Understanding and Forecasting Stock Price Changes^{*}

by

Pedro N. Rodríguez^{}**

Simon Sosvilla-Rivero^{*}**

DOCUMENTO DE TRABAJO 2006-03

January 2006

^{*} The authors would like to thank Amit Goyal, Ivo Welch, and Kenneth R. French for kindly providing the data set. Pedro N. Rodriguez thanks CONACY (Mexico) for financial support (Fellowship: 170328). All errors are solely ours.

^{**} Universidad Complutense de Madrid.

^{***} FEDEA and Universidad Complutense de Madrid.

Los Documentos de Trabajo se distribuyen gratuitamente a las Universidades e Instituciones de Investigación que lo solicitan. No obstante están disponibles en texto completo a través de Internet: <http://www.fedea.es>.

These Working Paper are distributed free of charge to University Department and other Research Centres. They are also available through Internet: <http://www.fedea.es>.

Abstract

Previous empirical studies have shown that predictive regressions in which model uncertainty is assessed and propagated generate desirable properties when predicting out-of-sample. However, it is still not clear (*a*) what the important conditioning variables for predicting stock returns out-of-sample are, and (*b*) how composite weighted ensembles outperform model selection criteria. By comparing the unconditional accuracy of prediction regressions to the conditional accuracy (conditioned on specific explanatory variables masked), we find that cross-sectional premium and term spread are robust predictors of future stock returns. Additionally, using the bias-variance decomposition for the 0/1 loss function, the analysis shows that lower bias, and not lower variance, is the fundamental difference between composite weighted ensembles and model selection criteria. This difference, nevertheless, does not necessarily imply that model averaging techniques improve our ability to describe monthly up-and-down movements' behavior in stock markets.

JEL Classification Numbers: G11, G15, C11

Keywords: Stock return predictability, Model averaging, Bias-variance decomposition

1. Introduction

Since 1970 a great deal of research has been devoted to examining the usefulness of publicly available information for predicting future stock returns [see Fama (1991) for a survey]. The goal of predictive regressions is to find a useful approximation $\hat{f}(X)$ to the function $f(X)$ that underlies the predictive relationship between some conditioning variables and future stock returns. To obtain and provide adequate and interpretable descriptions of how the explanatory variables affect future stock returns, researchers usually employ simple linear regression models [see, e.g., Chen *et al.* (1986), Campbell (1987), Campbell and Shiller (1988), Fama and French (1988), Kothari and Shanken (1997), Pontiff and Schall (1998), Baker and Wurgler (2000), and Rangvid (2005), among many others].

The capability of linear regression models to extract predictable components in stock returns, however, is heavily debated. Since when the most prominent variables proposed in the literature for predicting stock returns are evaluated singly or in an all-inclusive linear model, models tend to provide unsatisfactory out-of-sample results [see, e.g., Goyal and Welch (2006) and references therein].

However, Avramov (2002) and Cremers (2002) find that predictive regressions that subsume model uncertainty improve our ability to ‘describe’ the time-series behavior of security returns. The term ‘describe’ is emphasized, since one loses, with composite weighted ensembles of models, the simple and interpretable structure of linear regressions. What one gains, nevertheless, is increased accuracy.

This paper has two goals. The first is to provide a simple picture of how composite weighted ensembles, such as Bayesian model averaging, outperform model selection criteria. The second is to examine the relative contribution of each explanatory variable for predicting security returns out-of-sample.

The bias-variance decomposition is an important tool for understanding function approximating algorithms. In this paper, we use the notions of bias and variance to explain how composite weighted ensembles outperform model selection criteria. Since 0/1 loss function is usually the main criterion for classification problems, we use the bias-variance decomposition for the 0/1 loss function to gauge direction-of-change predictability (i. e., the ability of a learner to distinguish up from down movements) [see, e.g., Domingos (2000) and Valentini and Dietterich (2004)]. We find that the variance (i. e., the loss incurred by function’s fluctuations around the central tendency in response to different samples) on both

composite weighted ensembles and model selection criteria has an inversely related effect on error. In other words, the extent to which the function deviates from the incorrect predictions (unbiased variance) is higher than the extent to which the function deviates from the correct predictions (biased variance). However, composite weighted ensembles' biases (systematic loss incurred by functions) are generally lower to that of model selection criteria and correspond to the fundamental difference in ensembles and model selection criteria. This difference, nevertheless, does not necessarily imply that model averaging techniques improve our ability to describe monthly up-and-down movements' behavior.

In predictive regressions the explanatory variables are seldom equally relevant. Often only a few of them have small influence on the future stock returns. Even though Avramov (2002) and Cremers (2002) find that predictability is feasible if model uncertainty is assessed and propagated, they gauge the importance of the potential variables in-sample relevance. Therefore, it is not clear which variable helps to predict stock returns out-of-sample.

In this paper, we assess the relative influence of the most prominent variables for predicting stock returns by comparing the unconditional accuracy of prediction regressions to the conditional accuracy (conditioned on specific explanatory variables masked or omitted). We show that some variables have an important contribution on return predictability, while others can be considered as noise since they do not contain any useful information for predicting stock price returns. In particular, masking the term spread, or the cross-sectional premium, composite weighted ensembles obtain lower out-of-sample predictability accuracy. In contrast, variables such as dividend yield or book-to-market ratio are irrelevant for predicting future stock returns and could just as well have not been included, since their omission increases the out-of-sample predictive accuracy.

Nevertheless, it is a widely acknowledged fact that conditioning variables lose their predictive power after their discovery (see, e. g., Schwert, 2003). Consequently, we also examine the relative contribution of each conditioning variable before and after their discovery. Consistent with Schwert (2003) and Avramov and Chordia (2006), we find that some conditioning variables attenuate their predictive power after their discovery. However, our results suggest that the term spread remains a robust predictor of future stock returns after its discovery, while the book-to-market ratio, even before its discovery, does decrease our ability to understand future stock returns.

The remainder of the paper proceeds as follows. Section 2 explains, in the context of the existing sample evidence on return predictability, how composite weighted ensembles outperform model selection criteria. Section 3 examines the relative importance of the most prominent variables proposed in the literature for predicting stock returns. Section 4 concludes.

2. Another look at the sample evidence on return predictability: How composite weighted ensembles outperform model selection criteria?

Avramov (2002) and Cremers (2002) find that Bayesian model averaging's out-of-sample performance is superior to that of model selection criteria. This finding, however, prompts an important (and unanswered) question: How ensembles obtain higher predictive accuracy than model selection criteria?

Consider monthly returns on the value-weighted CRSP index over the sample period 1953:04 through 2002:12 using the following $p = 11$ conditioning variables (taking one lag):

1. Dividend yield on the S&P 500 index (d/y).
2. Size Premium (SMB).
3. Value Premium (HML).
4. Earnings price ratio on the S&P 500 index (e/p).
5. Stock variance of the S&P 500 index (svar).
6. Cross-sectional premium (csp).
7. Book-to-market ratio (b/m).
8. Net equity expansion of NYSE stocks (ntis).
9. Term spread, defined as the difference between the long term yield on government bond and the 3-month T-bill (tms).
10. Default yield spread, defined as the difference between the BAA- and AAA-rated corporate yields (dfy).
11. Default return spread, defined as the difference between on long-term corporate bonds and the returns on long-term government bonds (dfr).

The data set was kindly provided by Amit Goyal, Ivo Welch, and Kenneth R. French.

As Avramov (2002), we perform a fixed-size rolling windows analysis, in which model parameters are first estimated with data from 1 to T (our T corresponds to 180 observations), next with data from 2 to $T + 1, \dots$, and finally with data from $N - T$ to $T - 1$. At each iteration, one forecasts one-step ahead.

Table 1 reports several statistics examining the properties of out-of-sample monthly forecasts generated by several models and composite weighted ensembles.

Table 1: Out-of-sample results

The table displays several statistics examining the properties of out-of-sample forecast errors generated by models selected by statistical criteria and by composite weighted ensembles of models. The former set includes the i.i.d model and six models selected by adjusted R-squared (r2a), AIC, AICc, SIC, FIC, PIC, and gMDL. We examine three composite weighted ensembles: Ave, Median, and BMA. Ave represents the collection of all 2^p models (where p denotes the number of explanatory variables in the study) in which each model is equally-weighted. Med forecasts the median of all 2^p models. BMA stands for Bayesian Model Averaging. BMA computes posterior probabilities for the collection of all 2^p models. The posterior probability for each model was obtained via the BIC approximation. The forecasts of each model or composite weighted ensembles were evaluated with several regression-based test of prediction accuracy, such as MPE, Efficiency, and Serial correlation, all of which are described by Avramov (2002). Additionally, direction-of-change predictability was assessed via the 0/1 loss function.

	Ave	Med	BMA	AIC	AICc	SIC	PIC	i.i.d	R2a	gMDL
<i>The rolling scheme- monthly sample</i>										
MPE	0.0017	0.0013	0.0018	0.0060	0.0056	0.0048	0.0060	-0.0020	-0.0010	0.0052
<i>t-statistic</i>	0.7850	0.5730	0.7910	2.5790	2.4330	2.0890	2.5620	-0.9070	-0.4290	2.2770
Serial correlation	-0.0276	-0.0319	-0.0277	-0.0068	-0.0064	0.0106	0.0019	0.0510	0.0517	0.0045
<i>t-statistic</i>	-0.5633	-0.6510	-0.5653	-0.1388	-0.1306	0.2163	0.0388	1.0408	1.0551	0.0918
Efficiency	-0.3113	-0.2544	-0.3122	-0.5740	-0.5911	-0.5327	-0.6540	-6.6695	-1.3005	-0.5473
<i>t-statistic</i>	-1.4740	-1.1510	-1.4810	-5.4210	-5.6850	-5.1380	-6.5470	-2.7080	-4.1340	-5.3170
MSE(%)	0.2150	0.2141	0.2151	0.2294	0.2309	0.2245	0.2387	0.2206	0.2277	0.2264
0-1 Loss	0.4254	0.4302	0.4278	0.4519	0.4495	0.4230	0.4615	0.4375	0.4519	0.4278

Following Avramov (2002), we make use of three regression-based tests of predictive accuracy. Namely, forecasts errors' mean equal to zero, zero correlation between forecasts errors and predictive returns (Efficiency), and of zero first-order serial correlation. As Boothe and Glassman (1987) observe, a further test is the accuracy in predicting the direction of change, since getting the sign right in the prediction matters in markets with low transaction costs, like stock markets. Therefore, we also use a directional-based accuracy measure: the 0/1 loss. This loss function evaluates the ability to discriminate up from down movements, and corresponds to number of observations misclassified divided by the total number of observations.

We use ten forecasting models: First, we consider five models selected by adjusted R-squared, AIC, SIC, FIC, and PIC, all of which are described by Bossaerts and Hillion (1999). Second, we use models selected by the corrected AIC (denoted by AICc), and the minimum description length criteria (denoted by gMDL). Third, we examine the i.i.d model predicting the then-prevailing mean in stock returns. Finally, we generate three composite weighted ensembles by considering all linear data-generating processes in the presence of 11 conditioning variables (2^{11} models). In particular, the model denoted by Ave (Med) forecasts the average (median) of the 2^{11} models, whereas the model denoted by BMA computes posterior probabilities for the collection of all 2^{11} models. The posterior probability for each model was obtained via the BIC approximation (see Raftery, 1995).

The results in Table 1 indicate that model averaging techniques tend to outperform model selection criteria in terms of regression-based tests of predictive accuracy. Indeed, the prediction errors have zero mean and are essentially uncorrelated. In addition, the prediction errors are uncorrelated with predicted returns. However, in terms of classification error, or direction-of-change predictability, the results shown in Table 1 indicate that the composite weighted ensembles do not outperform model selection criteria. However, we do not know whether model averaging techniques still exhibit a different predictive structure.

In the machine learning literature, the bias-variance decomposition is widely used as key tool for understating function approximation algorithms. Although the bias-variance decomposition was originally proposed for the square loss (see, e. g., Geman *et al.*, 1992), this paper uses the 0/1 loss function for one main reason: level accuracy is not as strongly correlated with profits with a trading strategy based on a set of predictions as directional accuracy [see, e.g., Leitch and Tanner (1991) and Pesaran and Timmermann (1995)].

Following Domingos (2000) and Valentini and Dietterich (2004), bias and variance can be defined in terms of two quantities: the optimal prediction and the main prediction. The optimal prediction $y_*(\mathbf{x})$ is equal to the movement that is observed more often in the test sample, where $\mathbf{x}$ is a fixed point in the explanatory variable space. The main prediction $y_m(\mathbf{x})$ can be defined as the movement that is predicted more often in the test sample. Thus, the bias (systematic loss incurred by the function) can be computed as,

$$B(\mathbf{x}) = \begin{cases} 1 & \text{if } y_m \neq t \\ 0 & \text{if } y_m = t \end{cases}, \quad (1)$$

where t equals to up-movement if the observed stock return is higher than zero, down-movement otherwise.

To distinguish between the two different effects of the variance on the loss function, Domingos defines the unbiased variance, $V_u(\mathbf{x})$, to be the variance when $B(\mathbf{x}) = 0$, and can be calculated as,

$$V_u(\mathbf{x}) = \|(y_m = t) \text{ and } (y_m \neq \hat{y})\|, \quad (2)$$

where $\|s\| = 1$ if s is true, 0 otherwise. $\hat{y}$ denotes the predicted class, which equals to up-movement if the predicted return is higher than zero, down-movement otherwise. The unbiased variance evaluates the extent to which the estimated function deviates from the correct predictions. The biased variance, $V_b(\mathbf{x})$, occurs when $B(\mathbf{x}) = 1$, and evaluates the extent to which the estimated function deviates from the incorrect predictions. The biased variance can be estimated as,

$$V_b(\mathbf{x}) = \|(y_m \neq t) \text{ and } (y_m \neq \hat{y})\|. \quad (3)$$

To obtain the loss associated with a given explanatory variable $\mathbf{x}$ [denoted by $E_D(\mathbf{x})$], we simply compute the algebraic sum of bias, unbiased and biased variance as,

$$E_D(\mathbf{x}) = B(\mathbf{x}) + V_u(\mathbf{x}) - V_b(\mathbf{x}). \quad (4)$$

In order to compute the aforementioned variables in a test set, we simply obtain the average for each variable. Clearly, if we want a good function that distinguishes between up-and-down movements, we want the bias and the unbiased variance to be small. The results for the fixed-size rolling windows scheme are presented in Table 2.

Table 2 shows that the three of the five lowest 0/1 loss correspond to models that assess and propagate model uncertainty. Moreover, the biases associated with the composite weighted ensembles are generally lower than that of model selection criteria. It is worth noting that the i.i.d model and the model selected by the adjusted R-squared criteria can also be considered as low-bias learners. However, the variance in both models has a positive effect on the error, in clear contrast to the composite weighted ensembles, in which the variance has a negative effect on error (i.e., the unbiased variance is higher than the biased variance). Interestingly, the

analysis reveals that the bias, and not variance, plays a significant role in its contribution to the error rate.¹ Thus, a promising direction of future research is to consider different approaches in the function approximation techniques. In this sense, iterative bagging (Breiman, 2001) could be an interesting alternative, since it is a data-intensive methodology focusing in reducing bias. Such alternative may yield important improvements in the ability of functions to describe the time-series behavior of stock returns.

Table 2: Bias-variance decomposition

The table displays several statistics examining the properties of out-of-sample forecast errors generated by models selected by statistical criteria and by composite weighted ensembles of models. The former set includes the i.i.d model and six models selected by adjusted R-squared (r2a), AIC, AICc, SIC, FIC, PIC, and gMDL. We examine three composite weighted ensembles: Ave, Median, and BMA. Ave represents the collection of all 2^p models (where p denotes the number of explanatory variables in the study) in which each model is equally-weighted. Med forecasts the median of all 2^p models. BMA stands for Bayesian Model Averaging. BMA computes posterior probabilities for the collection of all 2^p models. The posterior probability for each model was obtained via the BIC approximation. The forecasts of each model or composite weighted ensembles were evaluated with 0/1 loss function, which evaluates the usefulness of the estimated model to distinguish up from down movements. 0/1 Loss, Bias, Net Variance, Unbiased Variance, and Biased Variance are all described by Domingos (2000) and Valentini and Dietterich (2004).

	0-1 Loss	Bias	Net Variance	Unbiased Variance	Biased Variance
Ave	0.4254	0.4375	-0.0120	0.1995	0.2115
Med	0.4302	0.4375	-0.0072	0.1875	0.1947
BMA	0.4278	0.4375	-0.0096	0.2019	0.2115
AIC	0.4519	0.5625	-0.1105	0.1947	0.3052
AICc	0.4495	0.5625	-0.1129	0.1899	0.3028
SIC	0.4230	0.5625	-0.1394	0.1875	0.3269
PIC	0.4615	0.5625	-0.1009	0.1995	0.3004
i.i.d	0.4399	0.4375	0.0024	0.0048	0.0024
R2a	0.4519	0.4375	0.0144	0.1033	0.0889
gMDL	0.4278	0.5625	-0.1346	0.1923	0.3269

3. Contribution of conditioning variables for predicting stock returns out-of-sample

In this section, we perform the relative contribution analysis of the eleven explanatory variables. The contribution analysis is based on the comparison between the original out-of-sample results (Table 1) and several reruns in which one explanatory is masked (or omitted) from the input space. The results for the (mean) squared loss function are show in Table 3.

¹ We have also performed bias-variance decomposition for the squared loss, and find no additional insights (i.e., the analysis produces virtually the same results).

Table 3: Conditioning variables importance using the (mean) squared loss function

Table 3A reports the Mean Square Error (MSE) generated by several models and composite weighted ensembles when the variable in the first column was omitted from the analysis. Table 3B provides the variation in the MSE when the variable in the first column was omitted from the analysis. The last column (Rel) shows an overall relevance measure, as described in Equation (5).

A) MSE

	Ave	Med	BMA	AIC	AICc	SIC	PIC	R2a	gMDL
SMB	0.2141	0.2135	0.2141	0.2297	0.2313	0.2246	0.2357	0.2328	0.2265
HML	0.2143	0.2136	0.2143	0.2291	0.2307	0.2246	0.2356	0.2284	0.2264
d/y	0.2150	0.2142	0.2150	0.2292	0.2256	0.2271	0.2362	0.2281	0.2246
e/p	0.2147	0.2150	0.2147	0.2279	0.2291	0.2253	0.2364	0.2293	0.2275
svar	0.2135	0.2128	0.2135	0.2291	0.2305	0.2246	0.2373	0.2219	0.2265
csp	0.2203	0.2199	0.2203	0.2342	0.2342	0.2271	0.2369	0.2275	0.2321
b/m	0.2124	0.2120	0.2124	0.2205	0.2224	0.2170	0.2263	0.2309	0.2185
ntis	0.2177	0.2160	0.2177	0.2303	0.2313	0.2292	0.2427	0.2278	0.2292
tms	0.2211	0.2230	0.2211	0.2336	0.2334	0.2277	0.2417	0.2274	0.2335
dfy	0.2150	0.2138	0.2150	0.2310	0.2318	0.2239	0.2344	0.2275	0.2294
dfr	0.2163	0.2148	0.2163	0.2291	0.2288	0.2228	0.2402	0.2273	0.2257

B) ΔMSE

	Ave	Med	BMA	AIC	AICc	SIC	PIC	R2a	gMDL	Rel
SMB	-0.4471	-0.2955	-0.4434	0.0847	0.1455	-0.0194	-1.2741	2.1744	-0.0189	-0.0891
HML	-0.3934	-0.2114	-0.3904	-0.1741	-0.1210	-0.0194	-1.3134	0.2500	-0.0583	-1.9387
d/y	-0.0335	0.0362	-0.0320	-0.1167	-2.3476	1.1003	-1.0633	0.1168	-0.8449	-2.5347
e/p	-0.1939	0.4066	-0.1937	-0.7032	-0.8222	0.3011	-0.9794	0.6400	0.4577	-0.8580
svar	-0.7629	-0.6193	-0.7569	-0.1703	-0.1974	-0.0067	-0.5860	-2.5811	-0.0171	-4.5691
csp	2.4054	2.7103	2.4053	2.0550	1.3726	1.1374	-0.7332	-0.1247	2.4694	10.9333
b/m	-1.2501	-0.9875	-1.2519	-3.9038	-3.7384	-3.3768	-5.1774	1.3718	-3.5164	-17.5302
ntis	1.2087	0.8724	1.2134	0.3389	0.1433	2.0580	1.6699	0.0074	1.1720	6.9186
tms	2.7909	4.1730	2.7932	1.7771	1.0588	1.3807	1.2743	-0.1559	3.1024	14.4927
dfy	-0.0516	-0.1615	-0.0475	0.6605	0.3343	-0.3068	-1.8134	-0.1382	1.2755	-0.1968
dfr	0.5694	0.3173	0.5735	-0.1801	-0.9572	-0.8154	0.6438	-0.1992	-0.3547	-0.3161

Table 3A provides the (mean) squared loss function when the explanatory variable reported in the first column is omitted. It is worth noting that each rerun was made with ten explanatory variables. Table 3B provides the variation in the (mean) squared loss function. A negative variation indicate that the MSE decreased when the corresponding conditioning variables was omitted from the analysis. Since each variable contribute in a different way to each model, we compute an overall relevance measure for each explanatory variable as,

$$\text{Rel}_j = \sum_{i=1}^m \Delta_{ij} [\exp(-L_{ij})], \quad (5)$$

where m denote the number of models, Δ_{ij} corresponds to the variation in the loss function of the conditioning variable j when it was omitted from the i model, and L_{ij} indicate the loss function of the conditioning variable j when it was omitted from the i model.

Table 3B shows that the term spread and cross-sectional premium provide incremental information to predict future stock returns. In particular, their overall relevance measure is considerable higher than that of the rest of variables. Note that variables such as book-to-market ratio, stock variance, and dividend yield can be considered as noise: their omission actually increases the out-of-sample predictability accuracy.

Table 4: Conditioning variables importance using the 0/1 loss function

Table 4A documents the 0/1 loss function (classification error) generated by several models and composite weighted ensembles when the variable in the first column was omitted from the analysis. Table 4B provides the variation in the 0/1 loss when the variable in the first column was omitted from the analysis. The last column (Rel) shows an overall relevance measure, as described in Equation (5).

0/1 Loss									
	Ave	Med	BMA	AIC	AICc	SIC	PIC	R2a	gMDL
SMB	0.4279	0.4279	0.4279	0.4519	0.4495	0.4231	0.4423	0.4567	0.4279
HML	0.4231	0.4303	0.4231	0.4495	0.4495	0.4231	0.4543	0.4519	0.4279
d/y	0.4375	0.4279	0.4375	0.4183	0.4135	0.4207	0.4351	0.4447	0.4087
e/p	0.4351	0.4183	0.4351	0.4543	0.4519	0.4303	0.4519	0.4519	0.4351
svar	0.4423	0.4375	0.4375	0.4495	0.4495	0.4231	0.4615	0.4712	0.4279
csp	0.4567	0.4519	0.4567	0.4832	0.4760	0.4712	0.4808	0.4567	0.4663
b/m	0.4327	0.4327	0.4303	0.4447	0.4471	0.4087	0.4447	0.4567	0.4279
ntis	0.4375	0.4447	0.4375	0.4519	0.4543	0.4495	0.4639	0.4471	0.4543
tms	0.4471	0.4567	0.4471	0.4615	0.4567	0.4567	0.4856	0.4495	0.4567
dfy	0.4399	0.4327	0.4399	0.4519	0.4471	0.4207	0.4327	0.4495	0.4279
dfr	0.4255	0.4375	0.4255	0.4567	0.4447	0.4279	0.4639	0.4519	0.4375

$\Delta 0/1$ Loss										
	Ave	Med	BMA	AIC	AICc	SIC	PIC	R2a	gMDL	Rel
SMB	0.5649	-0.5586	-0.0001	0.0000	0.0001	0.0000	-4.1666	1.0638	-0.0001	-1.9994
HML	-0.5650	0.0001	-1.1237	-0.5319	0.0001	0.0000	-1.5624	0.0000	-0.0001	-2.4374
d/y	2.8248	-0.5586	2.2471	-7.4468	-8.0213	-0.5682	-5.7291	-1.5957	-4.4945	-15.3864
e/p	2.2598	-2.7932	1.6853	0.5319	0.5348	1.7045	-2.0832	0.0000	1.6853	2.2664
svar	3.9547	1.6761	2.2471	-0.5319	0.0001	0.0000	0.0001	4.2553	-0.0001	7.3914
csp	7.3446	5.0280	6.7415	6.9149	5.8824	11.3636	4.1668	1.0638	8.9887	36.0240
b/m	1.6949	0.5588	0.5617	-1.5957	-0.5347	-3.4091	-3.6457	1.0638	-0.0001	-3.4662
ntis	2.8248	3.3521	2.2471	0.0000	1.0696	6.2500	0.5209	-1.0638	6.1797	13.6601
tms	5.0847	6.1454	4.4943	2.1277	1.6043	7.9545	5.2084	-0.5319	6.7415	24.5482
dfy	3.3898	0.5588	2.8089	0.0000	-0.5347	-0.5682	-6.2499	-0.5319	-0.0001	-0.7540
dfr	-0.0001	1.6761	-0.5619	1.0638	-1.0695	1.1363	0.5209	0.0000	2.2471	3.2224

Tables 4A-4B provide a similar analysis using the 0/1 loss function. As can be seen in these tables, the results once again suggest that the term spread and the cross-section premium are robust predictors of monthly up-and-down movements. Furthermore, the contribution of variables such as dividend yield, value premium and book-to-market ratio in the out-of-sample prediction of stock returns are found to be irrelevant.

Nevertheless, the overall relevance measure can largely be influenced by pre- and post-discovery periods [see, e. g., Schwert (2003) and Avramov and Chordia (2006)]. To evaluate whether or not the predictive power of the conditioning variables disappear, reverse or attenuate, we compute the overall relevance measure for each variable in pre- and post-discovery samples. The results are shown in Table 5.

Table 5: Pre- and post-discovery analysis

The table displays the overall relevance measure as described in Equation (5) in the pre- and post-discovery periods. We use Fama and French (1992) for size and value premium, Fama and French (1988) for dividend yield, Campbell and Shiller (1988) for earnings price ratio, Baker and Wurgler for net equity expansion, Pontiff and Schall (1998) for book-to-market ratio, Campbell (1987) for term spread, and finally, Chen *et al.* (1987) for default yield and default return spread. Variables such as stock variance and cross-sectional premium were omitted from the analysis since the data ends before their “discovery”. The number inside square brackets “[]” indicate the number of variables that had higher relevance measures.

	<i>Squared Loss function</i>		<i>0/1 Loss function</i>	
	Pre-publication	Post-publication	Pre-publication	Post-publication
SMB	1.23 [4]	-3.57 [5]	-1.48 [7]	-3.88 [7]
HML	-1.85 [5]	-3.89 [6]	1.09 [5]	-11.59 [8]
d/y	5.19 [3]	-18.61 [9]	-0.17 [8]	-37.20 [10]
e/p	-3.06 [6]	2.55 [4]	6.01 [5]	-2.84 [6]
b/m	-25.46 [10]	-14.89 [8]	-3.20 [9]	-4.43 [7]
ntis	11.40 [2]	-16.75 [7]	14.68 [2]	0.51 [4]
tms	14.65 [0]	23.66 [0]	35.32 [0]	10.60 [2]
dfy	-3.71 [7]	3.94 [4]	7.88 [6]	-8.82 [8]
dfr	-2.24 [4]	1.62 [6]	-2.78 [9]	9.17 [4]

Consistent with the literature, we find that the predictive powers of some variables do disappear. In fact, variables such as dividend yield and earnings price ratio loose entirely their predictive power. Interestingly, the analysis reveals that the book-to-market ratio, a popular variable in the literature, has, in fact, been irrelevant for predicting stock returns. However, the term spread remains a robust predictor of stock returns even in its post-discovery period.²

² Note that the cross-sectional premium and stock variance were omitted from the analysis, given the fact that the sample ends before their discovery.

4. Concluding remarks

In this paper, we have examined (*a*) how composite weighted ensembles outperform model selection criteria, and (*b*) what the important conditioning variables for predicting stock returns out-of-sample are. We obtain the following general results. First, we show that the main difference between model selection criteria and composite weighted ensembles that propagate model uncertainty is lower bias, and not lower variance. However, the evidence indicates that in terms of direction-of-change predictability composite weighted ensembles are not superior to model selection criteria.

Second, the results suggest that the predictive power of term spread and cross-section premium is superior to that of other predictors. In particular, their omissions from the analysis considerably decrement the out-of-sample predictive accuracy. In contrast, predictors such as book-to-market ratio and dividend yield, for example, can be considered as noise or irrelevant, since masking their values considerably increases the predictive accuracy.

The information supplied by bias-variance analysis suggests two promising approaches for designing predictive regressions. One approach is to employ different low-bias learning algorithms. The other approach is to use out-of-bag estimates to gauge the generalization performance of all competing regression specifications. In doing so, the *raison d'être* is not only to obtain models that are the best predictors with respect to a set of ex-ante observable economic variables but also to extend the reach of data-intensive techniques in the context of prediction regressions. In view of the mildly encouraging results of the present study, some optimism about the benefits from implementing these approaches seems justified.

References

- Avramov, A., 2002, "Stock return predictability and model uncertainty," *Journal of Financial Economics*, 64, 423-458.
- Avramov, A., and T. Chordia, 2006, "Predicting stock returns," *Journal of Financial Economics*, forthcoming.
- Baker, M. and J. Wurgler, 2000, "The equity share in new issues and aggregate stock returns," *Journal of Finance*, 55, 2219-2257.
- Bauer, E., and R. Kohavi, 1999, "An empirical comparison of voting classification algorithms: Bagging, boosting and variants," *Machine Learning*, 36, 525-536.
- Boothe, P. and D. Glassman, 1987, "Comparing exchange rate forecasting models: Accuracy versus profitability," *International Journal of Forecasting*, 3, 65-79.
- Bossaerts, P. and P. Hillion, 1999, "Implementing statistical criteria to select return forecasting models: What do we learn?," *Review of Financial Studies*, 12, 405-428.
- Breiman, L., 2001, "Using iterative bagging to debias regressions," *Machine Learning*, 45, 261-277.
- Campbell, J. Y., 1987, "Stock returns and the term structure," *Journal of Financial Economics*, 18, 373-399.
- Campbell, J. Y. and R. J. Shiller, 1988, "Stock prices, earnings, and expected dividends," *Journal of Finance*, 43, 661-676.
- Campbell, J. Y. and S. B. Thompson, 2005, "Predicting the equity premium out-of-sample: Can anything beat the historical average?," Working Paper 11468, National Bureau of Economic Research, July, <http://www.nber.org/papers/w11468>.
- Chen, N.F., R. Ross, and S. Ross, 1986, "Economic forces and the stock market," *Journal of Business*, 59, 383-404.
- Cremers, K. J. M., 2002, "Stock return predictability: A Bayesian model selection perspective," *Review of Financial Studies*, 15, 1223-1249.
- Domingos, P., 2000, A unified bias-variance decomposition for zero-one and squared loss, in *Proceedings of the Seventeenth National Conference on Artificial Intelligence* (Austin, TX: AAAI Press) 564-569.
- Fama, E., 1991, "Efficient capital markets: II," *Journal of Finance*, 46, 1575-1617.
- Fama, E., and K. R. French, 1988, "Dividend yields and expected stock returns," *Journal of Financial Economics*, 22, 3-25.
- Fama, E., and K. R. French, 1992, "The cross-section of expected stock returns," *Journal of Finance*, 47, 427-465.
- Geman, S., E. Bienenstock, and R. Doursat, 1992, "Neural networks and the bias-variance dilemma," *Neural Computation*, 4, 1-58.

- Goyal, A., and I. Welch, 2006, “A comprehensive look at the empirical performance of equity premium predictions,” Working Paper 04-11, International Center for Finance, Yale School of Management, January, <http://ssrn.com/abstract=517667>.
- Kothari, S. P. and J. Shanken, 1997, “Book-to-market time series analysis,” *Journal of Financial Economics*, 44, 169-203.
- Leitch, G. and J. E. Tanner, 1991, “Economic forecast evaluation: Profits versus the conventional error measures,” *American Economic Review*, 81, 580-590.
- Pesaran, M. H., and A. Timmermann, 1995, “Predictability of stock returns: Robustness and economic significance,” *Journal of Finance*, 50, 1201-1228.
- Polk, C., S. Thompson, and T. Vuolteenaho, 2006, “Cross-section forecasts of the equity premium,” *Journal of Financial Economics*, forthcoming.
- Pontiff, J., and L. D. Schall, 1998, “Book-to-market ratios as predictors of market returns,” *Journal of Financial Economics*, 49, 141-160.
- Rangvid, J., 2005, “Output and expected returns,” *Journal of Financial Economics* (forthcoming).
- Raftery, A. E., 1995, “Bayesian model selection in social research (with Discussion),” *Sociological Methodology*, 25, 111-196.
- Schwert, G. M., 2003, Anomalies and market efficiency, in G. M. Constantinides, M. Harris, and R. Stulz (eds.), *Handbook of the Economics of Finance*, Vol. 1, Part. 2 Amsterdam: North-Holland, , 937-972.
- Valentini, G., and T. G. Dietterich, 2004, “Bias-variance analysis of support vector machines for the development of SVM-based ensemble methods,” *Journal of Machine Learning Research*, 5, 725-775.

RELACION DE DOCUMENTOS TRABAJO

- 2006-03: "Understanding and Forecasting Stock Price Changes", **Pedro N. Rodríguez y Simón Sosvilla-Rivero**.
- 2006-02: "A Macro and Microeconomic Integrated Approach to Assessing the Effects of Public Policies", **Xavier Labandeira, José M. Labeaga y Miguel Rodríguez**.
- 2006-01: "The dynamics of regional inequalities", **Salvador Barrios y Eric Strobl**.
- 2005-28: "New European Member States and the dependent elderly", **Corinne Mette**.
- 2005-27: "Efectos del Programa Operativo Integrado de Castilla-La Mancha, 2000-2006: Un análisis basado en el modelo Hermin", **Emma García, Simón Sosvilla-Rivero**.
- 2005-26: "It's a Small Small Welfare Cost of Fluctuations", **Franck Portier y Luis A. Puch**.
- 2005-25: "Obsolescence and Productivity", **Fernando del Rio y Antonio R. Sampayo**.
- 2005-24: "EU Structural Funds and Spain's Objective 1 Regions: An Analysis Based on the Hermin Model", **Simón Sosvilla-Rivero**.
- 2005-23: "A sequential model for older workers' labor transitions after a health shock", **Sergi Jiménez-Martín, José M. Labeaga y Cristina Vilaplana Prieto**.
- 2005-22: "Price Convergence in the European Car Market", **Salvador Gil-Pareja y Simón Sosvilla-Rivero**.
- 2005-21: "Implicit regimes for the Spanish Peseta/Deutschmark exchange rate", **Francisco Ledesma-Rodríguez, Manuel Navarro-Ibáñez, Jorge Pérez-Rodríguez y Simón Sosvilla-Rivero**.
- 2005-20: "A Projection of Spanish Pension System under Demographic Uncertainty", **Namkee Ahn, Javier Alonso-Meseguer y Juan Ramón García**.
- 2005-19: "The Dynamic of temporary jobs: Theory and Some Evidence for Spain (The Role of Skill)", **Elena Casquel y Antoni Cunyat**.
- 2005-18: "The Welfare Cost of Business Cycles in an Economy with Nonclearing Markets", **Franck Portier y Luis A. Puch**.
- 2005-17: "Life Satisfaction among Spanish Workers: Importance of Intangible Job Characteristics", **Namkee Ahn**.
- 2005-16: "Persistence and ability in the innovation decisions", **José M. Labeaga y Ester Martínez-Ros**.
- 2004-15: "Measuring Changes in Health Capital", **Néboa Zozaya, Juan Oliva y Rubén Osuna**.
- 2005-14: "Discrete choice models of labour Supply, behavioural microsimulation and the Spanish tax reforms", **José M. Labeaga, Xisco Oliver y Amedeo Spadaro**.
- 2005-13: "A Closer Look at the Comparative Statics in Competitive Markets", **J. R. Ruiz-Tamarit y Manuel Sánchez-Moreno**.
- 2005-12: "Wellbeing and dependency among European elderly: The role of social integration", **Corinne Mette**.
- 2005-11: "Demand for life annuities from married couples with a bequest motive", **Carlos Vidal-Meliá y Ana Lejárraga-García**.
- 2005-10: "Air Pollution and the Macroeconomy across European Countries", **Francisco Álvarez, Gustavo A. Marrero y Luis A. Puch**.
- 2005-09: "The excess burden associated to characteristics of the goods: application to housing demand", **Amelia Bilbao, Celia Bilbao y José M. Labeaga**.
- 2005-08: "La situación laboral de los inmigrantes en España: Un análisis descriptivo", **Ana Carolina Ortega Masagué**.
- 2005-07: "Demographic Uncertainty and Health Care Expenditure in Spain", **Namkee Ahn, Juan Ramón García y José A. Herce**.
- 2005-06: "EL NO-MAGREB. Implicaciones económicas para (y más allá de) la región", **José A. Herce y Simón Sosvilla Rivero**.