

Bayesian Strategy for Improved Forecasts of the Equity Premium

Siddhartha Chib*

Xiaming Zeng[†]

May 24, 2016

Abstract

In this paper we develop an easily implemented Bayesian strategy for improved forecasts of the market excess returns. This strategy relies on 1) the use of multiple predictors in the predictive regression; 2) zero lower bound constraints on the Bayesian predictive mean (the mean marginalized over the parameters); and 3) conjugate prior distributions for smooth sequential updating and calculation of needed posterior quantities. This strategy produces striking results in the prediction of the market excess returns and outperforms the results from 8 other predictive methods, such as univariate and model combined univariate predictive regressions. We show that with a set of 11 common market return predictors used in the literature, and for an investor with power utility, the utility gain from our strategy exceeds that from the other predictive methods by a wide margin.

JEL classification: C11, C22, C53, G11, G12.

Keywords: Return prediction; Non-negative equity premium; Multiple regression; Bayesian analysis

*Corresponding author. Olin Business School, Washington University in St. Louis, Campus Box 1133, 1 Brookings Drive, St. Louis, MO 63130. e-mail: chib@wustl.edu.

[†]Olin Business School, Washington University in St. Louis, Campus Box 1133, 1 Brookings Drive, St. Louis, MO 63130. zengxiaming@wustl.edu.

1 Introduction

The question - how should the equity premium be forecast? - has spawned a large literature in finance that has revealed the relevance of different predictors, the pros and cons of working with univariate and multiple predictive regressions, and the value of different forecasting techniques (such as model combined forecasting methods, frequentist and Bayesian approaches). Despite much progress, few definitive strategies stand out. For instance, in a wide-ranging study, Goyal and Welch (2008) show that, in out-of-sample tests, univariate predictive regressions with many common predictors fail to outperform a simple benchmark model where historical average excess returns are used as forecasts. Goyal and Welch (2008) also show that a multiple predictive regression with a set of commonly used predictors does not outperform either the univariate predictive regressions or the simple benchmark model in out-of-sample comparisons. Rapach, Strauss, and Zhou (2010) show that a weighted average of point forecasts from individual univariate predictive regressions is less volatile than point forecasts from a multiple predictive regression and performs better out-of-sample, while Ludvigson and Ng (2007), Neely, Rapach, Tu, and Zhou (2014), and Kelly and Pruitt (2013, 2015) discuss the performance of predictive regressions with constructed factors, where the factors are constructed by principal component methods, or other means.

In this paper, based on the insights gleaned from the preceding literature, we provide a new, easily implemented, Bayesian strategy that produces strikingly improved forecasts of market excess returns. This strategy relies on 1) the use of multiple predictors in the predictive regression; 2) zero lower bound constraints on the Bayesian predictive mean (the mean marginalized over the parameters); and 3) conjugate prior distributions for smooth sequential updating and calculation of needed posterior quantities.

The motivation for enforcing the lower bound constraint comes from Merton (1980) where it has been argued that *“in estimating models of the expected market return, the non-negativity restriction of the expected excess return should be explicitly included as part of the specification.”* In the empirical literature, this point of view has been implemented in the frequentist context by Campbell and Thompson (2008) and Li and Tsiakas (2015), where a negative equity premium forecast is truncated from below at zero, and in the Bayesian context by Pettenuzzo, Timmermann, and Valkanov (2014),

where the expected excess market return, *conditioned on parameters*, is constrained to be non-negative for each time period in the sample. However, the demand that the lower bound has to hold for each value of the parameters is very strong and, as a result, the multiple predictive regression of market excess returns under such constraints is difficult to estimate.

In our strategy we impose a weaker lower bound restriction, that the Bayesian predictive mean (the mean marginalized over the parameters) is non-negative. Because these restrictions are weaker than restricting the predictive mean conditioned on the parameters, a multiple predictive regression for market excess returns become estimable. In effect, under our constraints, the data can speak more loudly in the sense that the constraints do not interfere with the prior-posterior updates if the constraints are intrinsically satisfied by the data. This is not true of the approach in Pettenuzzo et al. (2014) where the posterior distribution is truncated by the constraints *even if* the predictive mean is non-negative.

We apply our approach in an empirical study where the constrained multiple predictive regression model is specified with 11 common predictors, the log dividend-price ratio, log earnings-price ratio, stock return volatility, book-to-market ratio, net equity expansion, Treasury bill rate, long-term yield, long-term return, term spread, default yield spread, default return spread, and inflation. The outcome variable is the market excess return computed as the log returns on the S&P 500 index (including dividends) minus the Treasury bill rate. The data frequency is monthly. We consider two samples of data. The first sample, which we refer to as Sample 1, spans the period from January 1927 to December 2014, while the second sample, Sample 2, which is taken from Rapach, Ringgenberg, and Zhou (2015), spans the period January 1973 to December 2014. We use an initial portion of each sample (the training sample) to construct an objective prior distribution of the parameters in the model. The model is then estimated sequentially for another (middle) portion of the sample. Finally, forecasts are generated sequentially for the remaining portion of the sample (the forecast sample). For instance, in the case of Sample 2, the training sample period runs from January 1973 to December 1975, the middle sample runs from January 1976 to December 1989 and the forecast sample consists of the 300 month time span from January 1990 to December 2014.

Our results show that the predictive likelihood (a measure of predictive performance that takes account of the whole predictive distribution and not just its mean and variance) from our constrained multiple predictive regression strategy is higher than that of other alternative predictive methods, including univariate and model combined predictive regressions, and regressions in which the non-negativity constraints are imposed conditioned on the parameters. In addition, for an investor with power utility, in both samples, the utility gain from our model and approach, relative to the historical average, exceeds the gain from that of the various alternative techniques by a wide margin.

The remainder of the paper proceeds as follows. Section 2 develops our Bayesian estimation approach for excess market return prediction with non-negativity constraints on the predictive mean of excess returns. Section 3 describes the data we use in the empirical study and presents the results. Section 4 concludes the paper. Finally, derivations of results in the text are collected in the Appendix.

2 Market excess return prediction

2.1 The unconstrained case

We consider a predictive regression model of the market excess return that is given by

$$r_{\tau+1} = \mathbf{z}'_{\tau}\boldsymbol{\beta} + \sigma\varepsilon_{\tau+1}, \quad \tau = 1, 2, \dots, \quad (2.1)$$

where $r_{\tau+1}$ is the market excess return at the end of the time interval $(\tau, \tau + 1]$; \mathbf{z}_{τ} is a $k > 1$ dimensional vector that contains predictors that are available at time period τ ; $\varepsilon_{\tau+1}$ is a standard normal error term distributed identically and independently across time, and $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma > 0)$ are the parameters which are unknown.

Now suppose that we are given data $D_t = \{r_{1:t}, \mathbf{z}_{1:t}\}$, where $y_{1:t} = (y_1, \dots, y_t)$, and the goal is to do inferences about r_{t+1} . Clearly, conditioned on $\boldsymbol{\theta}$, the density of r_{t+1} given the past data is simply

$$p(r_{t+1}|D_t, \boldsymbol{\theta}) = \mathcal{N}(r_{t+1}|\mathbf{z}'_t\boldsymbol{\beta}, \sigma^2) \quad (2.2)$$

so that

$$E(r_{t+1}|D_t, \boldsymbol{\theta}) = \mathbf{z}'_t \boldsymbol{\beta} \quad (2.3)$$

is the mean conditioned on $\boldsymbol{\theta}$. In the Bayesian context, predictive inferences are based on the predictive density, which is the density of r_{t+1} given the parameters marginalized over the posterior distribution of the parameters. This marginalization ensures that the uncertainty surrounding $\boldsymbol{\theta}$ is properly accounted for in the predictions. Otherwise, one would understate the risks perceived by investors, as discussed in, for example, Kandel and Stambaugh (1996) and Barberis (2000). In particular, the Bayesian prediction density of r_{t+1} is given by

$$p(r_{t+1}|D_t) = \int_{\boldsymbol{\theta}} \mathcal{N}(r_{t+1}|\mathbf{z}'_t \boldsymbol{\beta}, \sigma^2) \pi(\boldsymbol{\beta}, \sigma|D_t) d\boldsymbol{\beta} d\sigma, \quad (2.4)$$

where the posterior density by the sequential form of Bayes theorem is

$$\pi(\boldsymbol{\beta}, \sigma|D_t) \propto \mathcal{N}(r_t|\mathbf{z}'_{t-1} \boldsymbol{\beta}, \sigma^2) \pi(\boldsymbol{\beta}, \sigma|D_{t-1}), \quad (2.5)$$

$\pi(\boldsymbol{\beta}, \sigma|D_{t-1})$, the posterior distribution given D_{t-1} , is the prior distribution for period t , and the predictive mean is

$$E(r_{t+1}|D_t) = \mathbf{z}'_t E(\boldsymbol{\beta}|D_t) \quad (2.6)$$

For convenience, we set up our prior distribution of $\boldsymbol{\theta}$ so that quantities in Eq. (2.5) and Eq. (2.6) can be calculated analytically. In particular, suppose that $\pi(\boldsymbol{\beta}, \sigma|D_{t-1})$ is a normal-inverse-gamma type II distribution with parameters $(\boldsymbol{\beta}_{0,t-1}, \mathbf{B}_{0,t-1}, \nu_{0,t-1}, \delta_{0,t-1})$ and density

$$\mathcal{NIG}_{k,2}(\boldsymbol{\beta}, \sigma|\boldsymbol{\beta}_{0,t-1}, \mathbf{B}_{0,t-1}, \nu_{0,t-1}, \delta_{0,t-1}) = \mathcal{N}_k(\boldsymbol{\beta}|\boldsymbol{\beta}_{0,t-1}, \sigma^2 \mathbf{B}_{0,t-1}) \mathcal{IG}_2\left(\sigma \middle| \frac{\nu_{0,t-1}}{2}, \frac{\delta_{0,t-1}}{2}\right), \quad (2.7)$$

where

$$\mathcal{IG}_2(\sigma|\nu, \delta) = \frac{2\delta^\nu}{\Gamma(\nu)} \sigma^{-2\nu-1} \exp(-\delta\sigma^{-2}), \sigma > 0 \quad (2.8)$$

We refer to this distribution as the normal-inverse-gamma type II distribution because it arises from a standard normal-inverse-gamma distribution on $(\boldsymbol{\beta}, \sigma^2)$ by transformation

to $(\boldsymbol{\beta}, \sigma)$.

By conjugacy, the posterior distribution $\pi(\boldsymbol{\beta}, \sigma | D_t)$ is also a normal-inverse-gamma type II distribution $\mathcal{NIG}_{k,2}(\boldsymbol{\beta}, \sigma | \boldsymbol{\beta}_{0,t}, \mathbf{B}_{0,t}, \nu_{0,t}, \delta_{0,t})$ with the updated parameters given by

$$\nu_{0,t} = \nu_{0,t-1} + 1 \quad (2.9)$$

$$\mathbf{B}_{0,t} = (\mathbf{B}_{0,t-1}^{-1} + \mathbf{z}_{t-1}\mathbf{z}'_{t-1})^{-1} \quad (2.10)$$

$$\boldsymbol{\beta}_{0,t} = \mathbf{B}_{0,t}(\mathbf{B}_{0,t-1}^{-1}\boldsymbol{\beta}_{0,t-1} + \mathbf{z}_{t-1}r_t) \quad (2.11)$$

$$\delta_{0,t} = \delta_{0,t-1} + r_t^2 + \boldsymbol{\beta}'_{0,t-1}\mathbf{B}_{0,t-1}^{-1}\boldsymbol{\beta}_{0,t-1} - \boldsymbol{\beta}'_{0,t}\mathbf{B}_{0,t}^{-1}\boldsymbol{\beta}_{0,t} \quad (2.12)$$

as derived in the Appendix. Then, a direct calculation performed in the Appendix shows that the unrestricted predictive density

$$p(r_{t+1}|D_t) = \int_{\boldsymbol{\theta}} \mathcal{N}(r_{t+1} | \mathbf{z}'_t \boldsymbol{\beta}, \sigma^2) \pi(\boldsymbol{\beta}, \sigma | D_t) d\boldsymbol{\beta} d\sigma \quad (2.13)$$

is

$$p(r_{t+1}|D_t) \propto \left(1 + \frac{1}{\nu_{0,t}} \frac{(r_{t+1} - \mathbf{z}'_t \boldsymbol{\beta}_{0,t})^2}{\frac{\delta_{0,t}}{\nu_{0,t}} (1 + \mathbf{z}'_t \mathbf{B}_{0,t} \mathbf{z}_t)} \right)^{-\frac{\nu_{0,t}+1}{2}} \quad (2.14)$$

a student- t distribution with $\nu_{0,t}$ degrees of freedom and mean

$$E(r_{t+1}|D_t) = \mathbf{z}'_t \boldsymbol{\beta}_{0,t}, \quad (2.15)$$

which is, of course, different than $\mathbf{z}'_t \boldsymbol{\beta}$, the predictive mean conditional on $\boldsymbol{\theta}$.

2.2 Non-negativity constraints

As mentioned in the Introduction, Pettenuzzo et al. (2014) show how the non-negativity constraint on $E(r_{t+1}|D_t, \boldsymbol{\theta}) = \mathbf{z}'_t \boldsymbol{\beta}$ can be imposed. In fact, Pettenuzzo et al. (2014) have argued that one should impose the latter constraint for each time point τ , $\tau = 1, 2, \dots, t$, in the sample. In other words, they impose the following constraints

$$E(r_{\tau+1}|D_\tau, \boldsymbol{\theta}) = \mathbf{z}'_\tau \boldsymbol{\beta} \geq 0, \tau = 1, \dots, t, \quad (2.16)$$

which leads to a posterior distribution proportional to

$$\pi(\boldsymbol{\beta}, \sigma | D_t) \prod_{\tau=1}^t I[\mathbf{z}'_{\tau} \boldsymbol{\beta} \geq 0]$$

where $I[\cdot]$ is the indicator function. Thus, the restrictions truncate the support of the unrestricted posterior distribution $\pi(\boldsymbol{\beta}, \sigma | D_t)$ at time t . As mentioned above, producing a sample of draws from this truncated posterior raises several challenges. Many draws from the untruncated posterior distribution $\pi(\boldsymbol{\beta}, \sigma | D_t)$ do not satisfy these constraints and must be rejected. This problem gets more acute as t increases. In addition, the number of rejections generally also increase rapidly in models with more than one predictor variable thus making it almost impossible to estimate multiple predictive regressions.

Instead, we pursue a different formulation of the constraint problem and impose the (weaker) restrictions

$$E(r_{\tau+1} | D_{\tau}) = \mathbf{z}'_{\tau} \boldsymbol{\beta}_{0,\tau} \geq 0, \tau = 1, \dots, t \quad (2.17)$$

These restrictions are less restrictive than the constraints in (2.16). To see this, note that by the law of iterated expectation, $E(r_{\tau+1} | D_{\tau}) = E[\mathbf{z}'_{\tau} \boldsymbol{\beta}]$ where the expectation is over the distribution $\pi(\boldsymbol{\beta}, \sigma | D_t) \prod_{\tau=1}^t I[\mathbf{z}'_{\tau} \boldsymbol{\beta} \geq 0]$. Thus, if $\mathbf{z}'_{\tau} \boldsymbol{\beta}$ is non-negative, then $E(r_{\tau+1} | D_{\tau})$ is also non-negative. On the other hand, if $\mathbf{z}'_{\tau} \boldsymbol{\beta}_{0,\tau}$ is non-negative, then $\mathbf{z}'_{\tau} \boldsymbol{\beta}$ is not necessarily non-negative.

The restrictions in (2.17) require that each of the posterior distributions

$$\pi(\boldsymbol{\beta}, \sigma | D_{\tau}), \tau = 1, 2, \dots, t \quad (2.18)$$

be such that the implied posterior means

$$\boldsymbol{\beta}_{0,\tau}, \tau = 1, 2, \dots, t \quad (2.19)$$

satisfy the sequence of constraints $\mathbf{z}'_{\tau} \boldsymbol{\beta}_{0,\tau} \geq 0, \tau = 1, 2, \dots, t$. We have developed a rather elegant solution to incorporating these constraints. The method is straightforward to implement, even in the context of the multiple predictive regression, and produces rather striking results in the out-of-sample prediction of market excess returns.

2.3 Solution to problem

Our idea for solving the question just described is to adjust the posterior distribution $\pi(\boldsymbol{\beta}, \sigma | D_\tau)$ such that under the adjusted (equivalently, corrected) posterior distribution the implied posterior mean $\boldsymbol{\beta}_{0,\tau}$ satisfies the constraint $\mathbf{z}'_\tau \boldsymbol{\beta}_{0,\tau} \geq 0$. The adjustment is carried out in such a way that the corrected posterior distribution is as close as possible in the Kullback-Leibler (K-L) distance (Kullback and Leibler, 1951) to the unrestricted one. The rationale behind minimizing this distance is that information from the constraint should already be in the unrestricted posterior and, therefore, the corrected posterior should be the one that is closest to the unrestricted posterior as measured by the K-L distance. Robertson, Tallman, and Whiteman (2005) use a similar (but different) idea and modify the weights of draws from the predictive distribution in order that the predictive distribution satisfies a set of equality constraints. In our strategy, however, the posterior distribution is adjusted, rather than the predictive distribution. This is not only computationally more straightforward but, because the adjusted posterior is the prior for the next period, information about the lower bound constraint from the current period is carried forward to future periods, thus leading to a solution that could not be reproduced by adjusting the predictive distribution directly.

At the start of time t , let $\pi_c(\boldsymbol{\beta}, \sigma | D_{t-1})$ be the current normal-inverse-gamma type II distributed posterior distribution given data D_{t-1} as well as the non-negativity constraints that have been imposed up till time $t - 1$. Let $\pi_u(\boldsymbol{\beta}, \sigma | D_t)$ denote the unrestricted posterior distribution that emerges after we apply Bayes theorem on seeing the data (\mathbf{z}_{t-1}, r_t) and using the updates from Eq. (2.10) to (2.12). Let the implied posterior mean of this distribution be $\boldsymbol{\beta}_{0u,t}$. This may not satisfy the restriction $\mathbf{z}'_t \boldsymbol{\beta}_{0u,t} \geq 0$. To incorporate this constraint, we find an adjusted posterior distribution $\pi_c(\boldsymbol{\beta}, \sigma | D_t)$ with mean $\boldsymbol{\beta}_{0c,t}$ that satisfies the constraint $\mathbf{z}'_t \boldsymbol{\beta}_{0c,t} \geq 0$. Because the non-negativity constraints are on the implied posterior mean of $\boldsymbol{\beta}$, the new posterior distribution $\pi_c(\boldsymbol{\beta}, \sigma | D_t)$ is assumed to be only different from $\pi_u(\boldsymbol{\beta}, \sigma | D_t)$ in the mean of $\boldsymbol{\beta}$. If $\pi_u(\boldsymbol{\beta}, \sigma | D_t)$ is $\mathcal{NIG}_{k,2}(\boldsymbol{\beta}, \sigma | \boldsymbol{\beta}_{0u,t}, \mathbf{B}_{0,t}, \nu_{0,t}, \delta_{0,t})$ then $\pi_c(\boldsymbol{\beta}, \sigma | D_t)$ is supposed to be $\mathcal{NIG}_{k,2}(\boldsymbol{\beta}, \sigma | \boldsymbol{\beta}_{0c,t}, \mathbf{B}_{0,t}, \nu_{0,t}, \delta_{0,t})$, differing only in the mean of $\boldsymbol{\beta}$.

In the Appendix we show that the K-L divergence between the latter two distributions

has a simple analytical form,

$$\mathbb{KL}[\pi_c(\boldsymbol{\beta}, \sigma|D_t)||\pi_u(\boldsymbol{\beta}, \sigma|D_t)] = \frac{\nu_{0,t}}{2\delta_{0,t}}(\boldsymbol{\beta}_{0c,t} - \boldsymbol{\beta}_{0u,t})' \mathbf{B}_{0,t}^{-1}(\boldsymbol{\beta}_{0c,t} - \boldsymbol{\beta}_{0u,t}) \quad (2.20)$$

Thus, the problem of minimizing the latter distance, subject to meeting the required non-negativity constraint, reduces to the following optimization problem,

$$\begin{aligned} \min_{\boldsymbol{\beta}_{0c,t}} \quad & \frac{\nu_{0,t}}{2\delta_{0,t}}(\boldsymbol{\beta}_{0c,t} - \boldsymbol{\beta}_{0u,t})' \mathbf{B}_{0,t}^{-1}(\boldsymbol{\beta}_{0c,t} - \boldsymbol{\beta}_{0u,t}) \\ \text{s.t.} \quad & \mathbf{z}'_t \boldsymbol{\beta}_{0c,t} \geq 0 \end{aligned} \quad (\text{P})$$

where the objective function is the K-L divergence between the corrected and unrestricted posterior distributions and the constraint is the non-negativity constraint on the predictive mean of r_{t+1} under $\pi_c(\boldsymbol{\beta}, \sigma|D_t)$.

The optimization problem (P) is a simple quadratic problem with a linear constraint and has a closed-form solution.

Proposition 1 *The solution to problem (P) is*

$$\boldsymbol{\beta}_{0c,t} = \begin{cases} \boldsymbol{\beta}_{0u,t} & \text{if } \mathbf{z}'_t \boldsymbol{\beta}_{0u,t} \geq 0, \\ \boldsymbol{\beta}_{0u,t} - \frac{\mathbf{z}'_t \boldsymbol{\beta}_{0u,t}}{\mathbf{z}'_t \mathbf{B}_{0,t} \mathbf{z}_t} \mathbf{B}_{0,t} \mathbf{z}_t & \text{otherwise} \end{cases} \quad (2.21)$$

Proof. See the Appendix. ■

Therefore, the predictive distribution of the market excess return with the constraint imposed

$$p_c(r_{t+1}|D_t) = \int_{\boldsymbol{\theta}} \mathcal{N}(r_{t+1}|\mathbf{z}'_t \boldsymbol{\beta}, \sigma^2) \pi_c(\boldsymbol{\beta}, \sigma|D_t) \mathbf{d}\boldsymbol{\beta} \mathbf{d}\sigma \quad (2.22)$$

is

$$p_c(r_{t+1}|D_t) \propto \left(1 + \frac{1}{\nu_{0,t}} \frac{(r_{t+1} - \mathbf{z}'_t \boldsymbol{\beta}_{0c,t})^2}{\frac{\delta_{0,t}}{\nu_{0,t}} (1 + \mathbf{z}'_t \mathbf{B}_{0,t} \mathbf{z}_t)} \right)^{-\frac{\nu_{0,t}+1}{2}} \quad (2.23)$$

a student- t distribution with $\nu_{0,t}$ degrees of freedom and mean

$$E_c(r_{t+1}|D_t) = \mathbf{z}'_t \boldsymbol{\beta}_{0c,t} \quad (2.24)$$

Note that $\pi_c(\boldsymbol{\beta}, \sigma | D_t)$ is identical to $\pi_u(\boldsymbol{\beta}, \sigma | D_t)$ if $\mathbf{z}'_t \boldsymbol{\beta}_{0u,t} \geq 0$. Hence, in this approach one modifies the unrestricted posterior only when the non-negativity constraint is violated. Note also that if the implied posterior mean from $\pi_u(\boldsymbol{\beta}, \sigma | D_\tau)$ delivers non-negative $\mathbf{z}_\tau \boldsymbol{\beta}_{0u,\tau}$ for all time period τ in the sample, then no modifications occur at all and predictive inferences correspond to the unrestricted case as discussed in Section 2.1. Also note that since $\pi_c(\boldsymbol{\beta}, \sigma | D_t)$ reflects all the information that is available at time t it becomes the prior distribution for the next time period $t + 1$, and the process described above is applied anew.

Another thing worth noting is the similarity of our approach and the truncation approach in Campbell and Thompson (2008) where the forecast of next period's excess return is truncated to zero when the OLS estimates deliver a negative forecast. If the predictive mean is used as the point forecast of excess return for next period, then the point forecast in our approach is also truncated to zero when $\pi_u(\boldsymbol{\beta}, \sigma | D_t)$ generates a negative predictive mean, because $\mathbf{z}'_t \boldsymbol{\beta}_{0c,t} = 0$ when $\mathbf{z}'_t \boldsymbol{\beta}_{0u,t} < 0$.

We now study the empirical performance of our approach. An R package for implementing these calculations is available on request.

3 Empirical results

In this section, we first describe the data set we use and then report our empirical findings.

3.1 Data

Our vector of predictors \mathbf{z}_τ consists of 11 variables that are commonly used in the literature. They are selected from a pool of 14 predictors. These 14 predictors are those predictors that are originally analyzed in the comprehensive study of Goyal and Welch (2008) and are still available up till to year 2014. Three predictors, the log dividend yield, log dividend-payout ratio and term spread, are omitted because they are perfectly or almost perfectly correlated with the 11 predictors we use. This set of predictors has also been used in many other studies of market return predictability, for example, Dangl and Halling (2012), Neely et al. (2014), Rapach et al. (2015).

The response variable is the market excess return computed as the log returns on the S&P 500 index (including dividends) minus the Treasury bill rate. The data frequency is monthly and spans the period from 1927:01 to 2014:12. Denote this data sample as Sample 1 (1927:01-2014:12)¹. It is one of the longest samples in the extant literature to examine predictability of US equity premium.

In addition, Rapach et al. (2015) construct a new predictor, the short interest index (SII), and show that the short interest is the best of the univariate predictors of the market excess returns. Because of its superiority over other predictors, we also use the short interest index to make a benchmark univariate predictive model. However, the SII is only available from 1973:01 to 2014:12. Therefore, to relate our results to those reported in Rapach et al. (2015), we also consider a shorter data sample that spans from 1973:01 to 2014:12 and denote it as Sample 2 (1973:01-2014:12)². Note we do not include the short interest index in \mathbf{z}_τ in the multiple predictive regression, however, because it does not improve the performance of our constrained model. This is to be expected because the performance of the short interest predictor by construction is not supposed to benefit from a lower bound restriction.

The 11 predictors in our multiple predictive regression are

- Log dividend-price ratio (DP) is the difference between the log of a twelve-month moving sum of dividends paid on the S&P 500 index and the log of stock prices (S&P 500 index).
- Log earnings-price ratio (EP) is the difference between the log of a twelve-month moving sum of earnings on the S&P 500 index and the log of stock prices.
- Excess stock return volatility (RVOL) uses a twelve-month moving standard deviation estimator as in Mele (2007)³.
- Book-to-market ratio (BM) is the book-to-market value ratio for the Dow Jones Industrial Average.

¹All data variables are available from Amit Goyal's webpage at <http://www.hec.unil.ch/agoyal/>

²We thank Matthew Ringgenberg for providing the short interest index data.

³Goyal and Welch (2008) use the sum of squared daily excess returns on the S&P 500 index to measure stock return volatility. However, this measure produces a severe outlier in October of 1987, while the moving standard deviation estimator avoids this problem. Therefore, Neely et al. (2014), Rapach et al. (2015) suggest to use the latter variable. Here we follow their suggestion.

- Net equity expansion (NTIS) is the ratio of a twelve-month moving sum of net equity issues by NYSE-listed stocks to the total end-of-year market capitalization of NYSE stocks.
- Treasury bill rate (TBL) is the interest rate on a secondary market rate of three-month US treasury bills.
- Long-term yield (LTY) is the long-term government bond yield.
- Long-term return (LTR) is the return on long-term government bonds.
- Term spread (TMS) is the difference between the long-term yield and Treasury bill rate.
- Default yield spread (DFY) is the difference between BAA-rated and AAA-rated corporate bond yields.
- Default return spread (DFR) is the difference between returns on long-term corporate bonds and returns on long-term government bonds.
- Inflation (INFL) is the Consumer Price Index for all urban consumers from the Bureau of Labor Statistics, lagged by an extra month.

The other predictors that are not included in the multiple predictive regression but are used in the alternative methods for comparisons are

- Dividend yield (DY) is the difference between the log of dividends and log of one month-lagged prices.
- Dividend-payout ratio (DE) is the difference between the log of dividends and log of earnings.
- Term spread (TMS) is the difference between the long-term yield and Treasury bill rate.
- Short interest index (SII) is the standardized de-trended log of the equal-weighted mean of all asset-level short interest data reported in Compustat.

Table 1 reports the summary statistics for these data. As we can see from the table, several predictors are highly persistent. According to Cochrane (2008), the imposition of non-negativity constraints can be helpful in overcoming the problems that arise from highly persistent predictors.

3.2 Prior analysis

We construct the prior distribution of the parameters on the basis of a training sample. This is to ensure that our prior beliefs about the parameters are well founded and objective. For Sample 1 (1927:01-2014:12), the training sample consists of the data from 1927:01 to 1929:12. Let \mathbf{r}_0 denote the vector of excess market returns from 1927:02 to 1929:12 and let Z_0 denote the matrix of predictor variables \mathbf{z}_{ts} from 1927:01 to 1929:11. Then, the prior distribution for the estimation that starts in month 1930:01 is $\mathcal{NIG}_{k,2}(\boldsymbol{\beta}, \sigma | \boldsymbol{\beta}_{0,0}, B_{0,0}, \nu_{0,0}, \delta_{0,0})$ where

$$\boldsymbol{\beta}_{0,0} = (Z_0'Z_0)^{-1}Z_0'\mathbf{r}_0, \quad (3.1)$$

$$B_{0,0} = g(Z_0'Z_0)^{-1}, \quad (3.2)$$

$$\nu_{0,0} = 36, \quad (3.3)$$

$$\delta_{0,0} = \frac{(\mathbf{r}_0 - Z_0\boldsymbol{\beta}_{0,0})'(\mathbf{r}_0 - Z_0\boldsymbol{\beta}_{0,0})(\nu_{0,0} - 2)}{34}, \quad (3.4)$$

and g is a scaling factor that is model specific and given below. Johannes, Korteweg, and Polson (2014) also use observations in training sample that spans from 1927 to 1929 to generate “objective” priors.

In the comparison with short interest index where we use Sample 2 (1973:01-2014:12), the training sample also consists of data in the first three years, from 1973:01 to 1975:12.

3.3 Alternative methods

To better understand the performance of our constrained multiple predictive regression, we compare our constrained multiple predictive regression with a range of alternative methods proposed in the literature. To investigate the effects of constraints, we present results of the unconstrained multiple predictive regression. To investigate the effects of multiple predictors, we present results of the best univariate predictive regressions. We

Table 1

Full sample summary statistics, 1927:01-2014:12

		Mean	Median	Std.	Min.	Max.	$\rho(1)$
	r^m (%)	0.50	0.95	5.49	-33.93	34.55	0.088
Predictors used in the multiple predictive regression	DP	-3.36	-3.33	0.46	-4.52	-1.87	0.991
	EP	-2.72	-2.78	0.42	-4.84	-1.77	0.986
	RVOL	0.17	0.15	0.09	0.05	0.74	0.980
	BM	0.58	0.55	0.27	0.12	2.03	0.985
	NTIS	0.02	0.02	0.03	-0.06	0.18	0.979
	TBL (%)	3.51	3.07	3.10	0.01	16.30	0.992
	LTY (%)	5.21	4.30	2.79	1.82	14.82	0.996
	LTR (%)	0.49	0.32	2.43	-11.24	15.23	0.042
	DFY (%)	1.13	0.90	0.70	0.32	5.64	0.975
	DFR (%)	0.03	0.05	1.35	-9.75	7.37	-0.129
	INFL (%)	0.25	0.24	0.51	-2.08	5.74	0.567
Predictors not used in the multiple predictive regression	DY	-3.35	-3.32	0.46	-4.53	-1.91	0.991
	DE	-0.63	-0.62	0.33	-1.24	1.38	0.991
	TMS (%)	1.71	1.76	1.32	-3.65	4.55	0.961
	SII	0.00	-0.09	1.00	-2.28	2.94	0.950

The table displays monthly summary statistics for excess market return (r^m , the log return on the S&P 500 index in excess of the risk-free rate), 14 predictor variables from Goyal and Welch (2008), and the predictor short interest index (SII) from Rapach et al. (2015). DP is the log dividend-price ratio; EP is the log earning-price ratio; RVOL is the excess market return volatility; BM is the book-to-market ratio; NTIS is net equity expansion; TBL is three month Treasury bill interest rate; LTY is the long-term government bond yield; LTR is the long-term government bond return; DFY is the difference between Moody's BAA- and AAA-rated corporate bond yields; DFR is the long-term corporate bond return minus the long-term government bond return; INFL is inflation calculated from CPI for all urban consumers; DY is the log dividend yield; DE is the log dividend-payout ratio and TMS is the long-term government bond yield minus the Treasury bill rate. For each variable, the time-series average (Mean), median (Median), standard deviation (Std.), minimum (Min.), maximum (Max.), and first-order autocorrelation ($\rho(1)$) are reported. The sample period is from 1927:01 to 2014:12, with 1056 observations in total. The summary statistics of SII are based on data from 1973:01 to 2014:12, the period for which these data are available.

also present results from the other two prominent approaches of incorporating multiple predictors, combination of forecasts and combination of predictors, to see if our method is able to improve over the other existing methods that utilize multiple predictors. For each alternative, we consider two cases, a case where no constraints are imposed and a case where constraints are imposed conditioned on parameters, except for the multiple predictive regression, which cannot be estimated with constraints imposed conditioned on parameters.

In more detail, the predictive methods we consider are as follows.

Method 1: MPR-LB is the multiple predictive regression with the lower bound constraints, as developed in Section 2.3. The predictive factors in \mathbf{z}_τ are a constant and the full set of 11 economic predictors described above.

Method 2: MPR-NLB is the multiple predictive regression with no lower bound constraints as developed in Section 2.1. \mathbf{z}_τ is the same as that in model MPR-LB and contains a constant and 11 economic predictors.

Method 3: UPR-NLB is the univariate predictive regression with no lower bound constraints. In Sample 1 (1927:01-2014:12), \mathbf{z}_τ contains a constant and one of the 14 economic predictors. In Sample 2 (1973:01-2014:12), \mathbf{z}_τ contains a constant and one of the 14 economic predictors or contains a constant and the short interest index.

Method 4: UPR-LB-P is the univariate predictive regression with lower bound constraints imposed conditioned on parameters.

Method 5: UPR-MC-NLB has predictive distribution that is the equal-weighted average of predictive distributions from 14 univariate predictive regressions with no lower bound constraints, each using one of the 14 economic predictors. The log dividend yield, log dividend-payout ratio and term spread predictors are included because the combination approach is unaffected by collinearity. The results are little changed if we also incorporate the new predictor SII in Sample 2.

Method 6: UPR-MC-LB-P is the restricted version of UPR-MC-NLB. Each univariate predictive regression is now estimated with lower bound constraints conditioned

on parameters.

Method 7: UPR-PC-NLB is the univariate predictive regression with no lower bound constraints, in which the predictor is the first principal component extracted from the 14 economic predictors. Similar to the model combination approach used in UPR-MC-NLB, the principal component is unaffected by the collinearity issue and the results are little changed if we incorporate the new predictor SII in Sample 2.

Method 8: UPR-PC-NLB-P is the restricted version of UPR-PC-NLB. Now the univariate predictive regression of the principal component regressor is estimated with lower bound constraints conditioned on parameters.

Method 9: NoPred is the no-predictability benchmark and it is an unconstrained predictive regression with \mathbf{z}_τ only containing a constant.

All the unconstrained methods, including MPR-NLB, UPR-NLB, UPR-MC-NLB, UPR-PC-NLB, and NoPred, are estimated by the approach described in Section 2.1; and all the methods with constraints imposed conditioned on parameters, including UPR-LB-P, UPR-MC-LB-P, and UPR-PC-LB-P, are estimated as described in Pettenuzzo et al. (2014).

The initial training sample-based prior distribution is selected as described in Section 3.2. The scaling factor g is 2 for multiple predictive regressions MPR-LB and MPR-NLB, and 4 for all the univariate predictive regressions and the no-predictability benchmark NoPred.

3.4 Out-of-sample performance

Goyal and Welch (2008) have shown that the out-of-sample performance of most economic predictor variables is weak. Therefore, in this section, we examine the out-of-sample performance of our constrained multiple predictive regression, MPR-LB. In this analysis, we predict the market excess return by calculating for each t the Bayesian predictive density $p(r_{t+1}|D_t)$ in the unconstrained case and $p_c(r_{t+1}|D_t)$ in the constrained case, where, for comparability with existing literature, $t + 1$ runs from January 1947 to December 2014 in Sample 1 (1927:01-2014:12) and runs from January 1990 to December

2014 in Sample 2 (1973:01-2014:12). These 816/300 observations constitute the prediction sample of Sample 1/Sample 2. For each time period in the prediction sample, we evaluate the accuracy of the forecast and the financial gain that an investor can achieve under the constrained multiple predictive regression.

We start by providing in Fig. 1 the means of the predictive densities from multiple predictive regressions MPR-LB (solid line) and MPR-NLB (dashed line). The forecasts from MPR-LB are, of course, always non-negative. In addition, forecasts from MPR-LB are, in general, similar to those from MPR-NLB, except when the unconstrained forecasts are negative.

Next we formally evaluate the statistical and economic performance of all the contending predictive methods.

3.4.1 Statistical evaluation

We evaluate the statistical performance of the predictions by calculating the *predictive likelihood* (see Chib and Greenberg, 1995, Geweke and Amisano, 2012), which is the product of the predictive density evaluated at the realized values of the outcome. The predictive likelihood (on the log-scale) is a natural metric in the Bayesian context and uses information of the entire predictive distribution. As mentioned earlier, the one-step-ahead predictive distribution is a student- t distribution. Therefore, in Sample 1 (1927:01-2014:12), the log predictive likelihood (LPL) over the prediction sample from $m =$ January 1947 to $n =$ December 2014 is the sum of the one-step-ahead log predictive likelihoods,

$$LPL(m, n) = \sum_{t=m-1}^{n-1} \log f(r_{t+1}|D_t) \quad (3.5)$$

where $f(r_{t+1}|D_t)$ is $p(r_{t+1}|D_t)$ or $p_c(r_{t+1}|D_t)$ from Eq. (2.14) and Eq. (2.23) depending on whether we are considering the unconstrained or constrained case. For Sample 2 (1973:01-2014:12), the log predictive likelihood for the prediction sample is computed in the sample way with $m =$ January 1990 and $n =$ December 2014.

Table 2 presents the predictive likelihood results for the constrained multiple predictive regression as well as other alternatives. The table reports the log predictive likelihood ratio of each method against the no-predictability benchmark (NoPred). A

Table 2

Out-of-sample predictive performance, log predictive likelihood ratio

	Sample 1 (1927:01-2014:12)	Sample 2 (1973:01-2014:12)
MPR-LB	26.31	9.07
MPR-NLB	-9.25	-11.77
UPR-NLB	1.64	1.88
UPR-LB-P	1.34	0.69
UPR-MC-NLB	1.25	1.10
UPR-MC-LB-P	-0.13	-0.21
UPR-PC-NLB	1.15	-0.84
UPR-PC-LB-P	0.93	-0.44

The table reports the log predictive likelihood ratio of each method against the no-predictability benchmark. The second column reports the log predictive likelihood ratio for Sample 1 (1927:01-2014:12) in the prediction sample 1947:01-2014:12, and the third column reports the log predictive likelihood ratio for Sample 2 (1973:01-2014:12) in the prediction sample 1990:01-2014:12. MPR-LB is the constrained multiple predictive regression with 11 economic predictors. For descriptions of the predictors, see the notes to Table 1. MPR-NLB is the unconstrained version of MPR-LB. UPR-NLB is the unconstrained univariate predictive regression with the highest predictive likelihood among all 14/15 single predictors in Sample 1/Sample 2. In Sample 1, it is predictor RVOL, whereas it is predictor SII in Sample 2. UPR-LB-P is the univariate predictive regression with non-negativity constraints imposed conditioned on parameters and has the highest predictive likelihood among all 14/15 single predictors in Sample 1/Sample 2. In Sample 1, it is predictor EP, whereas it is predictor SII in Sample 2. UPR-MC-NLB represents equal-weighted density combination of all unconstrained univariate predictive regressions with 14 economic predictors. UPR-PC-NLB is the unconstrained univariate predictive regression with predictor as the first principal component extracted from all 14 economic predictors. UPR-MC-LB-P and UPR-PC-LB-P are restricted versions of UPR-MC-NLB and UPR-PC-NLB respectively, with non-negativity constraints imposed conditioned on parameters.

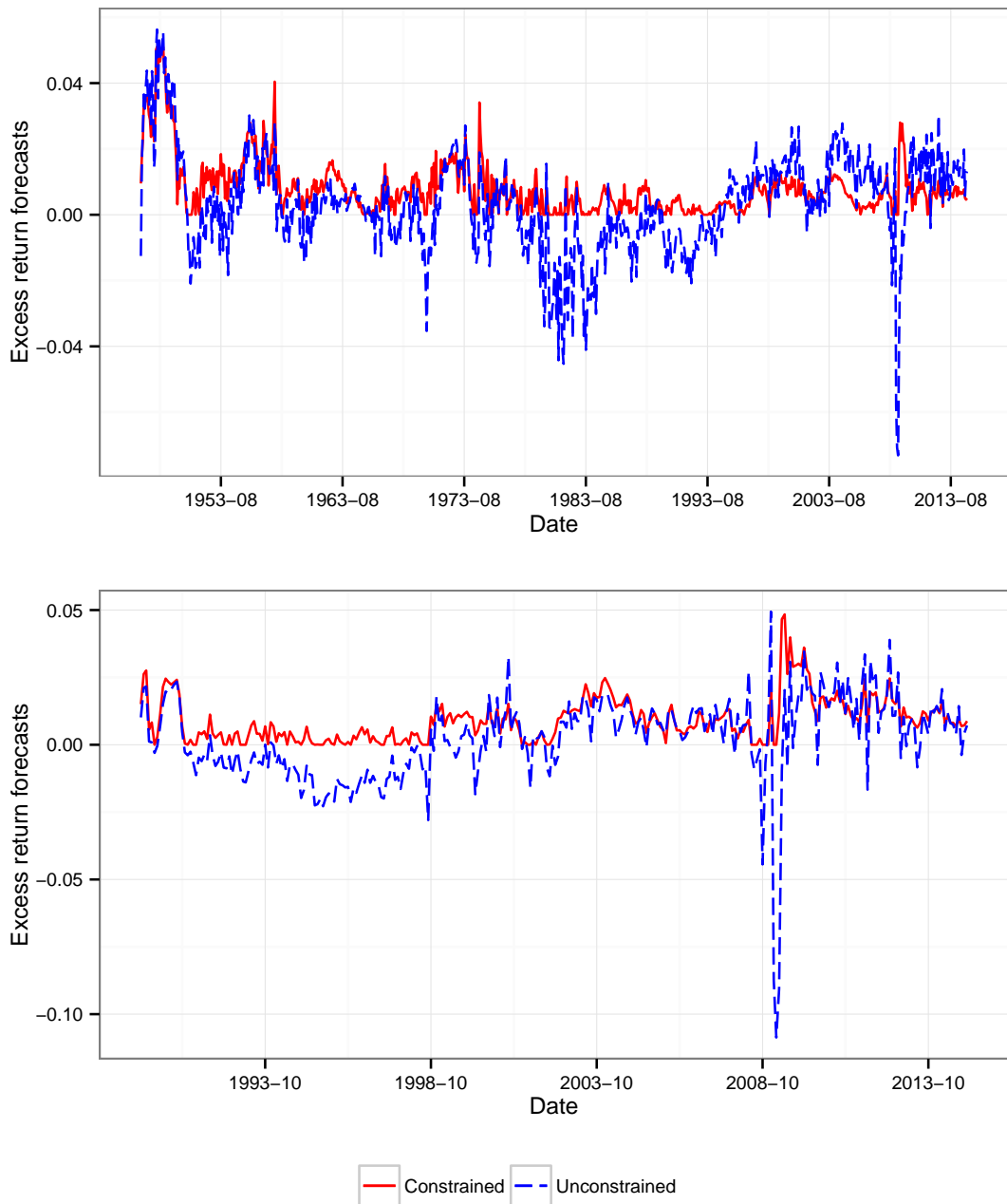


Figure 1: **Out-of-sample point forecasts of multiple predictive regressions**
 We compute predictive mean of constrained (solid) and unconstrained (dashed) multiple predictive regressions for Sample 1 (1927:01-2014:12) on the top and Sample 2 (1973:01-2014:12) on the bottom. The out-of-sample predictive means are calculated for each month from 1947:01 to 2014:12 in Sample 1 and from 1990:01 to 2014:12 in Sample 2.

value greater than 0 indicates that the method performs better than the NoPred in terms of the predictive likelihood. The second and third columns in Table 2 report log predictive likelihood ratio results for Sample 1 and Sample 2 respectively.

Next, we compare the predictive performance of the constrained multiple predictive regression with other alternatives. First, the economically motivated constraints significantly improve the out-of-sample performance. The unconstrained multiple predictive regression performs much worse than the no-predictability benchmark in terms of predictive accuracy. However, after imposing the non-negativity constraints, there is substantial improvement and the constrained multiple predictive regression outperforms the no-predictability benchmark by a wide margin. Fig. 2 plots the cumulative log predictive likelihood ratios of the multiple predictive regressions relative to the no-predictability benchmark. As we can see from the plot, the constrained multiple predictive regression (solid line) has a strong upward trend, except a slightly negative slope for the late 1990s, and is consistently greater than 0. The better performance of the constrained multiple predictive regression over the no-predictability benchmark is, therefore, quite consistent over time.

Second, the constrained multiple predictive regression improves on the performance of the univariate regressions UPR-NLB and UPR-LB-P. Both UPR-NLB and UPR-LB-P have 14 different specifications in Sample 1 (1927:01-2014:12) because there are 14 different economic predictors. This number increases to 15 in Sample 2 (1973:01-2014:12) as the newly constructed predictor SII becomes available at the beginning of 1973. We only report UPR-NLB and UPR-LB-P models with the highest predictive likelihood in the table. In Sample 1, they are UPR-NLB with predictor RVOL and UPR-LB-P with predictor EP. In Sample 2, we show that the SII predictor outperforms the other predictors in both UPR-NLB and UPR-LB-P, even after accounting for parameter uncertainty as we have done here, which lends further support for the findings of Rapach et al. (2015). Nevertheless there are quite a few univariate predictive regressions, there is few studies providing guidance on which predictor to use before Rapach et al. (2015). This makes multiple predictive regression an appealing alternative to univariate predictive regression as there is no need to select the single predictor that works best. Moreover, even though Rapach et al. (2015) show that they find the best single predictor SII, it

is demonstrated in Table 2 that the constrained multiple predictive regression performs better than the univariate predictive regression with SII in Sample 2. The superiority of the constrained multiple predictive regression over the best univariate predictive regression holds in Sample 1 as well. The implication of this result is consistent with Rapach et al. (2010), where they suggest there are more state variables that determine the equity premium instead of just one or a few.

Third, we look at if our constrained multiple predictive regression is able to improve over the other two approaches that utilize multiple predictors, the model combination approach (UPR-MC-NLB) and the predictor combination approach (UPR-PC-NLB). In Sample 1 (1927:01-2014:12), our result shows that these two approaches are able to generate more accurate prediction than the no-predictability benchmark, which is consistent with the previous findings (see, e.g., Rapach et al., 2010, Neely et al., 2014). In Sample 2 (1973:01-2014:12), UPR-MC-NLB beats the no-predictability benchmark whereas UPR-PC-NLB fails. Despite of the gains from these two approaches, our result shows that the constrained multiple predictive regression, a more direct way to incorporate multiple predictors, produces much higher predictive likelihood. Imposing the non-negativity constraints conditioned on parameters (UPR-MC-LB-P and UPR-PC-LB-P) does not improve these two approaches.

Finally, the constrained multiple predictive regression beats the other alternatives consistently over time in predictive likelihood. Table 3 reports the log predictive likelihood ratios for finer prediction samples. For Sample 1 (1927:01-2014:12), we split the prediction sample into two subsamples 1947:01-1978:12 and 1979:01-2014:12. For Sample 2 (1973:01-2014:12), we consider two subsamples 1990:01-2006:12 and 2007:01-2014:12. The break-up points years 1978 and 2006 are selected by following Pettenuzzo et al. (2014) and Rapach et al. (2015) respectively. Table 3 shows that the constrained multiple predictive regression achieve the highest predictive likelihoods in all subsamples.

3.4.2 Economic evaluation

In this section we examine the utility gain from each predictive method, relative to the no-predictability benchmark, in a portfolio optimization setting. Following previous studies on portfolio optimization from a Bayesian perspective, for example, Kandel and

Table 3

Log predictive likelihood ratios for subsamples

	Sample 1 (1927:01-2014:12)		Sample 2 (1973:01-2014:12)	
	1947:1-1978:12	1979:1-2014:12	1990:1-2006:12	2007:1-2014:12
MPR-LB	14.89	11.42	2.55	6.52
MPR-NLB	-0.20	-9.05	-8.47	-3.30
UPR-NLB	2.26	0.74	-0.10	1.98
UPR-LB-P	0.98	0.71	0.04	0.65
UPR-MC-NLB	1.02	0.23	-0.47	1.57
UPR-MC-LB-P	-0.17	0.03	-0.25	0.03
UPR-PC-NLB	1.24	-0.09	-0.64	-0.20
UPR-PC-LB-P	0.44	0.48	-0.49	0.05

The table reports the log predictive likelihood ratio of each method against the no-predictability benchmark in different subsamples. The second and third columns report the log predictive likelihood ratio for Sample 1 (1927:01-2014:12) in the prediction sample 1947:01-1978:12 and 1979:01-2014:12 respectively. The fourth and fifth columns report the log predictive likelihood ratio for Sample 2 (1973:01-2014:12) in the prediction sample 1990:01-2006:12 and 2007:01-2014:12 respectively. See the notes to Table 2 for method specifications of MPR-LB, MPR-NLB, UPR-MC-NLB, UPR-MC-LB-P, UPR-PC-NLB, and UPR-PC-LB-P listed in the first column. UPR-NLB is the unconstrained univariate predictive regression with the highest predictive likelihood among all single predictors in each prediction sample. In the prediction sample 1947:01-1978:12 of Sample 1, it is predictor NTIS; in the prediction sample 1979:01-2014 of Sample 1, it is predictor DFR; and in the prediction samples 1990:01-2006:12 and 2007:01-2014:12 of Sample 2, it is predictor SII. UPR-LB-P is the univariate predictive regression with non-negativity constraints imposed conditioned on parameters and has the highest predictive likelihood among all single predictors in each prediction sample. In the prediction sample 1947:01-1978:12 of Sample 1, it is predictor BM; in the prediction sample 1979:01-2014:12 of Sample 1, it is predictor EP; and in the prediction samples 1990:01-2006:12 and 2007:01-2014:12 of Sample 2, it is SII.

Stambaugh (1996) and Johannes et al. (2014), we consider the portfolio problem for a representative investor with a constant relative risk aversion (CRRA) utility function. For wealth weight ω on risky asset and excess return r on risky asset, the CRRA utility function is

$$U(\omega, r) = \frac{[(1 - \omega) \exp(r_f) + \omega \exp(r_f + r)]^{1-\gamma}}{1 - \gamma}, \quad (3.6)$$

where r_f is the continuously compounded Treasury bill rate and γ is the investor's coefficient of relative risk aversion, which we assume is five.

Suppose that at time t , the investor solves the optimal asset allocation problem using the Bayesian predictive density (Kandel and Stambaugh, 1996)

$$\omega_t^* = \arg \max_{\omega} \int [U(\omega, r_{t+1})] f(r_{t+1}|D_t) dr_{t+1} \quad (3.7)$$

where $f(r_{t+1}|D_t)$ in the current problem is $p(r_{t+1}|D_t)$ or $p_c(r_{t+1}|D_t)$ from Eq. (2.14) and Eq. (2.23). Approximating the integral by Monte Carlo draws from $f(r_{t+1}|D_t)$, the above problem can be solved as

$$\omega_t^* = \arg \max_{\omega} \sum_{j=1}^G \left\{ \frac{[(1 - \omega) \exp(r_{ft}) + \omega \exp(r_{ft} + r_{t+1}^{(j)})]^{1-\gamma}}{1 - \gamma} \right\}, \quad (3.8)$$

where $r_{t+1}^{(j)}$ is simulated from $f(r_{t+1}|D_t)$ and G is a large number, say 20,000. Following Kandel and Stambaugh (1996), we restrict ω_t^* to the interval $[0, 0.99]$.

Given the solution ω_t^* and the realized return r_{t+1} at time $t + 1$, we compute the hypothetical realized utility from each predictive model. Denote this realized utility as U_{t+1} where

$$U_{t+1} = U(\omega_t^*, r_{t+1}) \quad (3.9)$$

To evaluate the economic value of each model with monthly data, we compute the annualized certainty equivalent return (CER) in percent for the consecutive sample from month m to month n ,

$$\text{CER}(m, n) = 1200 \times \left\{ \left[(1 - \gamma) \frac{\sum_{t=m-1}^{n-1} U_{t+1}}{n - m + 1} \right]^{\frac{1}{1-\gamma}} - 1 \right\} \quad (3.10)$$

As in Section 3.4.1, we compare the performance of each method relative to the

no-predictability benchmark NoPred. The CER of each method minus the CER of the NoPred is defined as the CER gain of the method. It measures the willingness to pay for having access to the predictive method over the no-predictability benchmark. In Table 4, these gains are reported for the prediction samples of both Sample 1 (1927:01-2014:12) and Sample 2 (1973:01-2014:12). Similar to the statistical evaluation, UPR-NLB and UPR-LB-P are reported as the univariate predictive regressions with the highest certainty equivalent gains. In Sample 1, they are UPR-NLB with predictor DE and UPR-LB-P with predictor EP. In Sample 2, coinciding with the statistical evaluation, UPR-NLB and UPR-LB-P with predictor SII achieve the largest economic gain among all the predictive regressions with single predictor. Thus, the relevance of the short interest index in forming market portfolios, observed by Rapach et al. (2015) from a non-Bayesian perspective, is still very much valid from the Bayesian perspective.

Table 4 shows that our constrained multiple predictive regression delivers best economic performance among all the competing models we consider. Both measures of method performance, the predictive likelihood and portfolio gains, identify the constrained multiple predictive regression as the best performing specification. Therefore, the two metrics pick up the same method. However, if, for example, we are interested in using the unconstrained univariate predictive regressions, then in Sample 1 (1927:01-2014:12), the statistical metric picks up predictor RVOL and the economic metric picks up a different predictor DE. This creates further difficulties of which single predictor to use in univariate predictive regressions. It seems that predictor SII can ameliorate the problem because SII beats the other economic predictors in Sample 2 (1973:01-2014:12) under both metrics but it falls behind the constrained multiple predictive regression.

Fig. 3 offers a different perspective on the portfolio formed from our constrained multiple predictive regression model. It shows the log cumulative wealth for three portfolios, based on forecasts from the constrained multiple predictive regression (solid line), the unconstrained multiple predictive regression (dashed line), and the no-predictability model (dotted line), respectively. The constrained multiple predictive regression generates the highest cumulative wealth for nearly all time periods in both samples. In addition, for most of the time periods, the log cumulative wealth curve of the constrained multiple predictive regression has a steeper upward trend than the other two

Table 4

Out-of-sample certainty equivalent return gains (%)

	Sample 1 (1927:01-2014:12)	Sample 2 (1973:01-2014:12)
MPR-LB	0.896	2.912
MPR-NLB	-0.192	1.035
UPR-NLB	0.284	1.805
UPR-LB-P	0.440	0.688
UPR-MC-NLB	-0.053	-0.640
UPR-MC-LB-P	0.278	-0.040
UPR-PC-NLB	-0.285	-0.586
UPR-PC-LB-P	0.304	-0.195

The table reports the annualized certainty equivalent return gains (%) for investors who select market equity and Treasury bills every month based on forecasts from the predictive method relative to the no-predictability benchmark (NoPred). The second column reports the annualized certainty equivalent return gains for Sample 1 (1927:01-2014:12) in the prediction sample 1947:01-2014:12, and the third column reports the annualized certainty equivalent return gains for Sample 2 (1973:01-2014:12) in the prediction sample 1990:01-2014:12. The weight on equity is restricted from 0 to 0.99. Investors are assumed to have constant relative risk aversion utility with relative risk aversion coefficient of five. See the notes to Table 2 for method specifications of MPR-LB, MPR-NLB, UPR-MC-NLB, UPR-MC-LB-P, UPR-PC-NLB, and UPR-PC-LB-P listed in the first column. UPR-NLB is the unconstrained univariate predictive regression with the highest certainty equivalent returns among all 14/15 single predictors in Sample 1/Sample 2. In Sample 1, it is predictor DE, whereas it is predictor SII in Sample 2. UPR-LB-P is the univariate predictive regression with non-negativity constraints imposed conditioned on parameters and has the highest certainty equivalent returns among all 14/15 single predictors in Sample 1/Sample 2. In Sample 1, it is predictor EP, whereas it is predictor SII in Sample 2.

curves. To get a quantitative description of the economic performances over time, we report the CER gains over subsamples for both Sample 1 and Sample 2 in Table 5. The results show that the constrained multiple predictive regression consistently generates higher returns than the other alternatives over time as it produces the highest CER gains in all subsamples. In particular, the fifth column reports results for the subsample 2007:01-2014:12, which includes the period of the 2008-2009 Financial Crisis. For this subsample, the gain from the constrained multiple predictive regression is especially pronounced and amounts to 687 basis points per annum. This is consistent with findings of Rapach et al. (2010), Henkel, Martin, and Nardari (2011), and Rapach et al. (2015), who note that the market return predictability and its associated utility gain is particularly substantial over periods of severe economic downturns.

The above results assume that $\gamma = 5$. Table 6 shows that these results are robust to this choice. The second and fourth columns in Table 6 report CER gains for investors with $\gamma = 2$, while the third and fifth columns contain results for investors with $\gamma = 10$. For both values of γ , the constrained multiple predictive regression delivers highest CER values among all the methods and in both data samples.

4 Conclusion

In this paper, we provide a new strategy, from the Bayesian perspective, for predicting excess market returns. We show that the multiple predictive regression model can produce striking results if the Bayesian predictive mean is constrained to be non-negative. In our extensive empirical study we show that the Bayesian multiple predictive regression with our lower bound constraints can produce sizable statistical and economic gains, relative to various alternative methods and approaches. Importantly for practice, our new strategy is easy to implement. We believe that on account of these features, and clear successes in the empirical analysis, that the predictive strategy described in this paper will open a new fruitful avenue for dealing with the difficult and vital problem of predicting the excess market returns, one that is likely to attract considerable interest.

Table 5

Out-of-sample certainty equivalent return gains (%) for subsamples

	Sample 1 (1927:01-2014:12)		Sample 2 (1973:01-2014:12)	
	1947:1-1978:12	1979:1-2014:12	1990:1-2006:12	2007:1-2014:12
MPR-LB	0.872	0.917	1.043	6.868
MPR-NLB	-0.805	0.358	-0.506	4.289
UPR-NLB	0.234	0.329	0.457	5.399
UPR-LB-P	0.740	0.290	0.226	1.657
UPR-MC-NLB	0.129	-0.216	-0.917	-0.059
UPR-MC-LB-P	0.501	0.078	-0.083	0.049
UPR-PC-NLB	0.050	-0.586	-0.764	-0.214
UPR-PC-LB-P	0.660	-0.015	-0.225	-0.133

The table reports the annualized certainty equivalent return gains (%) in different subsamples for investors who select market equity and Treasury bills every month based on forecasts from the predictive method relative to the no-predictability benchmark (NoPred). The second and third columns report the annualized certainty equivalent return gains for Sample 1 (1927:01-2014:12) in the prediction sample 1947:01-1978:12 and 1979:01-2014:12 respectively. The fourth and fifth columns report the annualized certainty equivalent return gains for Sample 2 (1973:01-2014:12) in the prediction sample 1990:01-2006:12 and 2007:01-2014:12 respectively. The weight on equity is restricted from 0 to 0.99. Investors are assumed to have constant relative risk aversion utility with relative risk aversion coefficient of five. See the notes to Table 2 for method specifications of MPR-LB, MPR-NLB, UPR-MC-NLB, UPR-MC-LB-P, UPR-PC-NLB, and UPR-PC-LB-P listed in the first column. UPR-NLB is the unconstrained univariate predictive regression with the highest certainty equivalent returns among all single predictors in each prediction sample. In the prediction samples 1947:01-1978:12 and 1979:01-2014:12 of Sample 1, it is predictor DE; in the prediction sample 1990:01-2006:12 of Sample 2, it is DFR; and in the prediction sample 2007:01-2014:12 of Sample 2, it is predictor SII. UPR-LB-P is the univariate predictive regression with non-negativity constraints imposed conditioned on parameters and has the highest certainty equivalent returns among all single predictors in each prediction sample. In the prediction sample 1947:01-1978:12 of Sample 1, it is predictor DY; in the prediction sample 1979:01-2014:12 of Sample 1, it is predictor EP; and in the prediction samples 1990:01-2006:12 and 2007:01-2014:12 of Sample 2, it is SII.

Table 6

Risk aversion effects on certainty equivalent return gains

	Sample 1 (1927:01-2014:12)		Sample 2 (1973:01-2014:12)	
	$\gamma = 2$	$\gamma = 10$	$\gamma = 2$	$\gamma = 10$
MPR-LB	1.024	0.389	1.255	2.161
MPR-NLB	-1.533	-0.269	-0.734	0.652
UPR-NLB	0.663	0.150	0.912	1.225
UPR-LB-P	0.834	0.226	0.006	0.347
UPR-MC-NLB	-0.125	-0.028	-0.598	-0.318
UPR-MC-LB-P	0.682	0.142	0.006	-0.019
UPR-PC-NLB	-0.723	-0.150	-0.681	-0.382
UPR-PC-LB-P	0.639	0.151	-0.052	-0.098

The table reports the annualized certainty equivalent return gains for investors who select market equity and Treasury bills every month based on forecasts from the predictive method relative to the no-predictability benchmark (NoPred). The second and third columns report the annualized certainty equivalent return gains for Sample 1 (1927:01-2014:12) in the prediction sample 1947:01-2014:12. The fourth and fifth columns report the annualized certainty equivalent return gains for Sample 2 (1973:01-2014:12) in the prediction sample 1990:01-2014:12. The weight on equity is restricted from 0 to 0.99. Investors are assumed to have constant relative risk aversion utility with relative risk aversion coefficient γ of two (second and fourth columns) or ten (third and fifth columns). See the notes to Table 2 for method specifications of MPR-LB, MPR-NLB, UPR-NLB- MC, CUPR-MC-PC, UPR-PC-NLB, and UPR-PC-LB-P listed in the first column. UPR-NLB is the unconstrained univariate predictive regression with the highest certainty equivalent returns among all 14/15 single predictors in the Sample 1/Sample 2. In Sample 1, it is predictor DE, whereas it is predictor SII in Sample 2. UPR-LB-P is the univariate predictive regression with non-negativity constraints imposed conditioned on parameters and has the highest certainty equivalent returns among all 14/15 single predictors in Sample 1/Sample 2. In Sample 1, it is predictor EP, whereas it is predictor SII in Sample 2.

Appendix A

We first provide the prior-posterior updates with the normal-inverse-gamma type II distribution. Let $\pi(\boldsymbol{\beta}, \sigma | D_{t-1})$ be the $\mathcal{NIG}_{k,2}(\boldsymbol{\beta}, \sigma | \boldsymbol{\beta}_{0,t-1}, \mathbf{B}_{0,t-1}, \nu_{0,t-1}, \delta_{0,t-1})$ distribution with density

$$\pi(\boldsymbol{\beta}, \sigma | D_{t-1}) \propto \sigma^{-\nu_{0,t-1}-k-1} \exp \left\{ -\frac{1}{2\sigma^2} [(\boldsymbol{\beta} - \boldsymbol{\beta}_{0,t-1})' \mathbf{B}_{0,t-1}^{-1} (\boldsymbol{\beta} - \boldsymbol{\beta}_{0,t-1}) + \delta_{0,t-1}] \right\} \quad (4.1)$$

By Bayes' theorem,

$$\pi(\boldsymbol{\beta}, \sigma | D_t) \propto \mathcal{N}(r_t | \mathbf{z}'_{t-1} \boldsymbol{\beta}, \sigma^2) \pi(\boldsymbol{\beta}, \sigma | D_{t-1}) \quad (4.2)$$

$$\propto \sigma^{-\nu_{0,t-1}-k-2} \exp \left\{ -\frac{1}{2\sigma^2} [(\boldsymbol{\beta} - \boldsymbol{\beta}_{0,t-1})' \mathbf{B}_{0,t-1}^{-1} (\boldsymbol{\beta} - \boldsymbol{\beta}_{0,t-1}) + (r_t - \mathbf{z}'_{t-1} \boldsymbol{\beta})^2 + \delta_{0,t-1}] \right\} \quad (4.3)$$

$$\propto \sigma^{-\nu_{0,t-1}-k-2} \times \quad (4.4)$$

$$\exp \left\{ -\frac{\boldsymbol{\beta}' (\mathbf{B}_{0,t-1}^{-1} + \mathbf{z}_{t-1} \mathbf{z}'_{t-1}) \boldsymbol{\beta} - 2(\boldsymbol{\beta}'_{0,t-1} \mathbf{B}_{0,t-1}^{-1} + r_t \mathbf{z}'_{t-1}) \boldsymbol{\beta} + r_t^2 + \boldsymbol{\beta}'_{0,t-1} \mathbf{B}_{0,t-1}^{-1} \boldsymbol{\beta}_{0,t-1} + \delta_{0,t-1}}{2\sigma^2} \right\}, \quad (4.5)$$

which can be rewritten as

$$\pi(\boldsymbol{\beta}, \sigma | D_t) \propto \sigma^{-\nu_{0,t}-k-1} \exp \left\{ -\frac{1}{2\sigma^2} [(\boldsymbol{\beta} - \boldsymbol{\beta}_{0,t})' \mathbf{B}_{0,t}^{-1} (\boldsymbol{\beta} - \boldsymbol{\beta}_{0,t}) + \delta_{0,t}] \right\}, \quad (4.6)$$

where the updated parameters are given by

$$\nu_{0,t} = \nu_{0,t-1} + 1, \quad (4.7)$$

$$\mathbf{B}_{0,t} = (\mathbf{B}_{0,t-1}^{-1} + \mathbf{z}_{t-1} \mathbf{z}'_{t-1})^{-1}, \quad (4.8)$$

$$\boldsymbol{\beta}_{0,t} = \mathbf{B}_{0,t} (\mathbf{B}_{0,t-1}^{-1} \boldsymbol{\beta}_{0,t-1} + \mathbf{z}_{t-1} r_t), \quad (4.9)$$

$$\delta_{0,t} = \delta_{0,t-1} + r_t^2 + \boldsymbol{\beta}'_{0,t-1} \mathbf{B}_{0,t-1}^{-1} \boldsymbol{\beta}_{0,t-1} - \boldsymbol{\beta}'_{0,t} \mathbf{B}_{0,t}^{-1} \boldsymbol{\beta}_{0,t} \quad (4.10)$$

This is a normal-inverse-gamma type II distribution, with parameters $(\boldsymbol{\beta}_{0,t}, \mathbf{B}_{0,t}, \nu_{0,t}, \delta_{0,t})$.

Next, the predictive distribution of r_{t+1} is

$$p(r_{t+1}|D_t) = \int_{\theta} \mathcal{N}(r_{t+1}|\mathbf{z}'_t\boldsymbol{\beta}, \sigma^2)\pi(\boldsymbol{\beta}, \sigma|D_t)d\boldsymbol{\beta}d\sigma \quad (4.11)$$

$$= \int_{\theta} \mathcal{N}(r_{t+1}|\mathbf{z}'_t\boldsymbol{\beta}, \sigma^2)\mathcal{N}_k(\boldsymbol{\beta}|\boldsymbol{\beta}_{0,t}, \sigma^2\mathbf{B}_{0,t})\mathcal{IG}_2\left(\sigma\left|\frac{\nu_{0,t}}{2}, \frac{\delta_{0,t}}{2}\right.\right)d\boldsymbol{\beta}d\sigma \quad (4.12)$$

$$= \int_{\theta} \mathcal{IG}_2\left(\sigma\left|\frac{\nu_{0,t}}{2}, \frac{\delta_{0,t}}{2}\right.\right) \int \mathcal{N}(r_{t+1}|\mathbf{z}'_t\boldsymbol{\beta}, \sigma^2)\mathcal{N}_k(\boldsymbol{\beta}|\boldsymbol{\beta}_{0,t}, \sigma^2\mathbf{B}_{0,t})d\boldsymbol{\beta}d\sigma \quad (4.13)$$

$$= \int_{\theta} \mathcal{IG}_2\left(\sigma\left|\frac{\nu_{0,t}}{2}, \frac{\delta_{0,t}}{2}\right.\right) \mathcal{N}(r_{t+1}|\mathbf{z}'_t\boldsymbol{\beta}_{0,t}, \sigma^2(1 + \mathbf{z}'_t\mathbf{B}_{0,t}\mathbf{z}_t))d\sigma \quad (4.14)$$

$$\propto \Gamma\left(\frac{\nu_{0,t} + 1}{2}\right) \left(1 + \frac{1}{\nu_{0,t}} \frac{(r_{t+1} - \mathbf{z}'_t\boldsymbol{\beta}_{0,t})^2}{\frac{\delta_{0,t}}{\nu_{0,t}}(1 + \mathbf{z}'_t\mathbf{B}_{0,t}\mathbf{z}_t)}\right)^{-\frac{\nu_{0,t} + 1}{2}} \quad (4.15)$$

a student- t distribution with location parameter $\mathbf{z}'_t\boldsymbol{\beta}_{0,t}$, dispersion parameter $\frac{\delta_{0,t}}{\nu_{0,t}}(1 + \mathbf{z}'_t\mathbf{B}_{0,t}\mathbf{z}_t)$ and $\nu_{0,t}$ degrees of freedom.

Third, the Kullback-Leibler divergence between two normal-inverse-gamma type II distributions with parameters $(\boldsymbol{\beta}_{0c,t}, \mathbf{B}_{0,t}, \nu_{0,t}, \delta_{0,t})$ and $(\boldsymbol{\beta}_{0u,t}, \mathbf{B}_{0,t}, \nu_{0,t}, \delta_{0,t})$ is

$$E_{c,t} \left[\log \frac{\mathcal{N}_k(\boldsymbol{\beta}|\boldsymbol{\beta}_{0c,t}, \sigma^2\mathbf{B}_{0,t})}{\mathcal{N}_k(\boldsymbol{\beta}|\boldsymbol{\beta}_{0u,t}, \sigma^2\mathbf{B}_{0,t})} \right], \quad (4.16)$$

where $E_{c,t}$ is the expectation under $\mathcal{NIG}_{k,2}(\boldsymbol{\beta}, \sigma|\boldsymbol{\beta}_{0c,t}, \mathbf{B}_{0,t}, \nu_{0,t}, \delta_{0,t})$ and

$$\log \frac{\mathcal{N}_k(\boldsymbol{\beta}|\boldsymbol{\beta}_{0c,t}, \sigma^2\mathbf{B}_{0,t})}{\mathcal{N}_k(\boldsymbol{\beta}|\boldsymbol{\beta}_{0u,t}, \sigma^2\mathbf{B}_{0,t})} = -\frac{1}{2\sigma^2} [(\boldsymbol{\beta} - \boldsymbol{\beta}_{0c,t})'\mathbf{B}_{0,t}^{-1}(\boldsymbol{\beta} - \boldsymbol{\beta}_{0c,t}) - (\boldsymbol{\beta} - \boldsymbol{\beta}_{0u,t})'\mathbf{B}_{0,t}^{-1}(\boldsymbol{\beta} - \boldsymbol{\beta}_{0u,t})] \quad (4.17)$$

$$= (\boldsymbol{\beta}_{0c,t} - \boldsymbol{\beta}_{0u,t})'\mathbf{B}_{0,t}^{-1}\boldsymbol{\beta}\sigma^{-2} - \frac{1}{2}(\boldsymbol{\beta}'_{0c,t}\mathbf{B}_{0,t}^{-1}\boldsymbol{\beta}_{0c,t} - \boldsymbol{\beta}'_{0u,t}\mathbf{B}_{0,t}^{-1}\boldsymbol{\beta}_{0u,t})\sigma^{-2} \quad (4.18)$$

Therefore, to compute the right hand side quantity in Eq. (4.16), we just need to get $E_{c,t}[\boldsymbol{\beta}\sigma^{-2}]$ and $E_{c,t}[\sigma^{-2}]$. By definition,

$$\int_{\sigma>0} \sigma^{-2}\mathcal{IG}_2(\sigma|\nu, \delta)d\sigma = \int_0^{\infty} \frac{2\delta^{\nu}}{\Gamma(\nu)}\sigma^{-2\nu-3}\exp(-\delta\sigma^{-2})d\sigma \quad (4.19)$$

If we replace $\eta = \delta\sigma^{-2}$ and substitute it back to the above equation, then

$$\int_0^\infty \frac{2\delta^\nu}{\Gamma(\nu)} \sigma^{-2\nu-3} \exp(-\delta\sigma^{-2}) d\sigma = \int_0^\infty \frac{\eta^\nu}{\delta\Gamma(\nu)} \exp(-\eta) d\eta \quad (4.20)$$

$$= \frac{\Gamma(\nu+1)}{\delta\Gamma(\nu)} = \frac{\nu}{\delta} \quad (4.21)$$

from properties of the gamma function. Therefore,

$$E_{c,t}[\beta\sigma^{-2}] = \int_{\sigma>0} \sigma^{-2} \mathcal{IG}_2 \left(\sigma \left| \frac{\nu_{0,t}}{2}, \frac{\delta_{0,t}}{2} \right. \right) \int \beta \mathcal{N}_k(\beta | \beta_{0c,t}, \sigma^2 \mathbf{B}_{0,t}) d\beta d\sigma \quad (4.22)$$

$$= \int_{\sigma>0} \sigma^{-2} \mathcal{IG}_2 \left(\sigma \left| \frac{\nu_{0,t}}{2}, \frac{\delta_{0,t}}{2} \right. \right) \beta_{0c,t} d\sigma \quad (4.23)$$

$$= \beta_{0c,t} \int_{\sigma>0} \sigma^{-2} \mathcal{IG}_2 \left(\sigma \left| \frac{\nu_{0,t}}{2}, \frac{\delta_{0,t}}{2} \right. \right) d\sigma \quad (4.24)$$

$$= \frac{\nu_{0,t}}{\delta_{0,t}} \beta_{0c,t} \quad (4.25)$$

and

$$E_{c,t}[\sigma^{-2}] = \frac{\nu_{0,t}}{\delta_{0,t}} \quad (4.26)$$

Hence, the K-L divergence is

$$\frac{\nu_{0,t}}{2\delta_{0,t}} (\beta_{0c,t} - \beta_{0u,t})' \mathbf{B}_{0,t}^{-1} (\beta_{0c,t} - \beta_{0u,t}), \quad (4.27)$$

as stated in the text.

Appendix B

This appendix proves the result in Proposition 1.

Proof. For the optimization problem,

$$\begin{aligned} \min_{\beta_{0c,t}} & \frac{\nu_{0,t}}{2\delta_{0,t}} (\beta_{0c,t} - \beta_{0u,t})' \mathbf{B}_{0,t}^{-1} (\beta_{0c,t} - \beta_{0u,t}) \\ \text{s.t.} & \mathbf{z}'_t \beta_{0c,t} \geq 0 \end{aligned}$$

the objective function is quadratic and convex and the constraint is affine and satisfies Slater's condition. This implies that the Karush-Kuhn-Tucker (KKT) conditions are the

sufficient and necessary condition for optimality (Boyd and Vandenberghe, 2004).

Denote the solution to the problem as $\beta_{0c,t}$. Then the KKT conditions are

$$2(\beta_{0c,t} - \beta_{0u,t})' \mathbf{B}_{0,t}^{-1} - \lambda \mathbf{z}'_t = 0 \quad (4.28)$$

$$\mathbf{z}'_t \beta_{0c,t} \geq 0 \quad (4.29)$$

$$\lambda \mathbf{z}'_t \beta_{0c,t} = 0 \quad (4.30)$$

$$\lambda \geq 0 \quad (4.31)$$

First, it's easy to show that when $\mathbf{z}'_t \beta_{0u,t} \geq 0$, $\beta_{0c,t} = \beta_{0u,t}$ and the objective function value is 0. When $\mathbf{z}'_t \beta_{0u,t} < 0$, consider the following two cases,

1. $\mathbf{z}'_t \beta_{0c,t} > 0$. The KKT conditions then imply $\lambda = 0$. Hence $2(\beta_{0c,t} - \beta_{0u,t})' \mathbf{B}_{0,t}^{-1} = 0$. Matrix $\mathbf{B}_{0,t}$ is a covariance matrix and is always positive definite. Hence $\beta_{0c,t} = \beta_{0u,t}$. Therefore $\mathbf{z}'_t \beta_{0u,t} > 0$ and it is a contradiction with the assumption that $\mathbf{z}'_t \beta_{0u,t} < 0$.
2. $\mathbf{z}'_t \beta_{0c,t} = 0$. Multiply $\mathbf{B}_{0,t}$ from the right on both sides of Eq. (4.28) and we have

$$2(\beta_{0c,t} - \beta_{0u,t})' = \lambda \mathbf{z}'_t \mathbf{B}_{0,t} \quad (4.32)$$

Solve $\beta_{0c,t}$ from this equation and we have

$$\beta_{0c,t} = \beta_{0u,t} + \frac{\lambda}{2} \mathbf{B}_{0,t} \mathbf{z}_t \quad (4.33)$$

Substitute it into $\mathbf{z}'_t \beta_{0c,t} = 0$ and we have

$$\lambda = -\frac{2\mathbf{z}'_t \beta_{0u,t}}{\mathbf{z}'_t \mathbf{B}_{0,t} \mathbf{z}_t} \quad (4.34)$$

Note $\mathbf{z}'_t \mathbf{B}_{0,t} \mathbf{z}_t$ is strictly positive. As $\mathbf{B}_{0,t}$ is positive definite, $\mathbf{z}'_t \mathbf{B}_{0,t} \mathbf{z}_t$ can only be zero when $\mathbf{z}_t = 0$ and that will be a contradiction with $\mathbf{z}'_t \beta_{0u,t} < 0$. Therefore λ is positive. Substitute the λ solution in Eq. (4.34) back to Eq. (4.33) and we have

$$\beta_{0c,t} = \beta_{0u,t} - \frac{\mathbf{z}'_t \beta_{0u,t}}{\mathbf{z}'_t \mathbf{B}_{0,t} \mathbf{z}_t} \mathbf{B}_{0,t} \mathbf{z}_t \quad (4.35)$$

■

References

- Barberis, N. (2000), “Investing for the long run when returns are predictable,” *The Journal of Finance*, 55, 225–264.
- Boyd, S. and Vandenberghe, L. (2004), *Convex optimization*, Cambridge University Press.
- Campbell, J. Y. and Thompson, S. B. (2008), “Predicting excess stock returns out of sample: can anything beat the historical average?” *Review of Financial Studies*, 21, 1509–1531.
- Chib, S. and Greenberg, E. (1995), “Hierarchical analysis of SUR models with extensions to correlated serial errors and time-varying parameter models,” *Journal of Econometrics*, 68, 339–360.
- Cochrane, J. H. (2008), “The dog that did not bark: a defense of return predictability,” *Review of Financial Studies*, 21, 1533–1575.
- Dangl, T. and Halling, M. (2012), “Predictive regressions with time-varying coefficients,” *Journal of Financial Economics*, 106, 157–181.
- Geweke, J. and Amisano, G. (2012), “Prediction with misspecified models,” *American Economic Review*, 102, 482–486.
- Goyal, A. and Welch, I. (2008), “A comprehensive look at the empirical performance of equity premium prediction,” *Review of Financial Studies*, 21, 1455–1508.
- Henkel, S. J., Martin, J. S., and Nardari, F. (2011), “Time-varying short-horizon predictability,” *Journal of Financial Economics*, 99, 560–580.
- Johannes, M., Korteweg, A., and Polson, N. (2014), “Sequential learning, predictability, and optimal portfolio returns,” *The Journal of Finance*, 69, 611–644.

- Kandel, S. and Stambaugh, R. F. (1996), “On the predictability of stock returns: an asset-allocation perspective,” *The Journal of Finance*, 51, 385–424.
- Kelly, B. and Pruitt, S. (2013), “Market expectations in the cross-section of present values,” *The Journal of Finance*, 68, 1721–1756.
- (2015), “The three-pass regression filter: A new approach to forecasting using many predictors,” *Journal of Econometrics*, 186, 294–316.
- Kullback, S. and Leibler, R. A. (1951), “On information and sufficiency,” *The Annals of Mathematical Statistics*, 79–86.
- Li, J. and Tsiakas, I. (2015), “On equity premium and dividend growth predictability: economic fundamentals matter!” *Working Paper*.
- Ludvigson, S. C. and Ng, S. (2007), “The empirical riskreturn relation: a factor analysis approach,” *Journal of Financial Economics*, 83, 171–222.
- Mele, A. (2007), “Asymmetric stock market volatility and the cyclical behavior of expected returns,” *Journal of Financial Economics*, 86, 446–478.
- Merton, R. C. (1980), “On estimating the expected return on the market,” *Journal of Financial Economics*, 8, 323–361.
- Neely, C. J., Rapach, D. E., Tu, J., and Zhou, G. (2014), “Forecasting the equity risk premium: the role of technical indicators,” *Management Science*, 60, 1772–1791.
- Pettenuzzo, D., Timmermann, A., and Valkanov, R. (2014), “Forecasting stock returns under economic constraints,” *Journal of Financial Economics*, 114, 517–553.
- Rapach, D. E., Ringgenberg, M. C., and Zhou, G. (2015), “Short interest and aggregate stock returns,” *Journal of Financial Economics*, *forthcoming*.
- Rapach, D. E., Strauss, J. K., and Zhou, G. (2010), “Out-of-sample equity premium prediction: combination forecasts and links to the real economy,” *Review of Financial Studies*, 23, 821–862.

Robertson, J. C., Tallman, E. W., and Whiteman, C. H. (2005), “Forecasting using relative entropy,” *Journal of Money, Credit and Banking*, 37, 383–401.

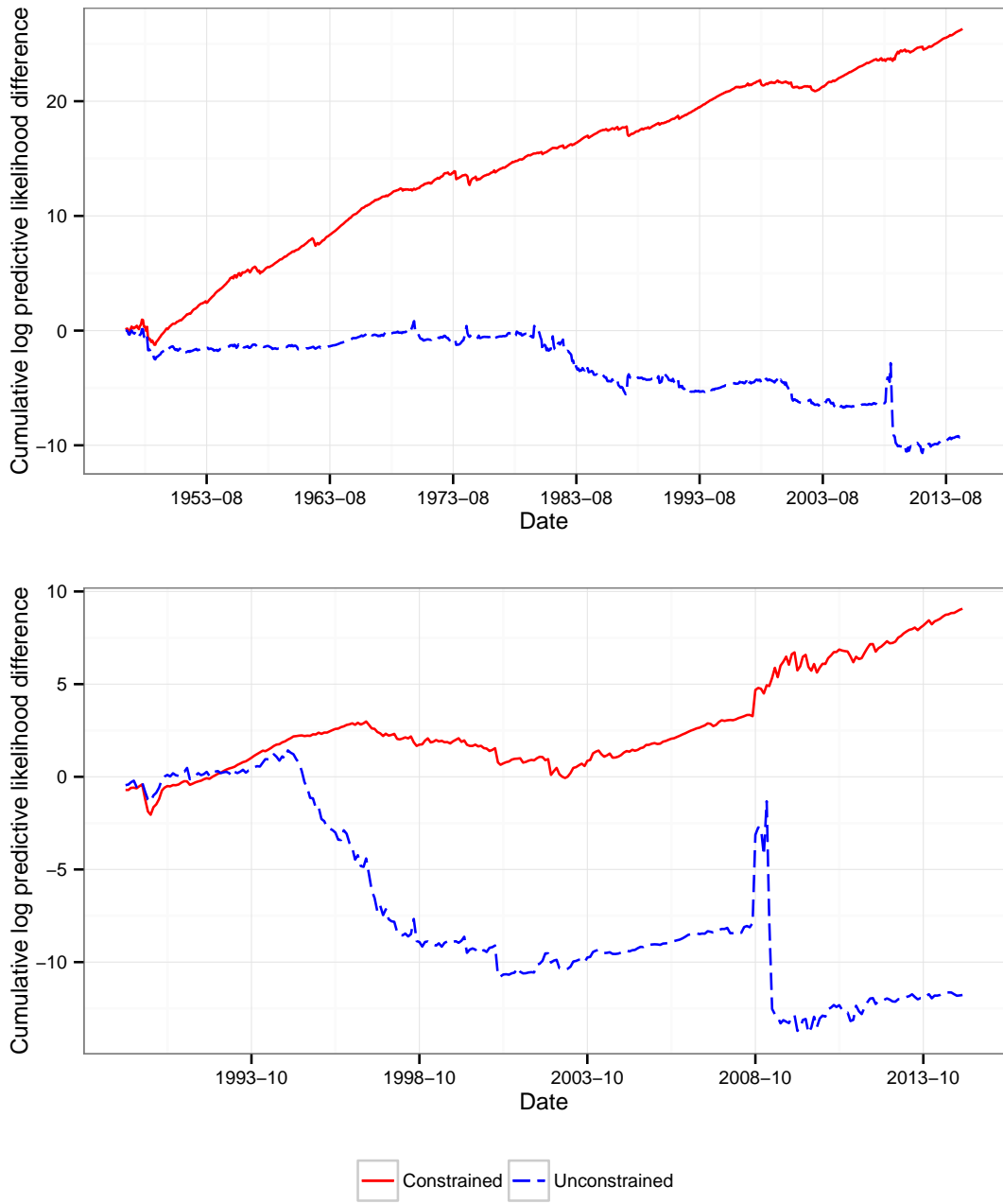


Figure 2: Cumulative log predictive likelihood difference of multiple predictive regressions

Cumulative log (base e) predictive likelihood difference of the constrained (solid) and the unconstrained (dashed) multiple predictive regression, relative to the no-predictability model. The cumulation is over 1947:01 to each month in 1947:01 to 2014:12 in Sample 1 (1927:01-2014:12) on the top. The cumulation is over 1990:01 to each month in 1990:01 to 2014:12 in Sample 2 (1973:01-2014:12) on the bottom.

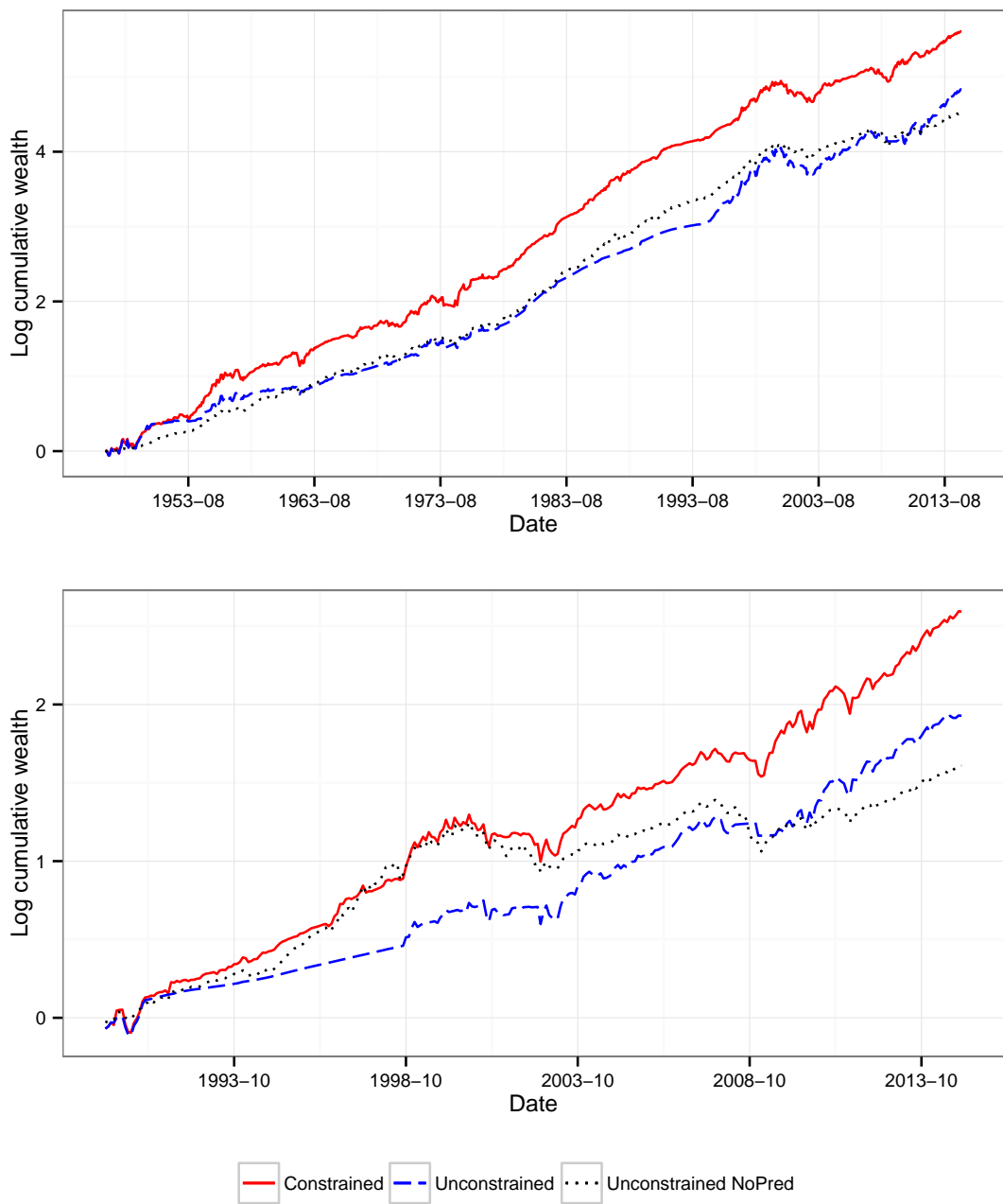


Figure 3: Cumulative wealth of investors forming portfolios by multiple predictive regressions

Investors form portfolios based on prediction from the constrained (solid) and the unconstrained (dashed) multiple predictive regression and the no-predictability model (dotted). They start with \$1 at the beginning of 1947:01 and reinvest all proceeds every month in the longer data sample (top), whereas start with \$1 at the beginning of 1990:01 in the shorter sample (bottom). Investors have relative risk aversion coefficient of five. The weight on the equity is restricted from 0 to 0.99.