Out-of-Sample Equity Premium Prediction: Consistently Beating the Historical Average

David E. Rapach
Saint Louis University

Jack K. Strauss
Saint Louis University

Guofu Zhou*
Washington University in St. Louis

July 23, 2007 (First Draft)
February 6, 2008 (Current Version)

Abstract

While a host of economic variables have been identified in the literature with the apparent in-sample ability to predict the equity premium, Goyal and Welch (2007) find that these variables fail to deliver consistent out-of-sample forecast gains relative to the historical average. Imposing theoretically motivated restrictions on individual predictive regression models, Campbell and Thompson (2007) provide improved forecasts, but the out-of-sample performance is still highly uneven over time. In this paper, we propose two approaches—combination forecasts and covariate estimation—to improve out-of-sample equity premium forecasts based on economic variables. These approaches accommodate structural instability and utilize additional information. We find that despite the failure of individual predictive regression model forecasts to outperform the historical average, combinations of individual model forecasts deliver statistically and economically significant out-of-sample gains relative to the historical average on a consistent basis over time. Forming combination forecasts from individual models estimated using covariates and with Campbell and Thompson (2007) restrictions imposed typically leads to further out-of-sample gains.

JEL classifications: C22, C53, G11, G12

Keywords: Out-of-sample equity premium predictability; Combination forecasts; Covariates

*Corresponding author. Send correspondence to Guofu Zhou, Olin School of Business, Washington University in St. Louis, St. Louis, MO 63130; e-mail: zhou@wustl.edu; phone: 314-935-6384. We are grateful to John Campbell, Hui Guo, Michael Halling, Yufeng Han, Ravi Jagannathan, Raymond Kan, Mike McCracken, Ľuboš Pástor, Matthew Spiegel, Jun Tu, and seminar participants at Washington University for many helpful comments. The usual disclaimer applies. We are also grateful to Amit Goyal for making the data from Goyal and Welch (2007) available on his web page. All of the results reported in this paper were generated using GAUSS 6.1.

1 Introduction

Forecasting stock returns is of great interest to both academics and practitioners in finance, and numerous economic variables have been proposed as predictors of stock returns in the literature. Examples include valuation ratios such as the dividend price (Dow, 1920; Fama and French, 1988), earnings price (Campbell and Shiller, 1988, 1998), and book to market (Kothari and Shanken, 1997; Pontiff and Schall, 1998), as well as nominal interest rates (Fama and Schwert, 1977; Campbell, 1987; Breen et al., 1989; Ang and Bekaert, 2007), the inflation rate (Nelson, 1976; Fama and Schwert, 1977; Campbell and Vuolteenaho, 2004), term and default spreads (Campbell, 1987; Fama and French, 1989), corporate issuing activity (Baker and Wurgler, 2000; Boudoukh et al., 2005), consumption-wealth ratio (Lettau and Ludvigson, 2001), and stock market volatility (Guo, 2006).[1] Most existing studies focus on in-sample tests and conclude that there is significant evidence of return predictability.

However, Goyal and Welch (2007) show that a long list of potential predictors from the literature are unable to deliver consistently superior out-of-sample forecasts of the U.S. equity premium relative to a simple forecast based on the historical average (constant expected equity premium model). They argue that individual predictive regression models thus fail to pass an important diagnostic test, so that despite in-sample evidence of predictability, "these models would not have helped an investor with access to available information to profitably time the market." Goyal and Welch's comprehensive study forcefully echoes the typically negative findings of the relatively few studies that consider out-of-sample tests of return predictability. For example, Bossaerts and Hillion (1999) fail to find significant evidence of out-of-sample predictive ability in a collection of industrialized countries for a number of variables for 1990–1995, and Goyal and Welch (2003) find that the dividend-price ratio is not a robust out-of-sample predictor of the U.S. equity premium. Campbell and Thompson (2007) find that placing theoretically motivated restrictions on individual predictive regression models improves their out-of-sample performance in both statistical and economic terms.[2] Unfortunately, the out-of-sample performance of their predictive regression models still deteriorates from 1980 onward.

In this paper, we propose two approaches—combination forecasts and covariate estimation—to improve out-of-sample equity premium forecasts based on economic variables. These approaches accommodate structural instability and utilize additional information. Goyal and Welch (2007)

[1] The list of studies above is not meant to be exhaustive; see Campbell (2000) and Goyal and Welch (2007) for more extensive surveys of the vast literature on return predictability.

[2] Also see Campbell (2008).

attribute the poor out-of-sample performance of individual predictive regression models to structural instability. This is consistent with Paye and Timmermann (2006) and Rapach and Wohar (2006a), who find extensive in-sample evidence of structural breaks in predictive regression models of U.S. stock returns, with individual economic variables exhibiting substantial predictive ability during certain periods and no predictive ability during others. Intuitively, numerous factors—including institutional change, policy shocks, advances in information technology, and investor learning—give rise to a highly complex and constantly evolving data-generating process for expected equity returns that is difficult to approximate with a single predictive regression model.[3] Hendry and Clements (2004) analyze an environment relevant to equity returns in which individual models provide only partial (perhaps overlapping) descriptions of the data-generating process and are subject to periodic structural breaks. They show that combining forecasts across individual models can lead to improved forecast accuracy, suggesting that combination forecasts are an effective tool for forecasting in the presence of structural breaks.[4] In an effort to generate improved equity premium forecasts based on economic variables when individual predictive regression models are characterized by instability, we consider combination forecasts of the equity premium.

We analyze combination forecasts formed as weighted averages of 14 individual predictive regression model forecasts, where each of the individual predictive regression models includes one of the 14 economic variables considered by Goyal and Welch (2007) for which monthly data are available starting at least in 1927. We consider a number of different methods for determining the combining weights, including simple averaging (mean, median, trimmed mean), discounting (Stock and Watson, 2004), clusters (Aiolfi and Timmermann, 2006), and approximate Bayesian model averaging (Garratt et al., 2003). Our results confirm Goyal and Welch's (2007) finding that individual predictive regression models fail to generate consistent out-of-sample forecast gains relative to the historical average. However, we also find that various combining methods consistently provide significant out-of-sample gains relative to the historical average. This is true using both statistical and economic criteria and holds across a number of historical periods, including more recent periods when the out-of-sample predictive ability of many individual variables is particularly poor. By combining forecasts across individual predictive regression models, we thus find that economic variables can be used to consistently beat historical average forecasts of the equity premium.

[3]Timmermann (2007) labels this "elusive predictability."

[4]In their recent extensive reviews of the forecasting literature, Clements and Hendry (2006) and Timmermann (2006) also observe that combination forecasts can improve forecast performance in the presence of structural breaks.

Covariate estimation is a method of utilizing additional information to more precisely estimate the parameters of predictive regression models, thereby potentially improving the forecast accuracy of predictive regression models. More precisely estimating the parameters of predictive regression models of stock returns is likely to be especially important given the noise inherent in stock returns. One way to utilize additional information is to incorporate information from other stock return series when estimating individual forecasting models. We use a covariate approach that is a simple add-in method to the standard predictive regression model in the spirit of Faust and Wright (2005), and we introduce extra information from a number of industry and international stock return series. We find that forming combination forecasts from individual forecasts generated by predictive regression models estimated using covariates typically provides further improvements in out-of-sample performance. Imposing Campbell and Thompson (2007) restrictions on the individual predictive regression models also often helps to improve the overall performance of the combining methods. In short, combination forecasts in conjunction with covariate estimation and theoretically motivated restrictions appear to offer a practical way of consistently beating historical average forecasts of the equity premium in real time.

The remainder of the paper is organized as follows. Section 2 outlines the predictive regression model framework, construction of combination forecasts, covariate estimation technique, and criteria we use to evaluate the out-of-sample forecasts. The out-of-sample forecast results for the individual predictive regression models and combining methods are reported in Section 3. Section 4 concludes.

2 Econometric Methodology

2.1 Predictive Regression Model

A standard predictive regression model for the equity premium is

$$r_{t+1} = \alpha + \beta x_{i,t} + \varepsilon_{t+1}, \tag{1}$$

where r_{t+1} is the return on a stock market index in excess of the risk-free interest rate, $x_{i,t}$ is a variable whose predictive ability is of interest, and ε_{t+1} is a disturbance term. As in Goyal and Welch (2007), we generate out-of-sample forecasts of the equity premium using a recursive (expanding) estimation window. More specifically, we first divide the total sample of T observations for r_t and $x_{i,t}$ into an in-sample portion composed of the first R observations and an out-of-sample portion composed of the last P observations. The initial out-of-sample forecast of the equity premium based on the predictor $x_{i,t}$ is given by $\hat{r}_{i,R+1} = \hat{\alpha}_R + \hat{\beta}_R x_{i,R}$, where $\hat{\alpha}_R$ and $\hat{\beta}_R$ are the OLS estimates of α and β, respectively, in (1) generated by regressing $\{r_t\}_{t=2}^{R}$ on a constant and $\{x_{i,t}\}_{t=1}^{R-1}$. The next out-of-sample forecast is given by $\hat{r}_{i,R+2} = \hat{\alpha}_{R+1} + \hat{\beta}_{R+1} x_{i,R+1}$, where $\hat{\alpha}_{R+1}$ and $\hat{\beta}_{R+1}$ are generated by regressing $\{r_t\}_{t=2}^{R+1}$ on a constant and $\{x_{i,t}\}_{t=1}^{R}$. Proceeding in this manner through the end of the out-of-sample period, we generate a series of P out-of-sample forecasts of the equity premium based on $x_{i,t}$ ($\{\hat{r}_{i,t+1}\}_{t=R}^{T-1}$). This out-of-sample forecast exercise mimics the situation of a forecaster in real time. In our empirical applications in Section 3 below, we generate out-of-sample forecasts of the equity premium using 14 individual predictive regression models ($i = 1, \ldots, N$ and $N = 14$), where each model is based on one of the 14 variables from Goyal and Welch (2007) for which monthly data are available for 1927:01–2005:12.

Following Goyal and Welch (2007) and Campbell and Thompson (2007), the historical average of the equity premium, $\bar{r}_{t+1} = \sum_{j=1}^{t} r_j$, serves as a natural benchmark forecast model corresponding to a constant expected equity premium. Intuitively, if $x_{i,t}$ contains information useful for predicting the equity premium, $\hat{r}_{i,t+1}$ should perform better than $\bar{r}_{t+1}$. Measures to compare $\hat{r}_{i,t+1}$ to $\bar{r}_{t+1}$ are provided in Section 2.4 below.

2.2 Combination Forecasts

The combination forecasts of r_{t+1} made at time t are all weighted averages (linear combinations) of the N individual forecasts based on (1):

$$\hat{r}_{c,t+1} = \sum_{i=1}^{N} \omega_{i,t} \hat{r}_{i,t+1}, \tag{2}$$

where $\{\omega_{i,t}\}_{i=1}^{N}$ ($N = 14$) are the *ex ante* combining weights formed at time t. Some of the combining methods require a holdout period to estimate the combining weights, and we use the first P_0 observations from the out-of-sample period as the initial holdout period. For each of the combining methods, we compute combination forecasts over the post-holdout out-of-sample period, leaving us with a total of $P - P_0$ combination forecasts available for evaluation. With one exception (the mean combination forecast described below), all of the combination forecasts allow the combining weights to change at each t.

The combining methods we consider differ in how the combining weights are determined and can be organized into four general classes. The first class uses simple averaging schemes: mean, median, and trimmed mean. The mean combination forecast sets $\omega_{i,t} = 1/N$ for $i = 1, \ldots, N$ in (2), the median combination forecast is the median of $\{\hat{r}_{i,t+1}\}_{i=1}^{N}$, and the trimmed mean combination forecast sets $\omega_{i,t} = 0$ for the individual forecasts with the smallest and largest values and $\omega_{i,t} = 1/(N-2)$ for the remaining individual forecasts in (2). Simple averaging schemes obviously do not require a holdout out-of-sample period.

The second class of combining methods is based on Stock and Watson (2004), where the combining weights formed at time t are functions of the historical forecast performance of the individual models over the holdout out-of-sample period. Their discount mean square prediction error (DMSPE) combining method uses the following weights:

$$\omega_{i,t} = m_{i,t}^{-1} / \sum_{j=1}^{N} m_{j,t}^{-1}, \tag{3}$$

where

$$m_{i,t} = \sum_{s=R}^{t-1} \theta^{t-1-s} (r_{s+1} - \hat{r}_{i,s+1})^2 \tag{4}$$

and θ is a discount factor. The DMSPE method thus assigns greater weights to individual predictive regression model forecasts that have lower MSPE values (better forecast performance) over the holdout out-of-sample period. When $\theta = 1$, there is no discounting, and (3) produces the optimal combination forecast derived by Bates and Granger (1969) for the case where the individual forecasts are uncorrelated. When $\theta < 1$, greater weight is attached to the recent forecast accuracy of the individual models. We consider the two values of 1.0 and 0.9 for θ.

The third class of combining methods is based on Aiolfi and Timmermann (2006) and is similar in spirit to Stock and Watson (2004). Aiolfi and Timmermann (2006) develop conditional combining methods that incorporate persistence in forecasting performance. We use their $C(K,PB)$ algorithm. The initial combination forecast is computed by grouping the individual forecasts over the initial holdout out-of-sample period, $\{\hat{r}_{i,s+1}\}_{s=R}^{R+P_0-1}$ $(i = 1,...,N)$ into K equal-sized clusters based on MSPE, with the first cluster containing the individual models with the lowest MSPE values, the second cluster containing the models with the next lowest MSPE values, and so on. The first combination forecast is the average of the individual forecasts of r_{R+P_0+1} for the models in the first cluster. In forming the second combination forecast, we compute the MSPE for the individual forecasts, $\{\hat{r}_{i,s+1}\}_{s=R+1}^{R+P_0}$, and again group the individual forecasts into K clusters. The second combination forecast is the average of the individual forecasts of r_{R+P_0+2} for the models included in the first cluster. We proceed in this manner through the end of the available out-of-sample period. Observe that the clusters are formed by computing MSPE using a rolling window. Following Aiolfi and Timmermann (2006), we consider $K = 2$ and $K = 3$ in our applications.

The final class of combining methods we consider is based on approximate Bayesian model averaging (ABMA), as in Garratt et al. (2003), in which functions of the SIC are used to approximate the posterior probabilities of the individual models (Draper, 1995). The combining weights are given by

$$\omega_{i,t} = \exp(\Delta_{i,t}) / \sum_{j=1}^{N} \exp(\Delta_{j,t}), \tag{5}$$

where $\Delta_{i,t} = SIC_{i,t} - \max_j(SIC_{j,t})$ and $SIC_{i,t}$ is the SIC corresponding to the estimated predictive regression model based on observations 1 through t used to generate $\hat{r}_{i,t+1}$. Intuitively, models with relatively high posterior probabilities over the model estimation period receive greater weights when forming the combination forecast. Garratt et al. (2003) also follow Burnham and Anderson (1998) and compute weights using the AIC, so that $\Delta_{i,t} = AIC_{i,t} - \max_j(AIC_{j,t})$ in (5). We consider ABMA combining weights based on both the SIC and AIC. Like the simple averaging schemes, the ABMA combining methods do not require a holdout out-of-sample period.[5]

2.3 Covariate Estimation

Similar to Faust and Wright (2005), we consider incorporating additional information to improve forecasts by estimating the parameters in (1) more efficiently. This can be accomplished by augmenting (1) with covariates when estimating α and β. As discussed by Faust and Wright (2005), we ideally would like to have a covariate (or vector of covariates), z_{t+1}, that is highly correlated with r_{t+1} and uncorrelated with $x_{i,t}$. Including such a covariate in (1) generates consistent estimates of α and β that can be substantially more precise than the conventional OLS estimates, which is likely in turn to improve forecasts based on (1). A challenge in using this procedure is to find a suitable z_{t+1} variable that is completely uncorrelated with $x_{i,t}$, as including a z_{t+1} variable in (1) that is correlated with $x_{i,t}$ produces inconsistent estimates of α and β. Nevertheless, even if z_{t+1} is correlated to a certain extent with $x_{i,t}$, there is a bias-efficiency tradeoff that can make it worthwhile to include z_{t+1} in (1) when estimating a forecast model based on (1).

We use a covariate estimation procedure that incorporates information from a large number (K) of additional stock returns series, $\{R_{k,t}\}_{k=1}^{K}$, including 41 U.S. industry-level stock returns and aggregate stock returns for eight industrialized countries (Australia, Canada, France, Germany,

[5] See Avramov (2002), Cremers (2002), and Dangl and Halling (2007) for Bayesian model averaging (BMA) in the context of stock return predictability. These studies focus either on in-sample analysis of predictive regression models or time-varying parameter models. In contrast, we focus on the ability of economic variables to generate equity premium forecasts capable of beating the historical average benchmark, as in Goyal and Welch (2007), in the standard predictive regression model framework. In addition, we focus on a classical approach and only use approximate Bayesian methods when forming combining weights using (5).

Japan, Sweden, Switzerland, and the U.K.), available on a monthly basis for 1927:1–2005:12 ($K = 49$).[6] These return series are generally highly correlated with aggregate U.S. returns. In order to use covariate estimates of α and β in (1) when forming a forecast of r_{t+1} at time t, we use the following procedure:

1. We compute the first q_1 principal components of $\{x_{i,s}\}_{i=1}^{N}$ for $s = 1, \dots, t$, where q_1 is selected using the IC_{p3} information criterion in Bai and Ng (2002) and a maximum q_1 value of five; denote the q_1-vector of principal components at time s by $z_{1,s}$.

2. For each k, we regress $\{R_{k,s}\}_{s=2}^{t}$ on a constant and $\{z_{1,s}\}_{s=1}^{t-1}$ and compute the residuals from this regression, which we denote by $\{\hat{u}_{k,s}\}_{s=2}^{t}$. By regressing $\{R_{k,s}\}_{s=2}^{t}$ on $\{z_{1,s}\}_{s=1}^{t-1}$, we are attempting to purge the correlation between $\{R_{k,s}\}_{s=2}^{t}$ and $\{x_{i,s}\}_{s=1}^{t-1}$ before including the return series as an augmenting covariate in (1).[7] Even if this is insufficient to purge all of the correlation between $\{R_{k,s}\}_{s=2}^{t}$ and $\{x_{i,s}\}_{s=1}^{t-1}$, it should help to reduce the bias in the covariate estimates of α and β in (1) and better exploit any efficiency gains.

3. We compute the first q_2 principal components of $\{\hat{u}_{k,s}\}_{k=1}^{K}$ for $s = 2, \dots, t$, where we again select q_2 using the IC_{p3} information criterion and a maximum q_2 value of five; denote the q_2-vector of principal components at time s by $z_{2,s}$.

4. We calculate covariate estimates of α and β ($\hat{\alpha}_t^{cov}$ and $\hat{\beta}_t^{cov}$, respectively) in (1) by regressing $\{r_s\}_{s=2}^{t}$ on a constant, $\{x_{i,s}\}_{s=1}^{t-1}$, and $\{z_{2,s}\}_{s=2}^{t}$. The forecast of r_{t+1} based on covariate estimation of (1) is given by $\hat{\alpha}_t^{cov} + \hat{\beta}_t^{cov} x_{i,t}$.[8]

In our empirical applications in Section 3 below, we examine whether covariate estimation of (1) helps to improve the out-of-sample forecast performance of individual predictive regression models. In addition, we analyze whether the combination forecasts described in Section 2.2 above

[6]The U.S. industry-level return series are from Kenneth French's web page at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/ and the international return series are from Global Financial Data.

[7]We could completely eliminate any correlation between $\{R_{k,s}\}_{s=2}^{t}$ and $\{x_{i,s-1}\}_{s=1}^{t-1}$ by using the residuals from a regression of $\{R_{k,s}\}_{s=2}^{t}$ on a constant and $\{x_{i,s}\}_{s=1}^{t-1}$. However, this would make the augmenting covariate in (1) orthogonal to $x_{i,t}$, and the resulting covariate estimates of α and β would be identical to the OLS estimates. This is tantamount to using seemingly unrelated regressions (SUR) to estimate a system of equations when all of the equations have the same set of regressors, so that the SUR and OLS estimates are identical.

[8]Faust and Wright (2005) suggest using macroeconomic survey forecast errors and the surprise components of macroeconomic news announcements as covariates in (1), and they find that these covariates can help to improve the forecast accuracy of predictive regression models for stock returns based on the dividend-price ratio and short-term nominal interest rate. We cannot use these covariates in our empirical applications in Section 3 below, as survey forecasts and macroeconomic news announcements are not available for the entire sample period we consider.

can be improved by basing them on individual forecasts generated from predictive regression models estimated using covariates.

Following Campbell and Thompson (2007), we also consider imposing theoretically motivated restrictions on the predictive regression model (1) when forming forecasts. First, if the sign of $\hat{\beta}_t$ does not agree with the theoretically expected sign, we restrict $\hat{\beta}_t$ to be zero when generating $\hat{r}_{i,t+1} = \hat{\alpha}_t + \hat{\beta}_t x_{i,t}$. For example, theory suggests that $\beta > 0$ in (1) when $x_{i,t}$ is a valuation ratio such as the dividend yield or earnings-price ratio, and we set $\hat{\beta}_t = 0$ when forming the forecast if $\hat{\beta}_t < 0$.[9] A second restriction sets $\hat{r}_{i,t+1}$ to zero if $\hat{\alpha}_t + \hat{\beta}_t x_{i,t} < 0$, so that the forecast of the equity premium based on (1) is restricted to be non-negative. These restrictions help to prevent some "aberrant" equity premium forecasts based on (1) from ruining the entire series of P out-of-sample forecasts. In our empirical applications in Section 3 below, we consider imposing restrictions on the individual predictive regression models estimated either with or without covariates and when the individual models are used to generate the combination forecasts described in Section 2.2 above.

2.4 Forecast Evaluation

We use the out-of-sample R^2 statistic, R^2_{OS}, suggested by Campbell and Thompson (2007) to compare the $\hat{r}_{j,t+1}$ and $\bar{r}_{t+1}$ forecasts, where j indexes either an individual forecast based on the predictive regression model (1) or a combination forecast. The R^2_{OS} statistic is akin to the familiar in-sample R^2 statistic and is given by

$$R^2_{OS} = 1 - \frac{\sum_{k=P_0+1}^{P} (r_{R+k} - \hat{r}_{j,R+k})^2}{\sum_{k=P_0+1}^{P} (r_{R+k} - \bar{r}_{R+k})^2}. \tag{6}$$

When $R^2_{OS} > 0$, the predictive regression model or combination forecasts have a lower MSPE than the benchmark forecasts based on the historical average. We also test whether the predictive regression model or combination forecasts have a significantly lower MSPE than the historical average benchmark forecasts; this is tantamount to testing the null hypothesis that $R^2_{OS} \leq 0$ against the alternative hypothesis that $R^2_{OS} > 0$.

The most popular method of testing for significant differences in MSPE is the Diebold and Mariano (1995) and West (1996) statistic, which has an asymptotic standard normal distribution when

[9] Following Campbell and Thompson (2007), restrictions on the slope coefficients are based on the full-sample estimates of the slope coefficient in (1). For the 1927:01–2005:12 full-sample period used in Section 3 below, this leads to restricting the slope coefficient to be non-negative for all variables, with the exceptions of the dividend payout ratio, net equity expansion, T-bill rate, long-term yield, and inflation (where the slope coefficient is restricted to be less than or equal to zero). These restrictions generally accord with theoretical priors.

comparing forecasts from non-nested models. However, Clark and McCracken (2001) and McCracken (2007) show that this statistic has a non-standard distribution when comparing forecasts from *nested* linear models, as is clearly the case when comparing predictive regression model forecasts of the equity premium to the historical average: setting $\beta = 0$ in (1) yields a model with a constant expected equity premium. Clark and West (2007) develop an adjusted version of the Diebold and Mariano (1995) and West (1996) statistic—what they label the *MSPE-adjusted* statistic—that can be used in conjunction with the standard normal distribution to generate asymptotically valid inferences when comparing forecasts from nested linear models.[10] The *MSPE-adjusted* statistic can be conveniently computed by first defining

$$f_{j,t+1} = (r_{t+1} - \bar{r}_{t+1})^2 - [(r_{t+1} - \hat{r}_{j,t+1})^2 - (\bar{r}_{t+1} - \hat{r}_{j,t+1})^2], \tag{7}$$

then regressing $\{f_{j,s+1}\}_{s=R+P_0}^{T-1}$ on a constant, and finally calculating the t-statistic corresponding to the constant. A p-value for a one-sided (upper-tail) test is then computed using the standard normal distribution. In Monte Carlo simulations, Clark and West (2007) show that the *MSPE-adjusted* statistic performs reasonably well in terms of size and power when comparing forecasts from nested linear models for a variety of sample sizes.[11]

Even if there is evidence that R^2_{OS} is statistically significant, R^2_{OS} values are typically small for predictive regression models, and this raises the issue of economic significance. Campbell and Thompson (2007) argue that even very small positive R^2_{OS} values, such as 0.5% for monthly data, can signal an economically meaningful degree of return predictability in terms of increased annual portfolio returns for a mean-variance investor. The main limitation to the R^2_{OS} measure is that it does not explicitly account for the risk borne by an investor over the out-of-sample period. To address this, following Marquering and Verbeek (2004), Campbell and Thompson (2007), Goyal and Welch (2007), and Wachter and Warusawitharana (2007), we also compute realized utility gains for a mean-variance investor on a real-time basis. More specifically, we first compute the average utility for a mean-variance investor with relative risk aversion parameter γ who allocates her portfolio monthly between stocks and risk-free bills using forecasts of the equity premium

[10] The Diebold and Mariano (1995) and West (1996) statistic can be severely undersized when comparing forecasts from nested linear models, leading to tests with very low power. Rapach and Wohar (2006b) show that there is stronger evidence of out-of-sample predictive ability for individual variables with respect to stock returns when tests with good size and power are used.

[11] The performance of the *MSPE-adjusted* statistic has not been analyzed for combination forecasts, so we know less about the reliability of the *MSPE-adjusted* when comparing the combination forecasts to the historical average forecasts. As described in Section 2.5 below, we also use a bootstrap procedure to assess the significance of differences in forecasting ability.

based on the historical average. This exercise requires the investor to forecast the variance of stock returns, and we follow Campbell and Thompson (2007) and assume that the investor estimates the variance using a five-year rolling window of monthly returns. A mean-variance investor who forecasts the equity premium using the historical average will decide at the end of period t to allocate the following share of her portfolio to equities in period $t+1$:

$$w_{0,t} = \left(\frac{1}{\gamma}\right)\left(\frac{\bar{r}_{t+1}}{\hat{\sigma}^2_{r,t+1}}\right), \tag{8}$$

where $\hat{\sigma}^2_{r,t+1}$ is the rolling-window estimate of the variance of stock returns.[12] Over the out-of-sample period, the investor realizes an average utility level of

$$\hat{v}_0 = \hat{\mu}_{0,p} - \left(\frac{1}{2}\right)\gamma\hat{\sigma}^2_{0,p}, \tag{9}$$

where $\hat{\mu}_{0,p}$ and $\hat{\sigma}^2_{0,p}$ are the sample mean and variance, respectively, over the out-of-sample period for the return on the benchmark portfolio formed using forecasts of the equity premium based on the historical average.

We then compute the average utility for the same investor when she forms forecasts of the equity premium using an individual predictive regression model or one of the combining methods. She will choose an equity share of

$$w_{j,t} = \left(\frac{1}{\gamma}\right)\left(\frac{\hat{r}_{j,t+1}}{\hat{\sigma}^2_{r,t+1}}\right), \tag{10}$$

and she realizes an average utility level of

$$\hat{v}_j = \hat{\mu}_{j,p} - \left(\frac{1}{2}\right)\gamma\hat{\sigma}^2_{j,p}, \tag{11}$$

where $\hat{\mu}_{j,p}$ and $\hat{\sigma}^2_{j,p}$ are the sample mean and variance, respectively, over the out-of-sample period for the return on the portfolio formed using forecasts of the equity premium based on an individual predictive regression model or combining method. We measure the utility gain as the difference between (11) and (9), and we multiply this difference by 1200 to express it in average annualized percentage return. The utility gain (or certainty equivalent return) can be interpreted as the portfolio management fee that an investor would be willing to pay to have access to the additional information available in a predictive regression model or combination forecast relative to the information in the historical equity premium alone. We report results for $\gamma = 3$; the results are qualitatively similar for other reasonable γ values.

[12]Following Campbell and Thompson (2007), we constrain the portfolio weight on stocks to lie between 0% and 150% (inclusive) each month, so that $w_{0,t} = 0$ ($w_{0,t} = 1.5$) if $w_{0,t} < 0$ ($w_{0,t} > 1.5$) in (8).

2.5 Data Mining and Robustness of Statistical Inferences

Statistically, data mining or snooping becomes an issue when considering multiple potential predictors or combining methods; see, for example, Lo and MacKinlay (1990), Foster et al. (1997), Inoue and Kilian (2004), and Rapach and Wohar (2006b). To help guard against data mining, we assess the statistical significance of the R^2_{OS} values and utility gains for the individual predictive regression models or combining methods using a version of the Inoue and Kilian (2004) bootstrap procedure based on maximal statistics. Intuitively, the bootstrap procedure "raises the bar" when assessing statistical significance by taking into account that researchers tend to focus on the "best" results when analyzing multiple potential predictors.

Consider first the R^2_{OS} and utility gain statistics corresponding to the $N = 14$ individual predictive regression models. We test the null hypothesis is that R^2_{OS}(utility gain) ≤ 0 against the alternative hypothesis that R^2_{OS}(utility gain) > 0 for at least one of the predictive regression models. We posit that the data are generated by the following process under the null hypothesis of no equity premium predictability (constant expected equity premium):

$$r_t = a_0 + e_{r,t}, \tag{12}$$

$$\begin{aligned} x_{1,t} &= b_{1,0} + b_{1,1}x_{1,t-1} + \ldots + b_{1,p_1}x_{1,t-p_1} + e_{1,t}, \\ &\vdots \\ x_{N,t} &= b_{N,0} + b_{N,1}x_{N,t-1} + \ldots + b_{N,p_N}x_{N,t-p_N} + e_{N,t}, \end{aligned} \tag{13}$$

where the disturbance vector $e_t = (e_{r,t}, e_{1,t}, \ldots, e_{N,t})'$ is independently and identically distributed with covariance matrix Σ. We first estimate (12) and each of the autoregressive (AR) processes in (13) using OLS and the full sample of data (with the lag order for each of the AR processes in (13) selected using the AIC) and compute the OLS residuals, $\{\hat{e}_t = (\hat{e}_{r,t}, \hat{e}_{1,t}, \ldots, \hat{e}_{N,t})'\}_{t=p+1}^{T}$, where $p = \max\{p_1, \ldots, p_N\}$. To generate a series of disturbances for the first pseudo-sample, we randomly draw (with replacement) $T + 100$ times from the OLS residuals, giving us a series of pseudo-series of disturbance terms, $\{e_t^*\}_{t=1}^{T+100}$. Observe that we draw from the OLS residuals jointly across (12) and each equation of (13), thereby preserving the contemporaneous correlation across all of the disturbances in the original sample. Using the pseudo-series of disturbance terms, OLS estimates of the coefficients in (12) and (13),[13] and setting the initial observations for each of the $x_{i,t}$ variables equal to zero in (13), we can build up a pseudo-sample of $T + 100$ observations,

[13] We actually use Shaman and Stine (1988, Table 1) bias-adjusted estimates of the slope coefficients for the AR processes in (13).

$\{r_t^*, x_{1,t}^*, \ldots, x_{N,t}^*\}_{t=1}^{T+100}$. We drop the first 100 transient start-up observations to randomize the initial observations, giving us a pseudo-sample of T observations, matching the original full-sample size. For the pseudo-sample, we compute the R^2_{OS} and utility gain values for each of the individual predictive regression models over the out-of-sample period of the pseudo-sample and store the maximal R^2_{OS} and utility gain values. We repeat this process 200 times, giving us empirical distributions for the maximal R^2_{OS} and utility gain statistics. After ordering the empirical distribution for each statistic, the 180th, 190th, and 198th values serve as the 10%, 5%, and 1% critical values, respectively, for each maximal statistic. If the R^2_{OS} or utility gain statistic corresponding to the original sample for an individual predictive regression model is greater than the bootstrap critical value for the maximal statistic, this provides evidence that the individual model forecast is significantly superior to the historical average forecast after controlling for data mining.

It is straightforward to use the data-mining bootstrap procedure to compute critical values for the combination forecasts by storing the maximal R^2_{OS} and utility gain values across all of the combining methods for each pseudo-sample. An augmented version of the bootstrap procedure can also be used to compute data-mining critical values for individual predictive regression models and combining methods based on covariate estimation.[14] In order to gain a broad sense of the statistical significance of the results in Section 3 below, we report the p-value for the Clark and West (2007) *MSPE-adjusted* statistic and indicate whether the R^2_{OS} or utility gain statistic is significant based on data-mining bootstrap critical values.

3 Empirical Results

3.1 Data

The monthly data are from Goyal and Welch (2007), who provide detailed descriptions of the data and their sources.[15] Stock returns are measured as continuously compounded returns on the S&P 500 index, including dividends, and the T-bill rate is used to compute the equity premium. With respect to the economic variables used to predict the equity premium, we consider the 14 variables from Goyal and Welch (2007) for which monthly data are available for 1927:01–2005:12.

[14] When computing bootstrap critical values for the individual predictive regression models and combining methods based on covariate estimation, we use 100 (instead of 200) bootstrap replications, as the computational costs are quite high for covariate estimation. For example, it took approximately 47 hours on a machine with a Pentium 4/3GHz processor to generate bootstrap critical values for covariate estimation using 100 replications for the 1947:01–2005:12 out-of-sample period analyzed in Section 3 below.

[15] The data are available at www.bus.emory.edu/AGoyal/Research.html.

- *Dividend-price ratio (log)*: difference between the log of dividends paid on the S&P 500 index and the log of prices, where dividends are measured using a twelve-month moving sum.
- *Dividend yield (log)*: difference between the log of dividends and the log of lagged prices.
- *Earnings-price ratio (log)*: difference between the log of earnings on the S&P 500 index and the log of prices, where earnings are measured using a twelve-month moving sum.
- *Dividend payout ratio (log)*: difference between the log of dividends and log of earnings.
- *Stock variance*: sum of squared daily returns on the S&P 500 index.
- *Book to market*: ratio of book value to market value for the Dow Jones Industrial Average.
- *Net equity expansion*: ratio of twelve-month moving sums of net issues by NYSE-listed stocks to total end-of-year market capitalization of NYSE stocks.
- *T-bill rate*: interest rate on a 3-month Treasury bill (secondary market).
- *Long-term yield*: long-term government bond yield.
- *Long-term return*: return on long-term government bonds.
- *Term spread*: difference between the long-term yield and the T-bill rate.
- *Default yield spread*: difference between BAA- and AAA-rated corporate bond yields.
- *Default return spread*: difference between long-term corporate bond and long-term government bond returns.
- *Inflation*: calculated from the CPI (all urban consumers); following Goyal and Welch (2007), since inflation rate data are released in the following month, we use $x_{i,t-1}$ in (1) for inflation.

We consider four different out-of-sample forecast evaluation periods. Three of the periods correspond to those analyzed in Goyal and Welch (2007): (i) a "long" out-of-sample period covering 1947:01–2005:12; (ii) a somewhat shorter out-of-sample period covering 1965:01–2005:12; (iii) a more recent out-of-sample period covering the last 30 years of the full sample, 1976:01–2005:12. Goyal and Welch's motivation for considering this last period is their finding that the out-of-sample predictive ability of a number of the economic variables deteriorates markedly after the Oil Shock of the mid-1970s. With this in mind, we also consider a very recent out-of-sample period covering the last six years of the full sample, 2000:01–2005:12, that allows us to analyze how the predictors fared over the recent market period characterized by the collapse of the "technology bubble." Overall, the consideration of multiple out-of-sample periods helps to provide us with a good sense of the robustness of the out-of-sample forecast results.[16]

Before reporting the complete results for each of the four out-of-sample periods, following Goyal and Welch (2007), we first present time-series plots of the differences between the cumulative square prediction error for the historical average benchmark forecasts and the cumulative square prediction error for the forecasts based on the individual predictive regression models (estimated using OLS) in Figure 1 for 1947:01–2005:12.[17] This is an informative graphical device that provides a visual impression of the consistency of an individual predictive regression model's out-of-sample forecast performance over time. When the curve in each panel of Figure 1 increases, the predictive regression model outperforms the historical average, while the opposite holds when the curve decreases.[18] The plots can be conveniently used to determine whether an individual predictive regression model has a lower MSPE than the historical average for any particular out-of-sample period by redrawing the horizontal zero line to correspond to the start of the out-of-sample period. Essentially, we can compare the height of the curve at the two points corresponding to the beginning and end of a given out-of-sample period: if the curve is higher (lower) at the end of the out-of-sample period than at the beginning, the predictive regression model (historical average) has a lower MSPE over the out-of-sample period. A predictive regression model that is always able to outperform the historical average for any out-of-sample period will thus have a curve with a slope that is always positive; the closer a predictive regression model is to this ideal, the greater its ability to consistently beat the historical average in terms of MSPE.

Figure 1 shows that none of the 14 individual economic variables consistently outperforms the historical average. Some of the panels have positively sloped curves during certain periods, but all panels also display extended periods where the curves are fairly flat or have negative slopes. It is interesting to note that a number of valuation ratios, such as the dividend-price ratio, dividend yield, and book to market, have distinctly positive slopes in the early 1970s; however, the curves either

[16] Note that the out-of-sample periods refer to the periods used to evaluate the out-of-sample forecasts. As indicated in Section 2.2 above, some of the combining methods require a holdout out-of-sample period, and we use the five years (60 months) before the start of the out-of-sample evaluation period as the initial holdout out-of-sample period.

[17] The plots for some of the individual predictive regression model forecasts are available in Figure 3 of Goyal and Welch (2007). We include them so that the results for all of the variables we consider are conveniently available in Figure 1.

[18] As pointed out by Goyal and Welch (2007), the units on the plots are not intuitive.

flatten considerably or become negatively sloped after the Oil Shock, and they become markedly negatively sloped during the 1990s. This erratic out-of-sample performance renders these valuation ratios unreliable out-of-sample predictors of the equity premium. Overall, Figure 1 visually conveys the primary message of Goyal and Welch (2007): due to instabilities, it is difficult to identify individual predictors that reliably outperform the historical average with respect to forecasting the equity premium.

Figure 2 plots the differences between the cumulative square prediction error for the historical average forecasts and the cumulative square prediction error for the combination forecasts we consider. In contrast to Figure 1, the curves in Figure 2 have slopes that are predominantly positive (with the exceptions of the cluster combining methods), indicating that the combination forecasts deliver out-of-sample gains on a much more consistent basis over time than the individual predictors. This suggests that combination forecasts are effective in mitigating the instabilities identified by Goyal and Welch (2007) that characterize individual predictive regression models.

The lines in Figures 1 and 2 have qualitatively similar shapes when we estimate the individual predictive regression models using covariates, either with or without imposing theoretically motivated restrictions. This indicates that combination forecasts are crucial in our investigation for dealing with the instabilities that plague individual predictive regression models and limit their usefulness to investors in real time. Nevertheless, detailed results for the four out-of-sample periods reported in Tables 1–4 below show that covariate estimation and theoretically motivated restrictions provide useful complements to combination forecasts, in that combination forecasts based on individual models estimated using covariates with theoretically motivated restrictions imposed typically provide additional out-of-sample gains relative to combination forecasts based on models estimated using OLS with no restrictions imposed. We turn next to the detailed results for the four out-of-sample periods.

3.2 1947:01–2005:12 Out-of-Sample Period

Table 1 presents results for the 1947:01–2005:12 out-of-sample period. The table reports p-values for the Clark and West (2007) *MSPE-adjusted* statistic, R^2_{OS} values, and average utility gains for each of the individual predictive regression models and combining methods relative to the historical average benchmark model. Results are reported for forecasts generated using OLS (Panels A and B) and covariate estimation (Panels B and D), as well as with and without imposing restrictions on the individual models.[19] We also indicate in each panel whether the R^2_{OS} values and utility gains are significant at conventional levels according to data-mining critical values generated using the bootstrap procedure described in Section 2.5 above. The complete bootstrap critical values for all of the out-of-sample periods are reported in Tables 5 and 6.

The second and third columns of Panel A of Table 1 show that for the unrestricted forecasts, only four of the 14 predictors have a positive R^2_{OS} value, none of which is above 0.33%. Just one (three) of the positive R^2_{OS} values is significant at the 5% (10%) level according to the *MSPE-adjusted* p-values, while none of the R^2_{OS} values are significant at conventional levels according to the data-mining critical values. Seven variables exhibit positive utility gains in the fourth column of Panel A, but none of the utility gains are significant based on the data-mining critical values. Restricting the sign of the slope coefficient in (1) provides a slight improvement in the overall performance of the individual models: one additional variable has a positive R^2_{OS} value that is significant at the 10% level according to the *MSPE-adjusted* p-value, and the utility gain increases in some cases in the seventh column of Panel A compared to those that are positive in the fourth column; however, none of the R^2_{OS} values are significant using the data-mining critical values, and the only utility gain that is significant is for the term spread at the 10% level. When the sign of the forecast is restricted, seven of the 14 state variables have a positive R^2_{OS} value, and four (six) of these values are significant at the 5% (10%) level based on the *MSPE-adjusted* p-values. Restricting the sign of the forecast has no effect on the average utility gains in the tenth column of Panel A relative to the unrestricted forecasts. None of the R^2_{OS} values and average utility gains in the ninth and tenth columns of Panel A are significant using the data-mining critical values.

Comparing the results in Panel B for the combination forecasts to those in Panel A, the most striking finding is how the various combining methods improve the accuracy of the out-of-sample forecasts. Most of the R^2_{OS} values in the third column of Table 1 are close to 1%. As shown in Campbell and Thompson (2007), a mean-variance investor can increase her monthly portfolio return by a proportional factor of R^2_{OS}/S^2, where S is the Sharpe ratio. For a squared monthly Sharpe ratio of 1.3% corresponding to our 1927:01–2005:12 full-sample period, an R^2_{OS} value of approximately 0.5% or greater corresponds to a sizable and economically meaningful proportional factor increase in monthly returns. For example, the R^2_{OS} value of 0.95% for the mean combination forecast in the third column of Panel B corresponds to a proportional factor increase in monthly portfolio returns of $0.95/1.3 = 71\%$. Seven of the R^2_{OS} values in the third column of Panel B are

[19] In situations where the p-values and R^2_{OS} values for the unrestricted forecasts in Tables 1–4 exactly match those for the restricted forecasts, this indicates that the restrictions are never binding over the out-of-sample period.

significant at the 1% level, and the other two (corresponding to the cluster combining methods) are significant at the 10% level, based on the *MSPE-adjusted p*-values; all of the R^2_{OS} values are significant at the 1% level according to the data-mining critical values that control for the fact that we consider multiple combining methods. Also note that all of the R^2_{OS} values for the combination forecasts in the third column of Panel B are greater than the largest R^2_{OS} value among the individual models in Panel A. Most of the utility gains in the fourth column of Table 1 are greater than 1%, and all are significant at the 5% or 10% level based on the data-mining critical values. Restricting the sign of the slope coefficient when estimating the individual predictive regression models used to form the combination forecasts often produces additional (albeit fairly small) gains, as evidenced by the increases in R^2_{OS} and average utility gains when we move from the third and fourth columns of Panel B to the sixth and seventh columns. In contrast, as shown in the last three columns of Panel B, with the exceptions of the cluster methods, restricting the sign of the individual model forecasts when forming the combination forecasts leads to a small deterioration in out-of-sample performance relative to the results in the second through seventh columns of Panel B.

Panel C of Table 1 reports results for covariate estimation of the individual predictive regression models. When no restrictions are imposed, covariate estimation leads to seven of the individual variables having positive R^2_{OS} values, and four (five) of these values are significant at the 5% (10%) level according to the *MSPE-adjusted p*-values (see the second and third columns of Panel C). Note that the positive R^2_{OS} values are larger for a number of variables when we use covariate estimation in Panel C compared to OLS estimation in Panel A, and two of the R^2_{OS} values in Panel C are near or slightly above 1%. There is little change in the R^2_{OS} values for the individual models when we restrict the sign of the slope coefficient (see the sixth column of Panel C). Restricting the sign of the forecast for the individual models is more helpful: from the ninth column of Panel C, we see that nine of the individual variables have positive R^2_{OS} values based on covariate estimation with the forecast sign restricted, and six (seven) are significant at the 5% (10%) level. Two of the R^2_{OS} values, for the long-term return and term spread, are significant in the third, sixth, and ninth columns of Panel C according to the data-mining critical values. There is little difference among the utility gains across the fourth, seventh, and tenth columns of Panel C, and the utility gains for the long-term return and/or term spread are significant at the 5% or 10% level in these columns based on the data-mining critical values.

The results in Panel D of Table 1 show that combination forecasts based on individual predictive regression models estimated using covariates generally perform very well over the 1947:01–

2005:12 out-of-sample period. The third column of Panel D shows that all of the R^2_{OS} values are sizable, and a number are substantially above unity. Seven of the R^2_{OS} values are significant at the 1% level using the *MSPE-adjusted p*-values, and the other two are significant at the 5% level; in addition, all of the R^2_{OS} values are significant at the 1% level according to the data-mining critical values. Eight of the nine combination forecasts have an annualized utility gain that is greater than 1% in the fourth column of Panel D, and all are significant at the 5% level or lower based on the data-mining critical values. As indicated in the sixth and seventh columns of Panel D, restricting the sign of the slope coefficient when using covariates typically provides further increases in R^2_{OS} values and utility gains for the combination forecasts. As shown in the last two columns of Panel D, imposing restrictions on the sign of individual model forecasts when forming the combination forecasts is less helpful. By comparing the results in Panels B and D of Table 1, we see that it is generally beneficial to employ covariates when estimating the individual predictive regression models that are used to generate the combination forecasts.

The results in Table 1 show that forecast combining methods, especially when complemented by covariate estimation and slope coefficient restrictions, are able to deliver meaningful out-of-sample gains. This indicates that practical methods are available for beating the historical average benchmark model by statistically and economically significant margins over the 1947:01–2005:12 out-of-sample period. The results in Tables 2–4 below show that similar results hold over other out-of-sample periods.

3.3 1965:01–2005:12 Out-of-Sample Period

Table 2 reports results for the 1965:01–2005:12 out-of-sample period. The results for the individual predictive regression models estimated using OLS in Panel A of Table 2 are similar to the corresponding results for the 1947:01–2005:12 out-of-sample period in Table 1. However, a few additional individual predictors offer significant utility gains according to the data-mining critical values in Panel A of Table 2 relative to Table 1. The combination forecasts based on individual models estimating using OLS in Panel B of Table 2 again perform consistently well, with R^2_{OS} values typically close to 1% and sizable utility gains. The R^2_{OS} values and utility gains for the combination forecasts are all significant at the 5% level or lower based on the data-mining critical values. Comparing the results in the third and fourth columns of Panel B to those in the sixth and seventh columns, restricting the sign of the slope coefficient before forming the combination forecasts is almost always beneficial; in contrast, restricting the sign of the forecast typically does not help (see the final two columns of Panel B).

Examining the results across Panels A and C of Table 2, we see that covariate estimation helps to increase the R^2_{OS} values and utility gains for a number of the individual models relative to OLS estimation. Most notable in Table 2 are the results in Panel D. Combination forecasts coupled with covariate estimation produce R^2_{OS} values that are consistently near or above 1% and utility gains that are often in the range of 1%–2%, and all of the R^2_{OS} values in Panel D are significant at the 5% level in Panel D using either the *MSPE-adjusted* p-values or the data-mining critical values. Almost all of the utility gains in Panel D are also significant according to the data-mining critical values. As in Table 1, forming combination forecasts using covariate estimation of the individual models with the sign of the slope coefficient restricted appears to be the best overall strategy.

3.4 1976:01–2005:12 Out-of-Sample Period

Goyal and Welch (2007) find that the out-of-sample predictive ability of many individual economic variables deteriorates markedly over the 1976:1–2005:12 period, and our results in Panel A of Table 3 confirm their findings. Only two of the R^2_{OS} values are positive in the third column of Panel A, but these values are only 0.14% and 0.06% and neither is significant at conventional levels based on the *MSPE-adjusted* p-values or data-mining critical values. Similar results obtain in Panel A when restrictions are imposed on the individual predictive regression models. While the individual models perform quite poorly overall in Panel A, combining methods based on individual models estimated using OLS again generally perform well in Panel B. With the exceptions of the cluster methods, all of the combining methods yield positive R^2_{OS} values ranging from approximately 0.30%–0.50% that are significant at the 10% level or lower according to the data-mining critical values. There are also fairly sizable utility gains associated with the positive R^2_{OS} values in Panel B, especially in the fourth and seventh columns, and these utility gains are also significant at the 10% level or less using the data-mining critical values.

Using covariates to estimate the individual predictive regression models leads to some improvements in Panel C of Table 3 relative to the results for OLS estimation in Panel A. When no restrictions are imposed, three of the R^2_{OS} values are positive in Panel C, but none are significant using the data-mining critical values. Four of the R^2_{OS} values are positive when either the sign of the slope coefficient or the sign of the forecast is restricted, but again none are significant according to the data-mining critical values. None of the utility gains are significant in Panel C based on the data-mining critical values. As shown in Panel D of Table 3—and similar to the results in Tables 1 and 2—combination forecasts based on individual models estimated using covariates typically outperform the historical average benchmark model. When no restrictions are placed on the individual predictive regression models or the slope coefficient is restricted, seven of the nine combination forecasts have positive R^2_{OS} values in Panel D, the exceptions being the two cluster forecasts; all nine of the combination forecasts are positive in Panel D when the sign of the forecast is restricted. The largest R^2_{OS} values and utility gains in Panel D typically result from restricting the sign of the slope coefficient in the individual predictive regression model, again in line with the results reported in Tables 1 and 2. There are a number of instances where the the R^2_{OS} values or utility gains are significant according to the data-mining critical values in Panel D.

3.5 2000:01–2005:12 Out-of-Sample Period

The final out-of-sample period we consider is 2000:01–2005:12, and the results are reported in Table 4. From Panel A of Table 4, we see that individual predictive regression models based on valuation ratios substantially outperform the historical average over this very recent period, with R^2_{OS} values of approximately 4%–6% that are significant using either the *MSPE-adjusted p*-values or data-mining critical values. The average utility gains are also quite large for the valuation ratios in Panel A, ranging from approximately 6%–7%, and are significant based on the data-mining critical values. The restrictions on the slope coefficients are never binding in Panel A (the results in the second through fourth and fifth through seventh columns are identical), and restricting the sign of the forecast usually does not help. Following the trend in Tables 1–3, Panel B shows that combination forecasts based on individual models estimated using OLS typically outperform the historical average by a sizable margin. With the exceptions of the cluster combining methods, all of the R^2_{OS} values are positive, and most are are significant using either the *MSPE-adjusted p*-values or data-mining critical values. The utility gains are also quite sizable in Panel B for the combining methods with positive R^2_{OS} values and most are significant based on the data-mining critical values.

Again following the trend in Tables 1–3, combination forecasts based on covariate estimation of individual models provide additional gains relative to OLS estimation (compare Panels B and D of Table 4). The results for the combination forecasts based on covariate estimation of individual models are very similar in Panel D whether or not the slope coefficient is restricted, and restricting the sign of the forecast is typically not beneficial relative to these cases. There are numerous instances of statistical significance using either the *MSPE-adjusted p*-values or data-mining critical values in Panel D.

3.6 Discussion

The key findings and implications reported in Sections 3.1–3.5 can be summarized as follows:

- The results in Figure 1 and Panel A of Tables 1–4 reinforce the findings of Goyal and Welch (2007) and demonstrate that it is very difficult to identify individual economic variables capable of generating reliable out-of-sample forecasts of the equity premium. Indeed, there is no single variable among the 14 we consider that can deliver a positive R^2_{OS} value over each of the four out-of-sample periods examined in Tables 1–4. As argued by Goyal and Welch (2007), this is not encouraging for an investor hoping to exploit return predictability in real time based on a single variable.

- Nevertheless, we show that a number of combination forecasts—especially combination forecasts used in conjunction with individual predictive regression models estimated using covariates and with the slope coefficient restricted to conform to theoretical priors—outperform the historical average for a variety of out-of-sample periods.[20] We have thus identified effective methods for forecasting the equity premium based on economic variables that consistently beat the historical average in real time.

The inconsistent forecast performance of individual economic variables appears to be the result of individual predictive regression model instability. We find that combination forecasts are a very useful tool for improving equity premium forecasts in the presence of individual model instability.[21] Another strategy involves pre-testing for structural breaks. In a recent paper, Lettau and Van Nieuwerburgh (2007) show that the in-sample evidence of stock return predictability based on valuation ratios is substantially stronger and stable over time if one allows for periodic structural breaks in the steady-state valuation ratio. However, they find that it is difficult to identify the size and timing of the breaks in the steady-state valuation ratio in real time, so that they typically cannot improve out-of-sample forecasts of the equity premium by pre-testing for breaks. In contrast, the combination forecasts we consider appear to offer improved methods for accommodating structural breaks of the type identified by Lettau and Van Nieuwerburgh (2007) when generating

[20] The cluster methods are the only combining methods that do not outperform the historical average for each of the out-of-sample periods examined in Tables 1–4.

[21] Following Goyal and Welch (2007), we also generated forecasts using a "kitchen sink" model that included all of the 14 individual economic variables jointly as regressors in (1). As in Goyal and Welch (2007), we find that the kitchen sink model is typically substantially outperformed by the historical average. The complete results for the kitchen sink forecasts are available upon request from the authors.

real-time equity premium forecasts.[22]

We also stress that we are concerned with providing practical methods for consistently beating the historical average when forecasting the equity premium in real time. This is a separate issue from determining the exact nature and extent of return predictability in population; see, for example, Inoue and Kilian (2004) and Cochrane (2007). While a number of recent studies such as Pástor and Stambaugh (2006) and Cochrane (2007) find that more powerful tests provide strong evidence of in-sample return predictability, it has remained difficult to beat the historical average with respect to out-of-sample forecasts of the equity premium. This has limited the relevance of return predictability based on economic variables for investors in real time. In contrast, our results show that economic variables can provide statistically and economically relevant forecasts of the equity premium for investors in real time.

Finally, we performed all of the out-of-sample forecasting exercises using quarterly and annual data. The results are qualitatively very similar to those reported in this section for monthly data, confirming our finding that forecast combination forecasts of the equity premium consistently beat the historical average in real time.[23]

4 Conclusion

While numerous economic variables have been identified in the literature with in-sample predictive ability for the equity premium, Goyal and Welch (2007) show that individual variables fail to deliver consistent out-of-sample forecast gains relative to the historical average. Although individual variables do not appear to possess consistent out-of-sample predictive power, by accommodating structural instability via forecast combining methods—as well as utilizing additional information and imposing theoretically motivated restrictions à la Campbell and Thompson (2007)—we provide convincing evidence of the out-of-sample predictive ability of 14 economic variables taken as a whole over a number of out-of-sample periods. Overall, we find that out-of-sample return predictability using economic variables does show up strongly in the data. In addition to being of

[22] Related to this approach, the estimation window for the predictive regression model (1) could be selected on the basis of a structural break test (Pesaran and Timmermann, 2002). Note that even when a break occurs, there is a bias-efficiency tradeoff, so that it can be optimal to include pre-break data when estimating the predictive regression model used to generate the forecast (Clark and McCracken, 2004; Pesaran and Timmermann, 2007). While beyond the scope of the present paper, it would be interesting to combine forecasts from individual models estimated using different window sizes to address the uncertainty surrounding the choice of the optimal window size (Pesaran and Timmermann, 2007; Timmermann, 2007).

[23] The quarterly and annual data are also available at www.bus.emory.edu/AGoyal/Research.html. The complete results for quarterly and annual data are available upon request from the authors.

practical interest, this has important theoretical implications, as recently emphasized by Cochrane (2007), who provides a strong theoretical rationale for the predictability of stock returns.

Our results also provide some guidance for the construction of future applied asset pricing models. For tractability, existing models are typically based on one or a few state variables to determine time-varying expected returns and the associated investment opportunities. In contrast to this traditional modeling approach, we find that a host of variables can collectively and consistently forecast stock returns. As we discussed in the introduction, it is likely that the data-generating process for expected stock returns is highly complex and constantly evolving, making it difficult to approximate the process well with a single model. Our results suggest that future applied asset pricing models should be less concerned with parsimony and instead allow for richer processes that better mimic time-varying fluctuations in expected returns.

References

Aiolfi, M. and A. Timmermann, 2006, "Persistence in Forecasting Performance and Conditional Combination Strategies," *Journal of Econometrics*, 135, 31–53.

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

Avramov, D., 2002, "Stock Return Predictability and Model Uncertainty," *Journal of Financial Economics*, 64, 423–458.

Bai, J. and S. Ng, 2002, "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 70, 191–221.

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

Bates, J.M. and C.W.J. Granger, 1969, "The Combination of Forecasts," *Operational Research Quarterly*, 20, 451–468.

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.

Boudoukh, J., R. Michaely, M.P. Richardson, and M.R. Roberts, 2005, "On the Importance of Measuring Payout Yield: Implications for Empirical Asset Pricing," *Journal of Finance*, forthcoming.

Breen, W., L.R. Glosten, and R. Jagannathan, 1989, "Economic Significance of Predictable Variations in Stock Index Returns," *Journal of Finance*, 64, 1177–1189.

Burnham, K.P. and D.R. Anderson, 1998, *Model Selection and Inference: A Practical Information-Theoretic Approach*, New York: Springer-Verlag.

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

Campbell, J.Y., 2000, "Asset Pricing at the Millennium," *Journal of Finance*, 55, 1515–1567.

Campbell, J.Y., 2008, "Viewpoint: Estimating the Equity Premium," *Canadian Journal of Economics*, 41, 1–21.

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

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

Campbell, J.Y. and S.B. Thompson, 2007, "Predicting the Equity Premium Out of Sample: Can

Anything Beat the Historical Average?" *Review of Financial Studies*, forthcoming.

Campbell, J.Y. and T. Vuolteenaho, 2004, "Inflation Illusion and Stock Prices," *American Economic Review*, 94, 19–23.

Clark, T.E. and M.W. McCracken, 2001, "Tests of Equal Forecast Accuracy and Encompassing for Nested Models," *Journal of Econometrics*, 105, 85–110.

Clark, T.E. and M.W. McCracken, 2004, "Improving Forecast Accuracy by Combining Recursive and Rolling Forecasts," Federal Reserve Bank of Kansas City Research Working Paper 04-10.

Clements, M.P. and D.F. Hendry, 2006, "Forecasting with Breaks," In G. Elliott, C.W.J. Granger, and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, Amsterdam: Elsevier, pp. 605–657.

Cochrane, J.H., 2007, "The Dog That Did Not Bark: A Defense of Return Predictability," *Review of Financial Studies*, forthcoming.

Cremers, K.J.M., 2002, "Stock Return Predictability: A Bayesian Model Selection Perspective," *Review of Financial Studies*, 15, 1223–1249.

Dangl, T. and M. Halling, 2007, "Predictive Regressions with Time-Varying Coefficients," Manuscript, University of Utah.

Diebold, F.X. and R.S. Mariano, 1995, "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics*, 13, 253–263.

Dow, C.H., 1920, "Scientific Stock Speculation," *The Magazine of Wall Street*.

Draper, D., 1995, "Assessment and Propagation of Model Uncertainty," *Journal of the Royal Statistical Society Series B*, 58, 45–97.

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

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

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

Faust, J. and J.H. Wright, 2005, "Efficient Predictive Regressions," Manuscript, Johns Hopkins University.

Foster, F.D., T. Smith, and R.E. Whaley, 1997, "Assessing Goodness-of-Fit of Asset Pricing Models: The Distribution of the Maximal R^2," *Journal of Finance*, 53, 591-607.

Garratt, A., K. Lee, M.H. Pesaran, and Y. Shin, 2003, "Forecast Uncertainties in Macroeconometric Modelling: An Application to the U.K. Economy," *Journal of the American Statistical Association*, 98, 829–838.

Goyal, A. and I. Welch, 2003, "Predicting the Equity Premium with Dividend Ratios," *Management Science*, 49, 639–654.

Goyal, A. and I. Welch, 2007, "A Comprehensive Look at the Empirical Performance of Equity Premium Prediction," *Review of Financial Studies*, forthcoming.

Guo, H., 2006, "On the Out-of-Sample Predictability of Stock Market Returns," *Journal of Business*, 79, 645–670.

Hendry, D.F. and M.P. Clements, 2004, "Pooling of Forecasts," *Econometrics Journal*, 7, 1–31.

Inoue, A. and L. Kilian, 2004, "In-Sample or Out-of-Sample Tests of Predictability: Which One Should We Use?" *Econometric Reviews*, 23, 371–402.

Kothari, S. and J. Shanken, 1997, "Book-to-Market, Dividend Yield, and Expected Market Returns: A Time-Series Analysis," *Journal of Financial Economics*, 44, 169–203.

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

Lettau, M. and S. Van Nieuwerburgh, 2007, "Reconciling the Return Predictability Evidence," *Review of Financial Studies*, forthcoming.

Lo, A.W. and A.C. MacKinlay, 1990, "Data-Snooping Biases in Tests of Financial Asset Pricing Models," *Review of Financial Studies*, 3, 431-467.

Marquering, W. and M. Verbeek, 2004, "The Economic Value of Predicting Stock Index Returns and Volatility," *Journal of Financial and Quantitative Analysis*, 39, 407–429.

McCracken, M.W., 2007, "Asymptotics for Out of Sample Tests of Granger Causality," *Journal of Econometrics*, 140, 719–752.

Nelson, C.R., 1976, "Inflation and the Rates of Return on Common Stock," *Journal of Finance*, 31, 471–483.

Pástor, L. and R.F. Stambaugh, 2006, "Predictive Systems: Living with Imperfect Predictors," Manscript, University of Chicago Graduate School of Business.

Paye, B. and A. Timmermann, 2006, "Instability of Return Prediction Models," *Journal of Empirical Finance*, 13, 274–315.

Pesaran, M.H. 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, 2007, "Selection of Estimation Window in the Presence of

Breaks," *Journal of Econometrics*, 137, 134–161.

Pontiff, J., and L.D. Schall, 1998, "Book-to-Market Ratios as Predictors of Market Returns," *Journal of Financial Economics*, 49, 141–160.

Rapach, D.E. and M.E. Wohar, 2006a, "Structural Breaks and Predictive Regression Models of Aggregate U.S. Stock Returns," *Journal of Financial Econometrics*, 4, 238–274.

Rapach, D.E. and M.E. Wohar, 2006b, "In-Sample vs. Out-of-Sample Tests of Stock Return Predictability in the Context of Data Mining," *Journal of Empirical Finance*, 13, 231–247.

Shaman, P. and R.A. Stine, 1988, "The Bias of Autoregressive Coefficient Estimators," *Journal of the American Statistical Association*, 83, 842-848.

Stock, J.H. and M.W. Watson, 2004, "Combination Forecasts of Output Growth in a Seven-Country Data Set," *Journal of Forecasting*, 23, 405–430.

Timmermann, A., 2006, "Forecast Combinations," In G. Elliott, C.W.J. Granger, and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, Amsterdam: Elsevier, pp. 135–196.

Timmermann, A., 2007, "Elusive Return Predictability," *International Journal of Forecasting*, forthcoming.

Wachter, J.A. and M. Warusawitharana, 2007, "Predictable Returns and Asset Allocation: Should a Skeptical Investor Time the Market?" Manuscript, University of Pennsylvania.

West, K.D., 1996, "Asymptotic Inference About Predictive Ability," *Econometrica*, 64, 1067–1084.

Notes to Tables 1–4: For Panels A and C, the first column indicates the economic variable included in the predictive regression forecast model; for Panels B and D, the first column indicates the method used to generate the combination forecasts based on the individual predictive regression forecast models. "OLS estimation" indicates that the individual predictive regression forecast models are estimated using OLS; "covariate estimation" indicates that the individual models are estimated using the covariate approach described in the text. "Unrestricted coefficients" means that the intercept and slope coefficients are unrestricted in the individual predictive regression forecast models; "Restricted slope coefficient" means that the sign of the slope coefficient is restricted in line with a theoretical prior in the individual predictive regression forecast models; "Restricted forecast sign" means that the equity premium forecasts are restricted to be positive for the individual predictive regression forecast models. "*MSPE-adj. p*-value" is the *p*-value for the Clark and West (2007) out-of-sample *MSPE-adjusted* statistic; the statistic corresponds to a one-sided test of the null hypothesis that the competing forecast model given in the first column has equal expected square prediction error relative to the historical average benchmark forecast model against the alternative hypothesis that the competing forecast model has a lower expected square prediction error than the historical average benchmark forecast model. The R^2_{OS} measure is the Campbell and Thompson (2007) out-of-sample R^2 statistic. Utility gain is the portfolio management fee (in annualized percentage return) that an investor with mean-variance preferences and risk aversion coefficient of three would be willing to pay to have access to the forecast model given in the first column relative to the historical average benchmark forecast model; the weight on stocks in the investor's portfolio is restricted to lie between zero and 1.5 (inclusive). 0.00 indicates less than 0.005 in absolute value. †, *, and ** indicate significance at the 10%, 5%, and 1% levels, respectively, according to the data-mining bootstrap procedure described in the text; critical values for the data-mining bootstrap procedure are reported in Tables 5 and 6.

Table 1: Equity premium forecast results, 1947:01–2005:12 out-of-sample evaluation period

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Unrestricted coefficients			Restricted slope coefficient			Restricted forecast sign		
Forecast model	*MSPE-adj.* p-value (%)	R^2_{OS} (%)	Utility gain (annual %)	*MSPE-adj.* p-value (%)	R^2_{OS} (%)	Utility gain (annual %)	*MSPE-adj.* p-value (%)	R^2_{OS} (%)	Utility gain (annual %)
A. Individual predictors, OLS estimation									
Dividend-price ratio	1.59	0.00	-1.19	1.59	0.00	-1.19	0.73	0.47	-1.19
Dividend yield	1.25	-0.74	-1.52	1.25	-0.74	-1.52	0.69	0.25	-1.52
Earnings-price ratio	4.75	-1.13	-0.56	4.75	-1.13	-0.56	6.75	-1.14	-0.56
Dividend payout ratio	80.08	-0.75	-0.19	76.94	-0.70	-0.13	80.08	-0.75	-0.19
Stock variance	78.21	-0.53	-0.31	78.21	-0.53	-0.31	78.21	-0.53	-0.31
Book to market	18.01	-2.50	-2.25	18.01	-2.50	-2.25	18.40	-1.80	-2.25
Net equity expansion	8.60	0.01	0.05	8.60	0.01	0.05	8.60	0.01	0.05
T-bill rate	3.73	0.25	1.24	3.73	0.25	1.24	3.15	0.50	1.24
Long-term yield	5.33	-0.52	0.81	5.33	-0.52	0.81	1.59	0.51	0.81
Long-term return	34.21	-0.51	0.19	8.97	0.21	0.71	32.50	-0.42	0.19
Term spread	5.39	0.33	1.58	5.36	0.34	$1.58^{\dagger}$	5.58	0.36	1.58
Default yield spread	26.26	-0.17	-0.68	26.26	-0.17	-0.68	26.04	-0.15	-0.68
Default return spread	83.75	-0.22	0.01	73.17	-0.14	0.08	83.75	-0.22	0.01
Inflation	15.07	0.20	0.41	15.07	0.20	0.41	12.43	0.26	0.41
B. Combining methods, OLS estimation									
Mean	0.09	0.95^{**}	1.13^{*}	0.06	1.01^{**}	1.15^{*}	0.14	0.81^{**}	$0.61^{\dagger}$
Median	0.02	0.81^{**}	$0.78^{\dagger}$	0.02	0.84^{**}	0.80^{*}	0.02	0.81^{**}	0.78^{*}
Trimmed mean	0.09	0.91^{**}	1.00^{*}	0.05	0.97^{**}	1.02^{*}	0.10	0.83^{**}	$0.70^{\dagger}$
DMSFE, $\theta = 1.0$	0.09	0.95^{**}	1.13^{*}	0.06	1.01^{**}	1.15^{*}	0.14	0.82^{**}	$0.61^{\dagger}$
DMSFE, $\theta = 0.9$	0.10	0.96^{**}	1.27^{*}	0.07	1.01^{**}	1.28^{**}	0.12	0.81^{**}	$0.69^{\dagger}$
$C(2,PB)$	5.17	0.46^{**}	$0.82^{\dagger}$	5.69	0.46^{**}	0.84^{*}	1.17	0.63^{**}	0.95^{*}
$C(3,PB)$	5.25	0.42^{**}	$0.70^{\dagger}$	5.80	0.42^{**}	$0.69^{\dagger}$	0.91	0.73^{**}	0.83^{*}
ABMA, SIC	0.09	0.95^{**}	1.13^{*}	0.06	1.01^{**}	1.15^{*}	0.14	0.81^{**}	$0.61^{\dagger}$
ABMA, AIC	0.09	0.95^{**}	1.13^{*}	0.06	1.01^{**}	1.15^{*}	0.14	0.81^{**}	$0.61^{\dagger}$
C. Individual predictors, covariate estimation									
Dividend-price ratio	1.56	0.02	-1.19	1.56	0.02	-1.19	0.73	0.47	-1.19
Dividend yield	1.28	-0.76	-1.52	1.28	-0.76	-1.52	0.69	0.25	-1.52
Earnings-price ratio	4.75	-1.13	-0.56	4.75	-1.13	-0.56	6.74	-1.14	-0.56
Dividend payout ratio	79.93	-0.75	-0.19	76.92	-0.70	-0.14	79.93	-0.75	-0.19
Stock variance	90.80	-0.20	-0.17	89.93	-0.20	-0.17	90.80	-0.20	-0.17
Book to market	18.02	-2.51	-2.25	18.02	-2.51	-2.25	18.40	-1.80	-2.25
Net equity expansion	5.23	0.38	0.12	5.23	0.38	0.12	5.23	0.38	0.12
T-bill rate	0.67	0.30	0.93	0.67	0.30	0.93	1.43	0.42	0.93
Long-term yield	0.91	-0.01	0.55	0.91	-0.01	0.55	1.13	0.29	0.55
Long-term return	0.19	1.02^{*}	$1.23^{\dagger}$	0.19	1.02^{*}	1.23	0.17	1.05^{*}	$1.23^{\dagger}$
Term spread	0.41	0.93^{*}	1.98^{*}	0.41	0.93^{*}	1.98^{*}	0.51	0.98^{*}	1.98^{*}
Default yield spread	11.47	0.16	0.15	12.80	0.15	0.15	11.47	0.16	0.15
Default return spread	34.47	0.02	-0.19	71.38	-0.01	0.00	34.47	0.02	-0.19
Inflation	97.06	-0.27	-0.13	11.82	0.04	0.04	97.06	-0.27	-0.13
D. Combining methods, covariate estimation									
Mean	0.02	1.20^{**}	1.28^{*}	0.02	1.22^{**}	1.34^{**}	0.05	0.97^{**}	$0.51^{\dagger}$
Median	0.00	0.89^{**}	0.57^{*}	0.00	0.87^{**}	$0.60^{\dagger}$	0.00	0.89^{**}	0.57^{*}
Trimmed mean	0.02	1.13^{**}	1.06^{*}	0.01	1.15^{**}	1.12^{**}	0.03	1.00^{**}	0.58^{*}
DMSFE, $\theta = 1.0$	0.02	1.20^{**}	1.29^{*}	0.02	1.22^{**}	1.34^{**}	0.05	0.97^{**}	$0.51^{\dagger}$
DMSFE, $\theta = 0.9$	0.02	1.21^{**}	1.41^{**}	0.01	1.23^{**}	1.46^{**}	0.04	0.98^{**}	0.58^{*}
$C(2,PB)$	1.42	0.79^{**}	1.15^{*}	1.12	0.79^{**}	1.32^{**}	0.36	0.83^{**}	1.17^{**}
$C(3,PB)$	1.37	0.75^{**}	1.16^{*}	1.55	0.75^{**}	1.05^{**}	0.13	1.09^{**}	1.41^{**}
ABMA, SIC	0.02	1.20^{**}	1.28^{*}	0.02	1.22^{**}	1.34^{**}	0.05	0.97^{**}	$0.51^{\dagger}$
ABMA, AIC	0.02	1.20^{**}	1.28^{*}	0.02	1.22^{**}	1.34^{**}	0.05	0.97^{**}	$0.51^{\dagger}$

Table 2: Equity premium forecast results, 1965:01–2005:12 out-of-sample evaluation period

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Unrestricted coefficients			Restricted slope coefficient			Restricted forecast sign		
Forecast model	*MSPE-adj.* p-value (%)	R^2_{OS} (%)	Utility gain (annual %)	*MSPE-adj.* p-value (%)	R^2_{OS} (%)	Utility gain (annual %)	*MSPE-adj.* p-value (%)	R^2_{OS} (%)	Utility gain (annual %)
A. Individual predictors, OLS estimation									
Dividend-price ratio	4.74	0.00	-0.34	4.74	0.00	-0.34	2.34	0.60	-0.34
Dividend yield	4.20	-0.48	0.04	4.20	-0.48	0.04	1.91	0.59	0.04
Earnings-price ratio	16.31	-0.77	0.17	16.31	-0.77	0.17	23.05	-0.84	0.17
Dividend payout ratio	99.98	-0.83	-0.40	99.92	-0.77	-0.32	99.98	-0.83	-0.40
Stock variance	70.21	-0.46	-0.17	70.21	-0.46	-0.17	70.21	-0.46	-0.17
Book to market	34.25	-3.48	-2.11	34.25	-3.48	-2.11	35.22	-2.54	-2.11
Net equity expansion	15.91	-0.29	0.13	15.91	-0.29	0.13	15.91	-0.29	0.13
T-bill rate	4.60	0.21	1.96†	4.60	0.21	1.96†	4.41	0.55	1.96†
Long-term yield	7.65	-0.70	1.81†	7.65	-0.70	1.81†	2.85	0.62	1.81†
Long-term return	14.55	0.02	0.91	11.10	0.17	0.91	13.22	0.14	0.91
Term spread	5.65	0.41	2.30*	5.65	0.41	2.30*	5.86	0.45	2.30*
Default yield spread	1.56	0.69	0.79	1.56	0.69†	0.79	1.56	0.69	0.79
Default return spread	63.75	-0.12	0.11	45.41	-0.01	0.22	63.75	-0.12	0.11
Inflation	18.30	0.26	0.77	18.30	0.26	0.77	14.74	0.32	0.77
B. Combining methods, OLS estimation									
Mean	0.29	0.99**	1.57**	0.25	1.01**	1.61**	0.41	0.82**	0.81*
Median	0.06	0.90**	0.96*	0.05	0.92**	0.98**	0.06	0.90**	0.96*
Trimmed mean	0.28	0.96**	1.38**	0.24	0.98**	1.42**	0.29	0.86**	0.95*
DMSFE, $\theta = 1.0$	0.30	0.98**	1.59**	0.26	1.00**	1.63**	0.39	0.82**	0.83*
DMSFE, $\theta = 0.9$	0.30	1.00**	1.75**	0.27	1.02**	1.79**	0.33	0.83**	0.89*
$C(2,PB)$	5.74	0.60**	1.07*	6.22	0.60**	1.07**	1.14	0.80**	1.31*
$C(3,PB)$	9.06	0.39**	0.94*	9.51	0.39*	0.94*	2.31	0.73**	1.18*
ABMA, SIC	0.29	0.99**	1.57**	0.26	1.01**	1.61**	0.41	0.82**	0.81*
ABMA, AIC	0.29	0.99**	1.57**	0.26	1.01**	1.61**	0.41	0.82**	0.81*
C. Individual predictors, covariate estimation									
Dividend-price ratio	4.67	0.01	-0.33	4.67	0.01	-0.33	2.33	0.60	-0.33
Dividend yield	4.26	-0.51	0.04	4.26	-0.51	0.04	1.91	0.58	0.04
Earnings-price ratio	16.31	-0.77	0.18	16.31	-0.77	0.18	23.05	-0.84	0.18
Dividend payout ratio	99.97	-0.82	-0.41	99.91	-0.77	-0.34	99.97	-0.82	-0.41
Stock variance	91.63	-0.26	-0.22	90.54	-0.25	-0.22	91.63	-0.26	-0.22
Book to market	34.25	-3.48	-2.11	34.25	-3.48	-2.11	35.23	-2.54	-2.11
Net equity expansion	11.13	0.20	0.22	11.13	0.20	0.22	11.13	0.20	0.22
T-bill rate	1.40	0.38	2.01*	1.40	0.38	2.01*	2.84	0.61	2.01*
Long-term yield	2.63	0.36	1.74†	2.63	0.36	1.74	2.97	0.58	1.74†
Long-term return	0.63	1.07*	1.76†	0.63	1.07*	1.76†	0.54	1.11*	1.76†
Term spread	0.51	1.24**	3.60**	0.51	1.24**	3.60**	0.61	1.30**	3.60**
Default yield spread	3.86	0.31	0.46	3.87	0.31	0.46	3.86	0.31	0.46
Default return spread	54.27	-0.07	-0.31	98.00	-0.02	-0.02	54.27	-0.07	-0.31
Inflation	97.19	-0.18	-0.10	5.96	0.07	0.07	97.19	-0.18	-0.10
D. Combining methods, covariate estimation									
Mean	0.14	1.15**	1.78**	0.12	1.17**	1.85**	0.35	0.87**	0.72†
Median	0.06	0.80**	0.67	0.03	0.82**	0.68	0.06	0.80**	0.67
Trimmed mean	0.15	1.05**	1.43*	0.13	1.07**	1.51*	0.21	0.91**	0.80†
DMSFE, $\theta = 1.0$	0.13	1.15**	1.81**	0.11	1.17**	1.88**	0.33	0.88**	0.74†
DMSFE, $\theta = 0.9$	0.12	1.16**	1.97**	0.10	1.19**	2.03**	0.26	0.89**	0.82†
$C(2,PB)$	1.47	0.98**	1.72**	1.23	0.98**	1.87**	0.64	0.93**	1.72**
$C(3,PB)$	2.55	0.74**	1.61**	3.24	0.74*	1.46**	0.53	1.09**	1.98**
ABMA, SIC	0.14	1.15**	1.78**	0.12	1.17**	1.85**	0.35	0.87**	0.72†
ABMA, AIC	0.14	1.15**	1.78**	0.12	1.17**	1.85**	0.35	0.87**	0.72†

Table 3: Equity premium forecast results, 1976:01–2005:12 out-of-sample evaluation period

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Unrestricted coefficients			Restricted slope coefficient			Restricted forecast sign		
Forecast model	*MSPE-adj.* p-value (%)	R^2_{OS} (%)	Utility gain (annual %)	*MSPE-adj.* p-value (%)	R^2_{OS} (%)	Utility gain (annual %)	*MSPE-adj.* p-value (%)	R^2_{OS} (%)	Utility gain (annual %)
A. Individual predictors, OLS estimation									
Dividend-price ratio	30.60	-1.61	-2.63	30.60	-1.61	-2.63	27.69	-0.78	-2.63
Dividend yield	31.92	-2.56	-3.07	31.92	-2.56	-3.07	31.37	-1.02	-3.07
Earnings-price ratio	27.41	-1.42	-0.19	27.41	-1.42	-0.19	39.38	-1.52	-0.19
Dividend payout ratio	89.66	-0.21	-0.34	74.19	-0.13	-0.23	89.66	-0.21	-0.34
Stock variance	79.43	-0.85	-0.15	79.43	-0.85	-0.15	79.43	-0.85	-0.15
Book to market	50.38	-4.94	-3.66	50.38	-4.94	-3.66	52.41	-3.62	-3.66
Net equity expansion	11.88	-0.05	-0.03	11.88	-0.05	-0.03	11.88	-0.05	-0.03
T-bill rate	26.22	-1.40	-0.68	26.22	-1.40	-0.68	43.30	-0.74	-0.68
Long-term yield	35.91	-1.85	-0.65	35.91	-1.85	-0.65	36.14	-0.53	-0.65
Long-term return	18.33	-0.12	0.42	18.33	-0.12	0.42	16.74	0.04	0.42
Term spread	22.68	-0.43	1.30	22.68	-0.43	1.30	24.35	-0.38	1.30
Default yield spread	23.15	0.14	0.15	23.15	0.14	0.15	23.15	0.14	0.15
Default return spread	35.25	0.0 6	0.28	33.39	0.07	0.32	35.25	0.06	0.28
Inflation	55.42	-0.30	-0.34	55.42	-0.30	-0.34	55.07	-0.29	-0.34
B. Combining methods, OLS estimation									
Mean	13.65	0.35^*	1.08^*	13.35	$0.36^\dagger$	1.09^*	17.15	$0.26^\dagger$	$0.65^\dagger$
Median	6.80	0.46^*	$0.75^\dagger$	6.42	0.47^*	$0.78^\dagger$	6.80	0.46^*	$0.75^\dagger$
Trimmed mean	12.85	0.37^*	$1.02^\dagger$	12.50	$0.37^\dagger$	$1.03^\dagger$	14.03	$0.31^\dagger$	$0.75^\dagger$
DMSFE, $\theta = 1.0$	13.77	0.35^*	1.07^*	13.46	$0.35^\dagger$	$1.08^\dagger$	16.98	$0.26^\dagger$	$0.66^\dagger$
DMSFE, $\theta = 0.9$	14.70	0.32^*	$1.01^\dagger$	14.36	$0.33^\dagger$	$1.02^\dagger$	18.43	$0.23^\dagger$	$0.59^\dagger$
$C(2,PB)$	58.26	-0.46	-1.32	61.61	-0.46	-1.34	34.42	-0.03	-0.35
$C(3,PB)$	53.00	-0.63	-1.58	54.79	-0.63	-1.59	24.24	0.04	-0.39
ABMA, SIC	13.66	0.35^*	1.08^*	13.36	$0.36^\dagger$	1.09^*	17.16	$0.26^\dagger$	$0.65^\dagger$
ABMA, AIC	13.67	0.35^*	1.08^*	13.37	$0.36^\dagger$	1.09^*	17.17	$0.26^\dagger$	$0.65^\dagger$
C. Individual predictors, covariate estimation									
Dividend-price ratio	30.59	-1.59	-2.62	30.59	-1.59	-2.62	27.66	-0.77	-2.62
Dividend yield	31.94	-2.60	-3.07	31.94	-2.60	-3.07	31.42	-1.03	-3.07
Earnings-price ratio	27.40	-1.42	-0.19	27.40	-1.42	-0.19	39.37	-1.52	-0.19
Dividend payout ratio	88.24	-0.21	-0.35	73.50	-0.13	-0.25	88.24	-0.21	-0.35
Stock variance	87.94	-0.31	-0.29	84.70	-0.27	-0.29	87.94	-0.31	-0.29
Book to market	50.38	-4.94	-3.66	50.38	-4.94	-3.66	52.42	-3.62	-3.66
Net equity expansion	6.65	0.61	0.11	6.65	0.61	0.11	6.65	0.61	0.11
T-bill rate	24.38	-1.48	-0.64	24.38	-1.48	-0.64	42.67	-0.66	-0.64
Long-term yield	40.28	-1.00	-0.70	40.28	-1.00	-0.70	41.41	-0.59	-0.70
Long-term return	9.01	0.54	1.00	9.01	0.54	1.00	8.08	0.59	1.00
Term spread	7.92	-0.09	1.29	7.92	-0.09	1.29	12.40	0.06	1.29
Default yield spread	20.75	0.15	0.18	20.75	0.15	0.18	20.75	0.15	0.18
Default return spread	97.05	-0.21	-0.45	61.18	0.00	-0.01	97.05	-0.21	-0.45
Inflation	53.25	-0.02	0.02	27.12	0.03	0.06	53.25	-0.02	0.02
D. Combining methods, covariate estimation									
Mean	11.08	$0.42^\dagger$	$1.12^\dagger$	10.09	$0.45^\dagger$	1.17	16.56	$0.27^\dagger$	0.60
Median	8.30	$0.31^\dagger$	0.49	4.83	0.37	0.57	8.30	$0.31^\dagger$	0.49
Trimmed mean	11.80	$0.38^\dagger$	0.96	10.48	$0.41^\dagger$	1.03	13.97	$0.31^\dagger$	0.59
DMSFE, $\theta = 1.0$	10.95	$0.42^\dagger$	$1.12^\dagger$	9.96	$0.44^\dagger$	1.18	16.22	$0.27^\dagger$	0.59
DMSFE, $\theta = 0.9$	11.55	$0.40^\dagger$	$1.04^\dagger$	10.42	$0.43^\dagger$	1.10	17.29	0.25	0.53
$C(2,PB)$	37.00	-0.02	-0.70	32.21	-0.02	-0.53	23.93	0.12	-0.06
$C(3,PB)$	32.52	-0.24	-0.81	37.26	-0.24	-0.93	15.67	0.25	0.05
ABMA, SIC	11.10	$0.42^\dagger$	$1.12^\dagger$	10.10	$0.45^\dagger$	1.17	16.58	$0.27^\dagger$	0.60
ABMA, AIC	11.10	$0.42^\dagger$	$1.12^\dagger$	10.11	$0.45^\dagger$	1.17	16.59	$0.27^\dagger$	0.60

Table 4: Equity premium forecast results, 2000:01–2005:12 out-of-sample evaluation period

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Unrestricted coefficients			Restricted slope coefficient			Restricted forecast sign		
Forecast model	*MSPE-adj.* p-value (%)	R^2_{OS} (%)	Utility gain (annual %)	*MSPE-adj.* p-value (%)	R^2_{OS} (%)	Utility gain (annual %)	*MSPE-adj.* p-value (%)	R^2_{OS} (%)	Utility gain (annual %)
A. Individual predictors, OLS estimation									
Dividend-price ratio	3.26	4.87^{*}	$6.12^{\dagger}$	3.26	4.87^{*}	$6.12^{\dagger}$	3.99	3.86^{*}	6.12
Dividend yield	3.18	5.18^{**}	$6.12^{\dagger}$	3.18	5.18^{*}	$6.12^{\dagger}$	4.00	3.86^{*}	6.12
Earnings-price ratio	2.63	5.94^{**}	7.19^{*}	2.63	5.94^{*}	$7.19^{\dagger}$	3.64	4.32^{*}	7.19
Dividend payout ratio	94.99	-0.78	-1.29	94.99	-0.78	-1.29	94.99	-0.78	-1.29
Stock variance	58.32	-0.07	-0.35	58.32	-0.07	-0.35	58.32	-0.07	-0.35
Book to market	3.79	4.55^{*}	$6.29^{\dagger}$	3.79	4.55^{*}	$6.29^{\dagger}$	3.94	3.94^{*}	6.29
Net equity expansion	87.98	-3.08	-1.49	87.98	-3.08	-1.49	87.98	-3.08	-1.49
T-bill rate	36.14	0.15	1.22	36.14	0.15	1.22	36.14	0.15	1.22
Long-term yield	1.20	0.44	0.89	1.20	0.44	0.89	1.20	0.44	0.89
Long-term return	69.35	-1.30	-2.08	69.35	-1.30	-2.08	67.69	-1.17	-2.08
Term spread	64.27	-1.67	0.02	64.27	-1.67	0.02	64.27	-1.67	0.02
Default yield spread	13.87	0.70	1.12	13.87	0.70	1.12	13.87	0.70	1.12
Default return spread	75.45	-0.67	-0.70	75.45	-0.67	-0.70	75.45	-0.67	-0.70
Inflation	89.57	-1.23	-1.62	89.57	-1.23	-1.62	89.57	-1.23	-1.62
B. Combining methods, OLS estimation									
Mean	3.05	1.99^{**}	3.29^{*}	3.05	1.99^{*}	$3.29^{\dagger}$	4.28	1.29^{*}	$2.14^{\dagger}$
Median	10.34	0.56	1.11	10.34	0.56	$1.11^{\dagger}$	10.34	0.56	1.11
Trimmed mean	2.67	1.94^{**}	3.21^{*}	2.67	1.94^{*}	$3.21^{\dagger}$	3.39	1.46^{*}	2.42^{*}
DMSFE, $\theta = 1.0$	3.36	1.82^{**}	2.99^{*}	3.36	1.82^{*}	$2.99^{\dagger}$	4.64	1.21^{*}	$2.01^{\dagger}$
DMSFE, $\theta = 0.9$	3.48	2.04^{**}	3.40^{*}	3.48	2.04^{*}	$3.40^{\dagger}$	4.59	1.33^{*}	$2.22^{\dagger}$
$C(2,PB)$	64.26	-1.04	-2.23	64.26	-1.04	-2.23	47.81	-0.28	-1.06
$C(3,PB)$	54.70	-1.33	-3.01	54.70	-1.33	-3.01	47.80	-0.66	-1.80
ABMA, SIC	3.05	1.99^{**}	3.29^{*}	3.05	1.99^{*}	$3.29^{\dagger}$	4.29	1.29^{*}	$2.14^{\dagger}$
ABMA, AIC	3.05	1.99^{**}	3.29^{*}	3.05	1.99^{*}	$3.29^{\dagger}$	4.29	1.29^{*}	$2.14^{\dagger}$
C. Individual predictors, covariate estimation									
Dividend-price ratio	3.28	4.85^{**}	6.13^{*}	3.28	4.85^{**}	6.13	3.99	3.86^{**}	6.13
Dividend yield	3.16	5.20^{**}	6.12^{*}	3.16	5.20^{**}	6.12	4.00	3.86^{**}	6.12
Earnings-price ratio	2.63	5.94^{**}	7.19^{**}	2.63	5.94^{**}	$7.19^{\dagger}$	3.64	4.32^{**}	7.19
Dividend payout ratio	94.86	-0.82	-1.34	94.86	-0.82	-1.34	94.86	-0.82	-1.34
Stock variance	48.51	-0.01	0.31	95.31	-0.59	-0.87	48.51	-0.01	0.31
Book to market	3.79	4.55^{**}	6.29^{*}	3.79	4.55^{**}	$6.29^{\dagger}$	3.94	3.94^{**}	6.29
Net equity expansion	88.17	-1.19	-1.86	88.17	-1.19	-1.86	88.17	-1.19	-1.86
T-bill rate	36.14	0.15	0.95	36.14	0.15	0.95	36.14	0.15	0.95
Long-term yield	1.20	0.46	0.94	1.20	0.46	0.94	1.20	0.46	0.94
Long-term return	66.45	-0.23	-0.46	66.45	-0.23	-0.46	66.45	-0.23	-0.46
Term spread	62.71	-0.18	0.00	62.71	-0.18	0.00	62.71	-0.18	0.00
Default yield spread	29.05	0.07	0.10	29.05	0.07	0.10	29.05	0.07	0.10
Default return spread	44.53	0.02	0.03	96.01	-0.02	-0.03	44.53	0.02	0.03
Inflation	12.92	0.21	0.28	5.07	0.27	0.40	12.92	0.21	0.28
D. Combining methods, covariate estimation									
Mean	2.76	2.21^{**}	3.60^{*}	2.74	2.18^{**}	$3.53^{\dagger}$	3.42	1.52^{*}	2.46^{*}
Median	4.82	0.27	0.56	3.54	0.24	0.47	4.82	0.27	0.56
Trimmed mean	2.43	1.97^{**}	3.23^{*}	2.39	1.93^{**}	$3.15^{\dagger}$	2.85	1.50^{*}	2.44^{*}
DMSFE, $\theta = 1.0$	2.99	2.05^{**}	3.32^{*}	2.97	2.01^{**}	$3.25^{\dagger}$	3.63	1.45^{*}	2.34^{*}
DMSFE, $\theta = 0.9$	3.17	2.24^{**}	3.68^{*}	3.15	2.21^{**}	$3.61^{\dagger}$	3.73	1.55^{*}	2.51^{*}
$C(2,PB)$	54.41	-0.76	-1.66	57.05	-0.76	-1.86	42.70	-0.10	-0.78
$C(3,PB)$	50.53	-1.24	-2.57	55.69	-1.24	-2.82	41.55	-0.30	-1.27
ABMA, SIC	2.76	2.21^{**}	3.60^{*}	2.74	2.18^{**}	$3.53^{\dagger}$	3.42	1.52^{*}	2.46^{*}
ABMA, AIC	2.76	2.21^{**}	3.60^{*}	2.74	2.18^{**}	$3.53^{\dagger}$	3.42	1.52^{*}	2.46^{*}

Notes to Tables 5 and 6: The tables report data-mining bootstrap critical values for the R^2_{OS} and utility gain statistics reported in Tables 1–4. The critical values are generated using the data-mining bootstrap procedure described in the text.

Table 5: Data-mining bootstrap critical values, OLS estimation

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
		Individual predictors				Combining methods			
Statistic	Significance level	1947:01–2005:12	1965:01–2005:12	1976:01–2005:12	2000:01–2005:12	1947:01–2005:12	1965:01–2005:12	1976:01–2005:12	2000:01–2005:12
A. Unrestricted coefficients									
R^2_{OS}	10%	0.49	0.71	0.85	2.31	0.20	0.21	0.24	0.67
	5%	0.62	0.90	0.98	2.96	0.27	0.27	0.32	0.88
	1%	0.89	1.11	1.76	5.00	0.41	0.41	0.67	1.82
Utility gain	10%	1.72	1.64	2.23	5.30	0.69	0.58	0.79	1.87
	5%	2.02	2.26	2.64	6.29	0.86	0.71	1.06	2.34
	1%	2.71	2.88	4.09	12.00	1.41	1.17	1.78	5.43
B. Restricted slope coefficient									
R^2_{OS}	10%	0.44	0.62	0.86	2.46	0.17	0.19	0.30	0.85
	5%	0.59	0.72	0.97	3.50	0.25	0.27	0.41	1.41
	1%	0.91	0.98	1.54	5.96	0.41	0.41	0.64	2.83
Utility gain	10%	1.58	1.52	2.25	5.95	0.65	0.61	0.87	2.01
	5%	1.92	1.93	2.60	8.06	0.80	0.77	1.08	3.54
	1%	2.72	2.74	4.09	15.42	1.17	0.96	1.91	8.59
C. Restricted forecast sign									
R^2_{OS}	10%	0.55	0.72	0.86	2.31	0.26	0.24	0.23	0.68
	5%	0.69	0.89	0.99	2.85	0.37	0.35	0.33	0.86
	1%	1.02	1.37	1.76	5.00	0.48	0.64	0.67	1.82
Utility gain	10%	1.72	1.64	2.23	5.30	0.61	0.60	0.75	1.87
	5%	2.02	2.26	2.64	6.29	0.75	0.75	0.95	2.36
	1%	2.71	2.88	4.09	12.00	1.28	1.37	1.60	5.43

Table 6: Data-mining bootstrap critical values, covariate estimation

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
		Individual predictors				Combining methods			
Statistic	Significance level	1947:01–2005:12	1965:01–2005:12	1976:01–2005:12	2000:01–2005:12	1947:01–2005:12	1965:01–2005:12	1976:01–2005:12	2000:01–2005:12
A. Unrestricted coefficients									
R^2_{OS}	10%	0.46	0.63	0.77	1.78	0.15	0.24	0.29	0.69
	5%	0.64	0.77	1.00	2.15	0.22	0.37	0.54	0.82
	1%	1.19	1.24	1.85	3.05	0.33	0.69	0.68	1.58
Utility gain	10%	1.23	1.67	2.05	3.94	0.49	0.69	1.00	1.49
	5%	1.57	1.97	2.50	4.81	0.56	0.98	1.21	1.82
	1%	2.24	2.65	2.96	6.66	1.41	1.57	1.63	4.28
B. Restricted slope coefficient									
R^2_{OS}	10%	0.48	0.62	0.77	2.55	0.17	0.31	0.39	1.02
	5%	0.62	0.77	0.95	3.04	0.20	0.41	0.52	1.31
	1%	1.03	1.24	1.85	3.83	0.57	0.80	0.74	1.83
Utility gain	10%	1.35	1.75	2.46	6.15	0.59	1.00	1.19	2.58
	5%	1.57	1.86	3.03	7.24	0.85	1.21	1.56	3.80
	1%	2.24	2.30	4.14	11.26	1.05	1.53	1.73	4.94
C. Restricted forecast sign									
R^2_{OS}	10%	0.56	0.63	0.77	1.78	0.17	0.26	0.26	0.69
	5%	0.74	0.77	0.95	2.15	0.21	0.37	0.58	0.86
	1%	1.28	1.24	1.88	3.05	0.41	0.69	0.75	1.58
Utility gain	10%	1.23	1.67	2.05	3.94	0.49	0.70	1.01	1.31
	5%	1.57	1.97	2.50	4.81	0.56	0.92	1.28	1.82
	1%	2.24	2.65	2.96	6.66	1.06	1.56	1.91	3.80

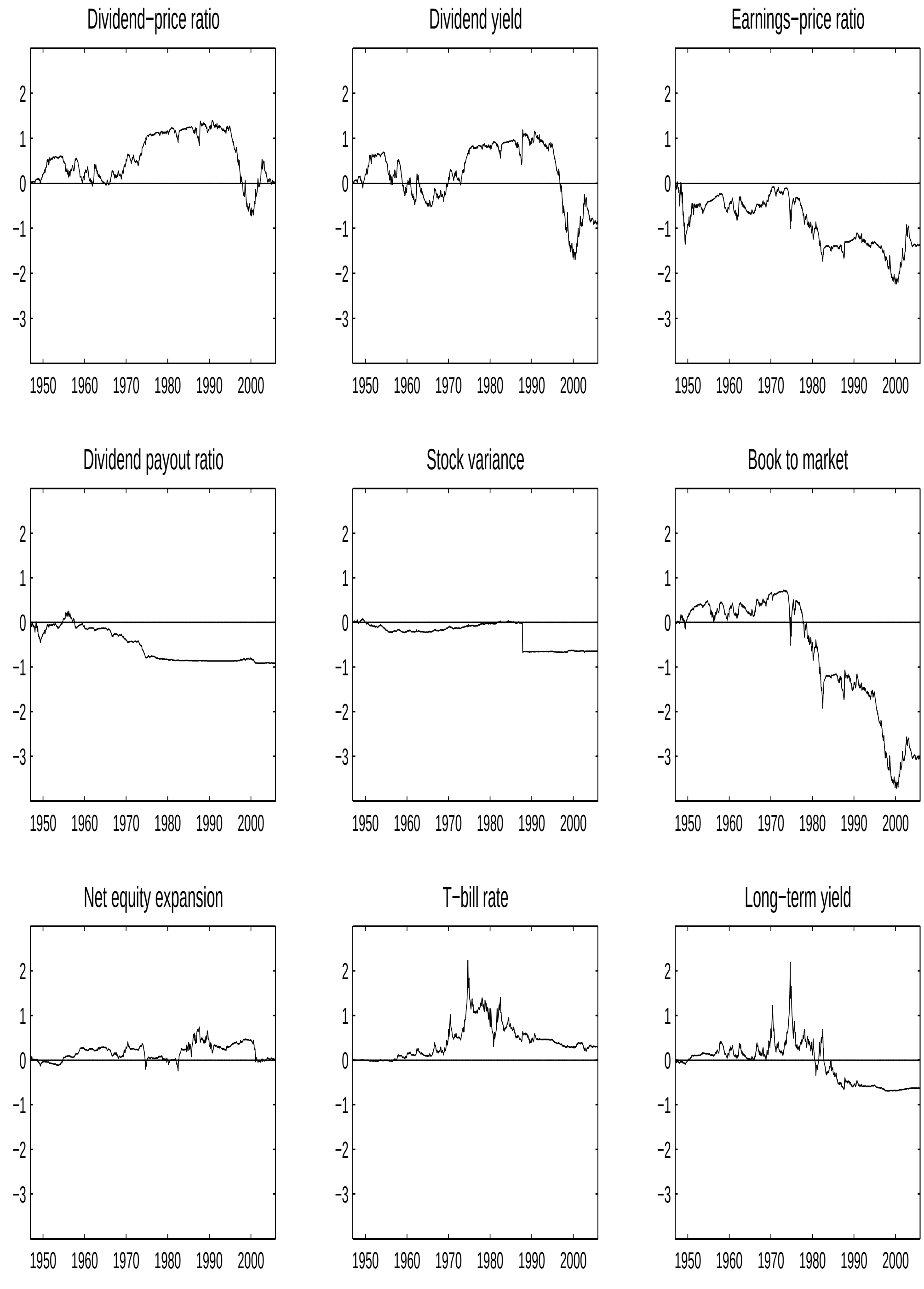

Figure 1: Historical average benchmark forecast model cumulative square prediction error minus individual predictive regression forecast model cumulative square prediction error (times 100), 1947:01–2005:12

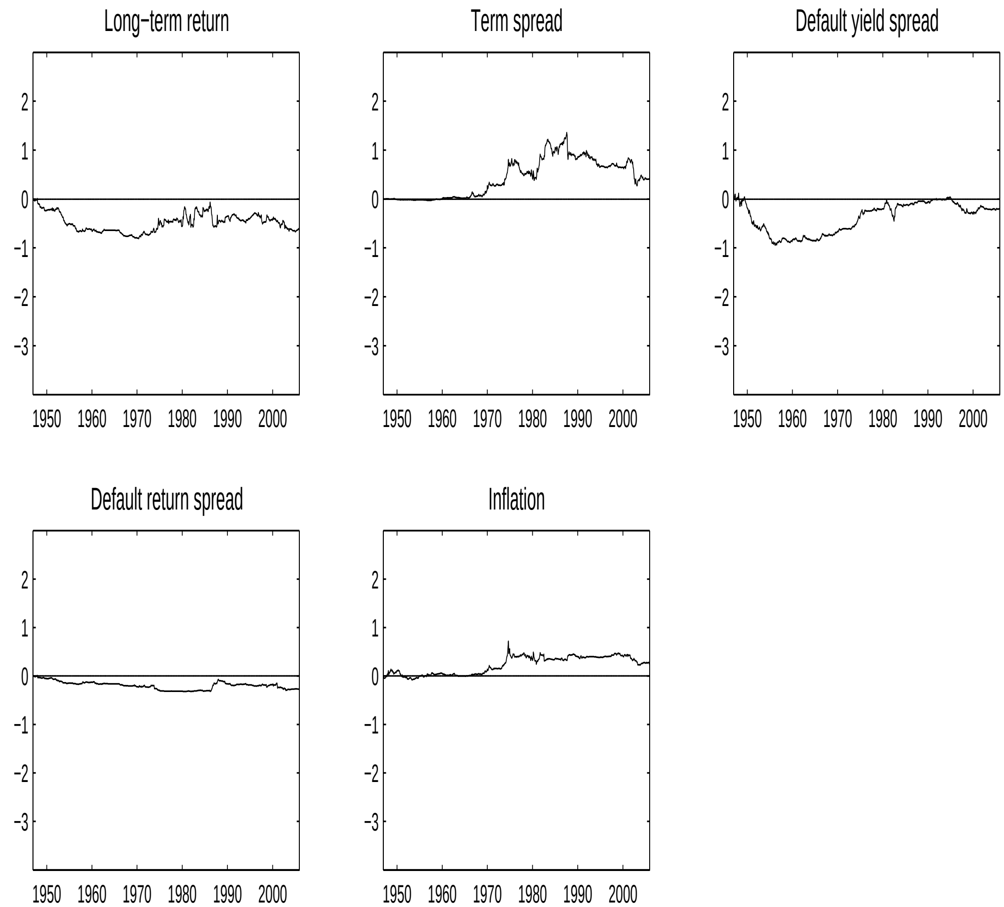

Figure 1 (continued)

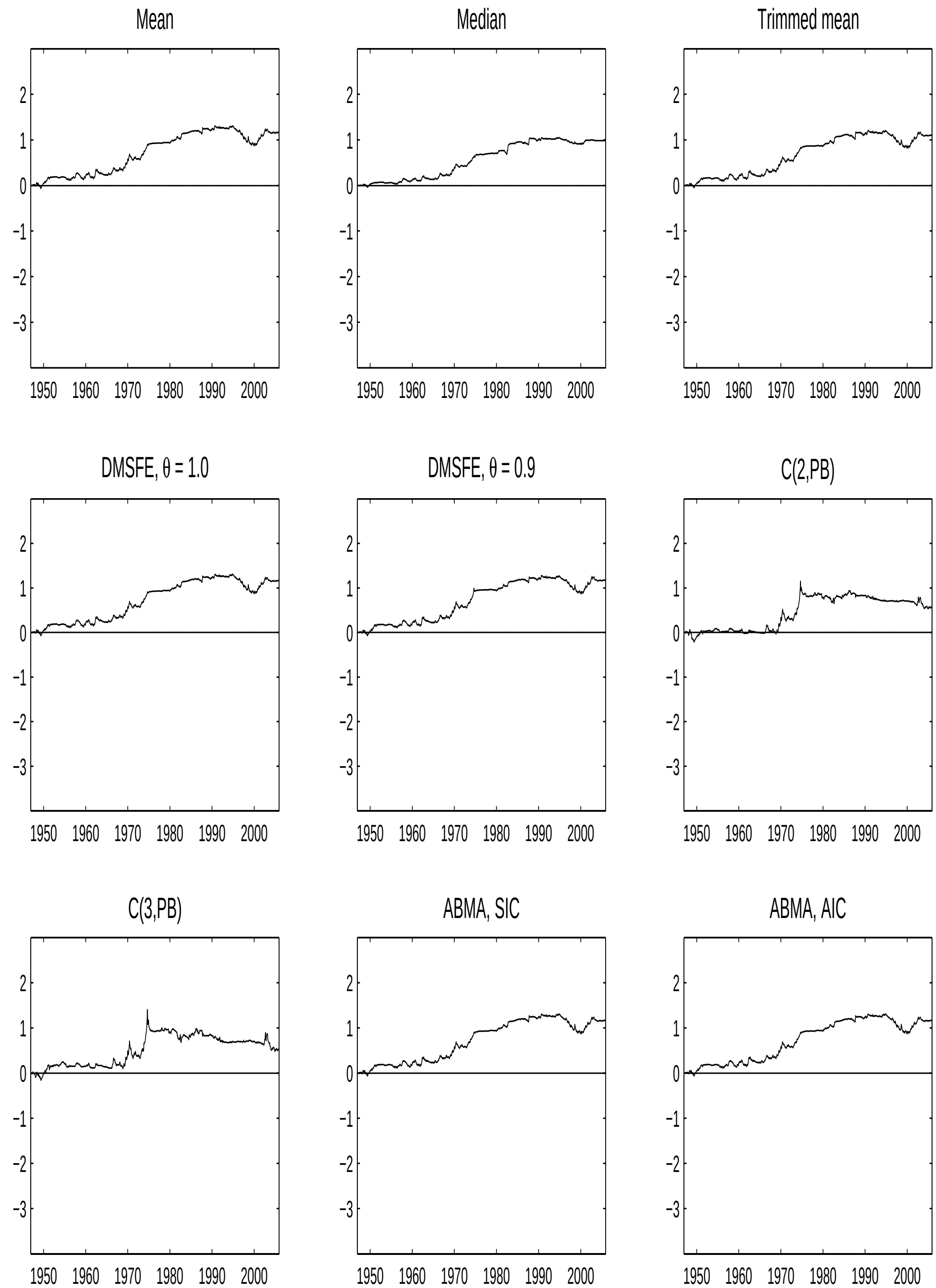

Figure 2: Historical average benchmark forecast model cumulative square prediction error minus combination forecast model square prediction error (times 100), 1947:01–2005:12