TEXTO PARA DISCUSSÃO

No. 604

Let´s do it again: bagging equity premium predictors

Eric Hillebrand
Tae-Hwy Lee
Marcelo C. Medeiros

DEPARTAMENTO DE ECONOMIA
www.econ.puc-rio.br

Let's Do It Again: Bagging Equity Premium Predictors

Eric Hillebrand* CREATES

Tae-Hwy Lee† UC Riverside

Marcelo C. Medeiros‡ PUC Rio de Janeiro

September 2012

Abstract

The literature on excess return prediction has considered a wide array of estimation schemes, among them unrestricted and restricted regression coefficients. We consider bootstrap aggregation (bagging) to smooth parameter restrictions. Two types of restrictions are considered: positivity of the regression coefficient and positivity of the forecast. Bagging constrained estimators can have smaller asymptotic mean-squared prediction errors than forecasts from a restricted model without bagging. Monte Carlo simulations show that forecast gains can be achieved in realistic sample sizes for the stock return problem. In an empirical application using the data set of Campbell, J., and S. Thompson (2008): "Predicting the Equity Premium Out of Sample: Can Anything Beat the Historical Average?", *Review of Financial Studies* 21, 1511-1531, we show that we can improve the forecast performance further by smoothing the restriction through bagging.

Keywords: Constraints on predictive regression function; Bagging; Asymptotic MSE; Equity premium; Out-of-sample forecasting; Economic value functions.

JEL Classification: C5, E4, G1.

*Department of Economics and Business, Aarhus University, Denmark. E-mail: ehillebrand@creates.au.dk. EH acknowledges support by the Danish National Research Foundation.

†Department of Economics, University of California, Riverside, CA 92521-0427, U.S.A. E-mail: taelee@ucr.edu

‡Department of Economics, Pontifical Catholic University of Rio de Janeiro, Rio de Janeiro, Brazil. E-mail: mcm@econ.puc-rio.br

1 Introduction

In this paper we study bootstrap aggregation (bagging) methods to improve the out-of-sample prediction of a linear model for the equity premium by imposing restrictions in the regression. In imposing restrictions on a small linear regression model, a hard-threshold indicator function is usually employed, which can be smoothed by bagging. In regularizing a large linear model (with many predictors), subset selection of the predictors is necessary, often by using a regularized (penalized) linear regression such as ridge or LASSO (Tibshirani, 1996). Regularization requires the choice of a penalty or tuning parameter, which is often difficult. This choice can be smoothed by averaging over a set of many regularization parameters. In this paper we show in theory, in simulations, and in an empirical application, that bagging hard-thresholding parameter and forecast restrictions can improve the predictive power of a linear model. Following Campbell and Thompson (CT, 2008), we focus on a small linear regression. Using the same data set as CT (2008), we find substantial forecast improvements from bagging the constraint.

Excess returns prediction has attracted academics and practitioners for many decades since the early 1920s, when Dow (1920) studied the role of dividend ratios as a possible predictor for returns. In the 1980s, a number of authors presented empirical evidence of ex-post (in-sample) return predictability. Fama and Schwert (1977), Fama and Schwert (1981), Rozeff (1984), Keim and Stambaugh (1986), Campbell (1987), Campbell and Shiller (1988a,b) and Fama and French (1988, 1989) showed that excess returns could be successfully predicted based on lagged values of variables such as dividend-price ratio and dividend yield, earnings-price ratio and dividend-earnings ratio, interest rates and spreads, inflation rates, book-to-market ratio, volatility, investment-capital ratio, consumption, wealth, and income ratio, and aggregate net or equity issuing activity.

Subsequent work, however, demonstrated that these results do not hold during the bull market period of the 1990s; see Lettau and Ludvigson (2001) or Schwert (2002). For example, during this period when stock prices soared, the dividend yield systematically drifted downwards, thus generating negative sample correlation between returns and dividend yield, contrary to the positive historical correlation. Furthermore, since early results concerned only ex-post predictability, they were of little practical interest. Studies of ex-ante (out-of-sample) return predictability have found either that previous successful results were restricted to particular sub-samples (Pesaran and Timmermann 1995) or that return predictability was a statistical illusion; see Bossaerts and Hillion (1999). In addition, several authors pointed out that the apparent predictability of stock returns might be spurious as many of the predictor variables were highly persistent, leading to possibly biased coefficients and incorrect t-tests in predictive regressions; see, for example, Nelson and Kim (1993), Cavanagh, Elliot, and Stock (1995), and Stambaugh (1999). These problems are exacerbated when large numbers of variables are considered and only results that are apparently statistically significant are reported; see Foster, Smith, and Whaley (1997) and Ferson, Sarkissian, and Simin (2003).

The inconclusive evidence has inspired the use of time-varying regression models. As pointed out by Pesaran and Timmermann (2002) and Timmermann (2007) *"forecasters of stock returns face a moving target that is constantly changing over time. Just when a forecaster may think that he has figured out how to predict returns, the dynamics of market prices will, in all likelihood, have moved on, possibly as a consequence of the forecaster's own efforts."* On the other hand, alternative econometric methods were advocated for correcting the above mentioned bias and conducting valid inference: for example Cavanagh, Elliot, and Stock (1995), Mark (1995), Kilian (1999), Ang and Bekaert (2006), Jansson and Moreira (2006), Lewellen (2004), Torous, Valkanov, and Yan (2004), Campbell and Yogo (2006), and Polk, Thompson, and Vuolteenaho (2006).

More recently, Goyal and Welch (2008) argued that none of the conventional predictor variables proposed in the literature seems capable of systematically predicting stock returns out-of-sample. Their empirical evidence suggests that most models were unstable or spurious, and most models are no longer significant even in-sample. The authors show that the earlier apparent statistical significance was especially confined to the years of the Oil Shock of 1973–1975; see also Butler, Grullon, and Weston (2006).

We motivate our approach from CT (2008), who show that many predictive regressions outperform the historical average return forecast once a restriction is imposed on the sign of the coefficient in the regression. Imposing a constraint on the coefficient amounts to applying shrinkage estimation. Shrinkage methods are designed to increase bias and reduce variance in return. CT (2008) find out-of-sample predictive power of the common stock return predictors over the historical average. The advantage is small (not statistically significant) but nonetheless economically meaningful for mean-variance investors. We impose these a-priori parameter restrictions (which we call CT restrictions) in a regression function of equity premium conditional on various predictors, then we smooth the restrictions by bagging (Bühlmann and Yu, 2002, and Inoue and Kilian, 2008). The resulting bagging forecast has lower variance than the forecast using the CT restricted estimator. Whether mean-squared forecast error (MSFE) is also reduced in the process depends on how much bagging increases bias. We explore this question in simulations and in an application to the same data set used in CT (2008).

The paper is organized as follows. In Section 2 we review bagging, present four motivating example of the use of bagging in forecasting with restrictions, and present the bagging approach to restricted parameter estimation. In Section 3, we present a Monte Carlo simulation. Section 4 describes the data set and presents empirical results. Section 5 concludes.

2 Bagging restrictions on regression functions

A linear model assumes that the regression function $\mathbb{E}(y|\mathbf{X})$ is linear in the predictors $\mathbf{X} = (x_1, \ldots, x_k)'$. When k is small, such as $k = 1$, we may have good reason to believe that the coefficient of X must be positive or must exceed some known value. In that case, we may use the a-priori belief to *shrink* the parameter space. Such a-priori beliefs are less intuitive when k is large, so we restrict the consideration to the case $k = 1$. A simple method is to use a hard-thresholding indicator function to define a constrained least squared estimator $\bar{\beta} = \max(\tilde{\beta}, 0) = \tilde{\beta} \cdot \mathbf{1}(\tilde{\beta} > 0)$ with $\tilde{\beta}$ being an unconstrained least squares estimator of β. By shrinking the parameter space for the slope coefficient of x_1 in the prediction model towards zero, one reduces the variance, and possibly the overall mean squared forecast error, at the cost of bias. While imposing such a constraint can improve the predictive power of the linear model if the constraint is correct, the restricted estimator $\bar{\beta}$ involves a discontinuous hard-threshold indicator (jump) function at the boundary of the constrained parameter space. We show that we can further improve the predictive ability of the constrained linear model by smoothing the indicator function using bagging. We consider two types of restrictions as considered in CT (2008), namely positivity of the forecast, motivated by the requirement that the mean of the equity premium should be positive, and positivity of the regression coefficient, motivated by simple insights from financial theory, for example that the dividend yield should have a positive influence on the equity premium.

2.1 Motivating examples

Example 1: Forecasting equity premium. Goyal and Welch (GW 2008) show that predictive regressions cannot beat the historical average. Campbell and Thompson (CT 2008), on the other hand, show that many predictive regressions beat the historical average, once constraints are imposed. Consider

$$y_{t+1} = \alpha + \beta x_t + u_{t+1}, \quad t = 1, \ldots, T, \tag{1}$$

where y is the excess return on the S&P500 over the 3-month T-bill interest rate. The regressor x_t stands for a predictor variable such as dividend yield, earnings yield, book-to-market ratio, return on equity, long-term government bond yield, term spread, default spread, inflation rate, equity share of new issues, etc. GW find that forecasts from the unrestricted model $\tilde{y}_{n+1} = \tilde{\alpha}_n + \tilde{\beta}_n x_n$ (with $\tilde{\alpha}_n, \tilde{\beta}_n$ unrestricted OLS) are worse than forecasts with the exclusion restriction $\beta = 0$, which amounts to the historical average (HA) $\frac{1}{n}\sum_{t=1}^{n} y_t$. CT show that the positivity constraint $\beta > 0$ produces a better forecast than the exclusion restriction. In this paper we aim to show that bagging can further improve the CT constrained forecast.

Example 2: Forecasting inflation. Stock and Watson (2007) use the linear Phillips-curve

$$y_{t+1} = \alpha + \beta x_t + \gamma(B) y_t + u_{t+1}, \tag{2}$$

where $y = \Delta\pi$, π is inflation measured as change in a log-price index such as GDP deflator, personal consumption expenditure (PCE) deflator for core items, PCE deflator for all items, or CPI. The regressor x captures change in macroeconomic activity, such as unemployment rate, logarithm of real GDP, capacity utilization rate, building permits, or the Chicago Fed National Activity Index. Stock and Watson (2007) show that the unrestricted forecast from the model is outperformed by the forecast generated from imposing the exclusion restriction $\beta = 0$ and the additional restriction $\gamma(B) = -\theta(1-\theta B)^{-1}$, which specifies an IMA(1,1) model $y_{t+1} = \mu + (1-\theta B)u_{t+1}$. Here, one may as well consider a strict positivity constraint $\beta > 0$.

Example 3: Forecasting output growth. Faust and Wright (2009) consider the model

$$y_{t+1} = \alpha + \beta x_t + \gamma(B) y_t + u_{t+1}, \tag{3}$$

where $y_{t+1} = \ln(Y_{t+k}/Y_t)$ is real GDP growth from time t to $t+k$, and x_t is the term spread between 10-year government bond yields and 3-month T-bill rates. Faust and Wright (2009) argue that the unrestricted forecast from the above model is dominated by the forecast with the exclusion restriction $\beta = 0$, which makes the model a simple AR model. Again, $\beta > 0$ is an interesting alternative constraint.

Example 4: Forecasting recession. Using the same variables for y and x as in Example 3, Estrella and Hardouvelis (1991) and Wheelock and Wohar (2009) consider the conditional probability model

$$\Pr(y_{t+1} < 0|x_t) = \Phi\left(\alpha + \beta x_t\right). \tag{4}$$

In line with CT (2008), a natural question is whether the constraint $\beta > 0$ improves the predictive ability of x.

2.2 Bagging

Bootstrap aggregating, or bagging, means to average an estimator over a number J of subsamples drawn from the original data set $\mathcal{D}$. As $J \to \infty$, the bagged estimator will differ from the estimator obtained from the entire data set $\mathcal{D}$ only if it is a nonlinear or adaptive estimator (Hastie, Tibshirani, and Friedman, 2001, p. 246). An estimator is said to be "unstable" if a small change in the training set will lead to a significant change in the estimator (Breiman, 1996). In our application to an indicator function, the bagged predictor smoothes the instability caused by estimation and model uncertainty and the hard threshold function.

The mechanism of bagging has been explained in various ways, for example Breiman (1996) under squared-error loss and Lee and Yang (2006) under convex loss (e.g., a tick function for quantiles). Bühlmann and Yu (2002) show that for a nonsmooth unstable predictor, bagging reduces variance of the first order term. In particular, they show that bagging can reduce the mean-squared forecast error by averaging over the randomness of variable selection. Buja and Stuetzle (2006) and Friedman and Hall (2007) expand a smooth unstable function into linear and higher order terms, and show bagging reduces the variance of the higher order terms. Grandvalet (2004) argues that bagging stabilizes prediction by equalizing the influence of training samples. Stock and Watson (2012) show that bagging is asymptotically equivalent to Bayesian shrinkage.

Applications of bagging include inflation (Inoue and Kilian 2008), financial volatility (Hillebrand and Medeiros 2010), equity premium (Huang and Lee 2010), short-term interest rates (Audrino and Medeiros 2011), and employment data (Rapach and Strauss 2007).

To fix notation, let

$$\mathcal{D}_t = \{(Y_s, \mathbf{X}_{s-1})\}_{s=t-R+1}^{t} \qquad (t = R, \ldots, T) \tag{5}$$

be a training set at time t and let $\varphi(\mathbf{X}_t, \mathcal{D}_t)$ be a forecast of Y_{t+1} or of the binary variable $\mathbf{1}(Y_{t+1} \geq 0)$ using this training set $\mathcal{D}_t$ and the explanatory variable vector $\mathbf{X}_t$. The optimal forecast $\varphi(\mathbf{X}_t, \mathcal{D}_t)$ for Y_{t+1} will be the conditional mean of Y_{t+1} given $\mathbf{X}_t$ if we have the squared error loss function, or the conditional quantile of Y_{t+1} on $\mathbf{X}_t$ if the loss is a tick function as in Koenker and Basset (1978).

Suppose each training set $\mathcal{D}_t$ consists of R observations generated from the underlying probability distribution $\mathbf{P}$. The forecast $\{\varphi(\mathbf{X}_t, \mathcal{D}_t)\}_{t=R+1}^{T}$ can be improved if more training sets can be generated from $\mathbf{P}$ and if the forecast can be formed from averaging the multiple forecasts obtained from the multiple training sets. Ideally, if $\mathbf{P}$ were known and multiple training sets $\mathcal{D}_t^{(j)}$ $(j = 1, \ldots, J)$ could be drawn from $\mathbf{P}$, an ensemble aggregating predictor $\varphi_A(\mathbf{X}_t)$ could be constructed by averaging of $\varphi(\mathbf{X}_t, \mathcal{D}_t^{(j)})$ with respect to $\mathbf{P}$, i.e.,

$$\varphi_A(\mathbf{X}_t) = \mathbb{E}_{\mathbf{P}} \varphi(\mathbf{X}_t, \mathcal{D}_t), \tag{6}$$

where $\mathbb{E}_{\mathbf{P}}(\cdot)$ denotes expectation with respect to $\mathbf{P}$, and the subscript A in φ_A denotes "aggregation."

In practice, $\mathbf{P}$ is not known. We may estimate $\mathbf{P}$ by its empirical distribution, $\hat{\mathbf{P}}(\mathcal{D}_t)$, for a given data set $\mathcal{D}_t$. Then, from the empirical distribution $\hat{\mathbf{P}}(\mathcal{D}_t)$, multiple subsamples $\mathcal{D}_t^*$ can be drawn by an appropriate bootstrap method. The question which bootstrap algorithms can provide consistent densities for moment estimators and quantile estimators in time series settings is addressed, for example, in Hall, Horowitz, and Jing (1995) and Fitzenberger (1997). Bagging predictors, $\varphi^B(\mathbf{X}_t, \mathcal{D}_t^*)$, can then be computed by averaging over the subsamples. More specifically, the bagging predictor $\varphi^B(\mathbf{X}_t, \mathcal{D}_t^*)$ can be obtained following the steps: (1) Given a training set of data at time t, $\mathcal{D}_t = \{(Y_s, \mathbf{X}_{s-1})\}_{s=t-R+1}^{t}$, construct the jth bootstrap sample $\mathcal{D}_t^{*(j)} = \{(Y_s^{*(j)}, \mathbf{X}_{s-1}^{*(j)})\}_{s=t-R+1}^{t}$, $j = 1, \ldots, J$, according to the empirical distribution of $\hat{\mathbf{P}}(\mathcal{D}_t)$ of $\mathcal{D}_t$. (2) Compute the bootstrap predictor $\varphi^{*(j)}(\mathbf{X}_t, \mathcal{D}_t^{*(j)})$ from the jth bootstrapped sample $\mathcal{D}_t^{*(j)}$. (3) Compute the bagging predictor $\varphi^B(\mathbf{X}_t, \mathcal{D}_t^*)$ by averaging over J bootstrap predictors

$$\varphi^B(\mathbf{X}_t, \mathcal{D}_t^*) = \frac{1}{J}\sum_{j=1}^{J} \varphi^{*(j)}(\mathbf{X}_t, \mathcal{D}_t^{*(j)}). \tag{7}$$

2.3 Bagging Restrictions

Returning to Example 1 we let

$$\varphi(x_t,\mathcal{D}_t) = \mathbb{E}(y_{t+1}|x_t) = \alpha + \beta x_t,$$

where $\mathcal{D}_t = \{(y_{s+1}, x_s)\}_{s=t-R+1}^{t}$ for $t = R, \ldots, T$.

The two types of restrictions considered in CT (2008) are positivity of the coefficient β (PC) and positivity of the forecast $\varphi(x_t,\mathcal{D}_t)$ (PF). We compare the following forecasts:

1. **HA** (Historical Average forecast with exclusion restriction $\beta = 0$):

$$y_{T+1}^{\text{HA}} = \frac{1}{T}\sum_{t=T-R+1}^{T} y_t. \tag{8}$$

2. **UF** (Unrestricted Forecast):

$$y_{T+1}^{\text{UF}} = \tilde{\alpha}_T + \tilde{\beta}_T x_T, \tag{9}$$

where $\tilde{\alpha}_T, \tilde{\beta}_T$ are unrestricted OLS estimators. UF is used in Goyal and Welch (2008).

3. **PC** (forecast with Positive Coefficient restriction $\beta > 0$):

$$y_{T+1}^{\text{PC}} = \bar{\alpha}_T + \bar{\beta}_T x_T, \tag{10}$$

where $\bar{\beta}_T = \max\{\tilde{\beta}_T, 0\} = \mathbf{1}\left(\tilde{\beta}_T > 0\right)\tilde{\beta}_T$, and $\bar{\alpha}_T = \mathbf{1}\left(\tilde{\beta}_T > 0\right)\tilde{\alpha}_T + \mathbf{1}\left(\tilde{\beta}_T \leq 0\right) y_{T+1}^{\text{HA}}$. PC is used in CT (2008).

4. **PC-GH** (forecast with Positive Coefficient restriction $\beta > 0$ using bagging as in Gordon and Hall (GH), 2009):

$$y_{T+1,J}^{\text{PC-GH}} = \frac{1}{J}\sum_{j=1}^{J}\left(y_{T+1}^{\text{PC},*(j)}\right) = \hat{\alpha}_{T,J} + \hat{\beta}_{T,J} x_T, \tag{11}$$

where $\hat{\beta}_{T,J} = \frac{1}{J}\sum_{j=1}^{J}\bar{\beta}_T^{*(j)}$ and $\hat{\alpha}_{T,J} = \frac{1}{J}\sum_{j=1}^{J}\bar{\alpha}_T^{*(j)}$. As $J \to \infty$, $\hat{\beta}_{T,J}$ converges to $\mathbb{E}^*\bar{\beta}_T$, $\hat{\alpha}_{T,J}$ converges to $\mathbb{E}^*\bar{\alpha}_T$, and $y_{T+1,J}^{\text{PC-GH}}$ converges to $\mathbb{E}^*(y_{T+1}^{\text{PC}})$, where $\mathbb{E}^*$ denotes expectation with respect to the empirical distribution $\hat{\mathbf{P}}(\mathcal{D}_T)$.

5. **PF** (forecast with Positive Forecast restriction $\varphi(x_t,\mathcal{D}_t) > 0$):

$$y_{T+1}^{\text{PF}} = \mathbf{1}\left(y_{T+1}^{\text{UF}} > 0\right) y_{T+1}^{\text{UF}}. \tag{12}$$

6. **PF-GH** (forecast with Positive Forecast restriction $\varphi(x_t,\mathcal{D}_t) > 0$ using bagging as in GH 2009):

$$y_{T+1,J}^{\text{PF-GH}} = \frac{1}{J}\sum_{j=1}^{J}\left(y_{T+1}^{\text{PF},*(j)}\right), \tag{13}$$

where again there is convergence $\lim_{J\to\infty} y_{T+1,J}^{\text{PF-GH}} = \mathbb{E}^*(y_{T+1}^{\text{PF}})$, $\mathbb{E}^*(\cdot)$ being expectation with respect to the empirical distribution $\hat{\mathbf{P}}(\mathcal{D}_T)$.

With this notation, we can rephrase the aim of this paper. Comparative statements can be understood in the mean-square error sense, but we also consider two alternative loss functions in the empirical exercise.

Objective 1. *GW (2008) find that unrestricted forecasts $y_{n+1}^{UF} = \tilde{\alpha}_n + \tilde{\beta}_n x_n$ (with $\tilde{\alpha}_n, \tilde{\beta}_n$ OLS estimators) are worse than HA forecasts (with exclusion restriction $\beta = 0$). CT (2008) show that PC and PF produce better forecasts than HA. We will show that PC-GH and PF-GH further improve PC and PF.*

The reason for this improvement is the following equivalence that is motivated by the second equality in (11).

Proposition 1. *Bagging the positive coefficient forecast y_{T+1}^{PC} is equivalent to computing the forecast $y_{T+1,J}^{PC\text{-}GH}$ from the Gordon and Hall (2009) bagging estimators $\hat{\alpha}_{T,J}$, $\hat{\beta}_{T,J}$. That is,*

$$\frac{1}{J}\sum_{j=1}^{J} y_{T+1}^{PC,*(j)} = y_{T+1,J}^{PC\text{-}GH}. \tag{14}$$

Proof.

$$\begin{aligned}\frac{1}{J}\sum_{j=1}^{J} y_{T+1}^{\text{PC},*(j)} &= \frac{1}{J}\sum_{j=1}^{J}(\bar{\alpha}_T^{*(j)} + \bar{\beta}_T^{*(j)} x_T) = \left(\frac{1}{J}\sum_{j=1}^{J}\bar{\alpha}_T^{*(j)}\right) + \left(\frac{1}{J}\sum_{j=1}^{J}\bar{\beta}_T^{*(j)}\right) x_T \\ &= \hat{\alpha}_T + \hat{\beta}_T x_T = y_{T+1,J}^{\text{PC-GH}}. \qquad \square\end{aligned}$$

Even though Proposition 1 is a very simple insight, it provides a powerful reason, in view of Breiman (1996), why bagging the coefficient estimates and obtaining $y_{T+1}^{\text{PC-GH}}$ can improve the PC constrained forecast y_{T+1}^{PC} used in CT (2008).

2.4 AMSE Comparison

With Objective 1 in view, we compare the asymptotic mean-square error (AMSE) of the unrestricted, the restricted, and the bagging estimator and forecast. Keeping the decomposition of MSE into variance and bias in mind, we show that under certain circumstances, bagging estimators and forecasts have a shrinkage advantage over simple constrained estimators and forecasts as used in CT (2008). That is, their reduction in variance outweighs their increase in bias. For this purpose, we collect some results from the literature, in particular from Bühlmann and Yu (2002) and from Gordon and Hall (2009), and express them in a unified framework.

Let θ denote either a prediction $\varphi(x_t,\mathcal{D}_t)$ from a regression or the parameter vector (α, β). The unrestricted estimator $\tilde{\theta}_T$ is, thus,

$$\tilde{\theta}_T = (\tilde{\alpha}_T,\, \tilde{\beta}_T),\ \text{or}\ \tilde{\theta}_T = \tilde{\alpha}_T + \tilde{\beta}_T x_T. \tag{15}$$

The restricted estimator for θ subject to a lower bound θ_1 is

$$\bar{\theta}_T = \max\{\tilde{\theta}_T, \theta_1\}. \tag{16}$$

The Gordon and Hall (GH, 2009) bagging estimator is

$$\hat{\theta}_T = \lim_{J\to\infty} \frac{1}{J} \sum_{j=1}^{J} \max\{\tilde{\theta}_T^{*(j)}, \theta_1\} = \lim_{J\to\infty} \frac{1}{J} \sum_{j=1}^{J} \bar{\theta}_T^{*(j)} = \mathbb{E}(\max\{\tilde{\theta}_T^{*}, \theta_1\}|\mathcal{D}_T) \tag{17}$$

for the situation where a lower bound θ_1 is known. As before, $\mathcal{D}_T$ is the available data set at time T, $\mathcal{D}_T^*$ is a bootstrap sample, and $\tilde{\theta}_T^*$ is a bootstrap replication of $\tilde{\theta}_T$ from $\mathcal{D}_T^*$. There are J such bootstrap replications; expectation statements hold for $J \to \infty$; in practice J is of course finite. We use the moving block bootstrap method for subsampling.

Consider the case where θ is the regression coefficient β. Let the data-generating process be

$$y_{t+1} = \alpha_0 + \beta_0 x_t + u_{t+1}, \tag{18}$$

where $\mathbb{E}(u_t) = 0$, $\mathbb{V}(u_t) = \sigma_u^2 < \infty$. Let the data-generating slope parameter be of a local-to-threshold form

$$\beta_0 = \beta_1 + b\sigma_\beta T^{-1/2}. \tag{19}$$

The estimated model is a simple linear regression of y_{t+1} onto x_t, and the estimators under consideration are the OLS estimator $\tilde{\beta}_T$ of UF, the simple PC constrained estimator

$$\bar{\beta}_T = \max\left\{\tilde{\beta}_T \mathbf{1}\left(\tilde{\beta}_T > \beta_1 + c\sigma_\beta T^{-1/2}\right),\ \beta_1\right\} = \max\left\{\tilde{\beta}_T \mathbf{1}\left(\sqrt{T}(\tilde{\beta}_T - \beta_1)/\sigma_\beta > c\right),\ \beta_1\right\}, \tag{20}$$

and the bagging estimator PC-GH

$$\hat{\beta}_T = \frac{1}{J}\sum_{j=1}^{J}\bar{\beta}_T^{*(j)} = \frac{1}{J}\sum_{j=1}^{J}\max\left\{\tilde{\beta}_T^{*(j)}\mathbf{1}\left(\sqrt{T}(\tilde{\beta}_T^{*(j)}-\beta_1)/\sigma_\beta > c\right),\ \beta_1\right\}. \tag{21}$$

In this setup, c is a critical value that the scaled and studentized difference between the estimator and the threshold value β_1 has to exceed before the estimator is adopted over the threshold value. In Bühlmann and Yu (2002) and in Gordon and Hall (2009) one finds the following results on the asymptotic distributions of the estimators and their dependence on the data-generating perturbation b and on the critical decision value c.

Proposition 2. *[Bühlmann and Yu (2002), Gordon and Hall (2009)]*
Let $Z \sim \mathsf{N}(0,1)$, $\Phi(\cdot)$ the cumulative distribution function and $\phi(\cdot)$ the probability density function of the standard normal distribution. Then, as $T \to \infty$,

$$\sqrt{T}\sigma_\beta^{-1}(\tilde{\beta}_T - \beta_1) \xrightarrow{d} Z + b. \tag{22}$$

$$\sqrt{T}\sigma_\beta^{-1}(\bar{\beta}_T - \beta_1) \xrightarrow{d} (Z+b)\mathbf{1}\left(Z+b-c>0\right). \tag{23}$$

$$\sqrt{T}\sigma_\beta^{-1}(\hat{\beta}_T - \beta_1) \xrightarrow{d} (Z+b)\Phi(Z+b-c)+\phi(Z+b-c). \tag{24}$$

The positive coefficient (PC) constraint in (10) means that $c = 0$ and $\beta_1 = 0$. For this case, we compare the asymptotic mean-squared error (AMSE) of the limiting random variables in (22), (23), and (24) of Proposition 2.

Asymptotic bias (Abias): For all $z \in \mathbb{R}$, note that

$$z + b \le (z+b)\,\mathbf{1}(z+b>0) < (z+b)\Phi(z+b)+\phi(z+b),$$

which results in the following order of the asymptotic biases:

$$\text{Abias of } \tilde{\beta}_T \le \text{Abias of } \bar{\beta}_T < \text{Abias of } \hat{\beta}_T,$$

where

$$\text{Abias of } \tilde{\beta}_T = \mathbb{E}(Z+b) - b = 0,$$

$$\text{Abias of } \bar{\beta}_T = \mathbb{E}\left[(Z+b)\,\mathbf{1}(Z+b>0)\right] - b,$$

$$\text{Abias of } \hat{\beta}_T = \mathbb{E}\left[(Z+b)\Phi(Z+b)+\phi(Z+b)\right] - b.$$

Therefore, Abias gets worse as we impose the restriction and as we add bagging. For example, when $b = 0$,

$$\begin{aligned}
&\text{Abias of } \tilde{\beta}_T = \mathbb{E}(Z) = 0, \\
&\text{Abias of } \bar{\beta}_T = \mathbb{E}\left[Z\mathbf{1}(Z > 0)\right] = 0.5\sqrt{2/\pi} \approx 0.3989, \\
&\text{Abias of } \hat{\beta}_T = \mathbb{E}\left[Z\Phi(Z) + \phi(Z)\right] = \pi^{-1/2} \approx 0.5642.
\end{aligned}$$

Asymptotic variance (Avar): While Abias increases, asymptotic variance is reduced by imposing a restriction and more so by bagging the restriction. The asymptotic distribution of the simple constrained estimator $\bar{\beta}_T$ is a standard normal truncated to the positive half-line and thus has asymptotic variance $\mathbb{V}(Z\mathbf{1}(Z > 0)) = (1 - 1/\pi)/2) = 0.3408$. The Avar of the bagging estimator is $\mathbb{V}(Z\Phi(Z) + \phi(Z)) = 1/3 + \sqrt{3}/(2\pi) - 1/\pi = 0.2907$.

$b = 0$	Abias	Avar	AMSE
$\tilde{\beta}_T$	0	1	1
$\bar{\beta}_T$	0.3989	0.3408	0.4998
$\hat{\beta}_T$	0.5642	0.2907	0.6088

AMSE($\hat{\beta}_T$) and AMSE($\bar{\beta}_T$) are substantially smaller than AMSE($\tilde{\beta}_T$). However, note that the ratio AMSE($\hat{\beta}_T$)/AMSE($\bar{\beta}_T$) = 1.2181. The AMSE of PC-GH is worse due to much larger Abias($\hat{\beta}_T$). However, the case $\beta = 0$ (i.e., $b = 0$) is not an interesting one. We impose the restriction $\beta > 0$ (i.e., $b > 0$). For other values of b, analytical evaluations are difficult and we use numerical evaluation.

AMSE Comparison (When shrinking just right): When $\beta > 0$ (i.e., $b > 0$), if we impose the *correct* restriction $\beta > 0$, we can improve AMSE. For example, when $b = 1$:

$b = 1$	Abias	Avar	AMSE
$\tilde{\beta}_T$	0.0000	1.0000	1.0000
$\bar{\beta}_T$	0.0834	0.7512	0.7581
$\hat{\beta}_T$	0.1997	0.6045	0.6444

Here, AMSE($\hat{\beta}_T$)/AMSE($\bar{\beta}_T$) = 0.85: PC-GH reduces AMSE by 15% from PC. When $b = 2$:

$b = 2$	Abias	Avar	AMSE
$\tilde{\beta}_T$	0.0000	1.0000	1.0000
$\bar{\beta}_T$	0.0084	0.9601	0.9602
$\hat{\beta}_T$	0.0502	0.8562	0.8587

Here, AMSE($\hat{\beta}_T$)/AMSE($\bar{\beta}_T$) = 0.8943, and thus PC-GH is about 11% better than PC.

AMSE Comparison (When shrinking too little): When $\beta \gg 0$ (i.e., $b \gg 0$), we impose the *obvious* restriction $\beta > 0$. For example, when $b = 3$:

$b = 3$	Abias	Avar	AMSE
$\tilde{\beta}_T$	0.0000	1.0000	1.0000
$\bar{\beta}_T$	0.0005	0.9976	0.9976
$\hat{\beta}_T$	0.0088	0.9674	0.9674

In this case, when the restriction is hardly binding, the gain is small. Bagging makes a minor contribution for $\text{AMSE}(\hat{\beta}_T)/\text{AMSE}(\bar{\beta}_T) = 0.9698$. When the restriction becomes even more obviously correct with $b = 4$:

$b = 4$	Abias	Avar	AMSE
$\tilde{\beta}_T$	0.0000	1.0000	1.0000
$\bar{\beta}_T$	0.0000	1.0000	1.0000
$\hat{\beta}_T$	0.0008	0.9954	0.9954

the gain becomes even smaller. PC is the same as UF in AMSE. PC-GH is only slightly better than PC with $\text{AMSE}(\hat{\beta}_T)/\text{AMSE}(\bar{\beta}_T) = 0.9954$.

AMSE Comparison (When shrinking too much): When $\beta < 0$ (i.e., $b < 0$), we impose the *wrong* restriction $\beta > 0$. In this case, the increase in Abias dominates the reduction in Avar. AMSE substantially deteriorates by imposing the wrong constraint and more so by bagging the constraint:

$b = -1$	Abias	Avar	AMSE
$\tilde{\beta}_T$	-0.0003	1.0007	1.0007
$\bar{\beta}_T$	1.0833	0.0684	1.2421
$\hat{\beta}_T$	1.1996	0.0840	1.5231

$b = -2$	Abias	Avar	AMSE
$\tilde{\beta}_T$	-0.0005	1.0007	1.0007
$\bar{\beta}_T$	2.0085	0.0057	4.0399
$\hat{\beta}_T$	2.0503	0.0136	4.2171

AMSFE Comparison We formulated Proposition 2 as a statement about an estimator, but the results in Bühlmann and Yu (2002) and Gordon and Hall (2009) apply to prediction of y as well, and the asymptotic mean-squared forecast error can be compared. Given a predictor x, denote a forecast of $\mathbb{E}(y_{T+1}|x_T = x)$ by $\theta_T(x)$. Then, the unrestricted forecast (UF) is

$$\tilde{\theta}_T(x) = \tilde{\beta}_T x; \tag{25}$$

the positive coefficient forecast (PC) is

$$\bar{\theta}_T(x) = \tilde{\beta}_T \, \mathbf{1}(\tilde{\beta}_T > \beta_1 + b\sigma_\beta T^{-1/2}) \; x. \tag{26}$$

The bagging forecast PC-GH is

$$\hat{\theta}_{T,J}(x) = \frac{1}{J}\sum_{j=1}^{J} \bar{\theta}_T^{*(j)}(x). \tag{27}$$

3 Simulation

Simulation design: In order to evaluate the performance of the restricted and bagging predictors, we construct a simulation experiment that is motivated by the stock-return prediction problem. First, we generate $\{w_t\}_{t=1}^{T=200}$ from

$$w_t = \rho w_{t-1} + e_t, \qquad e_t \sim NID\left(0, \sigma_e^2\right), \sigma_e = 0.2, \tag{28}$$

where

$$\rho \in \{0, 0.5, 0.9, 0.99\}, \tag{29}$$

setting different levels of persistence of the regressor. Next, we let

$$x_t = \frac{\exp(w_t)}{\mathrm{std}(\exp(w_t))}, \tag{30}$$

so that $\{x_t\}$ be positive to mimic the predictor variables that we will consider in the equity premium prediction in Section 4. The predictor x is normalized with the standard deviation $\mathrm{std}(\cdot)$. Then, we generate $\{y_t\}_{t=1}^{T=200}$ from

$$y_t = 0.02 + \beta x_{t-1} + u_t, \qquad u_t \sim NID\left(0, \sigma_u^2\right), \ \mathbb{E}(u_t e_s) = 0, \forall t, s, \tag{31}$$

where

$$\sigma_u \in \{0.01, 0.05, 0.10, 0.50, 1.00\}. \tag{32}$$

The data-generating value of β converges with T to the bound β_1, and the parameter b is a local perturbation away from the bound with rate $T^{-1/2}$:

$$\beta = \beta_1 + b\ \sigma_\beta\ T^{-1/2}.$$

We set the values of β_1, σ_β, and b as follows

$$\beta_1 = 0, \tag{33}$$

$$\sigma_\beta^2 = \sigma_u^2 \left(\sum_{t=1}^{100} (x_t - \bar{x})^2\right)^{-1}, \tag{34}$$

$$b \in \{1, 3, 5, 10, 15, 20, 30, 50, 100\}. \tag{35}$$

By different values of b (and therefore different β), we consider different signal-to-noise ratios.

We use the first half of the total 200 to estimate β using the unrestricted, restricted, and the bagging (PC-GH) estimators. The PC-GH estimator is computed over $J = 200$ bootstrap samples. Using each of the above estimators β_t at time $t = 100, \ldots, 199$, we compute the unrestricted, restricted, and bagging forecasts $\theta_{t+1}(x_t)$ of y_{t+1}. The forecasts over the second half of observations are compared with the actual value y_{t+1}. We compute the following out-of-sample predictive ability measure of each model

$$100 \cdot R^2_{\mathrm{OS}} = 100 \left(1 - \frac{\frac{1}{100}\sum_{t=101}^{200}(y_t - \theta_t(x_{t-1}))^2}{\frac{1}{100}\sum_{t=101}^{200}(y_t - \theta_t^{\mathrm{HA}})^2} \right), \tag{36}$$

where the historical average $\theta_t^{\mathrm{HA}} = \frac{1}{100}\sum_{s=t-100}^{t-1} y_s$ (for $t = 101, \ldots, 200$) is taken as a benchmark forecast as in CT (2008). The same statistic $\left(100 \cdot R^2_{\mathrm{OS}}\right)$ was used in CT (2008) to compare various forecast models. We repeat the steps above over 1000 Monte Carlo replications and compute theaverage of the out-of-sample $\left(100 \cdot R^2_{\mathrm{OS}}\right)$.

Simulation results: In this simulation, unlike in the empirical application to equity premium prediction, the unrestricted forecast UF always dominates HA by construction of the simulation design with $b > 0$. Hence, it is not interesting to compare the forecasts with HA for this simulation section as we do it in the empirical section. Here we present the gain in in $100 \cdot R^2_{\mathrm{OS}}$ over the unrestricted forecast UF from imposing a constraint and from bagging. Figures 1-5 report this gain, defined as

$$\text{Gain-in-}R^2 = \left(100 \cdot R^2_{\mathrm{OS}}\right)_{\text{model}} - \left(100 \cdot R^2_{\mathrm{OS}}\right)_{\mathrm{UF}}, \tag{37}$$

where *model* = PC, PC-GH, PF, or PF-GH. Each of Figures 1 through 5 reports the gain in R^2 for one of the five values of σ_u in (32). The four panels of each figure show the situation for one of the four different values of ρ in (29). In each panel, the abscissa shows the different values of b in (35), and the ordinate shows the gain as defined in (37) for that specific 3-tuple (σ_u, ρ, b). Summarizing, we make the following observations.

1. For a wide range of b, the gains over UF from imposing the PC constraint and PF constraint are positive.

2. For a wide range of b, bagging further improves the constrained forecasts. PC is further improved by PC-GH, and PF is further improved by PF-GH.

3. The PF constraint does not operate much when σ_u is small (Figures 1-2), indicating that the PF constraint may be more useful when the market becomes more volatile (Figures 3-5).

4. From (34), note that high persistence (high ρ) leads to a large value of $\sum_{t=1}^{100}(x_t - \bar{x})^2$ and thus reduces σ_β, which reduces the local drift $\left(b\ \sigma_\beta\ T^{-1/2}\right)$ from the bound. We do not observe pronounced changes in the effects of the constraint or of bagging by varying ρ.

4 Let's Do It Again: Equity Premium Prediction

4.1 Data and Constraints

Data: We use the data set of Campbell and Thompson (2008), which was kindly provided by Sam Thompson. The data frequency is monthly; the sample period is 1871-2005. Excess returns on the S&P 500 are calculated from the returns time series (1871M2 through 2005M12, CRSP since 1927) and the 3-month Treasury-Bill interest rate (denoted as r_t^f, 1920M1 through 2005M12, 1870M2 through 1919M12 calculated following Goyal and Welch (2008)). The predictor variables are the dividend yield ***d/p*** (1872M2 through 2005M12), earnings yield ***e/p*** (1872M2 through 2005M12), smoothed earnings yield ***se/p*** following Campbell and Shiller (1988b), Campbell and Shiller (1998) (1881M1 through 2005M12), book-to-market ratio ***b/m*** (1926M6 through 2005M12), smoothed return on equity ***roe*** as described in Campbell and Thompson (2008) (1936M6 through 2005M12), the 3-month Treasury-Bill ***tbl*** (1920M1 through 2005M12), long-term government bond yield ***lty*** (1870M1 through 2005M12), the term spread ***ts***, i.e. the difference between long-term and short-term treasury yields (1920M1 through 2005M12), the default spread ***ds***, i.e. the difference between corporate and Treasury bond yields (1919M1 through 2005M12), the lagged inflation rate ***inf*** (1871M5 through 2005M12), and the equity share of new issues ***nei*** proposed by Baker and Wurgler (2000). See Table 1 for a summary.

Comparing out-of-sample predictive ability: For the prediction exercise presented in this section, we use the data over the sample period 1965M1 through 2005M12. We take rolling windows of size $R = 120$ months (10 year estimation sample). The first forecast is made at 1974M12 using the estimation sample 1965M1-1974M12 to forecast 1975M1. We move the estimation window forward by one month to estimate the models using the estimation sample 1965M2-1976M1 to forecast 1976M2. We keep rolling the estimation sample forward until the last one-month ahead forecast is made for the month of 2005M12 using the estimation sample of 1994M12-2005M11. The total $P = 372$ (31 year evaluation sample) out-of-sample forecasts are computed for each model.

Constraints: We apply sign restrictions on the coefficients β depending on the predictor and a positivity restriction on the forecast y_{t+1} of the risk premium, as well as a combination of these two. The coefficient restrictions for the different predictors are listed in Table 1; they are the same as in CT (2008). We impose the hard constraint and we apply bagging to smooth it. This results in the following set of forecasts: UF, PC, PF, PCF (applying positivity on coefficient and forecast jointly), PC-GH, PF-GH, and PCF-GH (bagging the joint restriction).

4.2 Empirical results

We compare the prediction performance in terms of three criteria. Table 2 compare MSFE in $100 \cdot R^2_{\text{OS}}$ proposed by CT (2008). Table 3 reports the utility function of an investor with simple mean-variance preferences as proposed by CT (2008). Table 4 modifies the $100 \cdot R^2_{\text{OS}}$ with the adjustment in MSFE of Clark and West (2006) .

4.2.1 Comparing MSPE in $100 \cdot R^2_{\text{OS}}$ of CT (2008)

Table 2 present the results that are directly related to Tables 1 in Goyal and Welch (2008) and Campbell and Thompson (2008). The reported numbers are out-of-sample R^2 statistics R^2_{OS} multiplied by 100.

$$100 \cdot R^2_{\text{OS}} = 100 \left(1 - \frac{\frac{1}{P} \sum_{t=R+1}^{T} (y_t - \theta_t (x_{t-1}))^2}{\frac{1}{P} \sum_{t=R+1}^{T} (y_t - \theta_t^{\text{HA}})^2} \right), \quad R = 120,\ P = 372, \tag{38}$$

where $\theta_t (x_{t-1})$ is the prediction from the UF, PC, PF, PCF, PC-GH, PF-GH, and PCF-GH model, respectively. These models are organized in the columns of Tables 2 through 4. The rows of the tables show the different univariate predictors that are used for x, from dividend yield $\boldsymbol{d/p}$ through new issues $\boldsymbol{nei}$. There are 11 such predictors. The last row reports a simple, equally weighted, combined forecast from all 11 individual forecasts. We note the following observations from Table 2.

1. Positive values of $100 \cdot R^2_{\text{OS}}$ indicate that a model is better than HA. Many values of $100 \cdot R^2_{\text{OS}}$ in Table 2 are negative, indicating that it is not easy to beat the historical average. Every unrestricted forecast UF using each of the 11 predictors has a negative value.

2. The constraints work. All 11 PC forecasts are better than the unconstrained forecast. All 11 PFs are better than UFs. So are all 11 PCFs. In general PFs and PCFs are better than PCs.

3. Bagging works for all 11 cases when PF is compared to PF-GH. It works only for 3 out of 11 cases when PC is compared to PC-GH and only for 4 out of 11 when PCF is compared to PCF-GH.

4. CF produces the best forecast in each column, dominating all 11 individual forecasts in all of the 7 columns. This is observed in Table 2 (but not in Tables 3, 4). Bagging works, as can be seen from pair-wise comparing the numbers in the last row.

In summary, it is hard to beat HA with UF. But imposing the constraints and bagging can improve UF. A few constrained and bagged predictions outperform HA. Combined forecasts (CF) outperform HA for all constrained models and bagging further improves the forecast power.

4.2.2 How much R^2 is economically meaningful?

Although the gains from imposing constraints and bagging presented in Table 2 are small, they can be economically meaningful for mean-variance investors. As Barberis (2000) points out, *"...the evidence of predictability in asset returns affects optimal portfolio choice for investors with long horizons [...] even after incorporating parameter uncertainty, there is enough predictability in returns to make investors allocate substantially more to stocks."* To see how gains in R^2 may be translated into economic gain, following CT (2008), we consider an investor with single-period horizon and mean-variance preferences

$$\begin{aligned} U &= \text{expected portfolio return} - \frac{\gamma}{2}\,\text{portfolio variance}, \\ &= \mathbb{E}\left[wy_{t+1} + \left(1-w\right)r^f_{t+1}\right] - \frac{\gamma}{2}\mathbb{V}\left[wy_{t+1} + \left(1-w\right)r^f_{t+1}\right], \end{aligned} \tag{39}$$

where γ captures relative risk aversion. The excess return on a risky asset over the riskless interest rate is given by $y_{t+1} = \alpha + \beta x_t + u_{t+1}$ as in (1). Following CT (2008), we assume x_t has unconditional mean zero and unconditional variance σ_x^2, and the risk-free interest rate is constantly equal to zero. The random shock u_{t+1} has unconditional mean zero and unconditional variance $\sigma_u^2 > 0$. As a result, y_{t+1} has unconditional mean $\mathbb{E}\left(y_{t+1}\right) = \alpha$ and unconditional variance $\mathbb{V}\left(y_{t+1}\right) = \beta^2\sigma_x^2 + \sigma_u^2$, assuming independence of x and u.

Without observing x_t, the portfolio weight in the risky asset is

$$w_0 = \frac{1}{\gamma}\frac{\mathbb{E}\left(y_{t+1}\right)}{\mathbb{V}\left(y_{t+1}\right)} = \frac{1}{\gamma}\frac{\alpha}{\beta^2\sigma_x^2 + \sigma_u^2}, \tag{40}$$

and the equity premium (EP) is

$$EP_0 = \mathbb{E}\left[w_0 y_{t+1} + \left(1 - w_0\right) r_{t+1}^f\right] = \frac{1}{\gamma}\frac{\alpha^2}{\beta^2\sigma_x^2 + \sigma_u^2} = \frac{1}{\gamma}\frac{\mathbb{E}\left(y_{t+1}\right)^2}{\mathbb{V}\left(y_{t+1}\right)} = \frac{1}{\gamma}S^2, \tag{41}$$

where S is the Sharpe-ratio. Conditional on x_t, the portfolio weight in the risky asset becomes

$$w_t = \frac{1}{\gamma}\frac{\mathbb{E}\left(y_{t+1}|x_t\right)}{\mathbb{V}\left(y_{t+1}|x_t\right)} = \frac{1}{\gamma}\frac{\alpha + \beta x_t}{\sigma_u^2}, \tag{42}$$

with EP

$$EP_1 = \mathbb{E}\left[w_t y_{t+1} + \left(1 - w_t\right) r_{t+1}^f\right] = \frac{1}{\gamma}\frac{\alpha^2 + \beta\sigma_x^2}{\sigma_u^2} = \frac{1}{\gamma}\frac{S^2 + R^2}{1 - R^2}, \tag{43}$$

where $R^2 = \frac{\beta\sigma_x^2}{\beta\sigma_x^2 + \sigma_u^2}$. Note that w_0 is constant while w_t is time-varying. The increase in EP from observing x_t is

$$EP_1 - EP_0 = \left(\frac{1 + S^2}{1 - R^2}\right)\frac{1}{\gamma}R^2 > \frac{1}{\gamma}R^2. \tag{44}$$

The relative gain in EP is

$$\frac{EP_1 - EP_0}{EP_0} = \left(\frac{1 + S^2}{1 - R^2}\right)\frac{R^2}{S^2} > \frac{R^2}{S^2}. \tag{45}$$

If R^2 is large w.r.t. S^2, then an investor can use the information in the predictive regression to obtain a large proportional increase in return. CT (2008) report $S^2 = 0.0120$ for the CT data set (monthly 1871–2005).

For example, from the bagging results, the out-of-sample R^2_{OS} for dividend yield ($\boldsymbol{d/p}$) of PCF-GH is 0.0016. The relative EP gain is about 13% for dividend yield as a predictor compared to the HA forecast:

$$\frac{R^2}{S^2} = \frac{0.0016}{0.0120} = 0.13 \text{ or } 13\%.$$

Similarly, the out-of-sample R^2_{OS} for earnings yield ($\boldsymbol{e/p}$) of PCF-GH is 0.0024 and thus the relative EP gain is about 20% when earnings yield is used as predictor and compared to the HA forecast:

$$\frac{R^2}{S^2} = \frac{0.0024}{0.0120} = 0.20 \text{ or } 20\%.$$

4.2.3 Utility function of CT (2008)

As discussed in the previous subsection 4.2.2, CT (2008) show the economic significance of numerically small R^2's by interpreting them relative to the squared Sharpe ratio. Rapach, Strauss and Zhou (2010) use the investor utility value in (39) of CT (2008) to compare the economic values of different forecasts. We adopt the same utility function (39) to compare the seven conditional predictive regression models (UF, PC, PF, PCF, PC-GH, PF-GH, and PCF-GH) relative to HA. Table 3 reports the utility level (39) of an investor using the predictive regression over the utility level of the HA forecast.

The historical average (HA) forecast does not use x_t in forecasting y_{t+1}. The realized average utility level for the HA forecast over the out-of-sample period is

$$\hat{U}_0 = \widehat{\mathbb{E}}\left[w_0 y_{t+1} + \left(1 - w_0\right) r^f_{t+1}\right] - \frac{\gamma}{2}\widehat{\mathbb{V}}\left[w_0 y_{t+1} + \left(1 - w_0\right) r^f_{t+1}\right] \tag{46}$$

where $\widehat{\mathbb{E}}(\cdot)$ and $\widehat{\mathbb{V}}(\cdot)$ are the sample mean and sample variance over the out-of-sample period for the portfolio return $\left[w_0 y_{t+1} + \left(1 - w_0\right) r^f_{t+1}\right]$ that was formed using the HA forecasts of y_{t+1}.

The seven conditional forecasts use a predictor x_t in forecasting y_{t+1}. The realized out-of-sample average utility level for each of these forecasts is

$$\hat{U}_1 = \widehat{\mathbb{E}}\left[w_t y_{t+1} + \left(1 - w_t\right) r^f_{t+1}\right] - \frac{\gamma}{2}\widehat{\mathbb{V}}\left[w_t y_{t+1} + \left(1 - w_t\right) r^f_{t+1}\right], \tag{47}$$

where $\widehat{\mathbb{E}}(\cdot)$ and $\widehat{\mathbb{V}}(\cdot)$ are the out-of-sample sample-mean and sample-variance of the return $\left[w_t y_{t+1} + \left(1 - w_t\right) r^f_{t+1}\right]$ on the portfolio that was formed for each of the seven conditional forecasts for a given predictor. Table 3 then repeats this exercise for each of the 11 predictors considered and for the equally-weighted, combined forecast (CF) and reports the gains in utility $\left(\hat{U}_1 - \hat{U}_0\right)$ from the conditional forecasts over the utility $\hat{U}_0$ of the HA forecast. We set $\gamma = 3$; this value is commonly employed in the literature. The results with $\gamma = 2, 4$ are qualitatively similar. We note the following observations.

1. Positive values of the utility gain $\left(\hat{U}_1 - \hat{U}_0\right)$ indicate that a model is better than HA. Many values in Table 3 are negative, indicating that it is not easy to beat HA. The incidence of negative values is lower, however, than in Table 2. Even some unrestricted forecasts take positive values.

2. The PC constraint works with mixed success: 6 PC and 8 PC-GH out of 11 are better than UF. As in Table 2, the PF constraint works for all cases: 11 PFs are better than UFs. So are all 11 PCFs. In general, the PF constraint and the PCF constraint perform better than the PC constraint.

3. Bagging works especially well for the PF constraint, where in all 11 cases PF-GH outperforms PF. For the PC constraint, only 7 out of 11 cases have a higher value for PC-GH, and also only 7 out of 11 when PCF is compared to PCF-GH.

4. For CF in row 12, the PF constraint seems to be more effective than the PC constraint. Bagging improves on all the constrained forecasts PC, PF, and PCF.

In summary, in terms of the utility level, it is still hard for UF to beat HA. Imposing the constraints and bagging can improve the forecast. Combined forecasts (CF) outperform HA for all constrained models and bagging further improves their forecast power, just as in Table 2.

4.2.4 Adjusted $100 \cdot R^2_{\mathbf{OS}}$ of Clark and West (2006)

Campbell and Thompson write in CT (2008, p. 1515, footnote 5) that "*Clark and West (2006) point out that if the return series is truly unpredictable, then in a finite sample the predictive regression will on average have a higher mean squared prediction error because it must estimate an additional coefficient. Thus, the expected out-of-sample R^2 under the null of unpredictability is negative, and a zero out-of-sample R^2 can be interpreted as weak evidence for predictability. We do not pursue this point here because, like Goyal and Welch (2007), we ask whether predictive regressions or historical average return forecasts have delivered better out-of-sample forecasts, not whether stock returns are truly predictable.*" Following these lines, we have studied (in Table 2) whether the constrained or bagged predictive regressions can beat HA, but we have not studied whether stock returns are truly predictable by using various predictors x.

As suggested by Campbell and Thompson (2008) we use the out-of-sample R^2_{OS} in (38), which compares the MSFE $\frac{1}{P}\sum_{t=R+1}^{T}(y_t - \theta_t(x_{t-1}))^2$ of a predictive regression with the MSFE $\frac{1}{P}\sum_{t=R+1}^{T}(y_t - \theta_t^{\mathrm{HA}})^2$ of HA. To compare the MSFEs, the test statistics of Diebold and Mariano (1995) and West

(1996) use the MSFE differential

$$\frac{1}{P}\sum_{t=R+1}^{T}\left[(y_t-\theta_t^{\mathrm{HA}})^2-(y_t-\theta_t\left(x_{t-1}\right))^2\right] \tag{48}$$

to test the null hypothesis that $\mathbb{E}\left[(y_t-\theta_t^{\mathrm{HA}})^2-(y_t-\theta_t\left(x_{t-1}\right))^2\right]=0$. Note that R^2_{OS} in (38), reported in Table 2, is obtained from dividing (48) by $\frac{1}{P}\sum_{t=R+1}^{T}(y_t-\theta_t\left(x_{t-1}\right))^2$.

Clark and West (CW 2006) show that when a predictive regression model (using x) is compared with the HA, under the null hypothesis of no predictive ability of x, its MSFE is expected to be greater than the HA's MSFE. CW propose an adjustment to the MSFE differential in (48) in order to account for the disadvantage of the sample MSFE of the predictive model. The CW-adjusted MSFE differential is

$$\frac{1}{P}\sum_{t=R+1}^{T}\left[(y_t-\theta_t^{\mathrm{HA}})^2-\left\{(y_t-\theta_t\left(x_{t-1}\right))^2-(\theta_t^{\mathrm{HA}}-\theta_t\left(x_{t-1}\right))^2\right\}\right]. \tag{49}$$

We use the following "CW-adjusted-R^2_{OS}" defined by dividing (49) by $\frac{1}{P}\sum_{t=R+1}^{T}(y_t-\theta_t^{\mathrm{HA}})^2$:

$$\text{CW-adjusted-}100\cdot R^2_{\mathrm{OS}}=100\left(1-\frac{\frac{1}{P}\sum_{t=R+1}^{T}\left\{(y_t-\theta_t\left(x_{t-1}\right))^2-(\theta_t^{\mathrm{HA}}-\theta_t\left(x_{t-1}\right))^2\right\}}{\frac{1}{P}\sum_{t=R+1}^{T}(y_t-\theta_t^{\mathrm{HA}})^2}\right). \tag{50}$$

Table 4 presents the CW-adjusted-$100\cdot R^2_{\mathrm{OS}}$. We note the following observations.

1. Positive values of the CW-adjusted-$100\cdot R^2_{\mathrm{OS}}$ indicate that a predictive regression model using x is better than HA. Most values in Table 4 are positive.

2. The constraints work, with a few exceptions for PC and with no exceptions for PF. The PF constraint is better than the PC constraint. PFs and PCFs are better than PCs.

3. Bagging works very well. Bagging works for all 11 cases when PF is compared to PF-GH, 9 out of 11 when PC is compared to PC-GH, and 9 out of 11 when PCF is compared to PCF-GH.

4. While the combined forecast (CF) is nowhere the best, it is consistently better than HA across all constraints. Bagging works for CF as in the previous tables. The combined forecast with the PC constraint is further improved by bagging (PC-GH is better), the CF with PF is further improved by bagging (PF-GH is better), and the CF with PCF is also further improved by bagging (PCF-GH is better).

Positive gains in $100R^2_{\text{OS}}$ with the CW-adjustment indicate that a predictive regression model using a predictor x can beat the historical average when it is used together with the positivity constraints and even further with bagging.

5 Conclusions

The vast literature on equity return prediction has considered a wide array of models and methods. CT (2008) propose to impose certain sensible restrictions on the regression coefficient or on the return prediction. Their shrinkage approach reduces MSFE by increasing the forecast bias and reducing the forecast error variance.

In this paper, we apply bagging to potentially increase the bias even further than the CT constraints for the possible benefit of reducing the forecast error variance even more. We review the theory behind bagging, in particular Breiman (1996), Bühlmann and Yu (2002), and Gordon and Hall (2009), and explore the bias-variance trade-off and shrinkage properties of bagging in simulations. We show that for a large variety of signal-to-noise and regressor persistence scenarios, bagging can further improve predictive power as long as the imposed constraint is not completely obvious and far from binding, but, loosely speaking, true enough.

In the stock return prediction problem, we find that in particular the positivity constraint on the forecast itself improves prediction, more so than the constraint on the sign of the regression coefficient. Smoothing the hard constraint at zero for the return forecast by bagging over a large set of bootstrap replications, we further improve this edge in predictive power, which we measure by the out-of-sample R^2 as in CT (2008), by the utility function of CT (2008) as reported in Rapach, Strauss, and Zhou (2010), and by the adjusted out-of-sample R^2 of Clark and West (2006). In particular after accounting for the natural MSFE-disadvantage of a regressor model compared to the historical mean under the null, the advantage of bagging constraints becomes very clear. Simple combination forecasts do consistently well, but not always best. In our application, they could always be improved by bagging.

References

Ang, A., and G. Bekaert (2007): "Stock Return Predictability: Is It There?," *Review of Financial Studies* 20, 651–707.

Audrino, F. and M. C. Medeiros (2011): "Modeling and Forecasting Short-term Interest Rates: The Benefits of Smooth Regimes, Macroeconomic Variables, and Bagging," *Journal of Applied Econometrics*, 26, 999–1022.

Baker, M., and J. Wurgler (2000): "The Equity Share in New Issues and Aggregate Stock Returns," *Journal of Finance* 55, 2219–2257.

Barberis, N.C. (2000): "Investing for the Long Run When Returns Are Predictable," *Journal of Finance* 55, 225-264.

Bühlmann, P., and B. Yu (2002): "Analyzing Bagging," *Annals of Statistics* 30, 927–961.

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.

Breiman, L. (1996): "Bagging Predictors," *Machine Learning* 36, 105–139.

Buja, A. and W. Stuetzle (2002), "Observations on Bagging", University of Pennsylvania and University of Washington, Seattle.

Butler, A., G. Grullon, and J. Weston (2006): "Can Managers Forecast Aggregate Market Returns?," *Journal of Finance* 60, 963–986.

Campbell, J. (1987): "Stock Returns and the Term Structure," *Journal of Financial Economics* 18, 373–399.

Campbell, J., and R. Shiller (1988a): "The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors," *Review of Financial Studies* 1, 195–228.

Campbell, J., and R. Shiller (1988b): "Stock Prices, Earnings, and Expected Dividends," *Journal of Finance* 43, 661–676.

Campbell, J., and R. Shiller (1998): "Valuation Ratios and the Long-Run Stock Market Outlook," *Journal of Portfolio Management* 24, 11–26.

Campbell, J., and S. Thompson (2008): "Predicting the Equity Premium Out of Sample: Can Anything Beat the Historical Average?," *Review of Financial Studies* 21(4), 1511-1531.

Campbell, J., and M. Yogo (2006): "Efficient Tests of Stock Return Predictability," *Journal of Financial Economics* 81, 27–60.

Cavanagh, C., G. Elliot, and J. Stock (1995): "Inference in Models with Nearly Integrated Regressors," *Econometric Theory* 11, 1131–1147.

Clark, T.E. and K.D. West (2006): "Using out-of-sample mean squared prediction errors to test the martingale difference hypothesis," *Journal of Econometrics* 135, 155-186.

Estrella, A. and G.A. Hardouvelis (1991): "The Term Structure as a Predictor of Real Economic Activity," *Journal of Finance* 46(2), 555-576.

Fama, E., and K. French (1988): "Dividend Yields and Expected Stock Returns," *Journal of Financial Economics* 22, 3–25.

Fama, E., and K. French (1989): "Business Conditions and Expected Returns on Stocks and Bonds," *Journal of Financial Economics* 25, 23–49.

Fama, E., and G. Schwert (1977): "Asset Returns and Inflation," *Journal of Financial Economics* 5, 115–146.

Fama, E., and G. Schwert (1981): "Stock Returns, Real Activity, Inflation and Money," *American Economic Review* 71, 545–565.

Faust, J. and J. H. Wright (2009): "Comparing Greenbook and Reduced Form Forecasts Using a Large Real Time Dataset," *Journal of Business and Economic Statistics* 27(4): 468-479.

Ferson, W., S. Sarkissian, and T. Simin (2003): "Spurious Regressions in Financial Economics?," *Journal of Finance* 58, 1393–1413.

Fitzenberger, B. (1997): "The Moving Blocks Bootstrap and Robust Inference for Linear Least Squares and Quantile Regressions", *Journal of Econometrics* 82, 235-287.

Foster, F., T. Smith, and R.E. Whaley (1997): "Assessing Goodness-of-Fit of Asset Pricing Models: The Distribution of the Maximal R2," *Journal of Finance* 52, 591–607.

Friedman, J.H. and P. Hall (2007), "On Bagging and Nonlinear Estimation", *Journal of Statistical Planning and Inference* 137, 669-683.

Gordon, I., and P. Hall (2009): "Estimating a Parameter When It Is Known that the Parameter Exceeds a Given Value," *Australian and New Zealand Journal of Statistics* 51(4), 449–460

Goyal, A., and I. Welch (2008): "A Comprehensive Look at the Empirical Performance of Equity Premium Prediction," *Review of Financial Studies* 21(4), 1455-1508.

Hall, P., J. Horowitz, and B.-Y. Jing (1995): "On Blocking Rules for the Bootstrap with Dependent Data," *Biometrika* 82, 561–574.

Hastie, T., R. Tibshirani, and J. Friedman (2001): *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, New York: Springer.

Hillebrand, E. and M. Medeiros (2010): "Forecasting Realized Volatility Models: The Benefits of Bagging and Nonlinear Specifications," *Econometric Reviews* 29(5), 571-593.

Huang, H., and T.-H. Lee (2010): "To Combine Forecasts or to Combine Information," *Econometric Reviews* 29(5), 534-557.

Inoue, A., and L. Kilian (2008): "How Useful is Bagging in Forecasting Economic Time Series? A Case Study of U.S. CPI Inflation," *Journal of the American Statistical Association* 103(482), 511-522.

Jansson, M., and M. J. Moreira (2006): "Optimal Inference in Regression Models with Nearly Integrated Regressors," *Econometrica* 74, 681–715.

Keim, D., and R. F. Stambaugh (1986): "Predicting Returns in the Stock and Bond Markets," *Journal of Financial Economics* 17, 357–390.

Kilian, L. (1999): "Exchange Rates and Monetary Fundamentals: What Do We Learn from Long-Horizon Regressions?," *Journal of Applied Econometrics* 14, 491–510.

Koenker, R and G. Basset (1978): "Asymptotic Theory of Least Absolute Error Regression," *The Journal of the American Statistical Association* 73, 618-622.

Lee, T.-H., and Y. Yang (2006): "Bagging Binary and Quantile Predictors for Time Series," *Journal of Econometrics* 135, 465–497.

Lettau, M., and S. Ludvigson (2001): "Consumption, Aggregate Wealth, and Expected Stock Returns," *Journal of Finance* 56, 815–849.

Lewellen, J. (2004): "Predicting Returns with Financial Ratios," *Journal of Financial Economics* 74, 209–235.

Mark, N. (1995): "Exchange Rates and Fundamentals: Evidence on Long-Horizon Predictability," *American Economic Review* 85, 201–218.

Nelson, C., and M. Kim (1993): "Predictable Stock Returns: The Role of Small Sample Bias," *Journal of Finance* 48, 641–661.

Pesaran, M., and A. Timmermann (2002): "Market Timing and Return Prediction under Model Instability," *Journal of Empirical Finance* 9, 495–510.

Pesaran, M. H., and A. Timmermann (1995): "Predictability of stock returns: robustness and economic significance," *Journal of Finance* 50, 1201–1228.

Polk, C., S. Thompson, and T. Vuolteenaho (2006): "Cross-Sectional Forecasts of the Equity Premium," *Journal of Financial Economics* 81, 101–141.

Rapach, D., ans J. K. Strauss (2010): "Bagging or Combining (or Both)? An Analysis Based on Forecasting U.S. Employment Growth," *Econometric Reviews* 29(5), 511-533.

Rapach, D., J. Strauss, and G. Zhou (2010): "Out-of-Sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy" *Review of Financial Studies* 23(2), 821-862.

Rozeff, M. (1984): “Dividend Yields are Equity Risk Premiums,” *Journal of Portfolio Management* 11, 68–75.

Schwert, G. (2002): “Anomalies and market efficiency,” in *Handbook of the Economics of Finance*, edited by M. Harris, and R. Stulz. North Holland: Amsterdam.

Stambaugh, R. (1999): “Predictive Regressions,” *Journal of Financial Economics* 54, 375–421.

Stock, J.H. and Watson, M.W. (2007): “Why has U.S. inflation become harder to forecast?” *Journal of Money, Credit, and Banking* 39(1), 3-33.

Stock, J.H. and Watson, M.W. (2012), “Generalized Shrinkage Methods for Forecasting Using Many Predictors,” Harvard University and Princeton University.

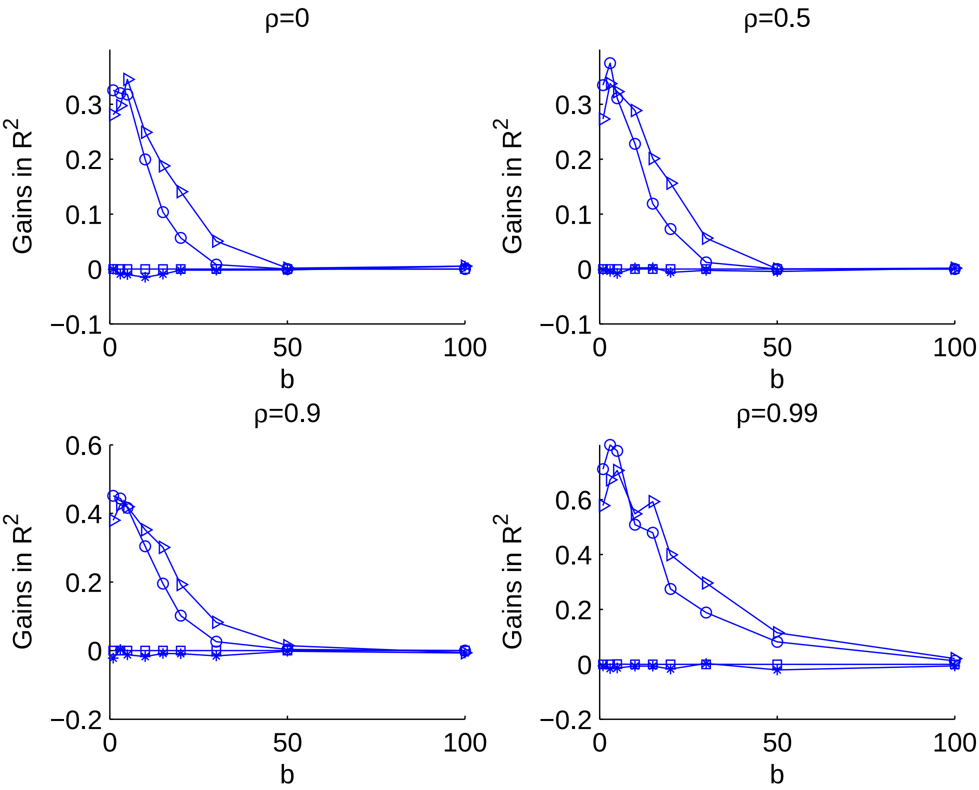
Tibshirani, R. (1996): “Regression shrinkage and selection via the lasso,” *Journal of the Royal Statistical Society B*, 58(1), 267–288.

Timmermann, A. (2008): “Elusive Return Predictability,” *International Journal of Forecasting*, 1-18.

Torous, W., R. Valkanov, and S. Yan (2004): “On Predicting Stock Returns with Nearly Integrated Explanatory Variables,” *Journal of Business* 77, 937–966.

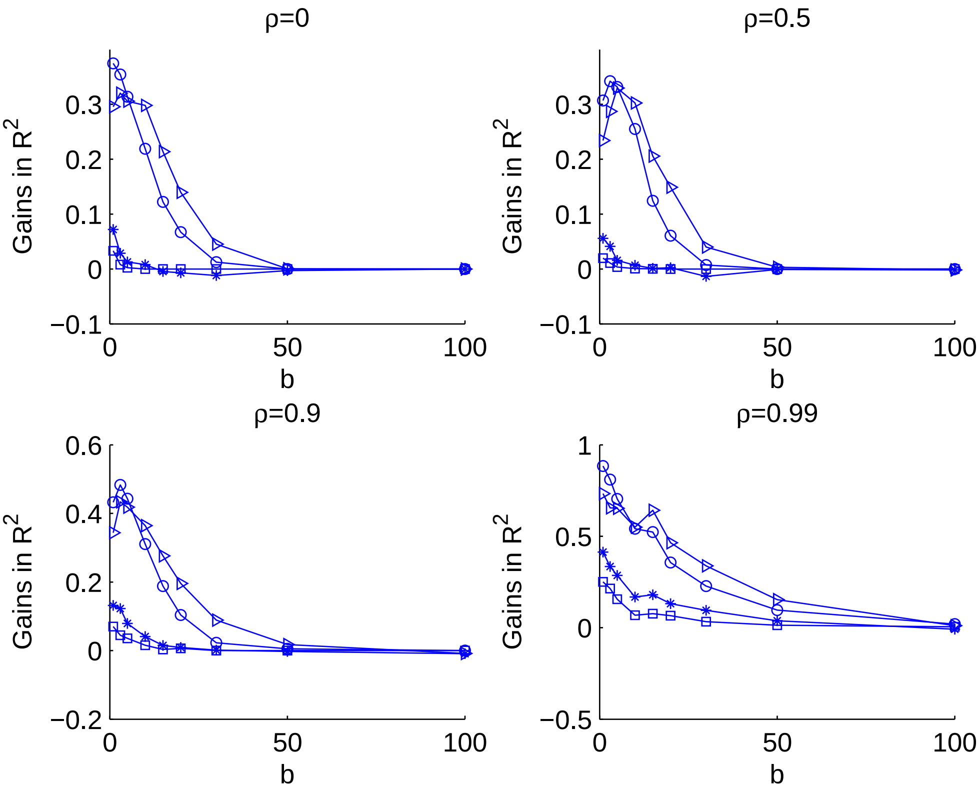
Wheelock, D.C. and M.E. Wohar (2009): “Can the Term Spread Predict Output Growth and Recessions? A Survey of the Literature,” *Federal Reserve Bank of St. Louis Review* 91(5) Part 1, 419-440.

Figure 1. Gains in $100 \cdot R^2_{\text{OS}}$ from imposing constraint and bagging over UF when $\sigma_u = 0.01$



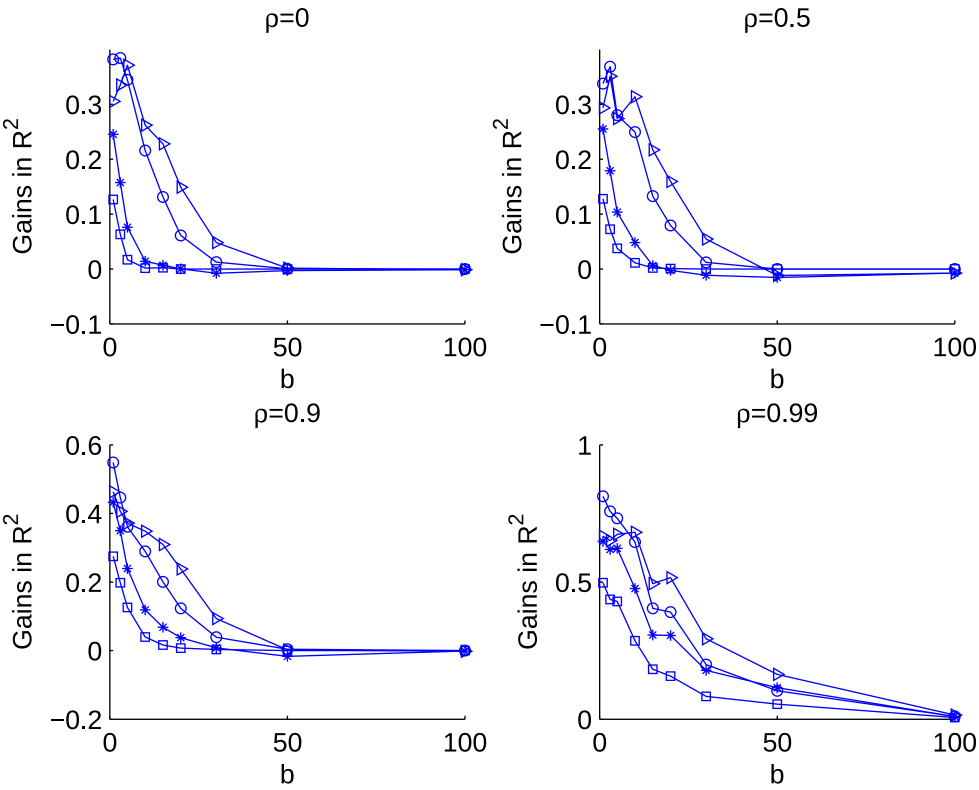
Notes: The gain of a model in $100 \cdot R^2_{\text{OS}}$ over UF is $\left(100 \cdot R^2_{\text{OS}}\right)_{\text{model}} - \left(100 \cdot R^2_{\text{OS}}\right)_{\text{UF}}$ for each of model = PC (line with circles o), PC-GH (line with triangles $\triangle$), PF (line with squares $\square$), or PF-GH (line with asterisks *).

Figure 2. Gains in $100 \cdot R^2_{\mathrm{OS}}$ from imposing constraint and bagging over UF when $\sigma_u = 0.05$



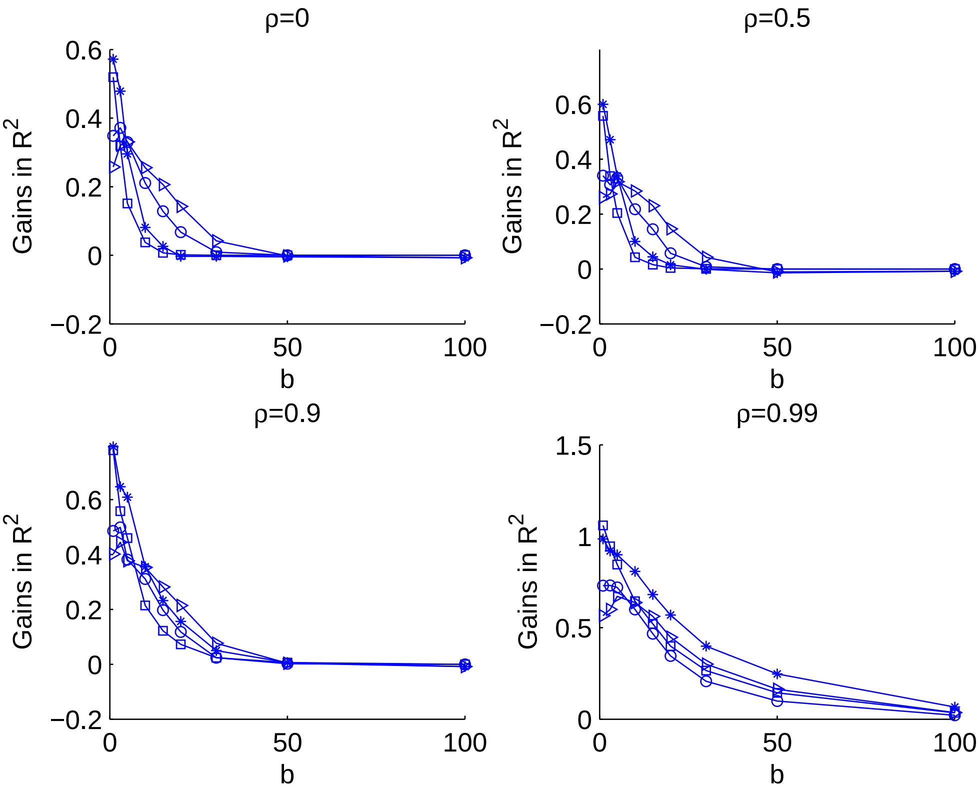
Notes: The gain of a model in $100 \cdot R^2_{\mathrm{OS}}$ over UF is $\left(100 \cdot R^2_{\mathrm{OS}}\right)_{\mathrm{model}} - \left(100 \cdot R^2_{\mathrm{OS}}\right)_{\mathrm{UF}}$ for each of model = PC (line with circles o), PC-GH (line with triangles △), PF (line with squares □), or PF-GH (line with asterisks *).

Figure 3. Gains in $100 \cdot R^2_{\mathrm{OS}}$ from imposing constraint and bagging over UF when $\sigma_u = 0.10$



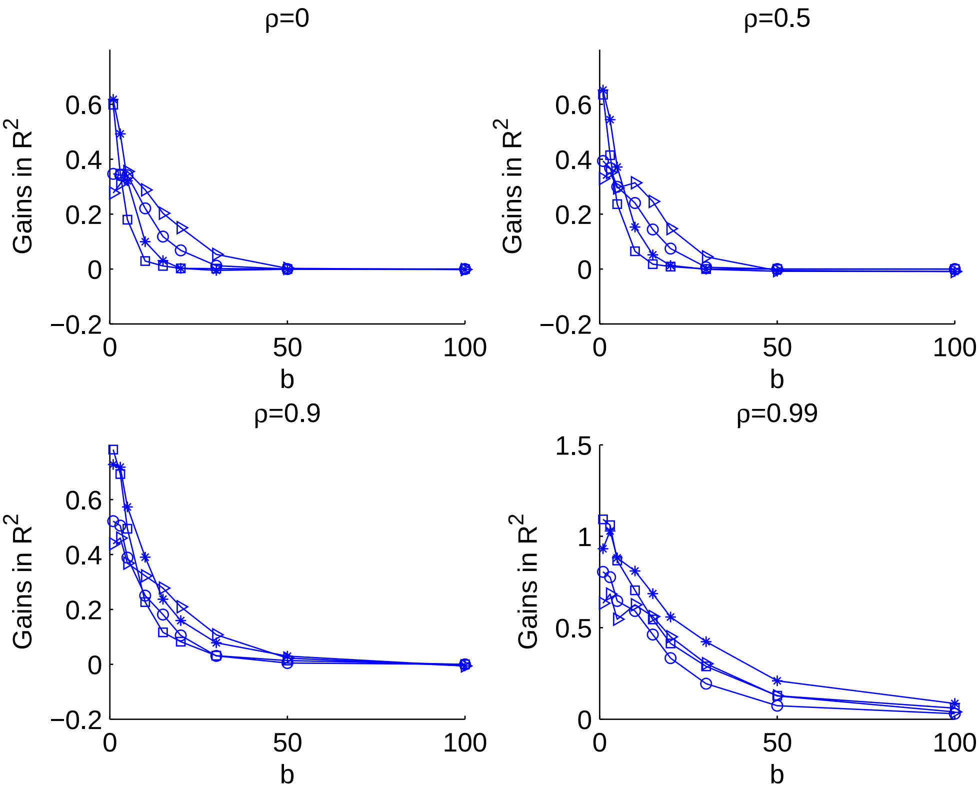
Notes: The gain of a model in $100 \cdot R^2_{\mathrm{OS}}$ over UF is $\left(100 \cdot R^2_{\mathrm{OS}}\right)_{\mathrm{model}} - \left(100 \cdot R^2_{\mathrm{OS}}\right)_{\mathrm{UF}}$ for each of model = PC (line with circles o), PC-GH (line with triangles $\triangle$), PF (line with squares $\square$), or PF-GH (line with asterisks *).

Figure 4. Gains in $100 \cdot R^2_{\mathrm{OS}}$ from imposing constraint and bagging over UF when $\sigma_u = 0.50$



Notes: The gain of a model in $100 \cdot R^2_{\mathrm{OS}}$ over UF is $\left(100 \cdot R^2_{\mathrm{OS}}\right)_{\mathrm{model}} - \left(100 \cdot R^2_{\mathrm{OS}}\right)_{\mathrm{UF}}$ for each of model = PC (line with circles o), PC-GH (line with triangles $\triangle$), PF (line with squares $\square$), or PF-GH (line with asterisks *).

Figure 5. Gains in $100 \cdot R^2_{\text{OS}}$ from imposing constraint and bagging over UF when $\sigma_u = 1.00$



Notes: The gain of a model in $100 \cdot R^2_{\text{OS}}$ over UF is $\left(100 \cdot R^2_{\text{OS}}\right)_{\text{model}} - \left(100 \cdot R^2_{\text{OS}}\right)_{\text{UF}}$ for each of model = PC (line with circles o), PC-GH (line with triangles △), PF (line with squares □), or PF-GH (line with asterisks *).

Table 1. Data and Restrictions

x	sign(β)	
d/p	$+$	dividend yield
e/p	$+$	earnings yield
se/p	$+$	smoothed earnings yield
b/m	$+$	book-to-market ratio
roe	$+$	smoothed return on equity
tbl	$-$	3-month Treasury-Bill
lty	$-$	long-term government bond yield
ts	$+$	term spread: long-term - short-term Treasury yields
ds	$+$	default spread: corporate - Treasury bond yields
inf	$-$	inflation rate
nei	$-$	equity share of new issues (Baker and Wurgler 2000)

Notes: We use the same data set of Campbell and Thompson (2008), which was kindly provided by Sam Thompson. The data frequency is monthly. See subsection 4.1 for details. The PC constraints of CT (2008) for each predictor is shown in column 2. When the sign on the coefficient β, sign(β) , is negative, the positive constraint PC should be understood as a negative constraint (or the positive constraint of the negative values of the predictor ($-x$).

Table 2. Relative Gains in MSFE over HA

	UF	PC	PF	PCF	PC-GH	PF-GH	PCF-GH
d/p	−2.62	−1.71	−0.82	0.09	−1.69	−0.17	0.16
e/p	−1.94	−0.52	−0.59	0.18	−0.47	0.31	0.24
se/p	−2.81	−1.61	−1.33	−0.55	−1.43	−0.43	−0.37
b/m	−2.29	−1.56	−1.15	−0.42	−1.73	−0.61	−0.52
roe	−2.30	−1.32	−1.36	−1.09	−1.80	−1.01	−1.34
tbl	−4.77	−3.71	−0.96	−0.45	−3.91	−0.77	−0.73
lty	−3.64	−2.11	−1.13	−0.09	−2.76	−1.00	−0.50
ts	−1.81	−1.26	−0.41	0.12	−1.47	−0.27	−0.08
ds	−1.31	−0.92	−0.20	−0.28	−1.12	−0.05	−0.47
inf	−3.53	−2.95	−1.37	−1.32	−3.03	−1.00	−0.78
nei	−2.79	−2.47	0.05	0.05	−2.71	0.30	−0.08
CF	−0.53	0.27	0.28	0.83	0.32	0.51	0.87

Notes: The reported numbers are out-of-sample statistics R^2_{OS} multiplied by 100, following Campbell and Thompson (2008), as defined in (38). This is to compare seven forecasts in seven columns using each of the 11 predictors (in 11 rows). The last row (row 12) presents the equally-weighted combined-forecast (CF) of the 11 forecasts in each column. R^2_{OS} measures the relative gain of a predictive regression over HA.

Table 3. Utility Gains over HA

	UF	PC	PF	PCF	PC-GH	PF-GH	PCF-GH
d/p	−1.65	−0.55	−1.65	−0.55	−0.92	−0.72	−0.92
e/p	−0.82	0.22	−0.82	0.22	0.30	0.75	0.30
se/p	−2.37	−1.34	−2.37	−1.34	−1.10	−0.54	−1.10
b/m	−2.15	−1.20	−2.15	−1.20	−1.31	−0.75	−1.31
roe	0.98	0.82	0.98	0.82	1.44	1.55	1.44
tbl	1.21	0.95	1.21	0.95	1.20	1.56	1.20
lty	−0.20	1.09	−0.20	1.09	1.10	0.45	1.10
ts	2.64	1.97	2.64	1.97	2.58	2.92	2.58
ds	1.37	0.86	1.37	0.86	0.64	1.58	0.64
inf	−0.55	−1.00	−0.55	−1.00	0.79	0.41	0.79
nei	1.17	1.27	1.17	1.27	1.25	1.76	1.25
CF	0.43	0.32	1.10	1.02	0.80	1.54	1.63

Notes: The reported numbers are out-of-sample statistics for the utility gain $\left(\hat{U}_1 - \hat{U}_0\right)$ of an investor with mean-variance preferences. It was discussed in CT (2008) and used by Rapach, Strauss and Zhou (2010). See subsection 4.2.3 for details. As in Table 2, we compare seven forecasts in seven columns using each of the 11 predictors (in 11 rows). The last row (row 12) presents the equally-weighted combined-forecast (CF) of the 11 forecasts in each column.

Table 4. Relative Gains in MSFE over HA with the Adjustment of Clark and West (2006)

	UF	PC	PF	PCF	PC-GH	PF-GH	PCF-GH
d/p	0.58	0.98	1.78	2.18	1.16	2.63	2.37
e/p	0.08	1.38	0.71	1.40	1.55	1.91	1.57
se/p	−0.62	0.19	0.63	1.04	0.59	1.84	1.40
b/m	−0.35	0.04	0.34	0.73	−0.00	1.03	0.72
roe	0.48	0.72	1.10	0.96	1.14	1.73	1.59
tbl	−0.74	−0.25	0.80	1.00	0.07	1.02	1.19
lty	−1.75	−0.56	0.41	1.23	−0.55	0.80	1.33
ts	3.29	3.21	3.39	3.28	3.40	3.58	3.49
ds	4.88	4.73	5.10	4.88	4.62	5.20	4.77
inf	−2.38	−2.20	−0.40	−0.70	−1.64	0.24	0.39
nei	0.39	0.32	1.28	1.05	0.54	1.56	1.31
CF	0.35	0.78	1.38	1.55	0.99	1.96	1.83

Notes: The reported numbers are out-of-sample statistics R^2_{OS} multiplied by 100, of Campbell and Thompson (2008), with the adjustment of Clark and West (2006), as discussed in subsection 4.2.4. As in Table 2, we compare seven forecasts in seven columns using each of the 11 predictors (in 11 rows). The last row (row 12) presents the equally-weighted combined-forecast (CF) of the 11 forecasts in each column.

Departamento de Economia PUC-Rio
Pontifícia Universidade Católica do Rio de Janeiro
Rua Marques de Sâo Vicente 225 - Rio de Janeiro 22453-900, RJ
Tel.(21) 35271078 Fax (21) 35271084
www.econ.puc-rio.br
flavia@econ.puc-rio.br