) is nested as a special case when $A = R^2$. Analytical results are available for the unconstrained model (1) which does not impose any restrictions on the parameter estimates and thus preserves the full (non-truncated) distribution so that $I(\theta \in A)$ is always *Identity*.

To derive similar results under the Normal-Jeffreys prior, we start from the joint kernel of the posterior distribution:

$$
\begin{aligned}
P(\theta, h | Z) &\propto P(Z | \theta, h) P(\theta, h) \times I(\theta \in A). \tag{15} \\
&\propto h^{T/2} \exp\{-\frac{h}{2}(Y - X\theta)'(Y - X\theta)\} \varphi^{-1/2} \exp(-\frac{\beta^2}{2\varphi}h) \times I(\theta \in A).
\end{aligned}
$$

---

[4] The priors are, as discussed earlier, $\underline{\theta} = (\bar{Y}, 0)$ and $\underline{H} = \psi I_2$.

If *we interpret the above formula* as a function of h only, then it is a posterior kernel for h conditional on θ:

$$
\begin{aligned}
P(h|\theta, Z) &\propto h^{T/2} \exp\{-\frac{h}{2}(Y - X\theta)'(Y - X\theta)\} \exp(-\frac{\beta^2}{2\varphi}h) \\
&= h^{((T+2)-2)/2} \exp\{-\frac{h}{2}[\frac{\beta^2}{\varphi} + (Y - X\theta)'(Y - X\theta)]\}.
\end{aligned}
$$

This is the kernel of a Gamma distribution:

$$
h|\theta, Z \sim G(\overline{v}, \overline{S^2}),
$$

where $\overline{v} = T + 2$ and $\overline{S^2} = \frac{\beta^2}{\varphi} + (Y - X\theta)'(Y - X\theta)$.

Conversely, if (15) is interpreted as a function of θ only, then it is a posterior kernel for θ conditional on h:

$$
\begin{aligned}
P(\theta|h, Z) &\propto h^{T/2} \exp\{-\frac{h}{2}(Y - X\theta)'(Y - X\theta)\} \exp(-\frac{\beta^2}{2\varphi}h) \times I(\theta \in A). \\
&= h^{T/2} \exp\{-\frac{1}{2}[(\theta - \underline{\theta})'\underline{H}(\theta - \underline{\theta}) + h(Y - X\theta)'(Y - X\theta)]\} \cdot I(\theta \in A),
\end{aligned}
$$

This is the kernel of a Normal distribution:

$$
\theta|(h, Z) \sim N(\overline{\theta}, \overline{H}^{-1}),
$$

whose posterior moments are $\overline{H} = \underline{H} + hX'X$ and $\overline{\theta} = \overline{H}^{-1}(\underline{H}\underline{\theta} + hX'Y) = \overline{H}^{-1}hX'Y$.[5]

Comparing the posterior distributions under the two sets of priors, we see that they could be unified under a single Normal-Gamma prior framework, but with different specifications of the priors *[θ,H]*.

2.5 The Gibbs Sampler

Unfortunately, closed-form expressions are not available for the constrained models (3) and (4). Estimating the parameters of these models requires evaluating the posterior distribution of the parameters given the data up to time t, $\mathcal{Z}$, denoted $\pi\left(\theta, \sigma^{-2} \middle| \mathcal{Z}\right)$. This in turn requires repeatedly drawing from the distribution π which is not always feasible in our context. Hence we cannot use Monte Carlo integration methods to simulate posterior moments of functions of the parameters. Instead we implement the Gibbs sampler and use importance sampling techniques which we next describe.

To implement the Gibbs sampler, we partition the parameters $\delta = (\theta, \sigma^{-2})$ into two blocks:

$$
\delta_{(1)} = \sigma^{-2}, \delta_{(2)} = \theta. \tag{16}
$$

---

[5] For this case the priors are $\underline{\theta} = [0, 0]$ and $\underline{H} = \begin{bmatrix} 0 & 0 \\ 0 & h\varphi^{-1} \end{bmatrix}$.

Given an initial draw $\delta^{(0)} = (\delta_{(1)}^{(0)}, \delta_{(2)}^{(0)})$ from $\pi(\delta | Z)$, we successively draw new parameters

$$\delta_{(b)}^{(s)} \sim \pi\left(\delta_{(b)} \big| \delta_{<(b)}^{(s)}, \delta_{>(b)}^{(s-1)}, Z\right), \quad b = 1, 2; \quad s = 1, 2, ..., . \tag{17}$$

The resulting sequence, $\{\delta^{(s)}\}$, is a realization of a Markov chain. Under well-known conditions the Markov chain converges (Roberts and Smith (1994)) and any single iterate $\delta^{(s)}$ retains the property that it is draw from the joint density $\pi\left(\theta, \sigma^{-2} \big| Z\right)$.

The Markov chain (17) requires sampling from the two conditional densities (13) and (14). Drawing from (14) is straightforward and can be carried out by many statistical packages. Drawing from (13), however, is non-standard and requires using importance sampling techniques. Suppose that random draws θ^s, $s = 1, ..., S$ can be generated from a density, $q(\theta)$, the so-called importance function. By appropriately weighting the random draws from $q(\theta)$, the moments computed from the draws of the importance function, θ^s, converge to the moments obtained from the (unknown) posterior distribution $\pi(\theta | Z_t)$.[6]

For importance sampling to work, $q(\theta)$ needs to approximate $\pi(\theta | Z_t)$ quite well. Otherwise cases can be found where $\omega(\theta^s)$ is equal to zero for virtually every draw and the weighted average involves very few draws. Thus, importance sampling may become inaccurate unless $q(\theta)$ is chosen carefully. Fortunately, the problem of finding an accurate importance function is easily resolved for the linear regression model that is subject to inequality constraints. By setting the importance function equal to the unconstrained posterior distribution, the weights can be computed as

$$\omega(\theta^s) = I(\theta^s \in A). \tag{18}$$

Hence the weights are either one (if $\theta^s \in A$) or zero (if $\theta^s \notin A$) and this strategy simply involves drawing from the unrestricted posterior distribution and discarding draws that violate the relevant inequality restrictions. Hence our approach is very simple to implement in practice.

3   Empirical Results

In this section we present empirical results from applying the methods described in the previous section to forecast stock returns.

---

[6] This property makes use of the result in Geweke (1989) that if θ^s, $s = 1, ..., S$ is a random sample from $q(\theta)$, then under weak conditions

$$\frac{\sum_{s=1}^{S} \omega(\theta^s) g(\theta^s)}{\sum_{s=1}^{S} \omega(\theta^s)} \to E[g(\theta) | \mathcal{F}_t],$$

where the weights of the importance function, $\omega(\theta^s)$, are given by

$$\omega(\theta^s) = \frac{\pi(\theta = \theta^s | \mathcal{F}_t)}{q(\theta = \theta^s)}.$$

3.1 Data

Our empirical analysis uses the data on monthly stock returns along with a set of sixteen predictor variables analyzed in Goyal and Welch (2007).[7] Stock returns are measured by the S&P500 index and include dividends. A short T-bill rate is subtracted from stock returns in order to capture excess returns. Data samples vary considerably across the individual predictor variables. To be able to compare *[[−and later combine−]](delete)* results across the individual predictor variables, we use the longest common sample which goes from 1940-2005.[8]

The identity of the predictor variables is listed in Table 1. Most variables fall into three broad categories, namely (i) valuation ratios capturing some measure of 'fundamentals' to market value such as the dividend price ratio, the dividend yield, the earnings-price ratio, the 10-year earnings-price ratio or the book-to-market ratio; (ii) measures of bond yields capturing level effects (the three-month T-bill rate and the yield on long term government bonds), slope effects (the term spread), and default risk effects (the default yield spread defined as the yield spread between BAA and AAA rated corporate bonds, and the default return spread defined as the difference between the yield on long-term corporate and government bonds); (iii) estimates of equity risk such as the cross-sectional equity premium (the relative valuations of high- and low-beta stocks), long term return and stock variance (a volatility estimate based on daily squared returns). Finally, two corporate finance variables, namely the dividend payout ratio (the log of the dividend-earnings ratio), and net equity expansion (the ratio of 12-month net issues by NYSE-listed stocks over the year-end market capitalization) and a macroeconomic variable, inflation (the rate of change in the consumer price index) are considered.

3.2 Effect of Constraints on coefficient estimates

Before turning to the forecasts of stock returns, we consider the posterior distribution of the coefficient estimates based on the full data sample available at the end of 2005. These contain interesting information about the economic significance of the various predictor variables.

Towards this end, the first two columns of Table 1 report OLS estimates of the slope coefficient β along with the associated t-statistics for the unconstrained model. Roughly half of the predictor variables generate coefficient estimates that are significant at the 5% level. However, the $t-$statistics for the valuation ratios should not be taken at face value given the well-known biases in their estimates (see, e.g., Stambaugh (1999)).

Columns 3-10 of Table 1 report the posterior means of β under the unconstrained and constrained Bayesian models using four combinations of uninformative Normal-Gamma priors with $\psi = 0.1$ or $\psi = 1$ and Normal-Jeffreys' priors with $\sigma_\beta = 0.02$ or $\sigma_\beta = 0.2$. Since the results are very similar under

---

[7]We are grateful to Amit Goyal for providing this data.

[8]One variable, the cross-sectional premium, only has data up to the end of 2003.

the two constrained models (3) and (4), we only report estimates for the latter. First consider the results for the unconstrained models. Under Normal-Gamma priors the posterior means of the slope coefficient tend to be very similar to the OLS estimates—more so as ψ is increased from 0.1 to 1 and less weight is put on the prior. Under the Jeffreys' priors, the posterior means of the slope coefficients remain quite close to the OLS estimates—although of course slightly closer to zero, the center of the priors—when σ_β is set to 0.2. Lowering σ_β to 0.02, and thus using a prior more strongly concentrated on zero, has the effect of shrinking the posterior mean of β more towards zero and so the absolute value of the posterior means are generally much smaller under this prior.

Turning to the constrained models, under either set of priors, the constraints have a clearly identifiable effect on the posterior means of the coefficients which tend to fall between the OLS estimates and the posterior means under the corresponding unconstrained models. The posterior means of the slope coefficients can vary significantly depending on whether the unconstrained or constrained model is adopted and on the choice of prior. For example, in the case of the net equity expansion variable, the OLS coefficient is -0.197 which is very close to the posterior mean of the unconstrained model under Jeffreys priors with $\sigma_\beta = 0.2$, but is somewhat smaller than the value (-0.12) obtained under the constrained model with $\sigma_\beta = 0.2$ and the values (-0.05 to -0.06) obtained when $\sigma_\beta = 0.02$.

One of the advantages of our methodology is that it treats the coefficients on the predictor variables as random variables. Hence we can study the entire distribution of the coefficients of the predictor variables (β) conditional on the data and any restrictions that may have been imposed on the forecasting model. This provides insights into the effect on the forecasting model of imposing constraints on the equity premium or on the sign of β. Figure 1 plots the posterior distribution of β *under Normal-Gamma prior with* $\varphi = 0.1$ for each of the models described in section 2 (i.e. the unconstrained model (1), the model (3) that imposes non-negative equity premia, $\hat{r}_1, ...\hat{r}_t \geq 0$ and the model (4) that further imposes a sign constraint on the slope coefficient).

Several points stand out from these plots. First, in many cases, imposing the constraint that the conditional equity premium cannot be negative has a very significant impact on the distribution of the slope coefficients in the return equation. In comparison, the sign constraint on β generally tends to have a much smaller additional effect on the posterior distribution of β.

Imposing the constraints on the forecasting model has separate effects on the location and dispersion of the slope parameter, β. First, the distribution of the slope coefficient, β, tends to be less dispersed with higher peaks under the constrained models. For example, whereas the distribution of β in the unconstrained return model based on the dividend-price ratio is concentrated between -0.01 and 0.025, it lies in a much more narrow band between zero and 0.01 under the constrained model. Even larger effects of imposing the constraints can be observed for the slope coefficients of variables such as the T-bill rate, long-term return, stock variance and inflation.

The second effect is related to the location or center of the distribution of the slope coefficient, β. For most variables (with the possible exception of

the default yield spread), Figure 1 shows that the distribution of β is shifted towards zero. For example, in the case of the dividend-price ratio, the mean of β changes from 0.008 under the unconstrained model to 0.004 under the constrained model. In general, predictor variables whose coefficient estimates are predominantly positive (such as the valuation ratios) therefore see their distributions shift to the left, while conversely variables such as the T-bill rate or the rate of inflation whose unconstrained coefficient estimates are centered on negative values are shifted to the right when the conditional equity premium is required to be non-negative.

Finally, Figure 1 shows the effect of imposing the double constraint that the conditional equity premium at each point in time is non-negative and that the sign of β must be positive for the valuation ratios, spread and interest rate variables and negative for inflation and the T-bill rate. The result of the sign constraint is to further narrow the distribution of the slope coefficients. Unsurprisingly the effect of the additional constraint is largest for those models whose slope coefficient has a distribution that is centered near zero with considerable probability mass on both positive and negative values. In practice the constraint has a much smaller effect for variables such as the T-bill rate or the dividend yield whose *unconstrained* slope coefficients are largely distributed on one side of zero. Because of the asymmetric nature of the slope restriction, imposing sign constraints on β generally pushes the mean of β further away from zero compared to when only the conditional equity premium is constrained to be non-negative.

4   Out-of-sample Forecasts of Stock Returns

The key question in the literature on return predictability is whether stock returns can be predicted ex ante. To address this issue, we next study the out-of-sample forecasting performance of the models under consideration here. Some studies (e.g. Pesaran and Timmermann (1995), Bossaerts and Hillion (1999), Goyal and Welch (2003, 2007)) have addressed out-of-sample or ex-ante predictability by accounting for the effect of parameter estimation error associated with investors' updating of their models based only on historically available ("real time") data.

Our analysis uses the first 10 years of data (1940-49) to obtain initial parameter estimates so the forecasts begin in 1950. More specifically, we use our approach to compute recursive parameter estimates both for the unconstrained model (1) and the constrained model (4). In order to avoid look-ahead bias in the parameter estimates we only use data up to the month prior to that for which return is being predicted.[9] For example, the forecast of returns in January 1950 is based on an estimate that uses data up to and including December 1949. To forecast excess returns for February 1950, we extend the data set by

---

[9]Forecasts from the model that only constrains the equity premium (3) are very similar to those from the model (4) that constrains both the equity premium and the sign of the slope coefficient β and are thus not reported separately.

one observation (i.e. up to January 1950), re-estimate the parameters of the forecasting model subject to any constraints and then compute a new forecast. We continue recursively with this estimation and forecasting procedure up to the end of the sample in 2005:12.

Figure 2 plots the out-of-sample forecasts associated with the unconstrained and constrained forecasting models over the period from 1950-2005. To preserve space we only show the prevailing mean and the forecasts based on the dividend yield and the T-bill rate. The unconstrained and constrained forecasts from the prevailing mean model are almost identical since the sample estimate of the prevailing mean is always positive so the equity premium constraint has the rather modest effect of truncating a very small part of the left tail of the distribution of the parameter μ that controls mean excess returns, .

In contrast, the sequence of return forecasts based on the T-bill rate (shown in the middle window) is an example where imposing constraints on the equity premium makes a very big difference. The unconstrained forecasts based on this model are very volatile and negative most of the time from 1970-1985, whereas the constrained forecasts are far smoother and, by construction, never take negative values. The constrained forecasts are mostly higher than the unconstrained values although this is not the case during the last three years of the sample.

The unconstrained and constrained forecasts based on the model that uses the dividend yield as a predictor variable (shown in the bottom window) clearly share a common trend. This is a reflecting of the persistent movements in the dividend yield. However, the unconstrained forecasts are generally smaller than those based on the constrained model and, moreover, turn negative in 1987 and from 1992 onwards. As we shall later see, this explains the poor forecasting performance of this model since on average stock returns were quite high during this period. Imposing that the equity premium is non-negative leads the dividend yield model to perform much better as the restricted model predicts large and positive excess returns that even drift slightly upwards after the mid-1990s.

We conclude from these findings that imposing basic equilibrium restrictions on the conditional equity premium can have large effects on the ex-ante predicted return. Comparing the forecasts from the unconstrained and constrained models, the predicted returns can in some cases differ by more than 200 basis points per month.

4.1 Evaluation of Forecasts

From the substantial difference in the time series of forecasts shown in Figure 2, we would expect that the unconstrained and constrained models produce quite different out-of-sample forecasting performance. Table 2 confirms that this is indeed the case. Following Campbell and Thompson (2007) this table reports

15

out-of-sample R^2-values computed as

$$R^2_{oos,i} = 1 - \frac{\sum_{t=1}^{T} e^2_{t,i}}{\sum_{t=1}^{T} \bar{e}^2_t}, \tag{19}$$

where $e_{t,i} = r_t - \hat{r}_{t,i}$ is the forecast error from the ith forecasting model, $\bar{e}_t = r_t - \bar{r}_t$ is the forecast error from the prevailing mean model (which assumes no predictability) and $t = 1, .., T$ is the out-of-sample period. The out-of-sample R^2-value is one minus the (squared) ratio of the root mean squared error (RMSE) of the ith forecasting model measured relative to that of the no predictability (prevailing mean) model with recursively updated parameter estimates. Forecasting models with smaller out-of-sample RMSE-values than the prevailing mean model generate positive values of $R^2_{oos,i}$, while conversely models with greater out-of-sample RMSE-values produce negative values of $R^2_{oos,i}$. Because of this one-to-one mapping between RMSE and out-of-sample R^2, there is no need for us to separately report the RMSE values.

As a benchmark the first column in the table shows the full-sample R^2-value obtained under the simple least-squares model. This varies significantly across predictor variables and is quite high (0.6%) for the dividend-price ratio, the dividend yield, the smoothed earnings price ratio, the T-bill rate and long-term return. It is even higher in the case of inflation (1.5%) and the cross-sectional premium (1.14%), but is quite low for many of the remaining predictor variables.

The rest of the table reports out-of-sample R^2 value under different methods. Consistent with the findings reported by Campbell and Thompson (2007) and Goyal and Welch (2007), the OLS results in Panel A show that there is only weak evidence of out-of-sample predictability based on the unconstrained forecasting models. In fact, nine of the sixteen forecasting models produce negative R^2-values and the average R^2-value (computed across all 16 models) is also negative. More troubling, perhaps, is that for many predictor variables the out-of-sample R^2 is quite large with negative values that exceed the corresponding positive in-sample R^2-values listed in the first column. Basing investment strategies on such forecasts would therefore in all likelihood lead to underperformance compared to a simple model with no predictability.

Similar findings hold for the unconstrained Bayesian models. Across all priors, close to half of these models produce negative out-of-sample R^2-values.

Following Campbell and Thompson (2007) we also considered the truncated OLS forecasts (2). *The results are listed in the first column under "Constrained Models".* For 14 of the 16 models under consideration, the out-of-sample R^2-value is improved by imposing the equity premium constraints. This is consistent with Campbell and Thompson's finding that such constraints can improve forecasting performance.

The truncated OLS forecasts proposed by Campbell and Thompson (2007) do not revise the parameter estimates in view of the constraints. In contrast, the Bayesian methodology which we propose *in this paper* incorporates this information in the parameter estimates. Table 2 shows that there are clearly considerable gains from adopting our methodology. For example, under the

Normal-Gamma prior with $\psi = 0.1$, the average R^2−value from the univariate forecasting models more than doubles from 0.15 under the constrained OLS forecasts to 0.33. This increase reflects improved forecasting performance under the Bayesian method for 11 of the 16 models.

Similar results are obtained under the Jeffreys priors. Imposing the constraints lead to an improvement in out-of-sample forecasting performance for 13 or 14 out of 16 models, depending on whether $\sigma_\beta = 0.02$ or $\sigma_\beta = 0.2$. This reflects an increase in the average R^2−value of 0.10 when $\sigma_\beta = 0.02$ and an increase in this statistic close to 0.40 when $\sigma_\beta = 0.2$. Clearly the constraints matter significantly to forecasting performance. Moreover, under the Jeffreys priors *on average* the constrained Bayesian models continue to outperform the truncated OLS forecasts.

We conclude from these results that imposing constraints on the equity premium leads to substantial improvements in the out-of-sample forecasting performance of the majority of univariate forecasting models with only four of 16 variables being unable to predict returns. Imposing such constraints cuts the number of cases with negative R^2−values roughly in half from 27 to 15 cases across all priors and univariate forecasting models (i.e. out of a total of 64 cases) and raises the R^2−value for twelve of the sixteen variables. In addition, we find significant improvements in the precision of the forecasts by exploiting these constraints to inform the posterior distribution of the coefficients by using our Bayesian methodology as opposed to simply truncating the forecasts at zero.

5   Stambaugh Bias

Many of the predictor variables used in the return forecasting literature are known to be highly persistent. Moreover, innovations to such variables and innovations to returns are in some cases strongly correlated. Such conditions are known to introduce a significant bias in the coefficient estimates and a resulting bias in the forecast. This phenomenon, which is generally known as the Stambaugh bias (Stambaugh (1986, 1999)) is particularly important for valuation ratios such as the dividend yield or the price-earnings ratio. These variables are highly persistent with innovations that are strongly negatively correlated with returns.

A large literature has attempted to address the small-sample bias problem by developing refined econometric estimators and bias-reduction methods. Examples include Ang and Bekaert (2007), Campbell and Yogo (2006), Cavanagh, Elliott and Stock (1995), Jansson and Moreira (2006), Lewellen (2004), Polk, Thompson and Vuolteenaho (2006) and Torous, Valkanov and Yan (2004).

Our approach provides an effective way to reduce the Stambaugh bias and, as a result, improve on the precision of the forecasts. A feature of our approach which distinguishes it from the more statistically-based methods for dealing with the problem is that it uses an estimator that is economically motivated by a constraint that holds as a consequence of simple equilibrium arguments. Imposing this constraint on the forecast in every period is particularly appropriate for

the simple forecasting models which, as pointed out by Ang and Bekaert (2007), can reasonably be viewed as representing approximations rather than an exact relationship.

Along with most studies in the literature, we consider the following simple model for returns (r_{t+1}) and the predictor variable (x_{t+1}):

$$
\begin{aligned}
r_{t+1} &= \mu_r + \beta_r x_t + u_{t+1} \\
x_{t+1} &= \mu_x + \beta_x x_t + v_{t+1}.
\end{aligned} \tag{20}
$$

Returns are predictable by means of past values of x whenever $\beta_r \neq 0$. Moreover, the predictor variable follows a first-order autoregressive process whose persistence is measured by $\beta_x \in [0, 1]$. Another key parameter is the correlation between shocks to returns and shocks to the state variable, $\rho_{uv} = \sigma_{uv}/(\sigma_v \sigma_u)$. As noted by Stambaugh (1999), OLS estimates of β_r, denoted by $\hat{\beta}_r$, are biased by approximately

$$
E[\hat{\beta}_r - \beta_r] \approx -\frac{\sigma_{uv}}{\sigma_v^2}\left(\frac{1 + 3\beta_x}{T}\right), \tag{21}
$$

where T is the sample size, σ_{uv} is the covariance between u and v and β_x is the persistence parameter for the predictor variable. Valuation ratios such as the earnings-price ratio or the dividend yield are highly persistent with estimates of β_x close to one and a correlation between u and v that is large and negative. This introduces a large upwards bias in the estimate of the slope coefficient β_r in (20) given $\sigma_{uv} < 0$. This has the dual effect of making inference on $\hat{\beta}_r$ very difficult and also leads to worse forecasting performance since the return forecasts become too sensitive to the value of x.

Figure 3 shows how the equity premium constraint works when $x \geq 0$ (top window) or $x < 0$ (bottom window). For a given value of x, $\mu + \beta x \geq 0$ is a joint constraint on both μ and β. When plotted against μ and β, the boundary of this constraint always goes through zero and has a slope $\mu/\beta = -x$ (for $\beta \neq 0$) that reflects the value of the predictor variable, x: The intercept-slope ratio μ/β must lie between $-\max(x)$ and $\min(x)$. Together with the objective of obtaining the "best fit" by trading off the intercept and the slope coefficient, the equity premium constraint tends to drag β towards zero.

5.1   Simulation Results

To see how the bias in the slope coefficient and the forecasting performance is affected by the equilibrium constraints, we conduct a series of simulation experiments. We simulate returns under the model (20) fitted to stock returns and the dividend yield, using OLS estimates of the coefficients $\{\hat{\mu}_r, \hat{\beta}_r, \hat{\sigma}_r, \hat{\mu}_x, \hat{\beta}_x, \hat{\sigma}_x, \hat{\sigma}_{rx}\}$ in place of their unknown values. In particular, we generate 5,000 random draws of time series of returns using sample sizes similar to those studied by Stam-

baugh (1999), namely $T = 200, 500, 800.$[10]

For each draw of returns and the predictor variable, we next obtain estimates of the slope coefficient, β_r, using OLS estimation as well as Bayesian estimates based on the unconstrained model (1), the model that constrains the equity premium (3) and the double-constrained model that restricts both the conditional equity premium and the sign (4).

Table 3 reports the outcome of these simulations. The true parameters for the simulation are set as:

$$
\begin{aligned}
r_{t+1} &= 0.033 + 0.008x_t + u_{t+1} \\
x_{t+1} &= -0.014 + 0.996x_t + v_{t+1} \\
var(u_{t+1}) &= 0.0018, var(v_{t+1}) = 0.0019, cov(u_{t+1}, v_{t+1}) - 0.0018
\end{aligned}
$$

The table shows the mean, skewness and kurtosis of the coefficient estimates as well as the root mean squared error (RMSE) of the one-step-ahead forecast. In common with other researchers we find that the OLS estimates are heavily biased. In the smallest sample, $T = 200$, the bias amounts to more than 100% of the true parameter value (0.8) assumed in the simulations. This bias is shared also with the unconstrained Bayesian estimates, which is perhaps unsurprising since we are using uninformative priors in the analysis. As the sample size increases to 500 observations, as expected the bias declines but is still quite substantial. Even in the largest sample with 800 observations some bias remains.

Very different results emerge for the constrained models. Imposing the constraints cuts the bias by more than half in the smallest sample, $T = 200$, and largely eliminates the bias in the two largest samples with 500 or 800 observations. Moreover, both the skewness and the kurtosis of the distribution of the coefficient estimates are significantly reduced under the constrained models.

Turning to the forecasting performance listed in the final column of Table 3, this is also significantly improved under the constrained models. In the smallest sample with 200 observations, the root mean squared forecast error gets reduced from 1.6% (using OLS) or 1.9 (using unconstrained Bayesian estimates) to 1.2 under the equity premium constraint and only 1.0 under the sign- and equity premium constrained model. Even in the largest samples with 500 or 800 observations, the constraints on the forecasting model continue to lead to a large reduction in RMSE-values both in relative and absolute terms.

---

[10] To simulate returns, we use the Cholesky decomposition

$$
\begin{aligned}
u_{t+1} &= \hat{\sigma}_r \varepsilon_{rt+1} \\
v_{t+1} &= \left(\frac{\hat{\sigma}_x^2 \hat{\sigma}_r^2 - \hat{\sigma}_{xr}^2}{\hat{\sigma}_r^2} \right)^{1/2} \varepsilon_{xt+1} + \frac{\hat{\sigma}_{xr}}{\hat{\sigma}_r} \varepsilon_{rt+1},
\end{aligned}
$$

where $\varepsilon_{rt+1}, \varepsilon_{xt+1}$ are IID and mutually uncorrelated random variables.

5.2 Forecasts Based on Valuation Ratios: Results from Longer Samples

As mentioned previously, the Stambaugh bias is a particular concern for the valuation ratios which are highly persistent with innovations that are strongly correlated with returns. For three of these, namely the dividend price ratio, the dividend yield and the earnings-price ratio we have particularly long data samples spanning the period 1871-2005:12, while for a fourth, the smoothed earnings price ratio, the data sample is 1881-2005:12. Recent results that have questioned that these variables can predict stock returns ex-ante (e.g. Goyal and Welch (2003, 2007)) have reported results going back to 1927. It is therefore of interest to consider the forecasting performance of these variables in the longer sample.

To this end, Table 4 reports the out-of-sample forecasting performance for these four predictor variables using the longest-available sample to estimate the parameters of the forecasting models. The out-of-sample forecast period begins in 1927 and the parameters are updated recursively through time using only data that was available prior to the date of the forecast. For each of the four priors under consideration we report the RMSE and out-of-sample R^2-values for the unconstrained and constrained models.

The forecasting performance of the unconstrained models is quite similar under the Normal-Gamma priors as well as under the Jeffreys prior with $\sigma_\beta = 0.2$. Under these priors the out-of-sample R^2 is negative for the forecasts based on the dividend-price ratio or the dividend yield, while it is around 0.25 and 0.08 under the earnings-price ratio and smoothed earnings-price ratio models. Forecasting performance improves for the dividend-based predictor variables under the Jeffreys prior with $\sigma_\beta = 0.02$, but it worsens for the earnings-price ratio predictor variable.

Turning to the results under the constrained forecasting models, the forecast precision improves significantly. The out-of-sample R^2-value is now always positive and it improves in 15 of 16 cases, in many cases more than doubling the R^2-value.

5.3 Parameter Estimation Error

Our results so far suggest that, for the valuation ratios, the improvement in forecasting performance under the equilibrium restriction on the return forecasting models can be understood in terms of its ability to reduce the finite-sample (Stambaugh) bias. For variables that are either less persistent or whose innovations are less correlated with returns—including managerial decision variables such as the dividend payout ratio and net equity expansion (Baker et al (2006))—this bias is less likely to be the explanation for the improved forecasting performance of the constrained models. This is also an issue for variables such as the T-bill rate or the inflation rate for which this bias is not a great concern.

To understand why the constraints work for these models, we turn to another explanation, namely the effect of parameter estimation error. It is widely

known that it is difficult to estimate the slope coefficients of the return forecasting models with much precision. As a result, the forecasting performance is likely to be improved by estimation methods that reduce parameter uncertainty. Imposing the constraint achieves this objective. As shown in Table *3*, the standard deviation—computed across the 5,000 Monte Carlo simulations—is much lower for the constrained than for the unconstrained forecasting models. Hence lower parameter estimation error is another way to understand why the constrained return forecasting models work better than the unconstrained ones.

6   Conclusion

We presented a new approach to forecasting stock returns out-of-sample that optimally incorporates information embedded in theoretical restrictions on the conditional equity premium and on the sign of the coefficient of the predictor variable. When implemented empirically, this approach was found to be highly successful at improving the precision of out-of-sample forecasts of stock returns.

References

[1] Aït-Sahalia, Y., and M., Brandt, 2001, Variable Selection for Portfolio Choice, *Journal of Finance*, 56, 1297-1351.

[2] Amihud, Y. and C. Hurvich, 2004, Predictive Regressions: A Reduced-Bias Estimation Method. Journal of Financial and Quantitative Analysis.

[3] Ang, A. and G., Bekaert, 2007, Stock Return Predictability: Is it There? *Review of Financial Studies*, 20, 651-707.

[4] Avramov, D., 2004, Stock Return Predictability and Asset Pricing Models. *Review of Financial Studies,* 17, 699-738.

[5] Avramov, D., and R. Wermers, 2006, Investing in Mutual Funds When Returns Are Predictable, Journal of Financial Economics, 81, 339-377.

[6] Barberis, N., 2000, Investing for the Long Run When Returns Are Predictable, *Journal of Finance*, 55, 225-264.

[7] Baker, M., and J. Wurgler, 2000, The Equity Share in New Issues and Aggregate Stock Returns. Journal of Finance 55, 2219-2257.

[8] Baker, M., R. Taliaferro and J. Wurgler, 2006, Predictive Regressions Based on Managerial Decision Variables: Is there a Small-sample Bias? Forthcoming in Journal of Finance.

[9] Bossaerts, P., and P., Hillion, 1999, Implementing Statistical Criteria to Select Return Forecasting Models: What Do We Learn?, *Review of Financial Studies,* 12, 405–428.

[10] Brandt, M., A., Goyal, P., Santa-Clara, and J., Stroud, 2002, A Simulation Approach to Dynamic Portfolio Choice with an Application to Learning about Predictability, *Review of Financial Studies,* forthcoming.

[11] Brennan, M., E., Schwartz, and R., Lagnado, 1997, Strategic Asset Allocation, *Journal of Economic Dynamics and Control*, 21, 1377-1403.

[12] Brennan, M., and Y., Xia, 2001, Assessing Asset Pricing Anomalies, *Review of Financial Studies,* 14, 905–942.

[13] Butler, A.W., G. Grullon and J.P. Weston, 2005, Can Managers Forecast Aggregate Stock Market Returns? Journal of Finance 60, 963-986.

[14] Campbell, J.Y., 1987, Stock Returns and the Term Structure, *Journal of Financial Economics,* 18, 373-399.

[15] Campbell, J.Y. and R., Shiller, 1988, The Dividend Price Ratio and Expectations of Future Dividends and Discount Factors, *Review of Financial Studies,* 1, 195-228.

[16] Campbell, J.Y. and S. Thompson, 2007, Predicting the Equity Premium

Out of Sample: Can Anything Beat the Historical Average? forthcoming *Review of Financial Studies.*

[17] Campbell, J.Y., and L., Viceira, 1999, Consumption and Portfolio Decisions when Expected Returns Are Time Varying, *Quarterly Journal of Economics*, 114, 433-495.

[18] Campbell, J.Y. and M. Yogo, 2006, Efficient Test of Stock Return Predictability. Journal of Financial Economics 81, 27-60.

[19] Cavanagh, C.L., G. Elliott, and J.H. Stock, 1995, Inference in Models with Nearly Integrated Regressors. Econometric Theory 11, 1131-1147.

[20] Fama, E.F., and K., French, 1988, Dividend Yields and Expected Stock Returns, *Journal of Financial Economics*, 22, 3-25.

[21] Fama, E.F., and K. French, 1989, Business Conditions and Expected Returns on Stocks and Bonds, *Journal of Financial Economics,* 29, 23-49.

[22] Fama, E.F., and G.W. Schwert, 1977, Asset Returns and Inflation. Journal of Financial Economics 5, 115-146.

[23] Ferson, W., and C., Harvey, 1991, The Variation of Economic Risk Premiums, *Journal of Political Economy*, 99, 385-415.

[24] Ferson, W., and R., Schadt, 1996, Measuring Fund Strategy and Performance in Changing Economic Conditions, *Journal of Finance,* 51, 425-462.

[25] Ferson, W., Simin, T., and S., Sarkissian, 2003, Spurious Regressions in Financial Economics? *Journal of Finance* 58, 1393-1414.

[26] Geweke, J., 1989, Bayesian Inference in Econometric Models Using Monte Carlo Integration. *Econometrica*, 57, 1317–1339.

[27] Geweke, J. and C. H. Whiteman, 2005, Bayesian Forecasting. Forthcoming in Elliott, G., C.W.J. Granger and A. Timmermann (eds.), Handbook of Economic Forecasting. North Holland.

[28] Goyal, A. and I. Welch, 2003, Predicting the equity premium with dividend ratios. Management Science 49, 639-654.

[29] Goyal, A. and I. Welch, 2007, A Comprehensive Look at The Empirical Performance of Equity Premium Prediction. *Review of Financial Studies*, forthcoming.

[30] Jagannathan, R. and T., Ma, 2003, Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps. *Journal of Finance*, 58, 1651-1684.

[31] Jansson, M. and M.J. Moreira, 2006, Optimal Inference in Regression Models with Nearly Integrated Regressors. Econometrica 74, 681-715.

[32] Kandel, S., and R., Stambaugh, 1996, On the Predictability of Stock Returns: An Asset Allocation Perspective, *Journal of Finance*, 51, 385-424.

[33] Keim, D., and R., Stambaugh, 1986, Predicting Returns in the Stock and Bond Markets, *Journal of Financial Economics,* 17, 357-390.

[34] Lettau, M. and S. Ludvigsson, 2001, Consumption, aggregate wealth, and expected stock returns. *Journal of Finance* 56, 815-849.

[35] Lewellen, J., 2004, Predicting Returns with Financial ratios. *Journal of Financial Economics* 74, 209-235.

[36] Paye, B., and A. Timmermann, 2006, Instability of Return Prediction Models. *Journal of Empirical Finance*, 13, 274-315.

[37] Merton, R., 1973, An Intertemporal Capital Asset Pricing Model. *Econometrica*, 41, 867-87.

[38] Pesaran, M., and A., Timmermann, 1995, Predictability of Stock Returns: Robustness and Economic Significance, *Journal of Finance*, 50, 1201-1228.

[39] Polk, C., S. Thompson and T. Vuolteenaho, 2006, Cross-sectional Forecasts of the Equity Premium. Journal of Financial Economics 81, 101-141.

[40] Roberts, G., and A. Smith, 1994, Simple Conditions for the Convergence of the Gibbs Sampler and Metropolis-Hastings Algorithms. *Stochastic Processes and Their Applications* 49, 207-216.

[41] Schwert, G., 2003, Anomalies and Market Efficiency. *Chapter 15 in Handbook of the Economics of Finance*, 937-972.

[42] Stambaugh, R.F., 1986, Bias in Regressions with Lagged Stochastic Regressors. University of Chicago Working Paper.

[43] Stambaugh, R.F., 1999, Predictive Regressions. Journal of Financial Economics 54, 375-421.

[44] Torous, W., R. Valkanov and S. Yan, 2004, On Predicting Stock Returns with Nearly Integrated Explanatory Variables. Journal of Business 77, 937-966.

[45] Wachter, J.A. and M. Warusawitharana, 2006, Predictable Returns and Asset Allocation: Should a Skeptical Investor Time the Market? Mimeo, University of Pennsylvania.

[46] Wachter, J.A. and M. Warusawitharana, 2007, What is the Chance that the Equity Premium Varies over Time? Mimeo, University of Pennsylvania.

[47] Xia, Y., 2001, Learning About Predictability: The Effects of Parameter Uncertainty on Dynamic Asset Allocation, *Journal of Finance*, 56, 205-246.

Table 1: Full-sample estimates of slope coefficients for the individual return forecasting models

Variables	OLS	t-statistic	Unconstrained Model				Constrained Model			
			Normal-Gamma Priors		Jeffreys Priors		Normal-Gamma Priors		Jeffreys Priors	
			$\psi=0.1$	$\psi=1$	$\sigma_\beta=0.02$	$\sigma_\beta=0.2$	$\psi=0.1$	$\psi=1$	$\sigma_\beta=0.02$	$\sigma_\beta=0.2$
Dividend Price ratio	0.008	2.35	0.0080	0.0081	0.0019	0.0078	0.0039	0.0038	0.0022	0.0042
Dividend yield	0.008	2.37	0.0081	0.0081	0.0020	0.0078	0.0039	0.0039	0.0022	0.0041
Earnings Price ratio	0.008	2.17	0.0079	0.0080	0.0019	0.0076	0.0038	0.0038	0.0023	0.0040
Smooth Earnings Price ratio	0.011	2.40	0.0105	0.0104	0.0025	0.0101	0.0047	0.0048	0.0028	0.0051
Book-to-market	0.008	1.35	0.0086	0.0082	0.0019	0.0080	0.0074	0.0075	0.0032	0.0074
T-Bill rate	-0.117	-2.42	-0.1107	-0.1168	-0.0280	-0.1117	-0.0375	-0.0371	-0.0274	-0.0402
Long term yield	-0.082	-1.61	-0.0785	-0.0813	-0.0200	-0.0783	-0.0447	-0.0457	-0.0273	-0.0464
Term spread	0.256	2.16	0.2066	0.2488	0.0619	0.2516	0.0823	0.0837	0.0626	0.0875
Default yield spread	-0.304	-0.85	-0.0984	-0.2362	-0.0767	-0.2936	-0.2442	-0.3993	-0.1707	-0.3957
Default return spread	0.091	0.65	0.0700	0.0879	0.0223	0.0918	0.0695	0.0693	0.0523	0.0687
cross-sectional premium	2.153	3.12	0.2221	1.1578	0.4793	2.0247	0.3324	0.9424	0.5280	1.3250
Long term return	0.146	2.35	0.1382	0.1444	0.0345	0.1404	0.0434	0.0435	0.0341	0.0457
Stock variance	-0.391	-0.80	-0.0735	-0.2729	-0.0962	-0.3791	-0.0514	-0.0527	-0.0513	-0.0505
Dividend Payout ratio	0.006	0.75	0.0056	0.0058	0.0013	0.0056	0.0068	0.0067	0.0036	0.0064
Net equity expansion	-0.197	-1.97	-0.1700	-0.1901	-0.0474	-0.1922	-0.1146	-0.1191	-0.0606	-0.1243
Inflation	-1.311	-3.57	-0.3957	-1.0761	-0.3218	-1.2922	-0.2094	-0.2382	-0.2065	-0.2610

Note: This table presents estimates of the slope coefficients of the univariate forecasting models using monthly stock returns in excess of a 1-month T-bill rate over the period 1940:1 – 2005:12. The first and second columns show the ordinary least squares (OLS) estimates and their t-statistics. The next four columns show the posterior means of the slope coefficients for unconstrained models estimated using Bayesian methods under different priors. The final four columns show the posterior means of the slope coefficients obtained using Bayesian methods that constrain the equity premium to be non-negative and also constrain the sign of the slope coefficient. The parameters ψ and σ_β reflect the weights on the priors: larger values indicating a smaller weight on the prior.

Table 2: Forecasting performance of individual return forecasting models

Variables	in-sample R²	out-of-sample R²									
		Unconstrained Models					Constrained Models				
			Normal-Gamma Priors		Jeffreys Priors			Normal-Gamma Priors		Jeffreys Priors	
		OLS	$\psi=0.1$	$\psi=1$	$\sigma_\beta=0.02$	$\sigma_\beta=0.2$	OLS	$\psi=0.1$	$\psi=1$	$\sigma_\beta=0.02$	$\sigma_\beta=0.2$
Dividend Price ratio	0.569	0.208	0.234	0.104	0.200	0.258	0.424	0.972	0.960	0.3685	0.8834
Dividend yield	0.582	0.227	0.250	0.021	0.202	0.275	0.427	1.004	1.009	0.3796	0.9166
Earnings Price ratio	0.466	0.252	0.243	-0.005	0.135	0.275	0.197	0.454	0.435	0.278	0.4642
Smooth Earnings Price ratio	0.602	-0.724	-0.519	-1.325	0.231	-0.499	-0.108	0.689	0.702	0.3747	0.6484
Book-to-market	0.106	-1.418	-1.326	-1.438	0.018	-1.179	-1.231	-0.116	-0.115	0.1027	-0.2007
T-Bill rate	0.610	-0.089	0.315	0.295	0.363	0.099	0.675	0.863	0.465	0.5512	0.7156
Long term yield	0.202	-1.289	-0.142	-0.876	0.202	-0.965	0.557	0.777	0.498	0.3713	0.3482
Term spread	0.461	0.109	0.283	0.041	0.308	0.200	0.262	0.357	0.151	0.3332	0.3253
Default yield spread	0.092	-0.183	-0.023	-0.128	-0.028	-0.158	-0.088	0.030	-0.262	-0.004	-0.1224
Default return spread	0.053	-0.544	-0.142	-0.336	-0.046	-0.502	-0.279	-0.440	-0.664	-0.0981	-0.4041
cross-sectional premium	1.142	-0.028	0.131	0.585	0.322	0.231	0.491	0.398	0.816	0.4618	1.0801
Long term return	0.570	0.036	0.298	-0.273	0.325	0.212	0.196	0.393	0.381	0.4677	0.5182
Stock variance	0.081	-3.175	-0.108	-1.329	-0.418	-2.960	-0.205	-0.253	-1.160	-0.6857	-3.1138
Dividend Payout ratio	0.071	-0.293	-0.283	-0.357	-0.014	-0.275	-0.128	-0.158	-0.150	0.0584	-0.0068
Net equity expansion	0.365	0.032	0.193	-0.037	0.153	0.072	0.068	0.179	0.124	0.2541	0.229
Inflation	1.468	1.140	0.385	1.259	0.282	1.125	1.223	0.192	0.024	0.2674	0.2397

Note: This table presents forecasting results for monthly stock returns using univariate forecasting models. Column 1 shows in-sample R2-values while the other columns show out-of-sample R2-values computed over the period 1950:01 - 2005:12. The data goes back to 1940:01 and the first 10 years of observations are used to obtain initial estimates of the parameters. Subsequently an expanding window is used to estimate the models recursively over time. Positive R^2-values show that a forecasting model produces more precise forecasts (lower root mean squared forecast errors) than the prevailing mean model. Columns 2-6 show results for the unconstrained forecasting models using either OLS or Bayesian estimation methods. Columns 7-11 show results using either truncated OLS or Bayesian estimation methods that constrain the equity premium to be non-negative and also constrain the sign of the slope coefficient

Table 3: Estimates of slope coefficient and root mean squared error performance for the dividend yield forecasting model with and without constraints.

Panel A: T=200	σ_β=0.2					σ_β=2				
	mean	s.d.	skewness	kurtosis	rmse	mean	s.d.	skewness	kurtosis	rmse
OLS	2.092	1.659	1.979	9.601	4.303	2.092	1.659	1.979	9.601	4.303
Unconstrained	1.839	1.838	1.676	8.608	4.333	2.093	2.005	1.612	7.457	4.333
Equity Premium Constrained	1.281	1.181	0.646	4.410	4.309	1.348	1.189	0.512	4.426	4.305
Equity Premium and Sign Constrained	1.461	0.989	1.233	5.759	4.300	1.538	1.043	1.133	5.050	4.297

Panel B: T=500	σ_β=0.2					σ_β=2				
	mean	s.d.	skewness	kurtosis	rmse	mean	s.d.	skewness	kurtosis	rmse
OLS	1.226	0.598	1.785	8.967	4.262	1.226	0.598	1.785	8.967	4.262
Unconstrained	1.176	0.724	1.616	9.271	4.271	1.230	0.755	1.564	8.225	4.275
Equity Premium Constrained	0.852	0.458	0.616	4.402	4.264	0.875	0.464	0.668	4.691	4.266
Equity Premium and Sign Constrained	0.879	0.440	0.945	4.976	4.262	0.893	0.445	0.868	4.626	4.264

Panel C: T=800	σ_β=0.2					σ_β=2				
	mean	s.d.	skewness	kurtosis	rmse	mean	s.d.	skewness	kurtosis	rmse
OLS	1.071	0.394	1.619	7.625	4.177	1.071	0.394	1.619	7.625	4.177
Unconstrained	1.043	0.499	1.360	6.864	4.185	1.071	0.506	1.378	7.114	4.180
Equity Premium Constrained	0.725	0.303	0.612	4.334	4.183	0.734	0.302	0.720	4.668	4.178
Equity Premium and Sign Constrained	0.726	0.301	0.796	4.263	4.179	0.736	0.298	0.831	4.506	4.184

Note: This table shows the result of 5,000 Monte Carlo simulations for the dividend yield forecasting model with parameters set to match the full-sample estimates for US stock returns: $r_t = \mu + \beta x_{t-1} + \varepsilon_t$, $x_t = \mu + \gamma x_{t-1} + u_t$. The (posterior) mean of the slope coefficient on the dividend yield in the return equation is shown in the first column followed by the skew and kurtosis of this coefficient. The fourth column shows the root mean squared forecast error based on forecasts of returns for the following period. In all cases an uninformative prior is used with psi = 0.1.
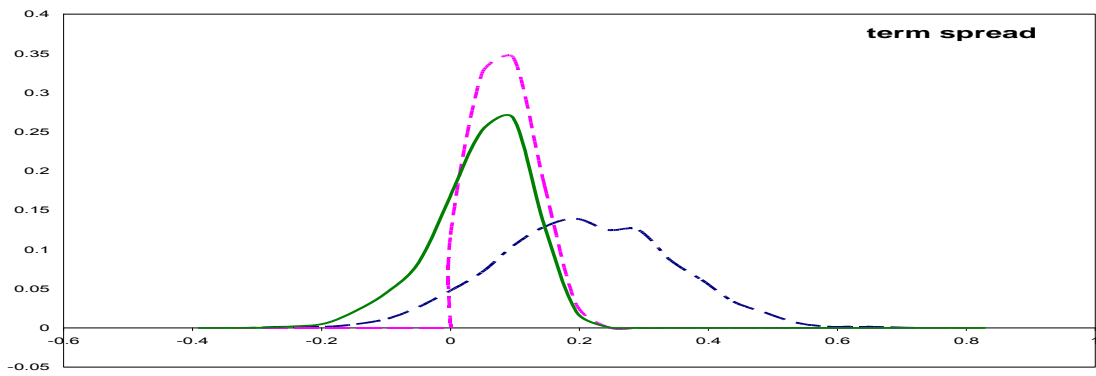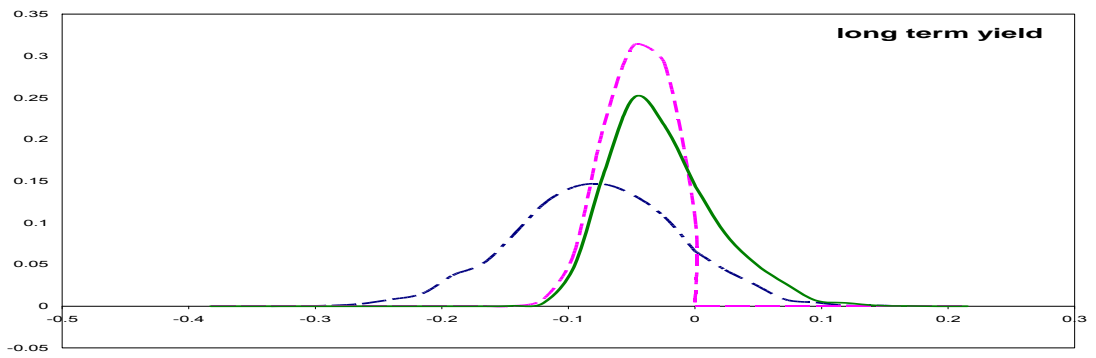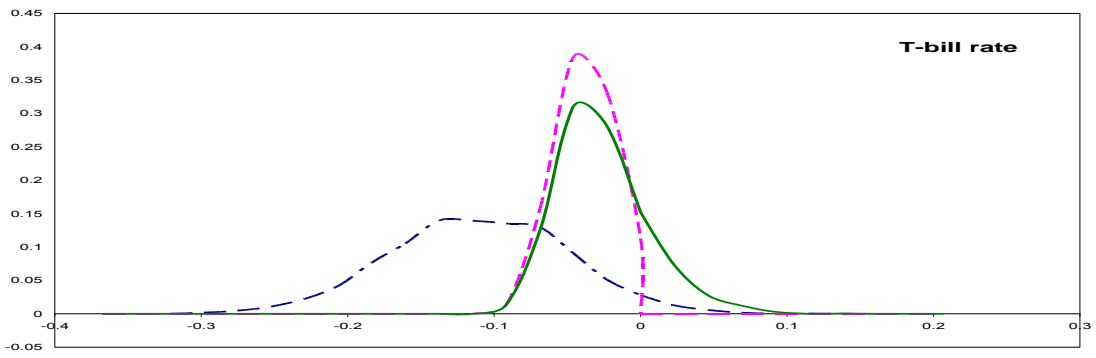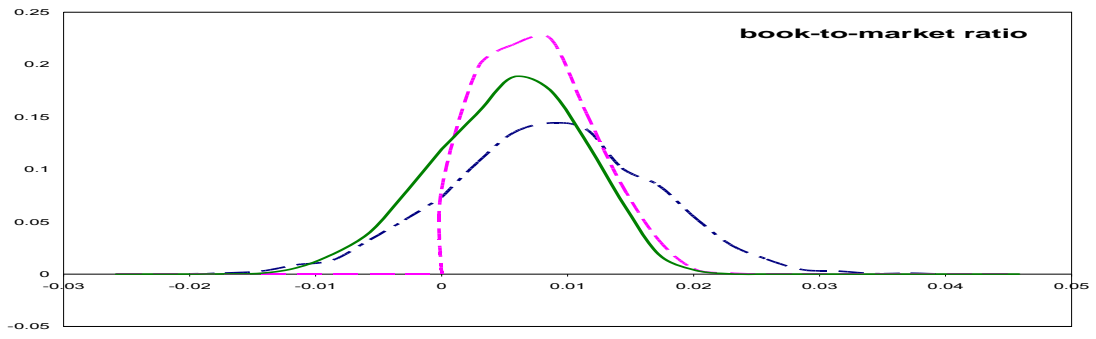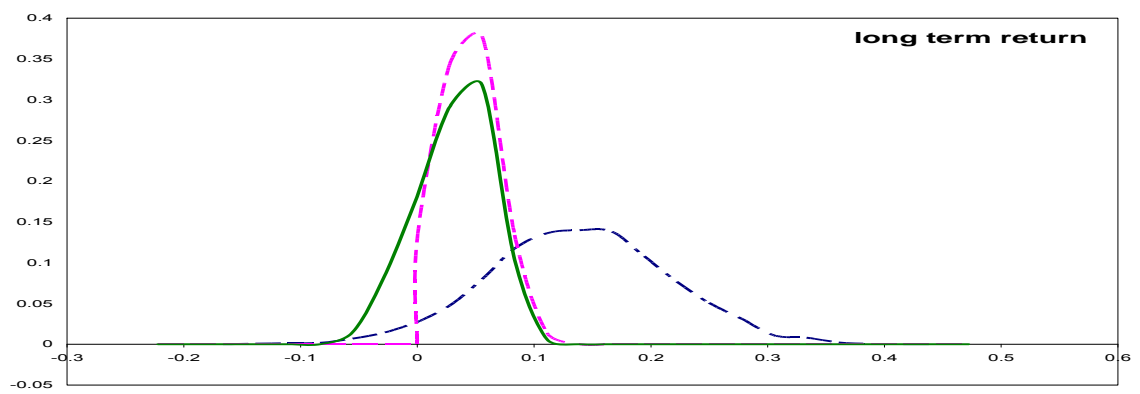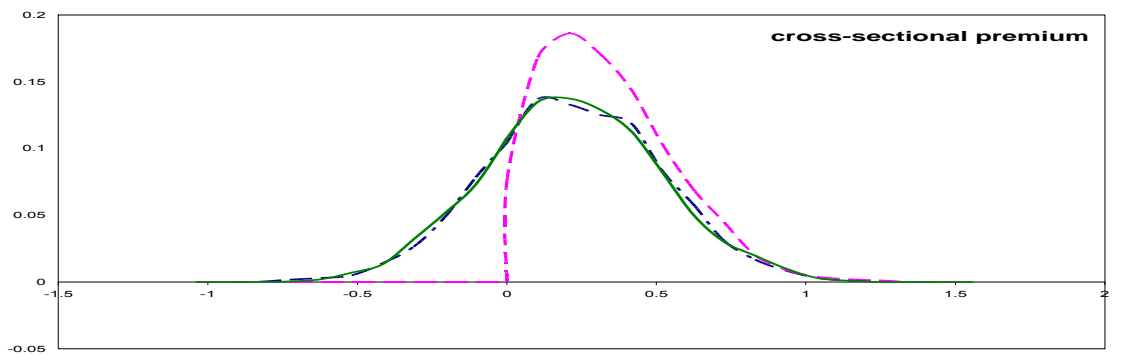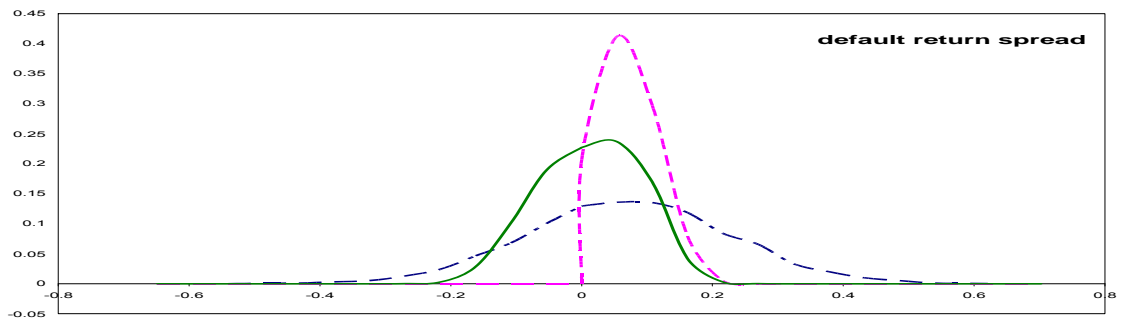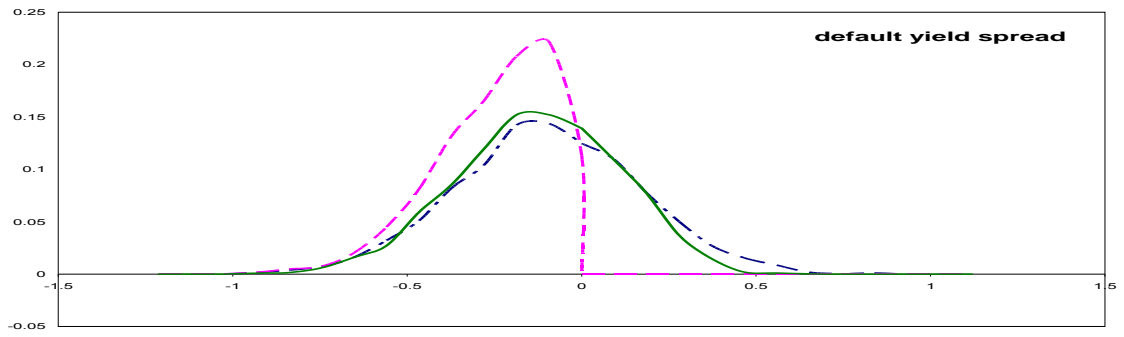
Table 4: Out-of-sample R^2 of individual return forecasting models

Panel A: Unconstrained Models	sample period	Normal-Gamma Priors		Jeffreys Priors	
		$\psi=0.1$	$\psi=1$	$\sigma_\beta=0.02$	$\sigma_\beta=0.2$
Dividend Price ratio	1871.2--2005.12	-0.332	-0.331	-0.036	-0.323
Dividend yield	1871.2--2005.12	-0.256	-0.242	0.048	-0.246
Earnings Price ratio	1871.2--2005.12	0.260	0.248	0.151	0.254
Smooth Earnings Price ratio	1881.1--2005.12	0.065	0.082	0.169	0.086
Panel B: Constrained Models (I)					
Dividend Price ratio	1871.2--2005.12	0.034	0.138	0.091	0.067
Dividend yield	1871.2--2005.12	0.031	0.176	0.142	0.023
Earnings Price ratio	1871.2--2005.12	0.031	0.166	0.195	0.300
Smooth Earnings Price ratio	1881.1--2005.12	0.049	0.178	0.241	-0.601
Panel C: Constrained Models (II)					
Dividend Price ratio	1871.2--2005.12	0.015	0.130	0.185	0.181
Dividend yield	1871.2--2005.12	0.062	0.242	0.350	0.349
Earnings Price ratio	1871.2--2005.12	0.070	0.273	0.364	0.371
Smooth Earnings Price ratio	1881.1--2005.12	0.065	0.267	0.191	0.189

Note: This table presents forecasting results for monthly stock returns using univariate forecasting models. Column 1 shows sample periods for each individual model. Columns 2-5 presents the forecasting performance in terms of out-of-sample R^2 using Bayesian estimation methods under different prior specifications. Panel A contains results of models without any constraint. Panel B contains results of models with non-negative return premium constraint. Panel C contains results of models with both non-negative return premium constraint and sign constraint on the slope of forecasting equation.

Figure1: Posterior distribution of slope coefficients in return prediction models

default yield spread

default return spread

cross-sectional premium

long term return

stock variance

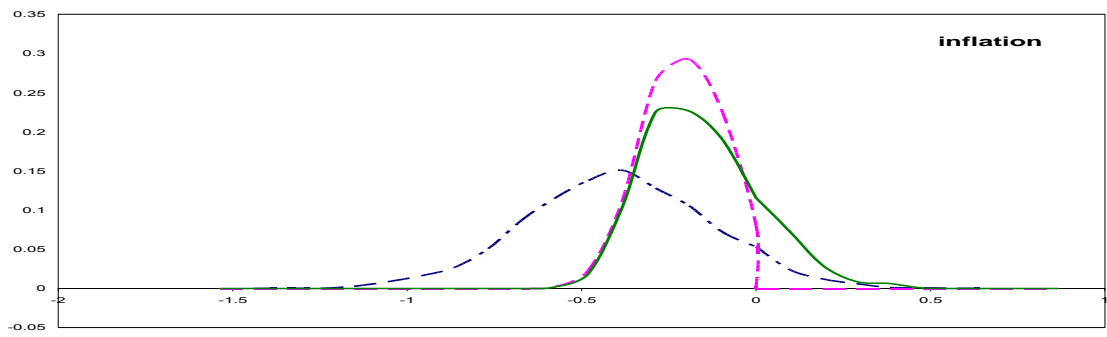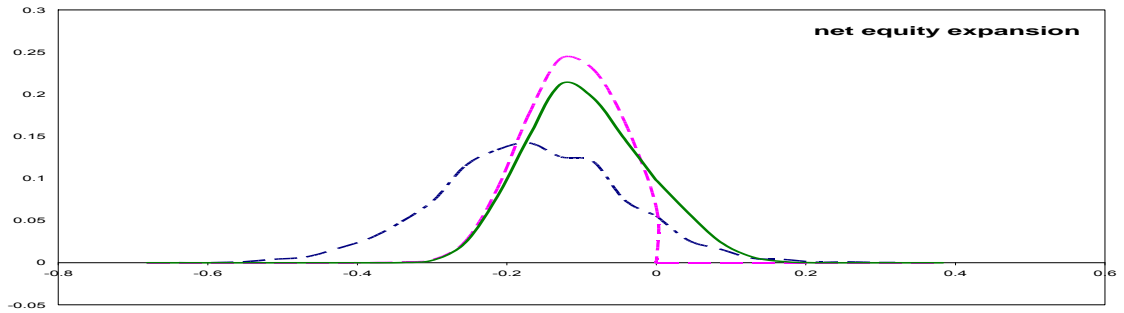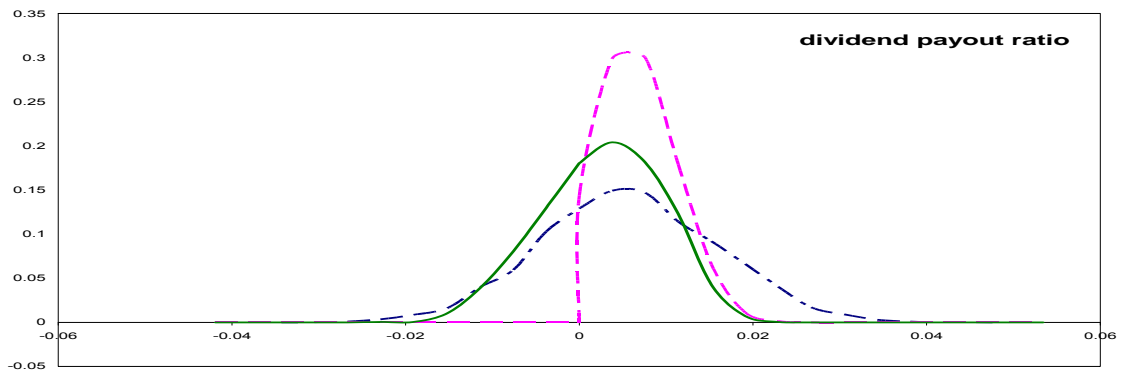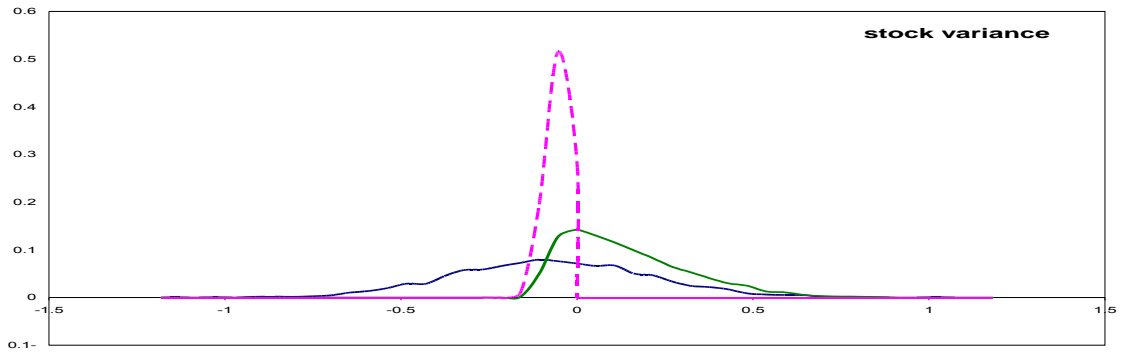dividend payout ratio

net equity expansion

inflation
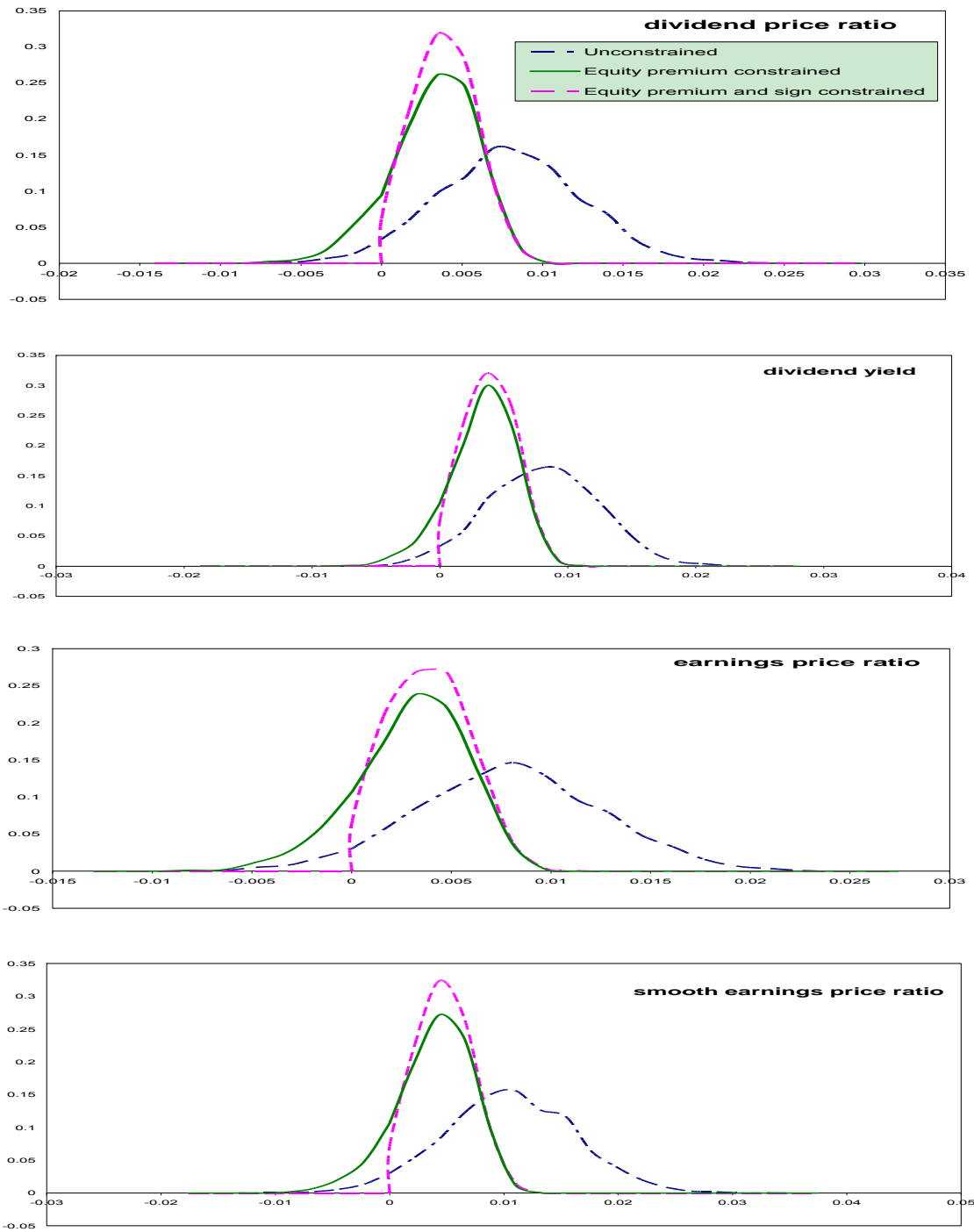
Figure 2: Out-of-sample forecasts of excess returns under unconstrained and constrained return prediction models (with uninformative priors and $\psi=0.1$)
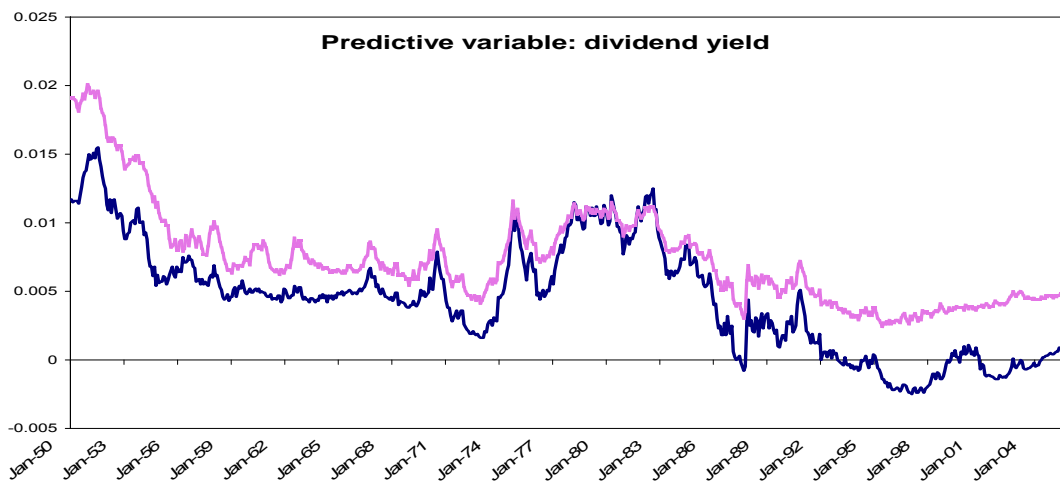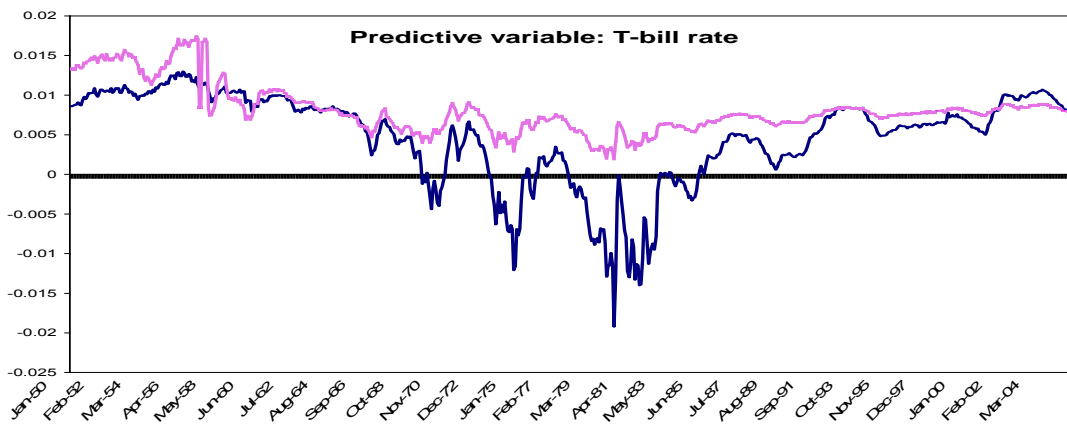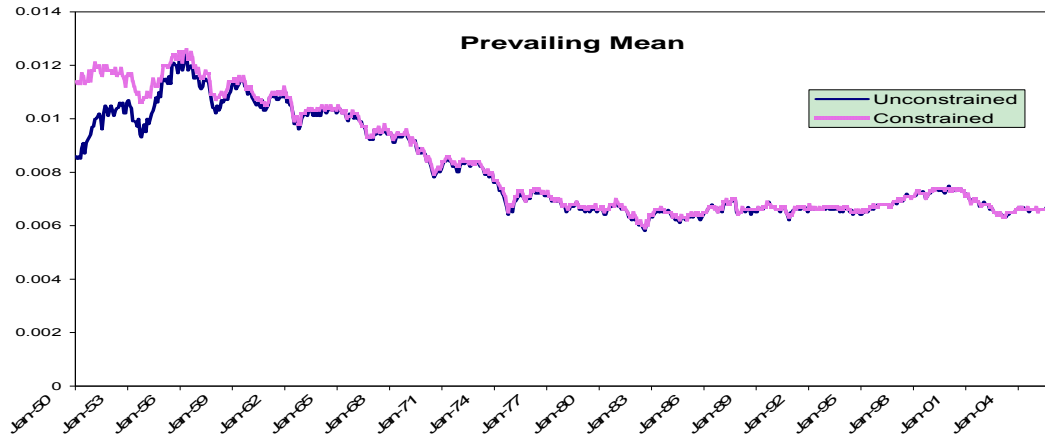
Figure 3.1: suppose x>0, shaded area stands for μ+βx >= 0

$$\text{Slope} = -\frac{1}{\min(x)}$$

$$\text{Slope} = -\frac{1}{\max(x)}$$
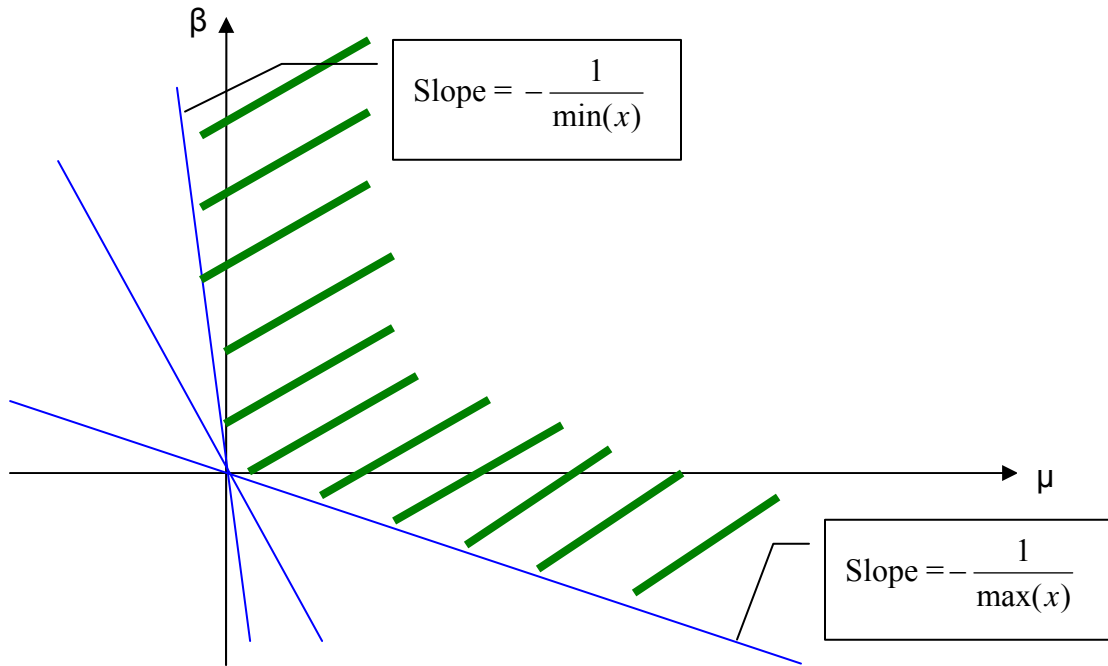
Figure 3.2: suppose x<0, shaded area stands for μ+βx >= 0

$$\text{Slope} = -\frac{1}{\max(x)}$$

$$\text{Slope} = -\frac{1}{\min(x)}$$