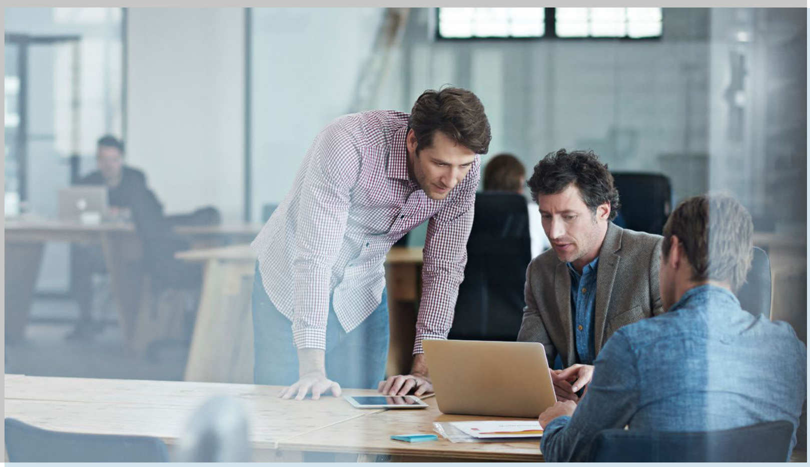
Swiss Finance Institute
Research Paper Series
N°19-15

The Keys of Predictability: a Comprehensive Study



Giovanni Barone-Adesi
Università della Svizzera italiana and Swiss Finance Institute

Antonietta Mira
Università della Svizzera italiana

Matteo Pisati
Università della Svizzera italiana

The Keys of Predictability: a Comprehensive Study. *

Giovanni Barone-Adesi[†]

Antonietta Mira [‡] Matteo M. Pisati[§]

March 20, 2019

Abstract

The problem of market predictability can be decomposed into two parts: predictive models and predictors. At first, we show how the joint employment of model selection and machine learning models can dramatically increase our capability to forecast the equity premium out-of-sample. Secondly, we introduce batteries of powerful predictors which brings the monthly S&P500 R^2_{OS} to a high level of 24%. Finally, we prove how predictability is a generalized characteristic of U.S. equity markets. For each of the three parts, we consider potential and challenges posed by the new approaches in the asset pricing field.

Keywords: Markets Predictability, Machine Learning, Model Selection.

[*] We are grateful to the Swiss National Foundation (SNF) project number: 100018 172892 Title: "Market Predictability and its Rationale: new insights in the Theoretical and Empirical Analysis of the Pricing Kernel"

[†]Universita' della Svizzera Italiana (USI) and Swiss Finance Institute, Institute of Finance, Via G. Buffi 13, CH-6900 Lugano, Switzerland. E-mail: giovanni.baroneadesi@usi.ch

[‡]Universita' della Svizzera Italiana (USI), Via G. Buffi 13, CH-6900 Lugano, Switzerland. E-mail: antonietta.mira@usi.ch

[§]Universita' della Svizzera Italiana (USI), Via G. Buffi 13, CH-6900 Lugano, Switzerland. E-mail: matteo.maria.pisati'@usi.ch

1 Introduction

The equity premium predictability literature typically introduces a new model or a new predictor and shows how it can rise the out-of-sample R^2 or Delta Utility. Differently, from the typical studies on the field, we acknowledge how stocks' pricing and predictability are intimately related[1]: being able to predict the market out-of-sample enrich our understanding of what ultimately the market prices. Consequently, we write this paper with the joint goal to provide both a comprehensive study of the out-of-sample predictability in equity markets and to trigger a fruitful discussion on the asset pricing implications of our findings. To gain a full understanding of the issues and potentials involved by a deeper understanding of financial market predictability we decompose the topic into two parts: predictive models and predictors. Each one provides new insight into our understanding of asset pricing and poses a variety of questions which aim at triggering a fruitful debate.

At first, we focus on predictive modeling. We re-examine the challenge posed by Welch and Goyal (2008) by employing the same predictors but combining model selection and machine learning predictive models. We observe how employing more and more powerful techniques our capability to accurately forecast out-of-sample increases steadily. Indeed, when model selection techniques are preventively adopted to alleviate multicollinearity, the results coming from the subsequent forecasts of machine learning techniques improve dramatically reaching monthly R^2_{OS} values above 5% for ensembles of Neural Networks. The remarkable results in terms of precision have substantial economic value for investors: delta Utility (against the traditional average mean return benchmark) rises by 2.5% with an even higher value of 4.5% during periods of recession when economic gains are more valuable. Our findings suggest that prices reflect inputs in a non-linear fashion. Consequently, the current research on the mathematical foundations of neural networks has a huge potential to widen our understanding of asset pricing[2]. Indeed, we stress the need for the identification of regime-dependent nonlinear pricing factors. Our results are also a direct challenge to the Efficient Market Hypothesis (Fama (1970)). This widely held assumption states that econometricians cannot systematically outperform the market using widely available information. In sharp contrast with this theory, it is becoming more and more evident how artificial intelligence is consistently overperforming the market. Even more surprisingly, with the progressing of technology, we observe steady

[1] A relevant exception comes from Campbell (1991) who first clearly states this relation.

[2] See, e.g., Shrikumar et al. (2016), Wei Koh and Liang (2017), Montavon et al. (2017), Montavon et al. (2018)

and, apparently, unbounded improvements. At this stage, a question naturally arises: how predictability originates? Indeed, it becomes apparent how some components of the pricing kernel are not fully reflected into prices, and this gives rise to the predictability phenomenon reported in our results. The identification of these predictable components and their dynamics is a major point in the financial literature agenda of the future. While the debate on the amount and the rationale of financial markets predictability is still in its infancy, on some points the consensus is broad:

- Equity premium predictability to some extent exists[3];
- It is linked to the the business cycle[4];
- It is linked to sentiment and liquidity[5].
- It is stronger in bear markets[6]
- It is time-varying and affected by financial research[7].
- it can be enhanced by imposing economically motivated constraints[8]

Our paper is also related to the data-science, and the machine learning approaches previously employed in the field of financial market predictability. Among the most remarkable machine learning approaches, we report the Kalman filter approach of Van Binsberg and Koijen (2010), the Markov Switching approach of Guidolin and Timmermann (2008), and the Bayesian system approach of Johannes et al. (2013). This last paper gave rise to a whole line of research which leverages the Bayesian statistic to make accurate financial forecasts. Among the most successful implementations in this area of study, we report the Bayesian latent threshold approach of Nakajima and West (2013), the dynamic dependent sparse factor model of Zhou et al. (2014), the dynamic dependence networks methodology of Yi et al. (2016), the simultaneous

[3]See, e.g., Dangl and Halling (2012), Rapach et al. (2010), Golez and Koudijs (2018)

[4]See, e.g., the seminal work of Fama and French (1989) and the recent works coming from Rapach et al. (2010) and Zhu (2015)

[5]Chen et al. (2018) show how to isolate a powerful liquidity predictor while Huang et al. (2015) propose a powerful sentiment one.

[6]Cujen and Hasler (2017) explain this phenomenon through the existence of a risk premium for uncertainty.

[7]Lo (2004) formulates a fascinating adaptive market hypothesis while Mclean and Pontiff (2015) proving how academic research reduce predictability implicitly confirm the hypothesis.

[8]Campbell and Thompson (2008) impose constraints on the regression coefficients and on the predicted returns (when the predicted returns are negative, they are replaced with zero) while Pettenuzzo et al. (2014) successfully introduces a constraint on the conditional Sharpe ratio.

graphical dynamic linear proposal of Gruber and West (2016), and the Bayesian predictive synthesis of Johnson and West (2018). After that, the papers most closely related to our one come from Gu et al. (2018) and Feng et al. (2018) who employ neural networks and machine learning techniques in the same framework. Differently, from their works, we combine model selection and machine learning techniques boosting the final predictive performance. Finally, in the ever-growing list of significant works on machine learning applied to financial forecasting we remark the stochastic neural network combination approach of Sermpinis et al. (2012), the adaptive evolutionary neural networks methodology of Georgios et al. (2015), the evolutionary support vector machines model of Karathanasopoulos et al. (2015), and the genetic programming approach of Karatahansopoulos et al. (2014)[9].

The second part of this paper focuses on predictors. At first, we consider as predictors for the $S\&P500$ the 12 different industries indexes. Accordingly, we perform an out-of-sample analysis of the study originally performed in-sample by Hong et al. (2007). We document big and rising monthly Delta Utility gains which, for the most recent 2001-2017 period, are well above 4% for the Money sector index and above 3% for the Chemical sector one. After that, we employ as predictors the returns coming from the long-short portfolio strategies commonly named in the literature factors (Fama and French (2015)) or anomalies (Stambaugh and Yuan (2017)). Here, for the 2001-2017 period, we document a record-high monthly R^2_{OS} of 28.6% for the Net Stock Issue matched by a 28% increase in terms of utility gains. Other return spreads like the Investment to Asset (Titman et al. (2003)), and the O-score (Ohlson (1980)) ones provide extremely powerful out-of-sample forecasts too.

Our results are related to the literature which proposes new predictors. Among them we report the Sentiment Index of Huang et al. (2015), the Trend Factor of Han et al. (2016), the short interest measure of Rapach et al. (2016), the Gold-Platinum ratio of Huang and Kilic (2018) and the aggregate Asset Growth indicator of Wen (2019). The studies more closely related to our one come from Hong et al. (2007) who employ industries and Greenwood and Hanson (2012) who employ the net issuance spread. For both cases, we extend their findings in an out-of-sample framework. More recently, even technical[10] and economic[11] indicators have been

[9]A comprehensive review of the existing literature on machine learning financial forecasting can be found in the works of Dunis et al. (2016) and de Prado (2018)

[10]Among the research on the value of technical analysis we report the seminal study of Lo et al. (2002) followed by the works of Neely et al. (2013) and Lin (2018)

[11]See, e.g.,Hong et al. (2007),Li and Tsiakas (2017) and Luo and Zhang (2017)

added to the list of powerful market predictors. Finally, powerful signals have been extrapolated from options. Martin and Wagner (2019) derive a formula for the expected return on a stock in terms of the risk-neutral variance of the market and the stock's excess risk-neutral variance relative to the average stock, Bakshi et al. (2011) build an option positioning that allows inferring forward variances from option portfolios, Bollerslev et al. (2015) build a fear measure from the left tail of the risk-neutral distribution, and Christoffersen and Pan (2017) show how oil option-implied information allows predicting stock market returns.

After having studied the predictive power of the listed predictors, we propose to employ an out-of-sample approach to identify the relevant pricing factors both for the S&P500 and for the French double-sorted portfolios. The identification of the most powerful predictors can shed new light on the drivers of the factorial profitability[12]. We observe how the predictive power of the different predictors is largely complementary through the spectrum of cross-sectional returns and while some stock are highly predictable by sentiment others are largely unaffected by it. This suggests that, contrary to the commonly held assumptions, a one-fits-all approach to the identification of the market pricing kernel could be misleading. Our simple approach is complementary to the blossoming literature on model selection in the asset pricing environment which, differently from our methodology, is entirely in-sample based. This line of literature has the goal to identify the relevant factors at the cross-sectional level. Among the newest approaches, we report the three pass method of Feng et al. (2017), the (Adaptive) Lasso methodology proposed by Messmer and Audrino (2017), the Tree-Based Conditional Portfolio Sorts of Moritz and Zimmermann (2016) and the deep learning methodology introduced by Feng et al. (2019). Other remarkable approaches to select a parsimonious amount of factors have been recently proposed by Fama and French (2018), Kozak et al. (2017b), and Stambaugh and Yuan (2017).

In the third part of the paper, we extend our analysis to include a broad sample of US stocks. We start by employing the French double-sorted portfolios: on the base of size and momentum, or size and the book-market ratio. We prove how predictability is not confined to small illiquid stocks only, but it is present even for the stocks of firms with big market capitalization. After that, we employ all the stocks available in the CRSP dataset to build 30 portfolios on the basis of a list of 10 characteristics (Stambaugh and Yuan (2017)). For each portfolio we individually

[12]This line of research stems from the seminal work of Fama and French (1993). Subsequently, a rich literature introduces a huge list of other anomalies (Campbell et al. (2016)). Among the most notorious works we list Frazzini and Pedersen (2014), Chan et al. (1996), Sloan (1996), and Novy-Marx (2013)

employ the Welch and Goyal (2008) and the anomalies spread returns predictors, the plus all the machine learning methodologies detailed before. We show how overall the Clark and West (2007) p-value for the R^2_{OS} out-of-sample statistics is less then 0.1 for the 20% of the cases considered and it is less then 0.05 for the 12% of the cases considered. In terms of economic value, the total average out-of-sample delta Utility is 4.84% with an average maximum delta utility for each portfolio of 9.96%. These results suggest that predictability is not only an attribute of the S&P500, but it is a generalized phenomenon of the US stock market.

The remaining of this paper is structured in the following way. Part ii) presents the data employed. Part iii) introduces the predictive modeling approaches employed and comments on the related empirical results inside the Welch and Goyal (2008) framework. Part iv) employs different sets of predictors and document their predictive performance. Part v) shows how predictability is a generalized feature of the US equity market. Part vi) concludes .

2 Data

In this section, we list all the data employed in our empirical analysis. We start from the Welch and Goyal predictors. Subsequently, we list data about industries and cross-sectional returns (anomalies and factors).

2.1 Welch and Goyal Predictors

The study of Welch and Goyal (2008) (W-G) is a benchmark and a challenge for the existing literature on market predictability. Consequently, we start with the fourteen predictors used in this provocative work[13]. The updated database is coming directly from the website of Goyal[14]. In more detail the predictors are:

- log Dividend-price ratio (DP): the 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.
- log Dividend yield (DY): the difference between the log of dividends and the log of lagged prices.

[13] Table A1 in the online appendix reports the correlation among the W-G predictors and the results for the autoregressive analysis of these predictors

[14] http://www.hec.unil.ch/agoyal/

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

In addition we employ the Sentiment Index of Huang et al. (2015). Data come directly from Zhou website[15].

2.2 Anomalies and Industries

In this section, we detail the factors and anomalies employed in this study. An anomaly is a statistically significant difference in cross-sectional average returns that persist after the adjustment for exposures to the Fama and French (1993) three factors model. Our empirical analysis makes use of i) the eleven anomalies proposed by Stambaugh et al. (2015), ii) the four

[15] http://apps.olin.wustl.edu/faculty/zhou/

factors of the extended Fama and French (2015) model, iii) the Long and Short term reversal anomalies. All data are monthly and span the period from 01-1965 to 12-2016 except the net operating assets, the accruals, the return on assets, and the distress anomaly for which data are available respectively only from 8-1965, 1-1970, 5-1976, and 1-1977. The considered factors-anomalies are:

- Financial distress. Campbell et al. (2008) show that firms with high failure probability have lower, not higher, subsequent returns (Distress). Another closely related measure of distress is the Ohlson (1980) O-score (O).
- Net stock issues and composite equity issues. Loughran and Ritter (1995) show that, in post-issue years, equity issuers under-perform non-issuers with similar characteristics (Net Stock Issues). Daniel and Titman (2006) propose an alternative measure, composite equity issuance (Comp eq Issue), defined as the amount of equity issued (or retired) by a firm in exchange for cash or services.
- Total accruals. Sloan (1996) demonstrates that firms with high accruals earn abnormal lower returns on average than firms with low accruals (Accruals).
- Net operating assets. Hirshleifer et al. (2004) find that net operating assets, computed as the difference between all operating assets and all operating liabilities divided by total assets is a negative predictor of long-run stock returns (NOA).
- Momentum. The momentum effect, proposed by Jegadeesh and Titman (1993) is one of the most widespread anomalies in asset pricing literature (Mom).
- Gross profitability premium. Novy-Marx (2013) shows that sorting on gross-profit-to-assets creates abnormal benchmark-adjusted returns, with more profitable firms having higher returns than less profitable ones (Gross Prof).
- Asset growth. Cooper et al. (2008) show how companies that grow their total assets more earn lower subsequent returns (Asset Growth).
- Return on assets. Chen et al. (2011) show that firms with higher past return on assets gain higher subsequent returns (ROA).
- Investment-to-assets. Titman et al. (2003) show that higher past investment predicts abnormally lower future returns (Inv to Assets).
- The four factors proposed by the extended model of Fama and French (2015): Small Minus Big (SMB), High Minus Low (HML), Robust Minus Weak (RMW), and Conservative

Minus Aggressive (CMA).

- The Short and Long Term reversal factors (ST, LT): as presented in the website of Professor Kenneth R. French.

Data for the four factors chosen by Fama and French (2015), the momentum, and the two short-long reversal factors comes from the website of Kenneth R. French[16] while anomalies are build matching CRSP and Compustat data following the approach detailed in Stambaugh and Yuan (2017). After that, we consider monthly data on the 12 industries indexes coming again from the website of Kenneth R. French. The time series span the period from January 1927 to December 2017. In detail, the indexes are: Consumer NonDurables, Consumer Durables, Manufacturing, Energy, Chemicals, Business Equipment, Telecommunications, Utilities, Shops, Healthcare, Finance and Others (Mines, constructions, Hotels, Entertainment, Business Services, Transportation).

3 Predictive Models

In this section, we first list the predictive models employed while we detail the methodology for each one of them in the appendix. After that, we present the performance metrics employed in our analysis. Finally, we report our empirical results, and we discuss them in light of the existing literature.

3.1 Econometric and Machine Learning Methodologies

To study the informative content which is possible to extrapolate from the predictors of Welch and Goyal (2008) we employ a wide list of models coming from the empirical financial literature and the Machine Learning one. While the list is far from being exhaustive, it is one of the first efforts to compare the predictive power of traditional econometric techniques with advanced machine learning ones in the field of empirical finance. Our approaches combines model selection with machine learning and statistical methodologies. The list of models considered includes:

1. Univariate OLS regressions for each predictor.

[16] $http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html$

2. A predictive OLS multivariate regression model (kitchen-sink) that incorporates all predictors jointly ("OLS" in the Tables).
3. A median combination forecasts approach which employ the median forecast among the ones generated by the univariate OLS regressions ("Pooled forecast: median", in the Tables).
4. The pooled DMSPE forecasts method proposed by Stock and Watson (2004) and successfully employed by Rapach et al. (2010) ("Pooled forecast: MDSFE" in the Tables).
5. Sum-of-the-parts forecast model of Ferreira and Santa-Clara (2011) (Sum-of-the-parts).
6. The Multivariate Adaptive Regression Splines approach Friedman (1991) for variable selection and a multivariate Support Vector Machine regression model (Boser et al. (1992) and Drucker et al. (1997)) to make out-of-sample forecasts ("MARS", in the Tables).
7. The SIC (Schwartz Information Criterion) for the variable selection and a multivariate Support Vector Machine regression model (Boser et al. (1992) and Drucker et al. (1997)) to make out-of-sample forecasts ("SVM SIC", in the Tables)
8. The Lasso for the variable selection and a multivariate Support Vector Machine regression model (Boser et al. (1992) and Drucker et al. (1997)) to make out-of-sample forecasts ("Lasso SVM" in the Tables).
9. The Regression Forest approach of Breiman (2001, 1996) built using regression trees (CART) Breiman and Friedman (1985) ("Random Forest", in the Tables).
10. The diffusion index approach employed by Ludvigson and Ng (2007) to filter the information and an univariate Support Vector Machine regression model (Boser et al. (1992) and Drucker et al. (1997)) to make out-of-sample forecasts ("Diffusion Index", in the Tables).
11. The Partial Least Squares approach of Kelly and Pruitt (2013) to filter the information and an univariate Support Vector Machine regression model (Boser et al. (1992) and Drucker et al. (1997)) to make out-of-sample forecasts ("PLS" in the Tables).
12. Variable selection made on the base of the MSFE performance of univariate OLS regressions. Out-of-sample forecasts generated though the Median, and the 40^{th} percentile of an ensemble of multi layer Neural Networks (Minsky and Papert (1969), Miller et al. (1995)) ("Neural Networks Median" and "Neural Networks 40^{th}" in the Tables).

3.2 Performance Metrics

To asses the out-of-sample predictive performance of the models and predictors considered in this study we follow the literature[17] and employ the R^2_{os}, and the Delta Utility metrics:

- The R^2_{os} statistic:
$$R^2_{os} = 1 - \frac{\sum_{t=1}^{T}(r_t - \hat{r}_t)^2}{\sum_{t=1}^{T}(r_t - \bar{r}_t)^2} \tag{1}$$
R^2_{os} measures the percent reduction in mean squared forecast error (MSFE) between the forecasts generated by the chosen predictive model, $\hat{r}$, and the historical average benchmark forecast, $\bar{r}$. To assess the statistical significance of R^2_{os} we employ the p-values coming from the Clark and West (2007) MSFE-adjusted statistic. This indicator tests the null hypothesis that the historical average MSFE is less than or equal to the forecasting method MSFE against the alternative that the historical average MSFE is greater than the forecasting method MSFE (corresponding to $H_0 : R^2_{os} <= 0$ against $H_1 : R^2_{os} > 0$).

- The Delta Utility measure. Following the literature (Campbell and Thompson (2008), Rapach et al. (2010)), we estimate the expected variance ($\hat{\sigma}^2_{t+1}$) using a ten-year rolling window of monthly returns. We consider a mean-variance investor who forecasts the equity premium using the historical averages. She will decide at the end of period t to allocate the following share of her portfolio to equity in the subsequent period t+1:
$$w_{0,t} = \frac{1}{\gamma}\frac{\bar{r}_{t+1}}{\hat{\sigma}_{t+1}} \tag{2}$$
where $\hat{\sigma}_{t+1}$ is the rolling-window estimate of the variance of stock returns. Over the out-of-sample period, she will obtain an average utility of:
$$\hat{\upsilon}_0 = \hat{\mu}_0 - \frac{1}{2}\gamma\hat{\sigma}^2_0 \tag{3}$$
where $\hat{\mu}_0$ and $\hat{\sigma}^2_0$ are the sample mean and variance, 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. Then we compute the average utility for the same investor when she forecasts the equity premium using one of the predictive approaches proposed

[17] Both these measures are introduced in the seminal work of Campbell and Thompson (2008) and subsequently employed in a number of studies among which Rapach et al. (2010), Strauss and Detzel (2017) and Rapach et al. (2016)

in this paper. In this case, the investor will choose an equity share of:

$$w_{j,t} = \frac{1}{\gamma} \frac{\hat{r}_{t+1}}{\hat{\sigma}_{t+1}} \tag{4}$$

and she will realize an average utility level of:

$$\hat{v}_j = \hat{\mu}_j - \frac{1}{2}\gamma\hat{\sigma}_j^2 \tag{5}$$

where $\hat{\mu}_j$ and $\hat{\sigma}_j$ are the sample mean and variance, over the out-of-sample period for the return on the portfolio formed using forecasts of the equity premium based on one of the methodologies proposed. In this paper, we measure the utility gain as the difference between $\hat{v}_j$ and $\hat{v}_0$, and we multiply this difference by 100 to express it in average annualized percentage return. In our analysis, following the existing literature,[18] we report results for $\gamma = 3$ and constraint the final weight for the risky asset in the range between -0.5 and 1.5.

3.3 Empirical Results and Discussion

Now we turn to the detailed results for the out-of-sample analysis, which are presented in Table 1 and Table 2. These tables report R^2_{OS} statistics and average utility gains for each of the individual predictive regression models and machine learning models relative to the historical average benchmark model. For the R^2_{OS} statistics statistical significance is assessed with the Clark and West (2007) MSPE-adjusted statistic. For brevity, the two tables included in the main text are the ones which report results for the longest out-of-sample period 1986:1-2017:12. For univariate OLS models, the restrictions imposed follows the approach of Campbell and Thompson (2008) while for all the machine learning approaches the only restriction is that if the forecasted return is negative, it is replaced with zero. The upper part of Table 1 based on the out-of-sample predictive performance of univariate linear regression largely confirms the results of Welch and Goyal (2008): out of 14 predictors, only 2 have a positive R^2_{OS}, and no one of them has a p-value under 0.1. Imposing the constraint of Campbell and Thompson (2008) results improve but again no positive R^2_{OS} statistics exhibit a p-value under 0.1. Our results diverge from the one originally presented by Campbell and Thompson (2008) confirming how the R^2_{OS}

[18] Among the most cited works on the subject Campbell and Thompson (2008) and Rapach et al. (2010) impose the same level of risk aversion

metric fluctuates dramatically changing the out-of-sample period. The average utility gains stemming from the same predictors confirms the previous conclusions: only the Earning Price ratio produces an increment of more than 1%, and this result is entirely due to its performance in Recession periods.

For predictive models, results are striking. While, as expected the R^2_{OS} value for the multivariate OLS model is negative, even the well-known Pooled forecast approach of Rapach et al. (2010) and the Sum-of-the parts methodology of Ferreira and Santa-Clara (2011) do not obtain statistically significant R^2_{OS} values. These results hold even for the restricted version of the previous predictive models. When yearly utility gains are considered only the Sum-of-the-Parts approach generates increments around 1% but again this performance arises almost entirely because of recession periods. Our newly introduced methodologies, which combine model selection (MARS, SIC, and Lasso) with support vector machines all produce positive R^2_{OS} values which are significant at the 10% level. The related utility gains range from 0.86% for the SIC Support Vector Machine approach to 2.29% for the MARS Support vector machine approach. These predictive approaches perform, as before, especially well during recession periods, but now the gains are positive even during expansions. After that, diffusion indexes produces a R^2_{OS} of 0.73 with a related p-value of 0.05 but fail to produce positive utility gains. Finally, our approach which employs neural networks is the winner of the horse race: R^2_{OS} values are above $3-4\%$ and are statistically significant at the 10% level while the delta utility gains are around 1.5 %, and for the 40^{th} percentile approach the performance is stable both in expansion and recession periods. In Figure 1 we compare the cumulated returns arising by a buy and hold strategy on the S&P500 and the returns generated by the median restricted forecast of our ensemble of neural networks. It is immediately apparent how the strategy is superior in terms of returns per unit of risk.

Insert Table 1

Insert Table 2

Insert Figure 1

To test the robustness of our findings, we repeat the analysis performed in Tables 1 and 2 using other out-of-sample periods: 2001:1-2017-12 and 2011:1-2017:12. Results for these robustness checks are reported in the online appendix in Tables A2-A4 and confirm our main results. Indeed, the R^2_{OS} statistics generated by our Neural Network approach are above 5% for the 2001-2017 window and remain above 3% for the shorter 2011-2017 period. The related Delta Utility gains are equally important and amount to an average yearly 2.7% for the 2001-2017 window and 2.3% for the 2011-2017 one. Interestingly, the performances of the constrained versions of the Neural Networks are weaker than the performance of their unconstrained counterparts suggesting how these models are successful in the timing of market declines. In these more recent periods, remarkably positive performances are generated even by MARS and Random Forest predictive approaches. For these algorithms, the R^2_{OS} statistics are above 1% with a p-value close or lower to 0.1 The related delta utility gains are approximately 2% for the MARS approach while they reach a recession dependent average 5% for the Random Forest approach.

The findings just recorded implies that some of the most influential papers published in the literature are sample dependent and unable to account for markets structural breaks[19] while our Neural Networks approach appears to remain mostly unaffected by them. These considerations are relevant in light of the highly competitive and fast-changing environment which characterizes stock markets nowadays. Indeed, some studied include in their out-of-sample window even remote periods when the understanding of financial markets was more limited (and market were less efficient) and consequently returns of that time are highly predictable with our state-of-the-art technology. Consequently, the results reported in those studies are biased and unlikely to hold in the current financial market environment.

The results just detailed confirm and augment the finding of Gu et al. (2018). Overall, these results pose a significant challenge to the Efficient Market Hypothesis, which states that prices incorporate all the information efficiently, and return in excess to the risk-free rate must be matched by higher risks. Now, it is becoming apparent how relatively simple machine learning techniques can consistently beat the market without incurring in risks (at least as long as higher risk can be mapped into market variance). While neural networks are black boxes, and the rationale underpinning the genesis of the predictability detected by these models remains largely unexplained, it is getting apparent how more and more powerful machine learning models continuously improve our capability to fore-

[19] Lettau and Van Nieuwerburgh (2007) identifies structural breaks in the dynamics of US stock markets in the early nineties. We prove the robustness of our result by choosing a hard out-of-sample window (1986:1-2017:12) which starts soon before the structural breaks.

cast returns out-of-sample. This poses two fundamental challenges:
1) Understand the genesis of the predictability: which are the factors (linear and non-linear) which the market is unable to reflect promptly and which ultimately generate the predictability detected by our models?
2) Does exist an upper bound to our capability to precisely time the market? Or the only limits are technological and informative? Whether this hypothesis holds financial markets are not efficient but adaptive (Lo (2004)).

4 Predictors

The identification of powerful stock predictors is the second pillar of market predictability. While well known, the predictors of Welch and Goyal (2008) are not necessarily the best predictors for the S&P500 index. We start by re-examining the results of Hong et al. (2007) who employs industry indexes as predictors. The authors perform extensive in-sample analysis and conclude that the returns of industry portfolios can predict the movements of the aggregate market. We re-examine their findings adopting an out-of-sample approach. The predictive models employed in our analysis are the same ones adopted in the previous part and include both univariate OLS regressions and machine learning methodologies. To make our results consistent with the ones of the previous literature we consider two windows of monthly returns: a long one spanning the period 1986:1-2017:12 and a shorter one considering the period 2001:1-2017:12. The most striking evidence is that predictability appears to be higher when we consider only the last seventeen years of monthly data. These results are against the hypothesis that the effect captured by Hong et al. (2007) is due to mispricing and consequently it is destinated to disappear. The R^2_{OS} metrics for the predictions of univariate regressions are positive and statistically significant at the 10% level only for three industries: Health care (1.27%), Money (1.97%) and Others (1.90%). Interestingly, the yearly delta utility percentage gains are positive for all the individual predictors in both the out-of-sample windows considered. These gains appear remarkable and, in the period 2001:1-2017:12, pick to 4.3% and 3.74% for the Money and the Chemical index.
Overall, our results confirm and augment the seminal findings of Hong et al. (2007): industries lead the stock market. After that, we focus on the performance generated

by combining the predictors through our machine learning techniques. Here the results are surprisingly disappointing: R^2_{OS} metrics are usually negative, and their p-value is never below the 0.05 threshold. The related yearly percentage of delta utility increments are consistently positive only for Pooled forecasts and for Random Forests, but these values are below the ones generated by the forecasts of the most performing univariate regressions employed before. In conclusion, when industry indexes are employed as predictors for the S&P500 there is no evidence that by combining predictors we obtain improvements relatively to employing individual predictors only. Our results imply that the results of Rapach et al. (2010) are linked to a specific set of predictors (the Welch and Goyal (2008) ones) and can not be generalized to all typologies of predictors.

Insert Table 3

The second alternative set of predictors which we employ to forecast the S&P500 is composed by the 17 spread returns of the factors-anomalies listed in section 2.2. These returns are the results of the difference between long and short factors-anomalies portfolios returns. As before, two out of sample windows are considered 1986:1-2016:12 and 2001:1-2016:12 and forecasts are performed both through univariate OLS regressions and machine learning techniques which consider all this set of predictors. The results which emerge from the univariate OLS out-of-sample forecasts are impressive. Out of 17 predictors, 3 generate high R^2_{OS} results with a p-value close to zero in both the out-of-sample evaluation windows. More precisely for the 1986:1-2016:12 window the R^2_{OS} resulting from the Asset Growth spread, the Net Stock Issue spread and of the Ohlson spread are respectively equal to 13.2%, 23,5%, and 6.4%. Remarkably the monthly 23,5% R^2_{OS} value stemming from the Net Stock Issue spread is record-high in the financial literature on out-of-sample forecasting. The related yearly delta utility spreads are equally impressive: for the 1986:1-2016:12 period the percentage gains generated by the Asset Growth return spread, the Investment to Asset return spread, the Net Stock Issue return spread and the Ohlson spread are respectively of 11.47%, 10.30%, 23.7%, and 4.53%. These results are confirmed by the 2001:1-2016:12 window. These results while novel and impressive are not entirely unexpected: a relatively unknown study, Greenwood and Hanson (2012), shows how the difference between the attributes of stock issuers and repurchasers can forecast characteristic factor returns.

While seminal and elegant, the analysis performed by the authors remains confined into the in-sample domain. We borrow and extend this intuition to forecast the S&P500 with a variety of spread returns coming from different firms characteristics. After that, the results of the employment of the spread return predictors in our machine learning approaches provide equally satisfactory results. Neural networks achieve an especially positive performance reaching a R^2_{OS} of 17.15% for the 1986:1-2016:12 out-of-sample period (delta utility of 17.15%) and a R^2_{OS} of 8.8% for the 2001:01-2016:12 one (delta utility of 15.85%). Remarkably, even OLS, Pooled MDSFE forecasts, and the Diffusion index approach are performing extremely well: for the 1986:1-2016:12 approach in terms of R^2_{OS} the performances are 17.69%, 12.36%, and 13.54%. While for the same out-of-sample window the yearly percentage Utility Gains of the three approaches are respectively 12.39%, 10.56%, and 7.11%. These results prove the economic value of our methodologies which is well beyond the ones commonly reported in the current academic literature.

Insert Table 4

Now we study a new, often ignored, feature of predictors. They are relevant not only because they allow to achieve better forecasts but even because they indirectly provide novel information on what ultimately the market prices. Indeed, as stressed by Campbell (1991) the field of asset pricing and predictability are intimately connected and represent two sides of the same issue. While, these ideas are largely accepted, nowadays persists a visible shortage of studies addressing this aspect of the problem[20]. We consider it by showing how the study of the predictive power of the different predictors allows us to gain a deeper understanding of the drivers of stock prices and spread returns. Our approach is closely linked to the vibrant literature which is currently employing powerful model selection techniques[21] to identify the key factors among the "factor zoo" denounced by Cochrane (2011) but our model selection technique is applied to identify the best predictors out-of-sample and only indirectly to measure their impact on the cross-section of stock returns.

[20]A recent exception comes from Cujen and Hasler (2017) who explain the higher predictability detected during recessions through the existence of an uncertainty risk premium.

[21]There is a vast blossoming literature on model selection both in therm of the identification of the relevant pricing kernel factors Feng et al. (2017) and on the cross-sectional ones Stambaugh and Yuan (2017),Fama and French (2018) Feng et al. (2017), Kelly et al. (2018), Barillas and Shanken (2018), Hwang and Rubesam (2018),Messmer and Audrino (2017), Kozak et al. (2017a)

Indeed, while predictors are pivotal components in any predictive approach, they can also be employed to gain a better understanding of equity markets both at the aggregate level (S&P500) and at the cross-sectional one (portfolios built on the base of sorting on size, financial ratios or firms characteristics). It is reasonable to believe that the predictors which are better able to forecast out-of-sample an index are somehow informative of the index or portfolio which they can consistently predict. Following this intuition, we propose an extremely simple, yet effective out-of-sample model selection approach, which can be complementary to the commonly employed in sample ones. For the S&P500, six portfolios sorted on the base of size and Momentum, and for six portfolios sorted on the base of size and the Book to Market ratio we identify the four predictors which in univariate regressions achieve the highest R^2_{OS} (Best Individuals in Table 5) AND the combination of four predictors which jointly provides the highest R^2_{OS} in a multivariate linear regression (Combination in Table 5). Our results are based on the monthly out-of-sample period 1998:1-2016:12.

Insert Table 5

Our results imply that the most relevant predictors for the S&P500 are Sentiment and Variance followed by measures extracted from the fix income market: the t-bill rate, the Long-Term yield, and the Long term return. After that, the most effective spread returns predictors for the S&P500 are the Net-Stock-Issue and the Ohlson ones followed by the Asset Growth and the Small minus Big Spread. Overall it appears how the most influential predictors for the S&P500 index can be clustered into two main categories: sentiment based and default risk-based. The first set of predictors do not include only the Sentiment index itself but even measures which are closely linked like Asset Growth and Net-stock-issue. The second set of predictors is linked to the default risk and include the Ohlson spread the Small minus Big spread, the long term yield, and the Long term return. In conclusion, our results agree with the broadly accepted view that risks and risks pricing are the driving forces of financial markets[22]. In conclusion, an effective

[22]The first seminal studies which address this topic comes from Campbell and Shiller (1988) and Campbell (1991) who introduce the key conceptual framework at the base of our understanding of financial markets. More recently Fuss et al. (2016) and Campbell et al. (2013) applied this framework to the study of the 2008 financial crisis. A line of studies which merges behavioural and neoclassical finance includes Shefrin (2008) and Shefrin and Statman (1994) for stocks and Barone-Adesi et al. (2016) for options.

predictor needs to be able to predict one of these two key features[23].

Looking at double-sorted portfolios, we report many novel findings. As before the upper panel employs the Welch and Goyal (2008) predictors while the lower panel employs the spread returns coming from factor-anomalies. First, we observe how, coherently with the results of Baker and Wurgler, sentiment is especially powerful in predicting stocks with a low book to market ratio: it is the most powerful predictors when individual regressions are considered and is one of the predictors included in the combination of the four most powerful predictors. Second, stocks with a high book to market value appear to be driven mostly by fundamentally driven predictors, like inflation and volatility. After that, default yield and long term bond returns are especially successful in forecasting the returns of the portfolios of stocks which experienced negative returns (low prior) while stocks which reported positive performances in the recent past (high prior) are better predicted by volatility, sentiment and long term yield. The lower panel (anomalies) reinforce the previous findings. The stocks which have a high book to market value are forecasted by the Net stock issue spread (a variable linked to the Baker-Wurgle Sentiment index formulation) while the Distress spread strongly predicts stocks which have a low book to market value. Now firms with relatively poor past stock returns are forecasted by the Distress spread while Net stock issue spread predicts stocks which experienced high past returns. Finally, the low-high Book/Market spread is strongly forecasted by the Asset Growth spread, and by the Net Operating Asset spread suggesting how profitability is linked to the relative performance of Book to Market sorted stocks. On the other hand, the High minus Low prior spread is better forecasted by the Ohlson, Distress, and Return on Asset spreads suggesting that default risk is the dominant issue here.

The results of this section while far from being conclusive aim at triggering fruitful discussions:

1. Market efficiency is challenged not only by more and more powerful models but even by more and more powerful predictors. This implies that we can address the challenge posed by Welch and Goyal (2008) even by using a simple univariate regression whether the predictor employed is powerful enough.

2. We suggest how commonly employed model selection techniques which perform well

[23] A first promising way to study the genesis of the out-of-sample predictability comes from Rapach et al. (2016) and Wen (2019)

in-sample should be backed by complementary out-of-sample ones.

3. More broadly, out-of-sample analyses are as informative as in-sample ones to gain a better understanding of financial market dynamics.

5 Predictability as a generalized phenomenon

In the previous sections, we proved how using more powerful predictive models or more powerful predictors it is possible to forecast out-of-sample the returns of the S&P500. These results are interesting because of the high efficiency of the U.S. stock market and suggest that in less efficient markets predictability should be even higher (Fama (1970), Lo (2012) and Barone-Adesi and Sala (2019)). In this section, we address this issue while remaining focused on the US equity stock market. Whether efficiency is directly linked to predictability, less efficient markets should be more predictable than more efficient ones. In the universe of US equities, small stocks are a natural candidate to test this hypothesis. Indeed, small caps are intrinsically less liquid, it is not always possible to short them and when it is possible this procedure is more expensive. After that, to reduce transaction costs, passive funds try to minimize their investments in them while they receive less attention from analysts and media. Even more importantly due to their illiquidity risk and high transaction costs some categories of institutional investors (like high-frequency trading funds and hedge funds) are less prone to invest in them. On the other hand, big capitalization stocks are the natural counterpart of small stocks to verify our hypothesis. Finally, our analysis allows us to test whether the predictability is driven by small stocks only.

To perform our empirical investigation, for each variable out-of-sample evaluation is based on the most recent 30% of the available monthly time series. To be consistent with our previous analyses we report the R^2_{OS} metric and the Δ Utility one. To anchor our result to the existing literature on the field we make use of the double sorted portfolio returns coming from the French data library: the six double-sorted portfolios formed on the base of size and the book to market ratio and the six double-sorted portfolios formed on the base of size and the previous returns performance (Momentum). For each portfolio, we report the average R^2_{OS} generated by the univariate OLS forecasts and all the machine learning methodologies detailed in section 2. We repeated these analyses twice: at first by making use of the Welch and Goyal (2008) predictors and subsequently by employing the 17 spread returns predictors introduced in section 2.2[24].

Insert Table 6

In the upper panel of the table, we present averages of R^2_{OS} for individual portfolios with the related subtotals R^2_{OS}. After that, we perform a difference between means hypothesis test between portfolios which diverge on size only (e.i. both have a low Book-Market ratio, but they diverge in terms of size). The null hypothesis is that the two means are equal against the general alternative hypothesis. The p-values generated by our analysis are reported in the lower panel of Table 6. Our results confirm that small stocks are more predictable than big ones confirming that efficiency and predictability are closely linked. Indeed, the bottom line of the SMALL-BIG columns shows that the difference in mean is statistically significant at the 0.01 level for the Delta Utility and at the 0.05/0.1 level for the R^2_{OS} metrics while the summary Delta Utility and R^2_{OS} metrics are always higher for small than for big caps. After that, the table shows how while Small stocks are more predictable, predictability is a broad phenomenon which includes even big caps: we observe how using return spreads predictors (Anomalies in the Table) it is possible to achieve, on average, utility gains for each portfolio of big stocks considered. These results are confirmed by the average R^2_{OS} values achieved using return spreads predictors: 0.68% and 1.16% for portfolios built on book to market and momentum sorting. Finally, some predictability patterns emerge: stocks with low Book-to-Market ratio are more predictable than stocks with a high Book-to-Market ratio while stocks with lower previous market returns are more predictable than stock with higher previous market returns. While a study on the genesis of this predictability is beyond the scope of this paper, our out-of-sample analysis in the previous section could provide some preliminary hints.

Having studied the dynamics of predictability inside the French double sorted portfolio framework, now we want to address the issue in a more general framework employing a broader set of stocks. Consequently, we study the predictability for the returns of

[24]The detailed results for both the R^2_{OS} and Δ Utility metrics are reported, for brevity, in the online appendix: Tables A5-A12

thirty equally weighted portfolios and ten related returns spreads[25] built following the method proposed by Stambaugh and Yuan (2017) who consider a set of eleven anomalies (we do not include momentum because we have already analyzed it in Table 6). The monthly out-of-sample period considered for each variable includes the most recent 30% of the available data. Forecasts are based on the Welch and Goyal (2008) predictors: we consider both univariate regressions and all the machine learning techniques detailed in the second part of this paper. For each portfolio and spread we report the average, median, maximum, and minimum value for the R^2_{OS} metric, for the related Clark and West (2007) p-value (Table 7) and, for the yearly percentage Delta Utility (Table 8). In Table 7 for each portfolio, we report even the percentage of forecasts which have a R^2_{OS} p-value lower than 0.1 and 0.05. Finally, in the lower panels of table 7 and 8, we report summary statistics from the panels above.

Insert Table 7

Insert Table 8

Our results document the existence of an extensive degree of predictability in financial markets: the 20% of the R^2_{OS} values is positive with a p-value under 0.1 while the average maximum R^2_{OS} documented for each portfolio is 1.9%. Even more remarkable are the results in terms of utility gains where the comprehensive average value for all the portfolios and spreads is 4.8%, and the average maximum delta utility is close to 10%. These results imply that on average it is possible to add value through predictive models. Even more importantly these results are not confined to the S&P500, to small stocks or a specific subset of the US equity market, but they are generalizable to the average U.S. equity stock. We want to stress how our results are conservative because they relay on the Welch and Goyal (2008) predictors which are less powerful than the spread return ones. The results of these last two table confirm and augment the ones coming from Gu et al. (2018) and Rasekhschaffe and Jones (2019) who used a broad set of machine learning predictors to identify the stocks which are more likely to perform relatively better or worse than the others. Differently, from them, we focused on portfolios built under a variety of criteria reaching similar conclusions: machine learning techniques can consistently time the market.

[25] A return spread for a given anomaly implies taking a long position on the portfolio which has the highest expected return and short on the portfolio which has the lowest expected return.

6 Conclusions

In this paper, we examine the key aspects of financial market predictability: predictive modelling and predictors.

At first, we focus on predictive models employing as inputs the well known Welch and Goyal (2008) predictors. We show how combining machine learning and model selection techniques the capability to forecast out-of-sample the S&P500 rises dramatically. Remarkably, when model selection techniques are combined with Ensembles of multilayer Neural Networks the monthly R^2_{OS} riches a statistically significant 4.4% for the period 1986-2017 and 6.14% for the shorter 2001-2017 period. The related annualized utility gains are of an equally relevant magnitude picking at 3% for the 2001-2017 interval. The implications of these findings for the theory of finance are twofold. From one side our results pose a significant challenge to the efficient market hypothesis proving how machine learning experts can build algorithms capable of consistently outperforming the market, on the other side they suggest how new asset pricing models should include nonlinear interdependencies in the formulation of the pricing kernel.

In the second section of the paper, we consider a variety of different predictors. We show how the returns generated from the long-short strategies (often addressed as anomalies in the literature) have a surprisingly strong out-of-sample predictive power for the S&P500. The most successful predictors are the spread returns based on Asset Growth, Net Stock Issue, Olshon and Investment to Asset characteristics which, for the 1986-2016 out-of-sample period reach a record high R^2_{OS} level of 13%, 23.5%, 6.4% and 11.1%. The Delta Utility gains generated by these predictors confirm their profitability. After that, we study the predictability of the double-sorted portfolios of Fama and French, and we find that while small stocks are on average more predictable then big ones, predictability is a generalized feature of financial markets when machine learning techniques are employed. Finally, we propose to employ an out-of-sample approach as a complement to the traditional in sample techniques to identify which characteristics are ultimately reflected into stock prices. Our simple method opens the ground to a much-needed study on the relationship between predictability and pricing.

7 Tables and Figures

Table 1: **Monthly equity premium out-of-sample forecasting results for individual forecasts, and machine learning methods**. The R^2_{OS} is the Campbell Thompson (2008) out-of-sample R^2 statistic. Statistical significance for the R^2_{OS} statistic is based on the p-value for the Clark and West (2007) out-of-sample MPSE-adjusted statistic; the statistic corresponds to a one-sided test of the null hypothesis that the competing forecasting model has equal expected square prediction error relative to the historical average benchmark forecasting model against the alternative hypothesis that the competing forecasting model has a lower expected square prediction error than the historical average benchmark forecasting model. The results refer to monthly forecasts for the out-of-sample period 1986-2017. For predictions based on univariate forecasts the restrictions are the ones suggested by Campbell and Thompson (2008) while for the machine learning methods when equity premium forecasts are negative they are replaced with zero. *,** and *** indicate significance level at the 10%, 5% and 1%. Bold indicates a pvalue for the R^2_{OS} statistic less than 0.1. .

Standard	**1986-2017**		**Restricted**	**1986-2017**	
Predictor	$R^2_{OS}(\%)$	**pval**	**Predictor**	$R^2_{OS}(\%)$	**pval**
DP	-1.34	0.52	DP	-1.00	0.50
DY	-1.99	0.48	DY	-1.17	0.52
EP	-1.41	0.32	EP	0.07	0.20
DE	-0.54	0.54	DE	-0.03	0.31
SVAR	0.39	0.16	SVAR	0.32	0.13
BM	-2.28	0.57	BM	-1.29	0.56
NTIS	-1.77	0.65	NTIS	-1.77	0.65
TBL	-0.21	0.47	TBL	-0.20	0.46
LTY	-0.06	0.44	LTY	-0.06	0.44
LTR	-0.31	0.40	LTR	-0.36	0.44
TMS	-0.83	0.64	TMS	-0.83	0.64
DFY	-0.20	0.92	DFY	-0.20	0.92
DFR	0.18	0.29	DFR	-0.19	0.43
INFL lag	-0.35	0.84	INFL lag	-0.35	0.84

Model	$R^2_{OS}(\%)$	**pval**	**Model**	$R^2_{OS}(\%)$	**pval**
OLS	-5.83	0.36	OLS	-1.83	0.24
Pooled forecast: median	0.08	0.32	Pooled forecast: median	0.08	0.32
Pooled forecast: DMSFE	-0.01	0.42	Pooled forecast: DMSFE	-0.01	0.42
Sum-of-the-parts	0.24	0.21	Sum-of-the-parts	0.47	0.12
MARS	**0.89****	**0.02**	MARS	**0.95*****	**0.01**
SVM SIC	**0.49***	**0.06**	SVM SIC	**0.71****	**0.02**
Lasso SVM	**0.37***	**0.10**	Lasso SVM	**0.58****	**0.05**
Random Forest	**0.52***	**0.10**	Random Forest	**0.59***	**0.08**
Diffusion index	**0.73****	**0.05**	Diffusion index	**0.73****	**0.05**
PLS	0.02	0.15	PLS	**0.31***	**0.09**
Neural Networks Median	**3.22***	**0.09**	Neural Networks Median	**0.03***	**0.06**
Neural Networks 40^{th}	**4.38***	**0.10**	Neural Networks 40^{th}	**0.85****	**0.03**

Table 2: **Monthly equity premium out-of-sample forecasting results for individual forecasts, and machine learning methods**. Utility gain (Δ Utility) 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 forecasting model considered relative to the historical average benchmark forecasting model; the weight on stocks in the investor's portfolio is restricted to lie between -0.5 and 1.5 (inclusive). The restriction imposed for the restricted case are the same of Table 1. The results refer to monthly forecasts for the out-of-sample period 1986-2017. The division between Recession and Expansion months comes from the NBER database. *,** and *** indicate a Δ Utility % increase above 1%, 5% and 10%. **Bold** indicates a Δ Utility above 1.00%.

Δ Utility	**1986-2017**			**Δ Utility**	**1986-2017**		
Standard	**Total**	**Expansion**	**Recession**	**Restricted**	**Total**	**Expansion**	**Recession**
DP	-2.59	-5.29	**23.72***	DP	-1.23	-3.78	**23.62***
DY	-2.66	-6.10	**31.22***	DY	-1.16	-4.44	**31.14***
EP	**1.83***	-1.50	**34.59***	EP	**2.01***	-1.17	**33.32***
DE	-0.26	-0.21	-0.43	DE	-0.05	-0.06	0.00
SVAR	-0.58	-0.42	-2.07	SVAR	0.05	0.03	0.35
BM	-2.67	-6.40	**34.25***	BM	-1.03	-4.56	33.86
NTIS	-0.66	-0.10	-6.53	NTIS	-0.66	-0.10	-6.53
TBL	0.00	0.45	-4.49	TBL	0.00	0.45	-4.49
LTY	0.06	0.23	-1.63	LTY	0.06	0.23	-1.63
LTR	-0.18	-0.63	**3.73***	LTR	-0.26	-0.60	**2.65***
TMS	-1.08	-0.73	-4.61	TMS	-1.08	-0.73	-4.61
DFY	-0.90	-0.58	-4.68	DFY	-0.90	-0.58	-4.68
DFR	0.96	0.26	**7.81****	DFR	0.87	0.21	**7.46****
INFL lag	-0.74	0.08	-8.41	INFL lag	-0.74	0.08	-8.41

Standard	**Total**	**Expansion**	**Recession**	**Restricted**	**Total**	**Expansion**	**Recession**
OLS	-1.23	-4.76	**33.46***	OLS	0.72	-2.22	**29.42***
Pooled forecast: median	0.07	0.28	-1.89	Pooled forecast: median	0.07	0.28	-1.89
Pooled forecast: DMSFE	-0.05	-0.24	**1.90***	Pooled forecast: DMSFE	-0.05	-0.24	**1.90***
Sum-of-the-Parts	0.60	-0.44	**10.77***	Sum-of-the-Parts	0.91	-0.31	**12.79***
MARS	**2.29***	**1.45***	**10.48***	MARS	**2.37***	**1.47***	**11.13***
SVM SIC	0.86	0.28	**6.86****	SVM SIC	**1.20***	0.28	**10.33***
Lasso SVM	**1.17***	0.47	**8.14****	Lasso SVM	**1.35***	0.47	**10.05***
Random Forest	**1.54***	-0.08	**17.39***	Random Forest	**1.54***	-0.08	**17.39***
Diffusion Index	-0.24	0.12	-4.32	Diffusion Index	-0.24	0.12	-4.32
PLS	0.22	0.60	-3.64	PLS	0.27	0.66	-3.64
Neural Networks Median	**1.38***	**1.80***	-1.34	Neural Networks Median	**1.18***	**1.52***	**2.10***
Neural Networks 40^{th}	**1.71***	**1.71***	**1.71***	Neural Networks 40^{th}	**1.64***	**1.62***	**1.92***

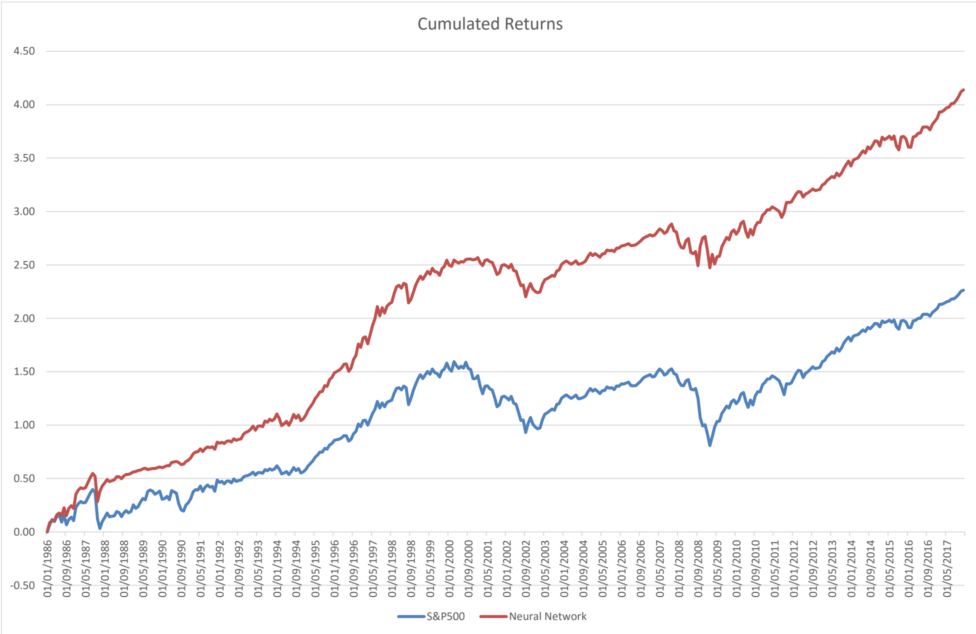


Figure 1: **Cumulated monthly returns for the 1986:1-2017:12 period**. The Blue line tracks cumulated returns for the S&P500 index (Blue) and for the strategy which employs the median (Restricted) forcast arising from an ensamble of Neural Networks to proxy for expected returns in the optimization process (Red).

Table 3: **Industry predictors**: monthly equity premium out-of-sample forecasting results for individual forecasts, and machine learning methods. We consider two monthly out-of-sample windows: 1986:1-2017:12 and 2001:1-2017:12. For the R^2_{OS} statistic *,** and *** indicate significance level at the 10%, 5% and 1%. For ΔUtility % * indicates an yearly increase above 1%. Bold indicates a ΔUtility above 1.00% or a R^2_{OS} with a p-value lower than 0.1.

	1986-2017		**2001-2017**	
Predictor	R^2_{OS}	**pval**	R^2_{OS}	**pval**
NoDur	-0.60	0.41	0.00	0.34
Durbl	-0.37	0.29	0.94	0.14
Manuf	-0.31	0.33	0.99	0.18
Enrgy	-0.42	0.67	0.08	0.36
Chems	-0.24	0.30	1.12	0.15
BusEq	-0.38	0.18	1.45	0.11
Telcm	-0.28	0.47	0.79	0.17
Utils	0.30	0.21	0.98	0.17
Shops	0.32	0.18	0.73	0.17
Hlth	-0.11	0.27	**1.27***	**0.09**
Money	0.07	0.17	**1.97***	**0.08**
Other	0.50	0.14	**1.90***	**0.09**

	1986-2017	**2001-2017**
Predictor	**Δ Utility**	**Δ Utility**
NoDur	0.46	**1.94***
Durbl	0.53	**2.45***
Manuf	0.40	**2.80***
Enrgy	0.81	**1.85***
Chems	**1.08***	**3.74***
BusEq	0.31	**3.21***
Telcm	0.13	**2.61***
Utils	**2.03***	**3.44***
Shops	**1.35***	**2.50***
Hlth	0.60	**2.91***
Money	**1.27***	**4.37***
Other	**1.75***	**3.90***

	1986-2017		**2001-2017**	
Model	R^2_{OS}	**pval**	R^2_{OS}	**pval**
OLS	-2.79	0.12	-1.20	0.14
Pooled forecast:median	0.21	0.23	1.07	0.13
Pooled forecast:MDSFE	0.18	0.24	1.29	0.12
MARS	-3.00	0.97	-3.00	0.89
SVM SIC	-0.60	0.38	-1.52	0.64
Lasso SVM	-0.64	0.39	-1.59	0.66
Radom Forest	0.06	0.30	**0.90***	**0.09**
Diffusion index	-0.35	0.26	-1.00	0.49
PLS	-1.29	0.56	-2.08	0.65
Neural Networks Median	-0.01	0.22	-1.21	0.52

	1986-2017	**2001-2017**
Model	**Δ Utility**	**Δ Utility**
OLS	0.09	**3.17***
Pooled forecast:median	**1.37***	**3.21***
Pooled forecast:MDSFE	**1.49***	**3.57***
MARS	-3.47	-1.61
SVM SIC	-0.93	-0.94
Lasso SVM	-0.94	-0.96
Radom Forest	**1.37***	**2.61***
Diffusion index	-0.41	-0.49
PLS	-0.96	-1.01
Neural Networks Median	0.02	-0.94

Table 4: **Factors-Anomalies spread return predictors**: monthly equity premium out-of-sample forecasting results for individual forecasts, and machine learning methods. We consider two monthly out-of-sample windows: 1986:1-2016:12 and 2001:1-2016:12. For the R^2_{OS} statistic *,** and *** indicate significance level at the 10%, 5% and 1%. For ΔUtility % *,** and *** indicate an increase above 1%, 5% and 10%. Bold indicates a ΔUtility above 1.00% or a R^2_{OS} with a p-value lower than 0.05.

	1986-2016		**2001-2016**	
Predictor	R^2_{OS}	**pval**	R^2_{OS}	**pval**
SMB	-0.45	0.29	-0.98	0.49
HML	-0.22	0.35	0.07	0.33
RMW	-0.40	0.43	-0.41	0.41
CMA	0.19	0.07	0.55	0.17
LT	-0.42	0.39	-0.87	0.54
ST	-0.76	0.82	-2.15	0.97
Mom	-0.78	0.90	-1.12	0.89
Asset Growth	**13.20***	**0.00**	**3.55***	**0.00**
Gross Prof	-0.15	0.00	-11.67	0.83
Inv to Assets	**11.13***	**0.00**	-1.17	0.01
Net Stock Issues	**23.54***	**0.00**	**28.67***	**0.00**
NOA	-2.95	0.36	-4.18	0.80
Accruals	-1.72	0.01	-10.18	0.56
O	**6.43***	**0.00**	**6.69**	**0.01**
ROA	-3.64	0.04	-12.88	0.95
Distress	0.71	0.08	0.14	0.25
Comp Eq Issue	-0.38	0.49	-0.19	0.32

	1986-2016	**2001-2016**
Predictor	**Δ Utility**	**Δ Utility**
SMB	-0.58	-1.44
HML	0.11	**1.23***
RMW	0.39	**1.20***
CMA	0.99	**1.51***
LT	0.45	0.46
ST	-0.68	-2.20
Mom	-0.61	-0.66
Asset Growth	**11.47***	**6.96**
Gross Prof	**1.06***	-6.68
Inv to Assets	**10.30***	**6.29**
Net Stock Issues	**23.56***	**28.53***
NOA	-2.31	-3.64
Accruals	**2.39***	-1.85
O	**4.53**	**5.43**
ROA	0.05	-5.50
Distress	**2.36***	**2.59***
Comp Eq Issue	-0.10	0.77

	1986-2016		**2001-2016**	
Model	R^2_{OS}	**pval**	R^2_{OS}	**pval**
OLS	**15.71***	**0.00**	**17.69***	**0.00**
Pooled Forecast median	**2.37***	**0.00**	**3.50***	**0.00**
Pooled Forecast MDSFE	**10.89***	**0.00**	**12.36***	**0.00**
MARS	**11.82**	**0.00**	**4.90***	**0.00**
SVM SIC	-20.13	0.34	-12.36	0.18
Lasso svm	-12.46	0.24	-9.03	0.18
Random Forest	-0.01	0.40	0.12	0.18
Diffusion Index	**4.16***	**0.01**	**13.54***	**0.00**
PLS	-30.75	0.92	-32.40	0.97
Neural Networks Median	**16.06***	**0.00**	**8.72**	**0.00**

	1986-2016	**2001-2016**
Model	**Δ Utility**	**Δ Utility**
OLS	**14.13***	**12.39***
Pooled Forecast median	**5.05**	**3.38***
Pooled Forecast MDSFE	**12.76***	**10.56***
MARS	**5.11**	**13.10***
SVM SIC	0.24	0.16
Lasso svm	0.00	0.86
Random Forest	0.08	0.47
Diffusion Index	**10.97***	**7.11**
PLS	-9.67	-4.99
Neural Networks Median	**7.96**	**15.64***

Table 5: **Out-of-sample model selection approach.** This table considers the S&P500 index, six portfolios built sorting on the base of size and the Book/Market ratio, six portfolios built sorting on the base of size and Momentum, and four portfolios spreads. The out-of-sample monthly evaluation period is 1998:1-2016:12. The upper panel of the table shows the results by employing the Welch and Goyal (2008) variables plus the Sentiment index of Huang et al. (2015) (W-G in the Table) while the lower panel considers the 16 spreads of factors-anomalies considered in this paper plus the sentiment index of Huang et al. (2015) (Anomalies in the Table). For each variable considered the table ranks the 4 predictors which individually exhibit the highest R^2_{OS} in univariate regression forecasting (Best Individual in the Table) and the four variables which jointly have the highest R^2_{OS} in multivariate OLS forecasting (Combination). Data are monthly and the R^2_{OS} metrics are based on the out-of-sample period from 1998 to 2016. Where for the W-G panel the numbers mean: DP(1), DY(2), EP(3), DE(4), SVAR(5), BM(6), NITIS(7), TBL(8), LTY(9), LTR(10), TMS(11), DFY(12), DFR(13), INFlag(14), SENT(15). While for the Anomalies panel the numbers means: SMB(1), HML(2), RMW(3), CMA(4), LT(5), ST(6), MOM(7), Asset Growth(8), Gross Profitability(9), Investment to Assets(10), Net Stock Issues(11), NOA(12), Accruals(13), Ohlson(14), ROA(15), Distress(16), Composite Equity Issue(17).

Combination				
W-G				
S&P 500	5	10	14	15
SMALL LoBM	5	9	10	15
ME1 BM2	5	10	11	12
SMALL HiBM	5	10	11	14
BIG LoBM	5	10	14	15
ME2 BM2	5	10	14	15
BIG HiBM	5	10	14	15
SMALL LoPRIOR	1	2	10	12
ME1 PRIOR2	5	10	11	12
SMALL HiPRIOR	5	8	9	10
BIG LoPRIOR	10	12	14	15
ME2 PRIOR2	5	10	14	15
BIG HiPRIOR	5	9	10	15
SMALL LoBM-SMALL HiBM	7	11	12	15
BIG LoBM-BIG HiBM	5	10	14	15
SMALL HiPRIOR- SMALL LoPRIOR	1	2	3	9
BIG HiPRIOR-BIGLoPRIOR	6	10	12	14

Best individual				
W-G	**(1)**	**(2)**	**(3)**	**(4)**
S&P 500	15	5	9	8
SMALL LoBM	15	5	10	9
ME1 BM2	5	9	15	8
SMALL HiBM	5	9	15	11
BIG LoBM	15	5	9	10
ME2 BM2	5	15	14	9
BIG HiBM	5	14	15	9
SMALL LoPRIOR	15	10	12	9
ME1 PRIOR2	5	10	9	15
SMALL HiPRIOR	5	9	15	10
BIG LoPRIOR	15	10	14	6
ME2 PRIOR2	15	5	8	14
BIG HiPRIOR	5	15	9	13
SMALL LoBM-SMALL HiBM	15	5	12	14
BIG LoBM-BIG HiBM	14	5	11	15
SMALL HiPRIOR- SMALL LoPRIOR	4	3	12	9
BIG HiPRIOR-BIGLoPRIOR	12	10	5	4

Anomalies				
S&P 500	1	11	12	14
SMALL LoBM	5	6	7	11
ME1 BM2	6	11	14	16
SMALL HiBM	6	11	14	16
BIG LoBM	1	5	11	14
ME2 BM2	6	11	12	14
BIG HiBM	9	11	12	15
SMALL LoPRIOR	5	6	11	16
ME1 PRIOR2	6	7	11	16
SMALL HiPRIOR	1	6	11	14
BIG LoPRIOR	11	12	14	16
ME2 PRIOR2	11	12	14	16
BIG HiPRIOR	1	5	6	11
SMALL LoBM-SMALL HiBM	8	11	12	16
BIG LoBM-BIG HiBM	6	8	12	15
SMALL HiPRIOR- SMALL LoPRIOR	12	14	15	16
BIG HiPRIOR-BIGLoPRIOR	12	14	15	16

Anomalies	**(1)**	**(2)**	**(3)**	**(4)**
S&P 500	11	14	8	10
SMALL LoBM	11	8	10	14
ME1 BM2	11	16	7	10
SMALL HiBM	11	16	7	6
BIG LoBM	11	14	8	10
ME2 BM2	11	16	14	10
BIG HiBM	11	16	10	14
SMALL LoPRIOR	11	16	10	8
ME1 PRIOR2	11	16	10	7
SMALL HiPRIOR	11	7	5	3
BIG LoPRIOR	16	11	10	8
ME2 PRIOR2	11	14	16	8
BIG HiPRIOR	11	14	3	17
SMALL LoBM-SMALL HiBM	11	8	14	10
BIG LoBM-BIG HiBM	14	15	8	12
SMALL HiPRIOR- SMALL LoPRIOR	16	15	10	12
BIG HiPRIOR-BIGLoPRIOR	16	15	14	12

Table 6: **Out-of-sample predictability of big vs small stocks.** In this table we compare the out-of-sample predictability of a set of 12 portfolios: six double sorted on the base of size and the Book to Market ratio and six sorted on the base of size and the previous performance (Momentum). In the first eight rows we report the average performance for each portfolios in terms of R^2_{OS} using the Welch and Goyal (2008) predictors (W-G) and the 17 factors-anomalies spread returns (Anomalies). In both cases we employ univariate regressions and all the machine learning methodologies previously detailed. The monthly out-of-sample period considered is 1:1986-12:2016. Total Average rows and SMALL and BIG columns (in bold in the table) compute the sub-total averages. In rows form 9 to 16 we repeat the same exercise with the Delta Utility metric.
In the last eight rows of the table we report the p-values for the difference in mean between portfolios which diverge only because of size. The Null Hypothesis is that there is no difference in means. ***,**, and * indicates a pvalue under 0.01, 0.05 and 0.1.

R^2_{OS}	SMALL LoBM	ME1 BM2	SMALL HiBM	SMALL	BIG LoBM	ME2 BM2	BIG HiBM	BIG
W-G	-0.65	-1.41	-2.08	**-1.38**	-0.73	-1.48	-1.70	**-1.30**
Anomalies	3.78	1.37	1.36	**2.17**	1.57	-0.20	0.67	**0.68**
Total Average	**1.56**	**-0.02**	**-0.36**	**0.39**	**0.42**	**-0.84**	**-0.51**	**-0.31**
R^2_{OS}	SMALL LoPRIOR	ME1 PRIOR2	SMALL HiPRIOR	SMALL	BIG LoPRIOR	ME2 PRIOR2	BIG HiPRIOR	BIG
W-G	-0.88	-1.59	-0.88	**-1.12**	-1.49	-1.12	-0.60	**-1.07**
Anomalies	5.43	2.01	1.49	**2.98**	3.46	0.33	-0.31	**1.16**
Total Average	**2.27**	**0.21**	**0.31**	**0.93**	**0.99**	**-0.40**	**-0.45**	**0.05**
Δ Utility	SMALL LoBM	ME1 BM2	SMALL HiBM	SMALL	BIG LoBM	ME2 BM2	BIG HiBM	BIG
W-G	-0.27	-0.50	-1.30	**-0.69**	0.53	0.25	-1.84	**-0.35**
Anomalies	7.22	4.56	3.88	**5.22**	3.85	2.64	2.93	**3.14**
Total Average	**3.48**	**2.03**	**1.29**	**2.26**	**2.19**	**1.45**	**0.55**	**1.40**
Δ Utility	SMALL LoPRIOR	ME1 PRIOR2	SMALL HiPRIOR	SMALL	BIG LoPRIOR	ME2 PRIOR2	BIG HiPRIOR	BIG
W-G	-1.66	-0.48	0.13	**-0.67**	-2.29	-0.02	0.92	**-0.47**
Anomalies	4.46	4.95	6.33	**5.24**	3.71	3.12	4.23	**3.68**
Total Average	**1.40**	**2.23**	**3.23**	**2.29**	**0.71**	**1.55**	**2.57**	**1.61**

p-values

R^2_{OS}	LoBM	ME1 BM2	HiBM	SMALL-BIG	LoPRIOR	PRIOR2	HiPRIOR	SMALL-BIG
W-G	0.58	0.67	**0.10****	0.45	**0.00********	**0.06****	0.33	0.74
Anomalies	0.23	0.21	0.41	**0.05****	**0.10***	0.12	0.15	**0.01*****
Total Average	0.21	0.19	0.69	**0.07***	**0.04****	0.25	0.22	**0.01*****
Δ Utility	LoBM	ME1 BM2	HiBM	SMALL-BIG	LoPRIOR	PRIOR2	HiPRIOR	SMALL-BIG
W-G	**0.04****	**0.00*****	**0.02****	**0.06***	**0.10***	0.32	**0.02****	0.37
Anomalies	**0.01***	**0.01***	**0.08***	**0.00*****	0.27	**0.00*****	**0.01*****	**0.00*****
Total Average	**0.07***	0.12	**0.01*****	**0.00*****	**0.08***	**0.08***	0.13	**0.00*****

Table 7: **Out-of-sample predictability of anomalies portfolios: R^2_{OS}.** In the Upper Panel we report the Average, Median, Maximum and Minimum R^2_{OS} values for 30 portfolios based on characteristics sorting and 10 spread portfolios returns. The monthly out-of-sample period considered is the most recent 30% for each variable. Forecasts are based on Welch and Goyal (2008) predictors: we consider both univariate regression and all the machine learning techniques detailed in the second part of this paper. Subsequently, we reported the related Clark and West (2007) p-values. Finally, for each portfolio and spread we report the % of p-values under 0.1 and under 0.05. In the Lower Panel we briefly summarize the results coming from the upper panel. We use bold to remark Maximum R^2_{OS} values above 1% and Minimum p-values under 0.1.

	R^2_{OS}	Average	Median	Max	Min	% Pval<0.1	% Pval<0.05		R^2_{OS}	Average	Median	Max	Min	% Pval<0.1	% Pval<0.05
Asset Growth	Low	0.09	0.42	**2.42**	-5.98			**Accruals**	Low	-0.33	-0.14	0.36	-2.81		
	pval	0.27	0.14	0.88	**0.01**	0.38	0.27		pval	0.35	0.31	0.73	0.14	0.00	0.00
	Medium	-0.48	0.15	**1.81**	-8.37				Medium	-0.89	-0.44	0.65	-3.42		
	pval	0.30	0.17	0.86	**0.03**	0.35	0.15		pval	0.46	0.47	0.76	0.11	0.00	0.00
	High	-1.41	-0.52	**1.42**	-12.24				High	-1.57	-0.47	**1.06**	-7.67		
	pval	0.45	0.48	0.83	**0.02**	0.08	0.08		pval	0.52	0.50	0.96	**0.04**	0.08	0.04
	Spread	-4.40	-4.49	**2.55**	-17.88				Spread	1.22	0.49	**8.51**	-4.78		
	pval	0.43	0.35	0.97	**0.01**	0.19	0.12		pval	0.38	0.06	1.00	**0.00**	0.54	0.50
Gross Prof	Low	-0.39	-0.20	**1.82**	-6.95			**O**	Low	-0.31	-0.01	**1.30**	-3.67		
	pval	0.38	0.32	0.94	**0.03**	0.27	0.15		pval	0.37	0.29	0.88	**0.02**	0.12	0.04
	Medium	-1.07	-0.41	**1.31**	-8.64				Medium	-0.73	-0.28	**1.74**	-5.32		
	pval	0.40	0.33	0.86	**0.01**	0.08	0.04		pval	0.34	0.26	0.79	**0.02**	0.04	0.04
	High	-0.91	0.08	**1.65**	-11.87				High	-0.63	-0.16	**1.65**	-4.72		
	pval	0.28	0.21	0.90	**0.00**	0.27	0.15		pval	0.32	0.29	0.78	**0.03**	0.19	0.04
	Spread	-2.80	-1.93	**0.94**	-21.68				Spread	-2.02	-1.01	**3.00**	-17.49		
	pval	0.58	0.65	0.84	**0.06**	0.08	0.00		pval	0.51	0.56	0.99	**0.00**	0.19	0.08
Inv to Assets	Low	0.26	0.73	**2.35**	-6.21			**ROA**	Low	-0.24	-0.28	**2.00**	-3.04		
	pval	0.25	0.09	0.87	**0.02**	0.50	0.31		pval	0.45	0.48	0.82	**0.01**	0.08	0.08
	Medium	-0.66	0.03	**1.74**	-8.01				Medium	-0.01	-0.01	**1.90**	-1.95		
	pval	0.33	0.31	0.87	**0.01**	0.19	0.15		pval	0.37	0.29	0.95	**0.01**	0.19	0.04
	High	-1.56	-0.60	**0.85**	-11.51				High	-0.34	-0.06	**2.05**	-3.70		
	pval	0.53	0.61	0.98	**0.02**	0.08	0.04		pval	0.47	0.34	0.98	**0.01**	0.08	0.04
	Spread	-0.73	0.34	**3.59**	-10.89				Spread	-2.45	-1.44	**0.33**	-11.41		
	pval	0.30	0.09	1.00	**0.00**	0.50	0.46		pval	0.81	0.92	1.00	**0.09**	0.04	0.00
Net Stock Issues	Low	-0.70	0.07	**2.82**	-11.69			**Distress**	Low	-0.33	-0.37	**1.67**	-4.00		
	pval	0.29	0.18	0.89	**0.01**	0.35	0.12		pval	0.41	0.30	1.00	**0.02**	0.31	0.15
	Medium	-0.39	0.31	**2.16**	-7.79				Medium	-0.35	-0.10	0.64	-2.96		
	pval	0.30	0.22	0.82	**0.01**	0.31	0.12		pval	0.49	0.45	0.96	0.12	0.00	0.00
	High	-0.64	0.00	**1.45**	-7.76				High	-0.86	-0.26	0.33	-7.65		
	pval	0.35	0.30	0.80	**0.02**	0.19	0.15		pval	0.45	0.37	0.88	0.20	0.00	0.00
	Spread	-1.55	-1.77	**0.25**	-4.07				Spread	-1.40	0.03	**2.51**	-25.32		
	pval	0.60	0.59	0.98	**0.13**	0.00	0.00		pval	0.20	0.06	0.99	**0.00**	0.58	0.46
NOA	Low	0.42	0.17	**3.58**	-2.05			**Comp Eq Issue**	Low	-1.15	-0.67	**1.67**	-16.04		
	pval	0.26	0.15	0.89	**0.00**	0.31	0.19		pval	0.30	0.20	0.89	**0.04**	0.27	0.12
	Medium	-0.10	-0.07	**3.20**	-3.20				Medium	-1.20	-0.54	**0.93**	-13.32		
	pval	0.31	0.28	0.68	**0.00**	0.19	0.12		pval	0.34	0.24	0.81	**0.09**	0.15	0.00
	High	-0.99	-0.14	**2.07**	-5.27				High	-1.20	-0.37	0.37	-12.22		
	pval	0.50	0.44	0.99	**0.01**	0.12	0.04		pval	0.49	0.40	0.93	0.20	0.00	0.00
	Spread	2.04	2.69	**5.50**	-2.04				Spread	-6.23	-5.83	**1.14**	-33.87		
	pval	0.20	0.00	0.85	**0.00**	0.69	0.65		pval	0.59	0.60	0.97	**0.01**	0.04	0.04

Totals	**% Pval<0.1**	0.2	**Average Max R^2_{OS}**	1.93	**% Max R^2_{OS} >2**	0.38
	%Pval <0.05	0.12	**% Max R^2_{OS} >1**	0.75	**Number of Variables**	40.00

Table 8: **Out-of-sample predictability of anomalies portfolios: Δ Utility.** In the Upper Panel we report the Average, Median, Maximum and Minimum yearly percentage Δ Utility values for 30 portfolios based on characteristics sorting and 10 spread portfolios returns. The monthly out-of-sample period considered is the most recent 30% for each variable. Forecasts are based on Welch and Goyal (2008) predictors: we consider both univariate regression and all the machine learning techniques detailed in the second part of this paper which jointly consider all these predictors. In the Lower Panel we briefly summarize the results coming from the upper panel. We use bold to remark yearly Δ Utility gains above 5%.

	Δ Utility	Average	Median	Max	Min		Δ Utility	Average	Median	Max	Min
Asset Growth	Low	**7.63**	**8.43**	**11.44**	1.88	**Accruals**	Low	**13.79**	**13.81**	**15.04**	**12.45**
	Medium	**5.46**	**6.01**	**8.53**	0.59		Medium	**5.47**	**6.20**	**8.26**	0.72
	High	1.87	2.33	4.97	-1.84		High	1.70	2.08	**5.28**	-2.19
	Spread	-0.74	1.47	1.69	-5.23		Spread	0.46	-0.79	**4.36**	-0.89
Gross Prof	Low	**11.56**	**10.57**	**17.87**	4.15	**O**	Low	**9.85**	**6.81**	**18.21**	3.28
	Medium	**10.49**	**10.39**	**15.44**	4.22		Medium	**7.64**	**8.11**	**11.71**	3.22
	High	**7.38**	**7.62**	**11.95**	1.75		High	**10.42**	**10.85**	**14.07**	3.73
	Spread	-0.72	-0.79	2.25	-3.75		Spread	-0.04	0.19	1.75	-1.89
Inv to Assets	Low	**10.63**	**10.85**	**15.44**	4.72	**ROA**	Low	1.94	1.84	**5.44**	-1.29
	Medium	**6.77**	**6.92**	**10.74**	1.94		Medium	2.83	-0.31	**11.77**	-5.61
	High	**7.99**	1.26	**23.03**	-3.29		High	2.62	-0.75	**10.42**	-3.96
	Spread	-0.70	0.59	0.81	-4.54		Spread	-2.42	0.49	2.99	-9.47
Net Stock Issues	Low	4.34	4.97	**7.89**	-0.63	**Distress**	Low	0.37	0.19	3.71	-0.59
	Medium	**8.50**	**8.76**	**13.31**	1.84		Medium	4.27	-0.96	**15.99**	-7.68
	High	**10.51**	**6.44**	**21.70**	1.82		High	4.68	1.81	**14.09**	-5.01
	Spread	-0.90	0.07	0.44	-3.42		Spread	-2.89	-0.85	-0.59	-7.90
NOA	Low	**12.41**	**11.17**	**18.08**	5.94	**Comp Eq Issue**	Low	4.01	4.80	**6.98**	-0.05
	Medium	**7.24**	**7.87**	**11.24**	1.72		Medium	**7.03**	**7.37**	**10.49**	1.42
	High	**7.05**	0.47	**21.27**	-2.39		High	**7.76**	2.03	**20.51**	-3.39
	Spread	-0.29	-0.20	0.17	-2.32		Spread	-2.52	-0.53	-0.20	-8.21

Totals			
Average Total	4.84	**Number Variables**	40
Average Max	9.96		

References

Aiolfi, M. and Timmermann, A. (2006). Persistence in forecasting performance and conditional combination strategies. *Journal of Econometrics*, 135(1):31 – 53.

Bakshi, G., Panayotov, G., and Skoulakis, G. (2011). Improving the predictability of real economic activity and asset returns with forward variances inferred from option portfolios. *Journal of Financial Economics*, 100(3):475 – 495.

Barillas, F. and Shanken, J. (2018). Comparing asset pricing models. *The Journal of Finance*, 73(2):715–754.

Barone-Adesi, G., Mancini, L., and Shefrin, H. (2016). Estimating sentiment, risk aversion, and time preference from behavioral pricing kernel theory. *Working Paper, Swiss Finance Institute Research Paper No. 12-21.*

Barone-Adesi, G. and Sala, C. (2019). Testing market efficiency with the pricing kernel. *The European Journal of Finance*, 0(0):1–28.

Bollerslev, T., Todorov, V., and Xu, L. (2015). Tail risk premia and return predictability. *Journal of Financial Economics*, 118(1):113–134.

Boser, B. E., Guyon, I. M., and Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory*, pages 144–152. ACM.

Breiman, L. (1996). Bagging predictors. *Machine learning*, 24(2):123–140.

Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.

Breiman, L. and Friedman, J. H. (1985). Estimating optimal transformations for multiple regression and correlation. *Journal of the American statistical Association*, 80(391):580–598.

Campbell, S, G., and C, P. (2013). Changing risk perception and the time-varying price of risk. *Review of Asset Pricing Studies*, 3:95–132.

Campbell, J. Y. (1991). A variance decomposition for stock returns. *The Economic Journal*, 101(405):157–179.

Campbell, J. Y., Hilscher, J., and Szilagyi, J. (2008). In search of distress risk. *The Journal of Finance*, 63(6):2899–2939.

Campbell, J. Y. and Shiller, R. J. (1988). The dividend-price ratio and expectations of future dividends and discount factors. *The Review of Financial Studies*, 1(3):195–228.

Campbell, J. Y. and Thompson, S. B. (2008). Predicting excess stock returns out of sample: Can anything beat the historical average? *The Review of Financial Studies*, 21(4):1509–1531.

Campbell, Harvey, R., Liu, Y., and Zhu, H. (2016). and the cross-section of expected returns. *The Review of Financial Studies*, 29(1):5–68.

Chan, L. K. C., Jegadeesh, N., and Josef, L. (1996). Momentum strategies. *The Journal of Finance*, 51(5):1681–1713.

Chen, L., Novy-Marx, R., and Zhang, L. (2011). An alternative three-factor model. *Working Paper, Cheung Kong Graduate School of Business, Simon Business School, University of Rochester and Ohio State University - Fisher College of Busines.*

Chen, Y., Eaton, G. W., and Paye, B. S. (2018). Micro(structure) before macro? the predictive power of aggregate illiquidity for stock returns and economic activity. *Journal of Financial Economics*, 130(1):48 – 73.

Christoffersen, P. and Pan, X. N. (2017). Oil volatility risk and expected stock returns. *Journal of Banking Finance.*

Clark, T. and West, K. (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics*, 138(1):291–311.

Cochrane, J. H. (2011). Presidential address: Discount rates. *The Journal of Finance*, 66(4):1047–1108.

Cooper, M. J., Gulen, H., and Schill, M. J. (2008). Asset growth and the cross-section of stock returns. *The Journal of Finance*, 63(4):1609–1651.

Cujen, J. and Hasler, M. (2017). Why does return predictability concentrate in bad times? *The Journal of Finance*, 72(6):2717–2758.

Dangl, T. and Halling, M. (2012). Predictive regressions with time-varying coefficient. *Journal of Financial Economics*, 106(1):157–181.

Daniel, K. and Titman, S. (2006). Market reactions to tangible and intangible information. *The Journal of Finance*, 61(4):1605–1643.

de Prado, M. L. (2018). *Advances in Financial Machine Learning*. Wiley, 1 edition.

Drucker, H., Burges, C. J., Kaufman, L., Smola, A., Vapnik, V., et al. (1997). Support vector regression machines. *Advances in neural information processing systems*, 9:155–161.

Dunis, C. L., Middleton, P. W., Karathanasopolous, A., and Theofilatos, K. (2016). *Artificial Intelligence in Financial Markets: Cutting Edge Applications for Risk Management, Portfolio Optimization and Economics*. Palgrave Macmillan, 1 edition.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2):383–417.

Fama, E. F. and French, K. R. (1989). Business conditions and expected returns on stocks and bonds. *Journal of Financial Economics*, 25(1):23 – 49.

Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of financial economics*, 33(1):3–56.

Fama, E. F. and French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, 116(1):1 – 22.

Fama, E. F. and French, K. R. (2018). Choosing factors. *Journal of Financial Economics*, 128(2):234 – 252.

Feng, G., Giglio, S., and Xiu, D. F.-M. (2017). Taming the factor zoo. *Working Paper Forthcoming; Chicago Booth Research Paper No. 17-04*.

Feng, G., He, J., and Polson, N. G. (2018). Deep learning for predicting asset returns. *Working Paper*.

Feng, G., Polson, N., and Xu, J. (2019). Deep learning in asset pricing. *Working Paper*.

Ferreira, M. A. and Santa-Clara, P. (2011). Forecasting stock market returns: The sum of the parts is more than the whole. *Journal of Financial Economics*, 100(3):514 – 537.

Frazzini, A. and Pedersen, L. H. (2014). Betting against beta. *Journal of Financial Economics*, 111(1):1 – 25.

Friedman, J. H. (1991). Multivariate adaptive regression splines. *The annals of statistics*, pages 1–67.

Fuss, R., Gehrig, T., and Rindler, P. B. (2016). Changing risk perception and the time-varying price of risk. *Review of Finance*, 20(4):15491585.

Georgios, S., Thanos, V., and Konstantinos, T. (2015). Adaptive evolutionary neural networks for forecasting and trading without a datasnooping bias. *Journal of Forecasting*, 35(1):1–12.

Golez, B. and Koudijs, P. (2018). Four centuries of return predictability. *Journal of Financial Economics*, 127(2):248 – 263.

Greenwood, R. and Hanson, S. G. (2012). Share issuance and factor timing. *The Journal of Finance*, 67(2):761–798.

Gruber, L. and West, M. (2016). Gpu-accelerated bayesian learning and forecasting in simultaneous graphical dynamic linear models. *Bayesian Analysis*, 11(1):125–149.

Gu, S., Kelly, B. T., and Xiu, D. (2018). Empirical asset pricing via machine learning. *Chicago Booth Research Paper No. 18-04*.

Guidolin, M. and Timmermann, A. (2008). International asset allocation under regime switching, skew, and kurtosis preferences. *The Review of Financial Studies*, 21(2):889–935.

Han, Y., Zhou, G., and Zhu, Y. (2016). A trend factor: Any economic gains from using information over investment horizons? *Journal of Financial Economics*, 122(2):352 – 375.

Hirshleifer, D., Hou, K., Teoh, S. H., and Zhang, Y. (2004). Do investors overvalue firms with bloated balance sheets? *Journal of Accounting and Economics*, 38:297–331.

Hong, H., Torous, W., and Valkanov, R. (2007). Do industries lead stock markets? *Journal of Financial Economics*, 83(2):367 – 396.

Huang, D., Jiang, F., Tu, J., and Zhou, G. (2015). Investor sentiment aligned: A powerful predictor of stock returns. *The Review of Financial Studies*, 28(3):791–837.

Huang, D. and Kilic, M. (2018). Gold, platinum, and expected stock returns. *Journal of Financial Economics*.

Hwang, S. and Rubesam, A. (2018). Searching the factor zoo. *IESEG Working Paper Series*.

Jegadeesh, N. and Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of finance*, 48(1):65–91.

Johannes, M., Korteweg, A., and Polson, N. (2013). Sequential learning, predictability, and optimal portfolio returns. *The Journal of Finance*, 69(2):611–644.

Johnson, M. C. and West, M. (2018). Bayesian predictive synthesis: Forecast calibration and combination. *Working Paper*.

Karatahansopoulos, A., Sermpinis, G., Laws, J., and Dunis, C. (2014). Modelling and trading the greek stock market with gene expression and genetic programing algorithms. *Journal of Forecasting*, 33(8):596–610.

Karathanasopoulos, A., Theofilatos, K. A., Sermpinis, G., Dunis, C., Mitra, S., and Stasinakis, C. (2015). Stock market prediction using evolutionary support vector machines: an application to the ASE20 index. *The European Journal of Finance*, 22(12):1145–1163.

Kelly, B. and Pruitt, S. (2013). Market expectations in the cross-section of present values. *The Journal of Finance*, 68(5):1721–1756.

Kelly, B. and Pruitt, S. (2015). The three-pass regression filter: A new approach to forecasting using many predictors. *Journal of Econometrics*, 186(2):294–316.

Kelly, B. T., Pruitt, S., and Su, Y. (2018). Characteristics are covariances: A unified model of risk and return. *SSRN working paper*.

Kozak, S., Nagel, S., and Santosh, S. (2017a). Shrinking the cross section. *SSRN Working Paper*.

Kozak, S., Nagel, S., and Santosh, S. (2017b). Tree-based conditional portfolio sorts: The relation between past and future stock returns. *Working Paper*.

Kuhn, M. and Johnson, K. (2013). *Applied Predictive Modeling*. Springer, 1 edition.

Lettau, M. and Van Nieuwerburgh, S. (2007). Reconciling the Return Predictability Evidence . *The Review of Financial Studies*, 21(4):1607–1652.

Li, J. and Tsiakas, I. (2017). Equity premium prediction: The role of economic and statistical constraints. *Journal of Financial Markets*, 36:56 – 75.

Lin, Q. (2018). Technical analysis and stock return predictability: An aligned approach. *Journal of Financial Markets*, 38:103 – 123.

Lo, A. W. (2004). The adaptive markets hypothesis. *The Journal of Portfolio Management*, 30(5):15–29.

Lo, A. W. (2012). Adaptive markets and the new world order. *Financial Analysts Journal*, 68(2):18–29.

Lo, A. W., Mamaysky, H., and Wang, J. (2002). Foundations of technical analysis: Computational algorithms, statistical inference, and empirical implementation. *The Journal of Finance*, 55(4):1705–1765.

Loughran, T. and Ritter, J. R. (1995). The new issues puzzle. *The Journal of finance*, 50(1):23–51.

Ludvigson, S. C. and Ng, S. (2007). The empirical riskreturn relation: A factor analysis approach. *Journal of Financial Economics*, 83(1):171 – 222.

Luo, X. and Zhang, J. E. (2017). Expected stock returns and forward variance. *Journal of Financial Markets*, 34:95 – 117.

Martin, I. W. R. and Wagner, C. (2019). What is the expected return on a stock? *Journal of Finance*, 0(0):1–28.

Mclean, R. D. and Pontiff, J. (2015). Does academic research destroy stock return predictability? *The Journal of Finance*, 71(1):5–32.

Messmer, M. and Audrino, F. (2017). The (adaptive) lasso in the zoo. *Working Paper*.

Miller, W. T., Werbos, P. J., and Sutton, R. S. (1995). *Neural networks for control*. MIT press.

Minsky, M. and Papert, S. (1969). Perceptrons. *MIT press*.

Montavon, G., Lapuschkin, S., Binder, A., Samek, W., and Mller, K.-R. (2017). Explaining nonlinear classification decisions with deep taylor decomposition. *Pattern Recognition*, 65:211 – 222.

Montavon, G., Samek, W., and Mller, K.-R. (2018). Methods for interpreting and understanding deep neural networks. *Digital Signal Processing*, 73:1 – 15.

Moritz, B. and Zimmermann, T. (2016). Tree-based conditional portfolio sorts: The relation between past and future stock returns. *Working Paper*.

Nakajima, J. and West, M. (2013). Dynamic factor volatility modeling: A bayesian latent threshold approach. *Journal of Financial Econometrics*, 11(1):116–153.

Neely, C. J., Rapach, D., Tu, J., and Zhou, G. (2013). Forecasting the equity risk premium: The role of technical indicators. *Federal Reserve Bank of St. Louis Working Paper No. 2010-008H*.

Novy-Marx, R. (2013). The other side of value: The gross profitability premium. *Journal of Financial Economics*, 108(1):1–28.

Ohlson, J. A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*, pages 109–131.

Pettenuzzo, D., Timmermann, A., and Valkanov, R. (2014). Forecasting stock returns under economic constraints. *Journal of Financial Economics*, 114(3):517 – 553.

Rapach, D. E., Ringgenberg, M., and Zhou, G. (2016). Short interest and aggregate stock returns. *Journal of Financial Economics*, 121(1):46–65.

Rapach, D. E., Strauss, J. K., and Zhou, G. (2010). Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. *The Review of Financial Studies*, 23(2):821.

Rasekhschaffe, K. and Jones, R. (2019). Machine learning for stock selection. *Financial Analysts Journal*.

Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Statist.*, 6(2):461–464.

Sermpinis, G., Dunis, C., Laws, J., and Stasinakis, C. (2012). Forecasting and trading the eur/usd exchange rate with stochastic neural network combination and time-varying leverage. *Decision Support Systems*, 54(1):316 – 329.

Shefrin, H. (2008). *A behavioral approach to asset pricing*. Academic Press.

Shefrin, H. and Statman, M. (1994). Behavioral capital asset pricing theory. *Journal of Financial and Quantitative Analysis*, 29(3):323–349.

Shrikumar, A., Greenside, P., Shcherbina, A., and Kundaje, A. (2016). Not just a black box: Learning important features through propagating activation differences. *Working Paper*.

Sloan, R. (1996). Do stock prices fully reflect information in accruals and cash flows about future earnings?(digest summary). *Accounting Review*, 71(3):289–315.

Stambaugh, R. F., Yu, J., and Yuan, Y. (2015). Arbitrage asymmetry and the idiosyncratic volatility puzzle. *The Journal of Finance*, 70(5):1903–1948.

Stambaugh, R. F. and Yuan, Y. (2017). Mispricing factors. *The Review of Financial Studies*, 30(4):1270–1315.

Stock, J. H. and Watson, M. W. (2004). Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting*, 23(6):405–430.

Strauss, J. and Detzel, A. (2017). Combination Return Forecasts and Portfolio Allocation with the Cross-Section of Book-to-Market Ratios*. *Review of Finance*, 22(5):1949–1973.

Titman, S., Wei, K.-C., and Xie, F. (2003). Capital investments and stock returns. *Working Paper, Hong Kong University of Science Technology, Southern Connecticut State University and University of Texas at Austin.*

Van Binsberg, J. and Koijen, R. S. (2010). Predictive regressions: A present-value approach. *The Journal of Finance*, 65(4):1439–1471.

Wei Koh, P. and Liang, P. (2017). Understanding black-box predictions via influence functions. *Working paper.*

Welch, I. and Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. *The Review of Financial Studies*, 21(4):1455–1508.

Wen, Q. (2019). Asset Growth and Stock Market Returns: A Time-Series Analysis*. *Review of Finance.*

Wright, S. J., Nowak, R. D., and Figueiredo, M. A. T. (2009). Sparse reconstruction by separable approximation. *Trans. Sig. Proc.*, 57(7):2479–2493.

Yi, Z. Z., Meng, X., and Mike, W. (2016). Dynamic dependence networks: Financial time series forecasting and portfolio decisions. *Applied Stochastic Models in Business and Industry*, 32(3):311–332.

Zhou, X., Nakajima, J., and West, M. (2014). Bayesian forecasting and portfolio decisions using dynamic dependent sparse factor models. *International Journal of Forecasting*, 30(4):963 – 980.

Zhu, X. (2015). Tug-of-war: Time-varying predictability of stock returns and dividend growth*. *Review of Finance*, 19(6):2317–2358.

8 Appendix

8.1 Predictive Models: Detailed Methodologies

8.1.1 Basic linear models

The Kitchen Sink Regression is a simple OLS multivariate regression which includes all the predictors at once. The estimation is performed employing all predictors up to time t-1 (the last available information) to perform the parameter estimation. After that, we use the estimated parameters to make inference for time t+1 employing regressors values at time t. In formulas this can be summarized in a two step procedure:

$$R_t = \alpha + \beta X_{t-1} + \epsilon_t \tag{6}$$

where R is the t*1 vector of the S&P500 returns and X is the t*N matrix of the N predictors considered in the analysis.

$$\hat{r}_{t+1} = \hat{\alpha}_t + \hat{\beta}_t X_t \tag{7}$$

where $\hat{r}_{t+1}$ is the univariate forecast produced by the model, $\hat{\alpha}_t$, $\hat{\beta}_t$ are the coefficient estimated in the previous step employing data up to time t-1, and X_t is the 1*N vector of predictors at time t. For univariate model N (the number of predictors) is equal to 1.

8.1.2 Combination Forecasts

Combination forecasts are common methodologies employed in the literature (Rapach et al. (2010), Aiolfi and Timmermann (2006), Strauss and Detzel (2017)). The DMSPE approach is based on a three-stages estimation.

1. At first for each date t, we run a separate univariate regression for each regressor, x_{t-1}, on the equity premium at time t using all data available up to that date.

$$R_t = \alpha + \beta x_{i,t-1} + \epsilon_t \tag{8}$$

2. After that, each univariate OLS model previously estimated is employed with predictors available at time x_t to make inference on the equity premium for the subsequent period, $\hat{R}_{t+1}$

$$\hat{R}_{t+1} = \hat{\alpha}_t + \hat{\beta}_t x_t \tag{9}$$

3. Finally, we combine the forecasts generated by univariate regressions via combination forecasts methods.

$$\hat{R}_{t+1,Comb} = \sum_{i=1}^{N} w_{i,t}\hat{R}_{t+1} \tag{10}$$

In the Pooled-DMSPE approach we computes the weights for the third step in the following way:

$$w_{i,t} = \frac{\phi_{i,t}^{-1}}{\sum_{k=1}^{K}\phi_{k,t}^{-1}} \tag{11}$$

where

$$\phi_{i,t} = \sum_{s=m}^{t-1}\theta^{t-1-s}(r_{s+1} - \hat{r}_{i,s+1}) \tag{12}$$

θ is a discount factor (equal to 0.5 in this study), m+1 is the start of the holdout period and K is the number of past periods considered to compute the weights (K=13 in this paper). The DMSPE method thus assigns greater weight to individual forecasts that had better forecasting performance in terms of lower mean-squared prediction errors.
The Pooled-Median, instead of using equation (5), simply employs the median of the univariate regression forecasts from equation (4).

8.1.3 Sum-of-the-Parts Method

The Sum-of-the-Parts Method has been proposed by Ferreira and Santa-Clara (2011). The authors start decomposing returns in the following manner:

$$R_{t+1} = \frac{P_{t+1} + D_{t+1}}{P_t} = CG_{t+1} + DY_{t+1} \tag{13}$$

where P_t is the stock price, D_t is the dividend, $CG_{t+1} = \frac{P_{t+1}}{P_t}$ is the gross capital gain, and $DY_{t+1} = \frac{D_{t+1}}{P_t}$ is the dividend yield. After that, the gross capital gain can be expressed as

$$CG_{t+1} = \frac{\frac{P_{t+1}}{E_{t+1}}}{\frac{P_t}{E_t}}\frac{E_{t+1}}{E_t} = \frac{M_{t+1}}{M_t}\frac{E_{t+1}}{E_t} = GM_{t+1}GE_{t+1} \tag{14}$$

where E_t denotes earnings, $M_t = \frac{P_t}{E_t}$ is the price-earnings multiple, $GM_{t+1} = \frac{M_{t+1}}{M_t}$, is the gross growth rate of the price-earnings multiple (earnings), and $GE_{t+1} = \frac{E_{t+1}}{E_t}$. Now the dividend yield can be written as

$$DY_{t+1} = \frac{D_{t+1}}{P_{t+1}}\frac{P_{t+1}}{P_t} = DP_{t+1}GM_{t+1}GE_{t+1} \tag{15}$$

where $\frac{D_t}{P_t}$ is the dividend-price ratio. Based on these results the gross return becomes

$$R_{t+1} = GM_{t+1}GE_{t+1}(1 + DP_{t+1}) \tag{16}$$

which for the log return can be expressed as

$$log(R_{t+1}) = gm_{t+1} + ge_{t+1} + dp_{t+1} \tag{17}$$

The authors argue that, since price-earnings multiples and dividend-price ratios are highly persistent and nearly random walks, reasonable forecasts for gm_{t+1} and dp_{t+1} based on information available at time t are zero and dp_t. Finally, a 20-year moving average of log earnings growth through t ge_t^{20}, is employed as a forecast of ge_{t+1}. The sum-of-the-parts equity premium forecast is then given by

$$\hat{r}_{t+1}^{SOP} = \bar{ge}_t^{20} + dp_t - r_{f,t+1} \tag{18}$$

where $r_{f,t+1}$ is the log risk-free rate for time t+1, which is known at the end of time t.

8.1.4 Multivariate Adaptive Regression Splines and Support Vector Machines for Regression

Given a set of predictors the MARS model (Friedman (1991)) selects and breaks a predictor into two groups and models linear relationships between the predictor and the outcome in each group. To determine the cut point each data point for each predictor is evaluated as a candidate cut-point by creating a linear regression model with the candidate features, and the corresponding model error is calculated. The predictor/cut point combination that achieves the smallest error is then used for the model. After the initial model is created with the first two features, the model conducts another exhaustive search to find the next set of features that, given the initial set, yield the best model fit. This process continues until a stopping point is reached. Once the full set of features has been created, the algorithm sequentially removes individual features that do not contribute significantly to the model equation. This pruning procedure assesses each predictor variable and estimates how much the error rate was decreased by including it in the model. MARS builds models of the form:

$$\hat{f}(x) = \sum_{i=1}^{m} c_i B_i(x) \tag{19}$$

where c_i is a fix coefficient and B_i can be equal to 1 or to a hinge function (a hinge function has the form max(0, x-const) or max(0, const-x)) or a product of hinge functions.

Our implementation of the algorithm builds the model in two phases: forward selection and backward deletion. In the forward phase, the algorithm starts with a model consisting of just the intercept term and iteratively adds reflected pairs of basis functions giving the largest reduction of training error (Mean Squared Error). We set the maximum number of basis functions to min(200, max(20,2d))+1, where d is the number of input variables. We do not allow for self-interaction. We impose no penalty for adding a new variable to a model in the forward phase, and we employ hinge functions only. The forward phase is executed until adding a new basis function changes R^2 by less than 1e-4.

At the end of the forward phase we have a large model which over-fits the data, and so a backward deletion phase is engaged. In the backward phase, the model is simplified by deleting one least important basis function (i.e., deletion of which reduces training error the least) at a time until the model has only the intercept term. At the end of the backward phase, from those best models of each size, the one with the lowest Generalized Cross-Validation (GCV) is selected and outputted as the final one. GCV, as an estimator for Prediction Mean Squared Error, for a MARS model is calculated as follows:

$$CVG = \frac{MSE_{train}}{(1 - \frac{enp}{n})^2} \tag{20}$$

where MSE_{train} is the Mean Squared Error of the model in the training data, n is the number of observations in the training data, and *enp* is the effective number of parameters:

$$enp = k + c * (k+1)/2 \tag{21}$$

where k is the number of basis functions in the model (including the intercept term), and c=3 is the Generalized Cross-Validation (GCV) penalty. We impose no further constraints on the Maximum number of basis functions (including the intercept term) in the final pruned model[26].

Once the model is built we perform variable importance assessment. The criterion counts the number of model subsets that include the variable. Where by "subsets" we mean the subsets of terms generated by the pruning pass. There is one subset for each model size (from 1 to the size of the selected model) and the subset is the best set of terms for that model size. Obviously, only subsets that are smaller than or equal in size to the final model are used for estimating variable importance. We select only variables with a score bigger than 12. After that, we use the selected variables to estimate a machine vector

[26] To boost computational performance, and following Friedman (1991), we employ piecewise-cubic modelling for the final model only after both the forward and the backward phases.

regression model.

The intuition of SVM for regression is to modify the traditional simple linear regression regularized error function

$$\frac{1}{2}\sum_{n=1}^{N}(y_n - t_n)^2 + \frac{\lambda}{2}||w||^2 \tag{22}$$

by introducing an ϵ insensitive error function.

$$E_\epsilon(y(x) - t) = \begin{cases} 0 & if \quad |y(x) - t| < \epsilon \\ |y(x) - t| - \epsilon & otherwise \end{cases} \tag{23}$$

This implies that we minimize a regularized error function given by

$$C\sum_{n=1}^{N} E_\epsilon(y(x_n) - t_n) + \frac{1}{2}||w||^2 \tag{24}$$

where C is a regularization parameter.

Now for each data point x_n, we now need two slack variables $\xi_n \geq 0$ and $\hat{\xi}_n > 0$, where $\xi_n > 0$ corresponds to a point for which $t_n > y(x_n) + \epsilon$ and $\hat{\xi}_n < 0$ correspond to a point for which $t_n < y(x_n) + \epsilon$. Consequently, a target point lies inside the ϵ tube whether $y_n - \epsilon \leq t_n \leq y_n + \epsilon$ where $y_n = y(x_n)$. The introduction of the two slack variables allows points to lie outside the tube provided the slack variables are different from zero:

$$t_n \leq y(x_n) + \epsilon + \xi_n \quad and \quad t_n \geq y(x_n) - \epsilon - \hat{\xi}_n \tag{25}$$

This implies that the error function for support vector regression can then be written as

$$C\sum_{n=1}^{N}(\xi_n + \hat{\xi}_n) + \frac{1}{2}||w||^2 \tag{26}$$

which should be minimized subject to the constraints $\xi_n \geq 0$ and $\hat{\xi}_n \geq 0$ plus the conditions $t_n \leq y(x_n) + \epsilon + \xi_n$ and $t_n \geq y(x_n) - \epsilon - \hat{\xi}_n$. Consequently, the problem can be solved optimizing the Lagrangian with multipliers $a_n \geq 0$, $\hat{a}_n \geq 0$, $\mu_n \geq 0$ and $\hat{\mu}_n \geq 0$

$$\begin{aligned} L = C\sum_{n=1}^{N}(\xi_n + \hat{\xi}_n) + \frac{1}{2}||w||^2 - \sum_{n=1}^{N}(\mu_n\xi_n + \hat{\mu}_n\hat{\xi}_n) \\ - \sum_{n=1}^{N} a_n(\epsilon + \xi_n + y_n - t_n) - \sum_{n=1}^{N} \hat{a}_n(\epsilon + \hat{\xi}_n - y_n + t_n) \end{aligned} \tag{27}$$

Computing the partial derivatives and replacing gives

$$\tilde{L}(a, \hat{a}) = -\frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N}(a_n - \hat{a}_n)(a_m - \hat{a}_m)k(x_n, x_m) - \epsilon\sum_{n=1}^{N}(a_n + \hat{a}_n) + \sum_{n=1}^{N}(a_n - \hat{a}_n) * t_n \tag{28}$$

where $k(x, x') = \phi(x)^T \phi(x')$ is the kernel.
Replacing $w = \sum_{n=1}^{N}(a_n - \hat{a}_n)\phi(x_n)$ in the general case $y(x) = w^T\phi(x) + b$ where $\phi(x)$ denotes a fixed feature-space transformation, $\phi(x) * \phi(x) = k(x, x_n)$, and b is the bias parameter, we see that predictions can be made using

$$y(x) = \sum_{n=1}^{N}(a_n - \hat{a}_n)k(x, x_n) + b \tag{29}$$

We implement the regularized support vector machines regression presented above in the following manner. The half width of the epsilon-insensitive band is set equal to the ratio of the interquartile range of the independent variable distribution and the scalar value 1.349. The regularization Lambda is set equal to one divided the training sample size. The objective function minimization technique chosen is SpaRSA (sparse reconstruction by separable approximation optimization, Wright et al. (2009)). Initial estimates of regression coefficients are all set to zero except the bias one which is initially fixed to the weighted median of the dependent variable in the training set. The criteria for convergence during the optimization process are[27]:

- Relative tolerance on linear coefficients and bias term: 1e-4
- Absolute gradient tolerance: 1e-6
- Size of history buffer for Hessian approximation: 15
- Maximal number of optimization iterations: 1000

For each date t, the model is estimated with predictors data up to t-1. Then the values of the regressors at time t are employed to make inference for date t+1.

8.1.5 Diffusion Indices and Partial Least Squares

The diffusion index approach assumes a latent factor model structure for the potential predictors:

$$x_{i,t} = \lambda_i' f_t + e_{i,t} \tag{30}$$

with (i=1,..., K) and f_t is a q-vector of latent factors, λ_i is a q-vector of factor loadings, and $e_{i,t}$ is a zero-mean disturbance term. Co-movements in the predictors are primarily governed by movements in the small number of factors (the number of factors is much

[27]Further details on the optimization procedure can be found looking at the details of the Matlab function "fitrlinear"

smaller than the number of predictors). The latent factors can be consistently estimated by principal components. To implement this approach we started standardizing all the predictors (standard deviation of 1 and zero mean). After that for each date t, we compute the first principal component employing all data available up to t-1. The first principal component is then employed as a regressor to estimate a support vector machine regression. Finally, the support vector machine regression previously estimated with data up to t-1 and the value f_t of the first principal component are used to make inference for time t+1. The approach employed for the estimation of the support vector machine regression is the same explained in the previous subsection: Multivariate Adaptive Regression Splines and Support Vector Machines for Regression.

The approach followed for the PLS is similar. At first, the PLS predictor is estimated following the approach of Kelly and Pruitt (2015) and Kelly and Pruitt (2013):

$$Y^{PLS} = XJ_NX'J_TR(R'J_TXJ_NX'J_TR)^{-1}R'J_TR \tag{31}$$

where X denotes the T x N matrix of predictors, $X = (x'_1, x'_2, ..., x'_T)$, and R denotes the T x 1 vector of excess stock returns as $R = (R_2, ..., R_{T+1})'$. The matrices J_T and J_N, $J_T = I_T - \frac{1}{T}i_Ti'_T$ and $J_N = I_N - \frac{1}{T}i_Ni'_N$ enter the formula because each regression is run with a constant. I_T is a T-dimensional identity matrix, and i_T is a T-vector of ones. The PLS predictor is then employed to estimate a univariate support vector machine regression. Finally, the support vector machine regression previously estimated with data up to t-1 and the value Y_t^{PLS} of the PLS predictor are used to make inference for time t+1. The approach employed for the estimation of the support vector machine regression is the same explained in the previous subsection: Multivariate Adaptive Regression Splines and Support Vector Machines for Regression.

8.1.6 Regression Trees and Regression Forest

Classification and regression trees or CART models (Breiman and Friedman (1985)), also called decision trees are defined by recursively partitioning the input space, and defining a local model in each resulting region of the input space.

$$f(x) = E[y|x] = \sum_{m=1}^{M} w_m I(x \in R_m) = \sum_{m=1}^{M} w_m \phi(x, v_m) \tag{32}$$

where R_m is the mth region, w_m is the mean response in this region, and v_m encodes the choice of the variables to split on, and the threshold values, on the path from the root to the mth leaf. Consequently, a CART is just an adaptive basis-function model, where the basic functions define the regions, and the weights specify the response value in each region. The split function chooses the best feature (j), and the best value for that feature (t), as follows:

$$(j^*, t^*) = arg \min_{j \in (1,\dots,D)} \min_{t \in T_j} cost(x_i, y_i : x_{i,j} \leq t) + cost(x_i, y_i : x_{i,j} > t) \tag{33}$$

Tree regressions extend the idea of CART but terminal nodes instead of providing the simple average employ a linear model to predict the outcome. Finally, Regression Forest follows an extension of the tree regression based on bootstrapping. The approach of the random forest consists of forecasting through the average (mean) of regression trees generated by bootstrapping the original data. At first, a number (m) of wanted regression trees is fixed. For each m a bootstrap sample of the original data is generated, and with them, trees regressions are trained. This approach introduces a change in the building of each tree: for each split, the model randomly selects k (less than P) of the total original predictors (P) and partitions the data selecting the best predictor among the k predictors. To calibrate this model we follow the suggestions of Kuhn and Johnson (2013). First, all trees are decision trees with binary splits for regression. Second, only 2% of data are employed (with replacement) for building each tree. After that, the number of predictor or feature variables to select at random for each decision split is set to three. We grow the tree using MSE (mean squared error) as the splitting criterion. The stopping criteria for the building of the tree are:

- The maximal number of decision splits (or branch nodes) per tree is equal to the number of observations-1
- Each leaf must have at least five observations
- Each splitting node in the tree must have at least ten observations.

No pruning is performed after the creation of the trees, and no cost function is imposed on errors. Finally, the forecasts generated by each tree are the result of the forecasts coming from leaves only, not from a weighted average of leaves and nodes. This procedure is employed to create 1000 different trees. Once every tree is grown we compute the average prediction from all individual trees and this mean is our forecast of market return at month t+1.

We repeat this procedure for each date t: the model is estimated with predictors up to t-1, then the values of the predictors at time t and the previously estimated parameters are employed to make inference for t+1.

8.1.7 SIC - Lasso Support Vector Machine

The joint employment of all the available predictors is likely to give rise to severe multicollinearity and poor out-of-sample performance. Consequently, employing variable selection is likely to boost the performance of the predictive model. Following this intuition, we consider two separate model selection approaches, and subsequently, we make use of the selected variables into a Support Vector Machine regression model. The first model selection approach considered is the Schwartz Information Criterion (SIC)(Schwarz (1978)).

We employ the SIC, imposing a maximum of 2 predictors for the model selection. For each date t, we use all data available up to that moment, we consider all individual regressors and all possible combinations among two regressors, and we compute the related SIC values

$$log(SIC) = log\left(\frac{SSR}{T}\right) + k * \frac{log(T)}{T} \tag{34}$$

where T is the number of observations, k is the number of predictors and SSR is the sum of squared residuals. After that, for each date t, we pick the model with the lowest SIC. Subsequently, we use the predictors of the chosen model to estimate a support vector machine regression model. Finally, we employ it to make inference using the values of predictors at time t to forecast the S&P500 returns at time t+1.

The alternative approach which we employ for model selection is Lasso. At each time t, we run a 10-fold Cross-validated Lasso.

$$\min_{\beta} RSS + \lambda \sum_{j=1}^{N} |\beta_j| \tag{35}$$

where N is the number of regressors, λ is the Lagrange multiplier, RSS is the sum of squared residuals. The value of lambda selected is the 95^{th} higher from a default geometric sequence of 100 values, with only the largest able to produce a model which exclude all predictors.

After that, the predictors selected by Lasso are employed to estimate the Linear Support Vector Machine. Finally, we employ it to make inference using the values of predictors at time t to forecast the S&P500 returns at time t+1.

8.1.8 Ensemble of Neural Networks

Feed-forward Network functions are extensions of classical models for regression and classification, which are based on linear combinations of fixed nonlinear basis functions $\phi(x)$ and take the form

$$y(x,w) = f(\sum_{j=1}^{M} w_j \phi_j(x)) \tag{36}$$

here f(.) is a nonlinear activation function in the case of classification and is the identity in the event of regression. Neural networks use basis functions that follow the same form so that each basis function is itself a nonlinear function of a linear combination of the inputs, where the coefficients in the linear combination are adaptive parameters. Consequently, the basic neural network model can be described as a series of functional transformations. At first we construct M linear combinations of the input variables $x_1, x_2, \ldots, x_D$ in the form

$$a_j = \sum_{i=1}^{D} w_{ji}^{(1)} x_i + w_{j,0}^{(1)} \tag{37}$$

where j=1,..., M and the superscript (1) indicates that the corresponding parameters are in the first layer of the network. The quantities a_j are known as activations. Each of them is then transformed using a differentiable, nonlinear activation function h(.) to give

$$z_j = h(a_j) \tag{38}$$

These quantities correspond to the outputs of the basis function y(x,w) above that in the context of neural networks are called hidden units. In our approach, the non-linear functions h(.) are sigmoid. Finally, these values are again linearly combined to give output unit activations.

$$a_k = \sum_{j=1}^{M} w_{kj}^{(2)} z_j + w_{k0}^{(2)} \tag{39}$$

where k=1,...,K, and K is the total number of outputs. We can combine these various stages to give the overall network function that takes the form

$$y_k(x,k) = \sum_{j=1}^{M} w_{kj}^{(2)} h(\sum_{i=1}^{D} w_{ji}^{(1)} x_i + w_{j,0}^{(1)}) + w_{k0}^{(2)} \tag{40}$$

Thus, the neural network model is simply a nonlinear function from a set of input variables x_i to a set of output variables y_k controlled by a vector w of adjustable parameters.

Our approach involves a preliminary variable selection step. Consequently, only the 4 variables which up to time t have the highest cumulated R^2_{OS} value in univariate predictive regressions are subsequently employed for the estimation of the neural networks. Our neural networks have a structure composed of six layers in which each higher layer has half the number of neurons of the subsequent one, and the first layer has 32 neurons.

Insert Figure 2

Inputs are connected to all the neurons of the first and fourth layer. All neurons of one layer are fully connected with the neurons of the subsequent layer. To train the network, we minimize the Mean Absolute Error changing the weights of the network. Training is performed through the Resilient backpropagation algorithm. To avoid overfitting issues we adopt the following procedures:

- We estimate an ensemble of 100 networks with different initialization points for parameters.
- We employ the Early Stopping approach.
- We adopt regularization.
- Before the training of each network we randomly divide the data available into three parts: training sample (60%), validation sample (30%) and test sample (10%).
- We include a network in the ensemble only whether it generates an R^2 above 20% in the training sample and 25% for the validation one.

After the estimation of the ensemble, we use the most updated predictors available at time t to forecast the equity premium at time t+1. Finally, we employ the median, and 40^{th} percentile forecasts generated by the ensemble (Neural Networks Median and Neural Networks 40^{th} in the Tables).

8.2 Toolboxes Employed

The making of this paper leveraged on many libraries. We list them both because we want to help the replicability of our results and because we want to express our genuine gratitude for all the people who worked to build and maintain them. The current paper makes use of Matlab only, and consequently, all the libraries which we will list are in this language. In detail the libraries employed are:

- The Statistic and Machine Learning Toolbox and the Deep Learning Toolbox of Matlab.

- The Optimization Toolbox and the Financial toolbox of Matlab.
- The ARESLab Toolbox by Gints Jekabsons.
- The website of by Professor Guofu Zhou
- The website of by Professor Grigory Vilkov
- The website of Attilio Meucci

8.3 Additional Tables

In the following pages we report the tables, which for brevity have been omitted from the main text, these include:

- Summary statistics for the Welch and Goyal (2008) predictors (A1).
- Robustness checks for the out-of-sample predictability of the S&P500 for different time horizons: 2001:1-2017:12, 2006:1-2017:12, 2011:1-2017:12 (A2-A4)
- The detail of the monthly out-of-sample predictability (1986:1-2016:12) both in term of R^2_{OS} and Δ Utility for:
  1. six double-sorted portfolios of French: on the basis of Size and the Book to Market ratio (A5-A8)
  2. six double-sorted portfolios of French: on the basis of Size and Momentum (A9-A12).
- The out-of-sample Δ Utility for the factor spreads (A13-A14).

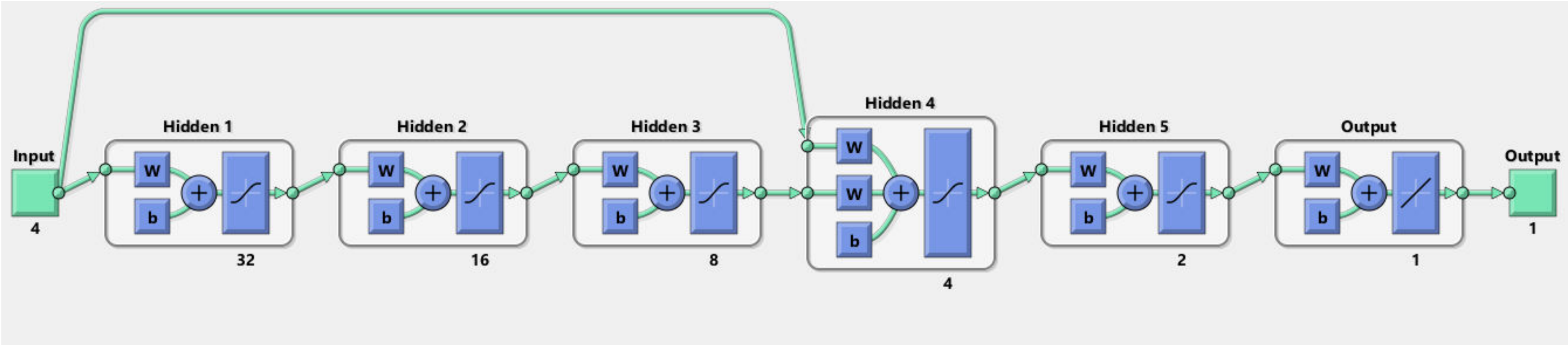


Figure 2: **Neural Network Structure**: 6 layers, 4 inputs and 1 output (the equity premium forecast). Activation all activation functions are sigmoids with the exception of the last one which is linear. Inputs are connected with the first and the fourth layer only. The number of neurons for each layer are 32,16,8,4,2 and 1. Training occurs through Resilient Backpropagation.

Table A1: **Welch-Goyal predictors: Summary Statistics**. In the upper panel we report the correlation matrix for the deltas of the W-G predictors. Correlations higher than 0.5 are reported in red while negative ones in blue. In the lower panel for each predictor we estimate the autoregressive coefficients up to the sixth lag and we report the related t-statistic.

Correlation	**DP**	**DY**	**EP**	**DE**	**SVAR**	**BM**	**NTIS**	**TBL**	**LTY**	**LTR**	**TMS**	**DFY**	**DFR**	**INFL lag**
DP	1.00													
DY	0.11	1.00												
EP	0.76	0.00	1.00											
DE	0.10	0.14	-0.57	1.00										
SVAR	0.23	0.01	0.20	-0.01	1.00									
BM	0.81	0.10	0.66	0.01	0.13	1.00								
NTIS	0.05	-0.12	0.05	-0.01	0.06	-0.03	1.00							
TBL	0.07	-0.02	0.11	-0.08	0.04	0.06	0.01	1.00						
LTY	0.10	-0.07	0.14	-0.09	0.00	0.12	0.01	0.34	1.00					
LTR	-0.02	0.12	-0.05	0.04	0.04	-0.05	-0.01	0.02	-0.65	1.00				
TMS	-0.01	-0.03	-0.02	0.02	-0.04	0.02	-0.01	-0.78	0.33	-0.45	1.00			
DFY	0.27	0.40	0.13	0.14	0.07	0.33	-0.14	-0.14	-0.14	0.04	0.05	1.00		
DFR	-0.07	0.02	-0.07	0.01	-0.05	-0.06	0.02	-0.01	0.27	-0.53	0.19	-0.01	1.00	
INFL lag	0.01	-0.05	0.02	-0.01	-0.05	-0.01	0.03	0.04	0.04	0.01	-0.01	-0.07	-0.03	1.00

Coefficients	**DP**	**DY**	**EP**	**DE**	**SVAR**	**BM**	**NTIS**	**TBL**	**LTY**	**LTR**	**TMS**	**DFY**	**DFR**	**INFL lag**
AR1	0.10	0.11	0.25	0.69	-0.46	0.19	0.13	0.38	0.07	-0.80	0.10	0.21	-0.98	-0.59
AR2	0.00	-0.01	0.08	0.10	-0.44	-0.10	0.02	-0.19	-0.08	-0.72	-0.10	-0.07	-0.88	-0.47
AR3	-0.09	-0.09	0.00	0.10	-0.33	-0.17	-0.04	0.03	-0.07	-0.58	0.01	-0.15	-0.69	-0.37
AR4	0.05	0.04	0.05	-0.15	-0.24	0.04	0.07	-0.07	0.02	-0.38	-0.05	-0.06	-0.48	-0.20
AR5	0.08	0.08	0.05	0.00	-0.20	0.09	0.11	0.11	0.02	-0.25	0.00	0.02	-0.26	-0.20
AR6	-0.05	-0.06	-0.07	-0.03	-0.09	-0.10	0.02	-0.22	0.02	-0.09	-0.09	0.00	-0.08	-0.15
t-stat	**DP**	**DY**	**EP**	**DE**	**SVAR**	**BM**	**NTIS**	**TBL**	**LTY**	**LTR**	**TMS**	**DFY**	**DFR**	**INFL lag**
AR1	5.54	5.79	14.45	64.42	-55.03	19.23	6.63	39.57	4.02	-37.68	7.81	15.12	-53.13	-28.19
AR2	-0.09	-0.30	4.03	5.89	-22.77	-7.34	0.88	-13.62	-4.46	-29.20	-4.46	-3.73	-30.60	-19.83
AR3	-4.40	-4.41	-0.13	11.05	-18.74	-13.69	-1.98	1.79	-4.32	-18.69	0.38	-10.48	-23.85	-14.37
AR4	2.38	1.65	1.95	-10.46	-13.45	2.45	3.66	-4.74	0.91	-11.31	-2.52	-6.25	-15.48	-7.10
AR5	3.62	3.27	1.95	0.14	-10.70	4.96	5.93	9.21	0.77	-8.15	0.23	0.86	-9.94	-7.62
AR6	-2.18	-2.71	-2.90	-1.48	-4.93	-6.31	0.85	-19.71	1.12	-3.80	-6.67	0.22	-4.07	-7.39

Table A2: **Monthly equity premium out-of-sample forecasting results for individual forecasts, and machine learning methods**. The R^2_{OS} is the Campbell Thompson (2008) out-of-sample R^2 statistic. Statistical significance for the R^2_{OS} statistic is based on the p-value for the Clark and West (2007) out-of-sample MPSE-adjusted statistic; the statistic corresponds to a one-sided test of the null hypothesis that the competing forecasting model has equal expected square prediction error relative to the historical average benchmark forecasting model against the alternative hypothesis that the competing forecasting model has a lower expected square prediction error than the historical average benchmark forecasting model. The results refer to monthly forecasts for the out-of-sample period 2001:01-2017:12. For predictions based on univariate forecasts the restrictions are the ones suggested by Campbell and Thompson (2008) while for the machine learning models when equity premium forecasts are negative they are replaced with zero. **Bold** indicates at least a significance level above 5%.

Standard	**2001-2017**		**Restricted**	**2001-2017**	
Predictor	$R^2_{OS}(\%)$	**pval**	**Predictor**	$R^2_{OS}(\%)$	**pval**
DP	0.13	0.20	DP	-0.02	0.24
DY	0.17	0.17	DY	-0.11	0.25
EP	-0.88	0.28	EP	1.23	0.08
DE	-1.34	0.69	DE	-0.31	0.51
SVAR	1.06	0.10	SVAR	0.73	0.10
BM	-0.10	0.24	BM	0.00	0.24
NTIS	-3.53	0.87	NTIS	-3.53	0.87
TBL	0.21	0.25	TBL	0.21	0.25
LTY	**0.49**	**0.03**	LTY	**0.49**	**0.03**
LTR	-0.01	0.34	LTR	-0.17	0.40
TMS	-1.15	0.76	TMS	-1.15	0.76
DFY	-0.28	0.92	DFY	-0.28	0.92
DFR	-0.33	0.43	DFR	-1.13	0.68
INFL lag	-0.86	0.93	INFL lag	-0.86	0.93

Model	$R^2_{OS}(\%)$	**pval**	**Model**	$R^2_{OS}(\%)$	**pval**
OLS	-6.63	0.36	OLS	-1.91	0.17
Pooled forecast: median	0.18	0.13	Pooled forecast: median	0.18	0.13
Pooled forecast: DMSFE	0.42	0.18	Pooled forecast: DMSFE	0.42	0.18
Sum-of-the-parts	0.89	0.10	Sum-of-the-parts	**1.35**	**0.03**
MARS	**1.18**	**0.04**	MARS	**1.29**	**0.01**
SVM SIC	0.16	0.17	SVM SIC	0.60	0.09
Lasso SVM	0.33	0.17	Lasso SVM	0.77	0.08
Random Forest	1.01	0.11	Random Forest	1.16	0.07
Diffusion index	0.36	0.27	Diffusion index	0.36	0.27
PLS	0.51	0.12	PLS	0.80	0.08
Neural Networks Median	5.77	0.13	Neural Networks Median	-2.19	0.29
Neural Networks 40^{th}	6.14	0.12	Neural Networks 40^{th}	-0.87	0.21

Table A3: **Monthly equity premium out-of-sample forecasting results for individual forecasts, and machine learning methods**. Utility gain (Δ Utility) 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 forecasting model considered relative to the historical average benchmark forecasting model; the weight on stocks in the investor's portfolio is restricted to lie between -0.5 and 1.5 (inclusive). The restriction imposed for the restricted case are the same of Table A2. The results refer to monthly forecasts for the out-of-sample period 2001:01-2017:12. The division between Recession and Expansion months comes from the NBER database. Bold indicates a ΔUtility above 1.00%.

Δ Utility	**2001-2017**		
Standard	**Total**	**Expansion**	**Recession**
DP	**1.84**	-1.90	**26.15**
DY	**2.38**	-2.43	**33.89**
EP	**4.75**	-1.20	**44.07**
DE	**0.82**	0.13	**5.31**
SVAR	**2.35**	**0.55**	**14.03**
BM	**2.57**	-2.82	**38.05**
NTIS	-1.23	**1.38**	-18.24
TBL	-0.71	**1.11**	-12.41
LTY	0.30	0.92	-3.65
LTR	-0.09	-0.21	0.25
TMS	-1.83	-0.02	-13.56
DFY	-0.94	-0.14	-6.13
DFR	**1.23**	-0.14	**9.95**
INFL lag	-1.66	0.15	-13.04

Δ Utility	**2001-2017**		
Restricted	**Total**	**Expansion**	**Recession**
DP	**1.70**	-2.04	**26.05**
DY	**2.13**	-2.68	**33.67**
EP	**3.89**	-1.18	**37.23**
DE	**1.24**	0.33	**7.18**
SVAR	**2.21**	**0.50**	**13.27**
BM	**2.55**	-2.75	**37.44**
NTIS	-1.23	**1.38**	-18.24
TBL	-0.71	**1.11**	-12.41
LTY	0.30	**0.92**	-3.65
LTR	-0.24	-0.17	-0.97
TMS	-1.83	-0.02	-13.56
DFY	-0.94	-0.14	-6.13
DFR	**0.96**	-0.45	**10.03**
INFL lag	-1.66	0.15	-13.04

Standard	**Total**	**Expansion**	**Recession**
OLS	**4.57**	**1.95**	**21.56**
Pooled forecast: median	0.33	0.21	**1.15**
Pooled forecast: DMSFE	**1.18**	0.24	**7.45**
Sum-of-the-parts	**2.22**	**2.23**	**2.63**
MARS	**1.78**	**1.93**	0.86
SVM SIC	**1.12**	**1.87**	-3.10
Lasso SVM	**1.15**	**1.64**	-1.48
Random Forest	**4.54**	0.22	**32.76**
Diffusion index	-1.25	**1.45**	-19.07
PLS	**1.11**	**2.81**	-9.32
Neural Networks Median	**2.93**	**4.20**	-4.97
Neural Networks 40^{th}	**2.75**	**3.81**	-3.73

Restricted	**Total**	**Expansion**	**Recession**
OLS	**4.56**	**2.05**	**21.04**
Pooled forecast: median	0.33	0.21	**1.15**
Pooled forecast: DMSFE	**1.18**	0.24	**7.45**
Sum-of-the-parts	**2.40**	**2.31**	**3.40**
MARS	**1.85**	**1.93**	**1.40**
SVM SIC	**1.53**	**2.03**	-1.19
Lasso SVM	**1.59**	**1.82**	0.47
Random Forest	**4.40**	0.27	**31.37**
Diffusion index	-1.25	**1.45**	-19.07
PLS	**1.40**	**2.91**	-7.88
Neural Networks Median	**2.49**	**3.82**	-5.62
Neural Networks 40^{th}	**2.44**	**3.6**	-4.35

Table A4: **Monthly equity premium out-of-sample forecasting results for individual forecasts, and machine learning methods**. This table reports R^2_{OS} and Δ Utility results for the out-of-sample period 2011:01-2017:12. The methodologies employed are the same explained in the tables above (A2-A3). **Bold** indicates a p-value for the R^2_{OS} under 0.05 or a ΔUtility above 1.00%

Standard	**2011-2017**		**Restricted**	**2011-2017**		**Δ Utility**	**2011-2017**	
Predictor	$R^2_{OS}(\%)$	**pval**	**Predictor**	$R^2_{OS}(\%)$	**pval**	**Predictor**	**Standard**	**Restricted**
DP	-4.42	0.93	DP	-4.42	0.93	DP	-4.01	-4.01
DY	-5.74	0.93	DY	-5.81	0.94	DY	-5.05	-5.07
EP	-2.24	0.86	EP	-2.24	0.86	EP	-3.27	-3.27
DE	0.78	0.10	DE	0.78	0.10	DE	0.78	0.78
SVAR	1.31	0.13	SVAR	1.09	0.13	SVAR	**1.02**	0.93
BM	-5.41	0.96	BM	-5.41	0.96	BM	-5.09	-5.09
NTIS	**2.86**	**0.03**	NTIS	**2.86**	**0.03**	NTIS	**1.57**	**1.57**
TBL	2.25	0.06	TBL	2.25	0.06	TBL	**1.71**	**1.71**
LTY	**1.82**	**0.03**	LTY	**1.82**	**0.03**	LTY	**1.37**	**1.37**
LTR	-0.42	0.45	LTR	-0.42	0.45	LTR	-0.96	-0.90
TMS	0.38	0.28	TMS	0.38	0.28	TMS	0.56	0.56
DFY	-0.20	0.88	DFY	-0.20	0.88	DFY	-0.38	-0.38
DFR	0.22	0.31	DFR	-0.56	0.45	DFR	0.15	-0.21
INFL lag	0.45	0.16	INFL lag	0.45	0.16	INFL lag	0.13	0.13

Model	$R^2_{OS}(\%)$	**pval**	**Model**	$R^2_{OS}(\%)$	**pval**	**Model**	**Standard**	**Restricted**
OLS	**2.12**	**0.04**	OLS	**1.96**	**0.04**	OLS	**4.02**	**3.94**
Pooled forecast: median	0.16	0.29	Pooled forecast: median	0.16	0.29	Pooled forecast: median	0.08	0.08
Pooled forecast: DMSFE	0.15	0.27	Pooled forecast: DMSFE	0.15	0.27	Pooled forecast: DMSFE	0.05	0.05
Sum-of-the-parts	2.23	0.07	Sum-of-the-parts	2.23	0.07	Sum-of-the-parts	**3.59**	**3.59**
MARS	**4.40**	**0.00**	MARS	**4.40**	**0.00**	MARS	**2.71**	**2.71**
SVM SIC	**3.22**	**0.04**	SVM SIC	**3.22**	**0.04**	SVM SIC	**2.99**	**2.99**
Lasso SVM	**3.19**	**0.05**	Lasso SVM	**3.19**	**0.05**	Lasso SVM	**2.63**	**2.63**
Random Forest	0.06	0.38	Random Forest	0.06	0.38	Random Forest	0.33	0.33
Diffusion index	**2.45**	**0.04**	Diffusion index	**2.45**	**0.04**	Diffusion index	**1.49**	**1.49**
PLS	**4.06**	**0.03**	PLS	**4.06**	**0.03**	PLS	**3.55**	**3.55**
Neural Networks Median	**3.08**	**0.05**	Neural Networks Median	**3.08**	**0.05**	Neural Networks Median	**2.34**	**2.34**
Neural Networks 40^{th}	**3.06**	**0.05**	Neural Networks 40^{th}	**3.08**	**0.05**	Neural Networks 40^{th}	**2.16**	**2.16**

Table A5: **R^2_{OS} of the six portfolios built double sorting on Size and the Book/Market ratio using the Welch and Goyal (2008) predictors.** The R^2_{OS} is the Campbell Thompson (2008) out-of-sample R^2 statistic. Statistical significance for the R^2_{OS} statistic is tested using the Clark and West (2007) p-value; the statistic corresponds to a one-sided test of the null hypothesis that the competing forecasting model has equal expected square prediction error relative to the historical average benchmark forecasting model against the alternative hypothesis that the competing forecasting model has a lower expected square prediction error than the historical average benchmark forecasting model. The results refer to monthly forecasts for the out-of-sample period 1986:01-2016:12. **Bold** indicates a p-value for the R^2_{OS} statistic less than 0.1.

R^2_{OS}

Predictor	**SMALL LoBM**	**pval**	**ME1 BM2**	**pval**	**SMALL HiBM**	**pval**	**BIG LoBM**	**pval**	**ME2 BM2**	**pval**	**BIG HiBM**	**pval**
DP	-1.11	0.22	-1.78	0.30	-2.97	0.16	-1.41	0.55	-1.75	0.43	-1.88	0.12
DY	-2.22	0.17	-3.67	0.23	-5.76	0.11	-2.11	0.53	-2.69	0.40	-2.99	0.11
EP	-0.51	0.27	-1.39	0.39	-2.18	0.36	-0.94	0.30	-1.82	0.27	-1.22	0.21
DE	-0.19	0.59	-0.32	0.84	-0.14	0.54	-0.29	0.59	-0.71	0.53	-0.58	0.70
SVAR	-1.02	0.91	-1.81	0.93	-3.10	0.95	-0.41	0.79	-1.03	0.89	-3.36	0.95
BM	-2.58	0.20	-4.91	0.28	-8.33	0.15	-2.50	0.64	-5.38	0.44	-6.04	0.15
NTIS	-3.48	0.77	-4.91	0.77	-8.10	0.90	-1.30	0.51	-3.59	0.75	-4.82	0.76
TBL	0.02	0.33	-0.08	0.56	-0.08	0.49	-0.29	0.92	-0.18	0.92	-0.24	0.91
LTY	0.05	0.29	-0.11	0.88	-0.07	0.90	-0.11	0.99	-0.07	0.88	-0.05	0.96
LTR	-0.04	0.25	-0.14	0.17	-0.44	0.21	-0.11	0.30	-0.53	0.30	-0.28	0.38
TMS	-1.48	0.73	-1.46	0.65	-1.65	0.62	-1.04	0.77	-0.91	0.77	-1.38	0.93
DFY	-0.26	0.37	-0.85	0.48	-1.85	0.55	-0.45	0.82	-0.89	0.78	-2.06	0.79
DFR	-0.18	0.54	-0.27	0.57	-1.40	0.97	0.12	0.28	-0.24	0.55	-0.71	0.90
INFL lag	-0.21	0.57	-0.74	0.69	-1.48	0.78	-0.32	0.79	-0.49	0.87	-1.52	0.95

R^2_{OS}

Model	**SMALL LoBM**	**pval**	**ME1 BM2**	**pval**	**SMALL HiBM**	**pval**	**BIG LoBM**	**pval**	**ME2 BM2**	**pval**	**BIG HiBM**	**pval**
OLS	-3.79	0.18	-8.14	0.21	-13.85	0.26	-4.66	0.51	-11.17	0.45	-14.51	0.66
Pooled forecast:median	0.14	0.20	**0.22**	**0.09**	0.16	0.19	0.03	0.39	0.08	0.23	0.03	0.35
Pooled forecast:MDSFE	0.18	0.25	-0.08	0.40	0.00	0.28	-0.11	0.54	-0.19	0.52	-0.14	0.39
Sum-of-the-parts	0.04	0.18	-2.12	0.28	-2.14	0.14	-1.81	0.57	-1.79	0.41	-1.36	0.17
MARS	0.60	0.09	-0.60	0.43	-0.70	0.46	-0.38	0.31	-1.21	0.45	-0.49	0.37
SVM SIC	-0.94	0.67	-1.51	0.81	-2.24	0.93	0.32	0.19	-1.56	0.79	-0.64	0.76
Lasso SVM	-0.94	0.80	-1.07	0.86	-0.53	0.88	-0.09	0.33	-1.13	0.91	-0.19	0.52
Radom Forest	**0.50**	**0.10**	0.23	0.16	-0.23	0.44	-0.16	0.47	0.05	0.31	0.42	0.12
Diffusion index	0.32	0.17	-0.03	0.37	0.07	0.36	0.06	0.35	-0.37	0.74	-0.29	0.58
PLS	-0.08	0.20	-1.47	0.46	-0.96	0.52	-0.52	0.48	-1.50	0.72	-0.62	0.41
Neural Networks Median	0.11	0.26	-0.06	0.05	**1.71**	**0.00**	-0.13	0.23	0.37	0.14	**0.69**	**0.09**

Table A6: **Δ Utility of the six portfolios built double sorting on Size and the Book/Market ratio using the Welch and Goyal (2008) predictors.** The Utility gain (Δ Utility) 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 forecasting model considered relative to the historical average benchmark forecasting model; the weight on stocks in the investor's portfolio is restricted to lie between -0.5 and 1.5 (inclusive). The restriction imposed for the restricted case are the same of Table 1. The results refer to monthly forecasts for the out-of-sample period 1986:01-2016:12. **Bold** indicates a ΔUtility above 1.00%.

Δ Utility

Predictor	**SMALL LoBM**	**ME1 BM2**	**SMALL HiBM**	**BIG LoBM**	**ME2 BM2**	**BIG HiBM**
DP	-0.58	-1.96	-3.52	-1.37	-1.62	-4.90
DY	-1.17	-3.41	-4.32	-2.43	-3.37	-4.95
EP	**1.78**	**1.40**	**1.76**	**1.74**	**2.76**	**2.82**
DE	-0.23	-0.63	-0.79	0.97	**1.23**	-0.75
SVAR	-1.67	-1.15	-1.21	**1.22**	0.64	-2.27
BM	-1.75	-4.61	-6.31	-2.11	-4.57	-6.10
NTIS	-2.79	-0.70	-0.64	**1.59**	0.98	-1.81
TBL	-0.10	-0.44	-0.46	**1.26**	0.83	-1.68
LTY	0.10	-0.73	-0.20	**1.21**	0.65	-1.38
LTR	-0.23	0.49	**1.00**	**1.12**	**1.02**	-1.09
TMS	-3.37	-1.09	-0.99	0.54	0.19	-2.54
DFY	-1.54	-1.23	-1.80	0.97	0.31	-3.43
DFR	0.76	0.61	-0.95	**2.20**	**1.65**	-2.00
INFL lag	-0.97	-0.79	-1.13	**1.22**	0.25	-2.93

Δ Utility

Model	**SMALL LoBM**	**ME1 BM2**	**SMALL HiBM**	**BIG LoBM**	**ME2 BM2**	**BIG HiBM**
OLS	0.04	**1.24**	-1.74	0.60	-0.26	-3.34
Pooled forecast:median	0.14	-0.43	-0.45	**1.40**	0.79	-1.43
Pooled forecast:MDSFE	0.83	-0.28	-0.75	**1.08**	0.83	-1.31
Sum-of-the-parts	**1.43**	-0.35	0.09	-0.79	0.28	-0.79
MARS	**1.96**	-0.16	0.27	**1.21**	**2.05**	**1.86**
SVM SIC	0.18	-3.65	-3.33	-4.58	-3.50	-3.83
Lasso SVM	-1.44	-0.58	-0.88	**1.72**	0.80	-1.30
Radom Forest	**1.45**	-0.32	-0.57	0.72	**1.32**	-0.58
Diffusion index	0.64	-0.52	-0.81	**1.49**	0.97	-1.78
PLS	-0.48	0.58	-1.21	**1.36**	0.66	-1.19
Neural Networks Median	-0.02	**2.96**	-3.48	0.60	0.78	-1.26

Table A7: R^2_{OS} **of the six portfolios built double sorting on Size and the Momentum using the Welch and Goyal (2008) predictors.** The R^2_{OS} is the Campbell Thompson (2008) out-of-sample R^2 statistic. Statistical significance for the R^2_{OS} statistic is tested using the Clark and West (2007) p-value; the statistic corresponds to a one-sided test of the null hypothesis that the competing forecasting model has equal expected square prediction error relative to the historical average benchmark forecasting model against the alternative hypothesis that the competing forecasting model has a lower expected square prediction error than the historical average benchmark forecasting model. The results refer to monthly forecasts for the out-of-sample period 1986:01-2017:12. **Bold** indicates a p-value for the R^2_{OS} statistic less than 0.1.

R^2_{OS}

Predictor	**SMALL LoPRIOR**	**pval**	**ME1 PRIOR2**	**pval**	**SMALL HiPRIOR**	**pval**	**BIG LoPRIOR**	**pval**	**ME2 PRIOR2**	**pval**	**BIG HiPRIOR**	**pval**
DP	-2.38	0.15	-2.51	0.16	-1.09	0.34	-2.88	0.45	-1.48	0.26	-0.50	0.33
DY	-4.14	0.10	-5.02	0.11	-2.24	0.29	-4.01	0.40	-2.35	0.25	-0.83	0.32
EP	-2.29	0.60	-1.68	0.41	-0.30	0.19	-3.07	0.59	-1.62	0.30	0.00	0.09
DE	**1.06**	**0.07**	-0.09	0.36	-0.31	0.81	-0.32	0.60	-0.54	0.72	-0.42	0.45
SVAR	-1.87	0.79	-3.16	0.92	-0.67	0.91	-1.40	0.89	-1.71	0.91	0.62	0.06
BM	-5.22	0.22	-6.71	0.19	-2.80	0.24	-6.37	0.60	-4.16	0.39	-1.22	0.33
NTIS	-3.50	0.61	-6.06	0.77	-4.96	0.93	-2.87	0.52	-2.22	0.52	-2.25	0.84
TBL	**0.55**	**0.08**	0.02	0.33	-0.17	0.74	-0.12	0.79	-0.22	0.71	-0.20	0.95
LTY	**0.39**	**0.04**	0.00	0.48	-0.08	0.91	-0.05	0.72	-0.08	0.85	-0.09	0.99
LTR	0.50	0.11	**0.59**	**0.07**	-0.13	0.30	0.49	0.10	-0.41	0.23	-0.21	0.51
TMS	-1.01	0.46	-2.08	0.65	-1.45	0.75	-0.72	0.51	-1.28	0.75	-0.69	0.81
DFY	1.15	0.11	-1.07	0.41	-0.99	0.72	-0.06	0.34	-0.43	0.57	-0.22	0.94
DFR	-0.63	0.89	-0.60	0.86	-0.09	0.44	-0.66	0.96	-0.28	0.59	0.46	0.18
INFL lag	-0.65	0.71	-1.12	0.79	-0.35	0.54	-0.43	0.93	-0.34	0.84	-0.29	0.68

R^2_{OS}

Model	**SMALL LoPRIOR**	**pval**	**ME1 PRIOR2**	**pval**	**SMALL HiPRIOR**	**pval**	**BIG LoPRIOR**	**pval**	**ME2 PRIOR2**	**pval**	**BIG HiPRIOR**	**pval**
OLS	-5.82	0.11	-9.22	0.19	-6.33	0.23	-8.87	0.80	-8.90	0.65	-2.80	0.23
Pooled forecast:median	**0.53**	**0.04**	0.20	0.17	0.20	0.11	0.00	0.44	0.02	0.39	0.12	0.13
Pooled forecast:MDSFE	0.47	0.12	0.04	0.27	-0.03	0.39	-0.24	0.52	-0.02	0.37	0.18	0.19
Sum-of-the-parts	-0.16	0.37	-2.04	0.23	-2.36	0.15	-0.63	0.65	-1.58	0.44	-2.34	0.25
MARS	0.20	0.21	-0.07	0.25	-0.83	0.39	-0.87	0.45	-0.90	0.52	-1.21	0.28
SVM SIC	-0.13	0.30	-0.41	0.37	-0.62	0.52	-0.87	0.56	0.75	0.11	-0.21	0.30
Lasso SVM	-0.03	0.25	-0.73	0.51	-0.60	0.85	-0.51	0.48	-0.57	0.74	-0.43	0.37
Radom Forest	**0.44**	**0.08**	-0.02	0.29	0.60	0.05	0.18	0.18	0.20	0.19	-0.05	0.35
Diffusion index	0.32	0.11	0.17	0.17	-0.22	0.63	-0.72	0.41	0.23	0.23	-1.68	0.80
PLS	**0.98**	**0.04**	-1.23	0.31	-1.81	0.75	-1.62	0.72	-0.49	0.40	0.02	0.17
Neural Networks Median	-0.77	0.41	**1.01**	**0.02**	**2.05**	**0.00**	-0.71	0.55	-0.16	0.33	-0.04	0.08

Table A8: **Δ Utility of the six portfolios built double sorting on Size and the Momentum using the Welch and Goyal (2008) predictors.** The Utility gain (Δ Utility) 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 forecasting model considered relative to the historical average benchmark forecasting model; the weight on stocks in the investor's portfolio is restricted to lie between -0.5 and 1.5 (inclusive). The restriction imposed for the restricted case are the same of Table 1. The results refer to monthly forecasts for the out-of-sample period 1986:01-2016:12. **Bold** indicates a ΔUtility above 1.00%.

Δ Utility						
Predictor	**SMALL LoPRIOR**	**ME1 PRIOR2**	**SMALL HiPRIOR**	**BIG LoPRIOR**	**ME2 PRIOR2**	**BIG HiPRIOR**
DP	-0.34	-3.53	0.28	-3.13	-2.20	0.33
DY	-0.10	-4.15	-1.10	-3.17	-3.28	0.17
EP	-0.34	**1.54**	**2.14**	-1.03	**2.57**	**2.59**
DE	-1.76	-1.13	-0.97	-1.84	0.44	0.33
SVAR	-3.95	-1.11	-1.03	-3.67	0.31	**2.28**
BM	-2.27	-5.30	-1.38	-5.35	-3.85	-0.03
NTIS	-10.36	-0.54	-1.10	-5.67	**1.17**	0.60
TBL	-1.57	-0.13	-0.59	-1.61	0.93	**1.05**
LTY	-0.95	-0.41	-1.03	-1.25	0.72	0.89
LTR	-1.55	**1.09**	0.19	-1.20	0.76	0.95
TMS	-4.91	-1.00	-0.89	-2.68	0.36	0.69
DFY	-6.11	-1.64	-0.94	-5.57	0.56	0.79
DFR	-1.69	-0.31	0.87	-1.90	**1.62**	**2.86**
INFL lag	-3.60	-0.79	-0.47	-1.94	0.60	0.93

Δ Utility						
Model	**SMALL LoPRIOR**	**ME1 PRIOR2**	**SMALL HiPRIOR**	**BIG LoPRIOR**	**ME2 PRIOR2**	**BIG HiPRIOR**
OLS	-2.06	0.23	0.91	-6.65	-1.60	**1.71**
Pooled forecast:median	-0.53	-0.49	-0.41	-1.16	0.75	**1.19**
Pooled forecast:MDSFE	-0.09	-0.11	0.02	-1.06	0.89	**1.25**
Sum-of-the-parts	0.75	-0.08	0.00	-0.85	-0.28	-0.83
MARS	0.84	**1.31**	-0.52	0.42	0.66	0.74
SVM SIC	-0.65	-0.55	0.76	-1.12	**1.85**	0.47
Lasso SVM	0.05	-0.67	-0.83	-0.33	0.48	0.42
Radom Forest	0.42	-0.30	0.65	-0.05	**1.11**	0.72
Diffusion index	0.46	-0.43	-0.55	-2.12	**1.32**	0.18
PLS	-1.60	0.52	0.08	-4.77	**1.57**	**1.45**
Neural Networks Median	-0.59	**2.90**	**2.97**	-0.98	-4.11	**1.31**

Table A9: R^2_{OS} **of the six portfolios built double sorting on Size and the Book/Market ratio using factor-anomalies returns spread as predictors.** The R^2_{OS} is the Campbell Thompson (2008) out-of-sample R^2 statistic. Statistical significance for the R^2_{OS} statistic is tested using the Clark and West (2007) p-value; the statistic corresponds to a one-sided test of the null hypothesis that the competing forecasting model has equal expected square prediction error relative to the historical average benchmark forecasting model against the alternative hypothesis that the competing forecasting model has a lower expected square prediction error than the historical average benchmark forecasting model. The results refer to monthly forecasts for the out-of-sample period 1986:01-2016:12. **Bold** indicates a p-value for the R^2_{OS} statistic less than 0.1.

R^2_{OS}

Predictor	**SMALL LoBM**	**pval**	**ME1 BM2**	**pval**	**SMALL HiBM**	**pval**	**BIG LoBM**	**pval**	**ME2 BM2**	**pval**	**BIG HiBM**	**pval**
SMB	-1.16	0.46	-0.68	0.26	-0.27	0.14	-0.61	0.46	-0.36	0.28	-0.29	0.33
HML	-0.32	0.14	-0.74	0.36	-0.67	0.65	0.30	0.12	-0.82	0.84	-0.63	0.80
RMW	-1.28	0.81	-0.47	0.68	-0.06	0.37	-0.83	0.56	-0.96	0.67	-0.49	0.53
CMA	**0.39**	**0.02**	**0.09**	**0.02**	-0.19	0.05	**0.40**	**0.04**	-0.27	0.11	0.17	0.11
LT	-0.21	0.38	-0.51	0.65	-0.63	0.96	0.05	0.17	-0.77	0.64	-0.65	0.70
ST	-0.69	0.47	**0.31**	**0.08**	**0.95**	**0.03**	-0.79	0.87	-0.58	0.66	-0.01	0.29
Mom	-0.32	0.52	**0.80**	**0.02**	**0.83**	**0.02**	-0.99	0.88	-0.50	0.78	0.07	0.30
Asset Growth	**13.47**	**0.00**	**5.09**	**0.00**	**1.19**	**0.00**	**19.15**	**0.00**	**5.92**	**0.00**	**3.89**	**0.00**
Gross Prof	-3.01	0.00	-0.57	0.00	-0.11	0.00	**0.72**	**0.00**	-0.98	0.03	-0.22	0.07
Inv to Assets	**12.83**	**0.00**	**7.00**	**0.00**	**3.61**	**0.00**	**13.50**	**0.00**	**7.90**	**0.00**	**4.72**	**0.00**
Net Stock Issues	**58.32**	**0.00**	**32.58**	**0.00**	**24.07**	**0.00**	**28.74**	**0.00**	**11.84**	**0.00**	**13.44**	**0.00**
NOA	-3.60	0.22	-2.72	0.36	-1.93	0.36	-2.85	0.30	-1.50	0.06	-0.72	0.16
Accruals	-1.90	0.00	-3.32	0.01	-2.83	0.05	-1.48	0.00	-2.28	0.04	-2.26	0.12
O	-1.15	0.28	-3.90	0.94	-3.93	0.74	**12.45**	**0.00**	**0.27**	**0.05**	-1.11	0.48
ROA	-7.85	0.13	-4.23	0.25	-2.30	0.35	-3.41	0.00	-4.01	0.14	-0.94	0.26
Distress	-0.86	0.45	**0.61**	**0.09**	**1.78**	**0.02**	-0.13	0.20	**1.51**	**0.03**	**3.14**	**0.00**
Comp Eq Issue	-0.48	0.16	-0.58	0.14	-0.29	0.10	-0.25	0.29	-0.92	0.91	-0.78	0.87
Model	**SMALL LoBM**	**pval**	**ME1 BM2**	**pval**	**SMALL HiBM**	**pval**	**BIG LoBM**	**pval**	**ME2 BM2**	**pval**	**BIG HiBM**	**pval**
OLS	**53.49**	**0.00**	**30.28**	**0.00**	**27.05**	**0.00**	**27.98**	**0.00**	**6.90**	**0.00**	**8.40**	**0.00**
Pooled forecast:median	**3.39**	**0.00**	**2.58**	**0.00**	**2.19**	**0.00**	**5.05**	**0.00**	**2.13**	**0.00**	**1.63**	**0.00**
Pooled forecast:MDSFE	**29.65**	**0.00**	**11.54**	**0.00**	**7.04**	**0.00**	**18.45**	**0.00**	**7.50**	**0.00**	**4.41**	**0.00**
MARS	**16.53**	**0.00**	**10.20**	**0.00**	**8.44**	**0.00**	-2.41	0.08	-0.52	0.01	**4.05**	**0.00**
SVM SIC	-19.89	0.05	-8.81	0.18	-3.67	0.17	-14.66	0.32	-8.57	0.37	-3.68	0.44
Lasso SVM	-22.86	0.10	-9.63	0.29	-3.36	0.21	-21.32	0.26	-5.15	0.34	-2.19	0.36
Radom Forest	0.11	0.21	-0.12	0.65	0.08	0.24	-0.15	0.63	-0.02	0.43	0.00	0.43
Diffusion index	**11.21**	**0.00**	**4.22**	**0.00**	**0.26**	**0.03**	19.27	0.00	**6.25**	**0.00**	**2.74**	**0.00**
PLS	-41.52	0.82	-25.67	0.78	-14.40	0.61	-48.62	0.95	-27.61	0.99	-14.38	0.94
Neural Network Median	**6.32**	**0.00**	-2.61	0.14	-2.45	0.43	-1.77	0.28	-0.04	0.25	**0.34**	**0.06**

Table A10: **Δ Utility of the six portfolios built double sorting on Size and the Book/Market ratio using factor-anomalies returns spread as predictors**.The Utility gain (Δ Utility) 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 forecasting model considered relative to the historical average benchmark forecasting model; the weight on stocks in the investor's portfolio is restricted to lie between -0.5 and 1.5 (inclusive). The restriction imposed for the restricted case are the same of Table 1. The results refer to monthly forecasts for the out-of-sample period 1986:01-2016:12. **Bold** indicates a Δ Utility above 1.00%.

Δ Utility

Predictor	**SMALL LoBM**	**ME1 BM2**	**SMALL HiBM**	**BIG LoBM**	**ME2 BM2**	**BIG HiBM**
SMB	0.35	0.16	0.11	-0.43	0.23	-0.15
HML	-0.53	0.21	-0.45	**1.64**	-0.24	-0.65
RMW	-1.00	0.07	0.51	-0.03	-0.18	0.05
CMA	**4.35**	**1.90**	0.30	**2.05**	0.72	0.05
LT	-0.47	0.37	-0.30	**1.66**	0.51	0.85
ST	0.42	**1.37**	**2.11**	-0.80	-0.28	0.81
Mom	0.08	**1.25**	**1.14**	-0.69	-0.21	0.25
Asset Growth	**17.89**	**11.57**	**7.82**	**14.52**	**7.99**	**7.68**
Gross Prof	**4.01**	**6.14**	**5.11**	**2.92**	**2.65**	**2.39**
Inv to Assets	**16.62**	**11.82**	**8.80**	**11.82**	**8.18**	**8.68**
Net Stock Issues	**38.85**	**25.75**	**22.81**	**23.00**	**17.61**	**18.96**
NOA	**3.92**	**3.00**	**1.21**	-0.36	**4.88**	**2.54**
Accruals	**4.06**	**3.68**	**3.31**	**4.61**	**3.03**	**1.66**
O	-2.97	-0.39	0.32	**7.58**	0.35	-0.71
ROA	0.84	**1.10**	-0.07	**3.33**	0.51	-0.98
Distress	0.03	0.32	**1.83**	-0.06	**3.36**	**4.20**
Comp Eq Issue	**1.71**	**1.20**	0.50	0.02	-0.46	-0.41
Predictor	**SMALL LoBM**	**ME1 BM2**	**SMALL HiBM**	**BIG LoBM**	**ME2 BM2**	**BIG HiBM**
OLS	**36.28**	**24.01**	**23.77**	**17.54**	**9.14**	**10.72**
Pooled forecast:median	**6.40**	**1.66**	**1.59**	**3.97**	**2.26**	**1.43**
Pooled forecast:MDSFE	**27.60**	**12.96**	**7.96**	**15.41**	**7.75**	**5.95**
MARS	**12.07**	**6.29**	**7.83**	-3.77	0.22	**4.23**
SVM SIC	**2.16**	**2.13**	**3.30**	-0.22	**1.63**	**3.02**
Lasso SVM	**3.12**	**2.00**	**3.28**	-0.12	**1.24**	**3.15**
Radom Forest	0.50	0.01	0.05	-0.32	0.08	0.36
Diffusion index	**16.54**	**12.22**	**7.29**	**13.19**	**9.10**	**8.41**
PLS	-12.19	-7.58	-5.65	-7.51	-7.26	-5.47
Neural Network Median	**10.63**	**2.17**	**2.06**	-0.63	0.67	**2.77**

Table A11: R^2_{OS} **of the six portfolios built double sorting on Size and the Momentum using factor-anomalies returns spread as predictors.** The R^2_{OS} is the Campbell Thompson (2008) out-of-sample R^2 statistic. Statistical significance for the R^2_{OS} statistic is tested using the Clark and West (2007) p-value; the statistic corresponds to a one-sided test of the null hypothesis that the competing forecasting model has equal expected square prediction error relative to the historical average benchmark forecasting model against the alternative hypothesis that the competing forecasting model has a lower expected square prediction error than the historical average benchmark forecasting model. The results refer to monthly forecasts for the out-of-sample period 1986:01-2016:12. **Bold** indicates a p-value for the R^2_{OS} statistic less than 0.1.

R^2_{OS}

Predictor	SMALL LoPRIOR	pval	ME1 PRIOR2	pval	SMALL HiPRIOR	pval	BIG LoPRIOR	pval	ME2 PRIOR2	pval	BIG HiPRIOR	pval
SMB	-0.42	0.28	-0.45	0.21	-0.58	0.19	-0.66	0.99	-0.72	0.68	-0.40	0.16
HML	-0.54	0.42	-0.70	0.46	-0.69	0.18	-0.74	0.95	-0.48	0.55	-0.14	0.17
RMW	0.22	0.20	-0.25	0.54	-0.38	0.36	-1.58	0.69	-1.37	0.66	**0.61**	**0.06**
CMA	**0.53**	**0.04**	**0.46**	**0.01**	-0.26	0.03	-0.38	0.20	**0.12**	**0.05**	**0.30**	**0.04**
LT	-0.65	0.90	-0.57	0.67	-0.60	0.72	-0.69	0.66	-0.48	0.39	-0.46	0.45
ST	-0.57	0.33	**0.37**	**0.06**	-0.61	0.46	-1.07	0.80	-0.68	0.85	-0.41	0.68
Mom	0.31	0.19	**0.84**	**0.02**	-0.19	0.42	-0.97	0.82	-0.74	0.89	-0.72	0.80
Asset Growth	**12.38**	**0.00**	**8.18**	**0.00**	**5.19**	**0.00**	**14.95**	**0.00**	**10.57**	**0.00**	**6.04**	**0.00**
Gross Prof	**0.92**	**0.00**	**0.43**	**0.00**	-3.37	0.00	**1.74**	**0.01**	**0.20**	**0.00**	-3.62	0.00
Inv to Assets	**14.77**	**0.00**	**8.81**	**0.00**	**5.60**	**0.00**	**15.84**	**0.00**	**8.46**	**0.00**	**4.20**	**0.00**
Net Stock Issues	**37.70**	**0.00**	**26.58**	**0.00**	**42.36**	**0.00**	**18.99**	**0.00**	**13.90**	**0.00**	**25.17**	**0.00**
NOA	-2.74	0.21	-2.28	0.11	-2.43	0.26	-1.45	0.11	-2.75	0.16	-3.14	0.44
Accruals	-1.42	0.03	-3.13	0.01	-3.19	0.01	-1.13	0.04	-1.91	0.01	-2.86	0.02
O	-0.65	0.26	-3.41	0.95	-4.44	0.93	**7.25**	**0.00**	**3.87**	**0.00**	**0.51**	**0.03**
ROA	-0.13	0.10	-4.42	0.22	-6.21	0.06	-0.77	0.31	-4.08	0.07	-5.61	0.00
Distress	**18.36**	**0.00**	**4.52**	**0.00**	-1.97	0.33	**22.00**	**0.00**	**3.13**	**0.00**	-2.19	0.04
Comp Eq Issue	**0.16**	**0.06**	-0.38	0.07	-0.44	0.10	-0.96	0.98	-0.74	0.86	-0.14	0.25
Model	**SMALL LoPRIOR**	**pval**	**ME1 PRIOR2**	**pval**	**SMALL HiPRIOR**	**pval**	**BIG LoPRIOR**	**pval**	**ME2 PRIOR2**	**pval**	**BIG HiPRIOR**	**pval**
OLS	**53.42**	**0.00**	**28.94**	**0.00**	**38.00**	**0.00**	**37.95**	**0.00**	**9.11**	**0.00**	**16.69**	**0.00**
Pooled forecast:median	**3.39**	**0.00**	**3.03**	**0.00**	**3.46**	**0.00**	**2.08**	**0.00**	**2.63**	**0.00**	**4.01**	**0.00**
Pooled forecast:MDSFE	**14.98**	**0.00**	**10.86**	**0.00**	**17.85**	**0.00**	**13.71**	**0.00**	**9.94**	**0.00**	**11.46**	**0.00**
MARS	**33.80**	**0.00**	**14.20**	**0.00**	**19.44**	**0.00**	**14.96**	**0.00**	-0.37	0.01	**1.12**	**0.00**
SVM SIC	-5.29	0.06	-4.51	0.02	-13.68	0.04	-8.13	0.36	-9.68	0.28	-12.61	0.21
Lasso SVM	-9.23	0.12	-5.33	0.07	-15.09	0.14	-8.21	0.35	-5.83	0.19	-15.38	0.27
Radom Forest	**0.25**	**0.06**	**0.34**	**0.02**	0.00	0.38	-0.07	0.55	-0.01	0.41	-0.01	0.39
Diffusion index	**11.80**	**0.00**	**7.79**	**0.00**	**1.81**	**0.00**	**13.50**	**0.00**	**10.42**	**0.00**	**7.02**	**0.00**
PLS	-31.55	0.95	-27.02	0.88	-33.90	0.80	-39.06	0.98	-31.33	0.97	-36.41	0.89
Neural Network Median	**1.38**	**0.02**	-3.28	0.41	-1.49	0.19	-0.02	0.21	-0.85	0.26	-0.64	0.08

Table A12: **Δ Utility of the six portfolios built double sorting on Size and the Momentum using factor-anomalies returns spread as predictors.** The Utility gain (Δ Utility) 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 forecasting model considered relative to the historical average benchmark forecasting model; the weight on stocks in the investor's portfolio is restricted to lie between -0.5 and 1.5 (inclusive). The restriction imposed for the restricted case are the same of Table 1. The results refer to monthly forecasts for the out-of-sample period 1986:01-2016:12. **Bold** indicates a Δ Utility above 1.00%.

Δ Utility

Predictor	**SMALL LoPRIOR**	**ME1 PRIOR2**	**SMALL HiPRIOR**	**BIG LoPRIOR**	**ME2 PRIOR2**	**BIG HiPRIOR**
SMB	0.74	0.29	0.54	-0.89	-0.47	0.66
HML	-1.86	-0.13	**1.00**	-0.85	0.48	**1.66**
RMW	-0.84	-0.14	**1.28**	-0.45	0.30	**1.94**
CMA	**2.98**	**2.95**	**2.01**	-0.61	**1.44**	**2.33**
LT	-1.50	0.40	-0.55	-0.10	**1.56**	0.42
ST	**1.10**	**1.55**	-0.39	-1.02	-0.60	0.04
Mom	0.30	0.90	-0.15	-1.10	-0.35	-0.30
Asset Growth	**9.98**	**12.12**	**15.20**	**11.72**	**10.25**	**11.74**
Gross Prof	-1.64	**6.42**	**6.68**	0.98	**2.95**	**2.67**
Inv to Assets	**11.70**	**12.19**	**14.00**	**12.03**	**8.97**	**9.18**
Net Stock Issues	**28.12**	**23.26**	**32.15**	**21.51**	**18.61**	**22.48**
NOA	0.87	**3.59**	**5.01**	-0.04	**2.32**	**1.02**
Accruals	**1.70**	**4.68**	**4.02**	**3.44**	**4.80**	**2.76**
O	-5.54	-0.38	-0.38	**3.10**	**2.81**	**2.15**
ROA	**3.57**	**1.21**	**4.33**	0.45	**1.21**	**5.34**
Distress	**8.62**	**4.91**	0.71	**12.73**	**4.46**	**2.85**
Comp Eq Issue	**1.16**	**1.52**	**2.20**	-0.92	-0.36	**1.11**
Model	**SMALL LoPRIOR**	**ME1 PRIOR2**	**SMALL HiPRIOR**	**BIG LoPRIOR**	**ME2 PRIOR2**	**BIG HiPRIOR**
OLS	**30.50**	**22.12**	**29.36**	**21.38**	**9.38**	**18.31**
Pooled forecast:median	**6.97**	**2.59**	**4.15**	**2.49**	**2.42**	**3.53**
Pooled forecast:MDSFE	**17.67**	**12.46**	**18.55**	**14.17**	**9.47**	**11.10**
MARS	**16.44**	**9.22**	**13.77**	**10.96**	0.40	**2.13**
SVM SIC	-2.39	**2.87**	**8.01**	-1.03	**2.00**	**2.30**
Lasso SVM	-0.96	**3.45**	**6.09**	-3.75	**2.12**	**1.57**
Radom Forest	0.50	0.61	0.37	-0.23	0.00	0.02
Diffusion index	**8.62**	**12.95**	**15.23**	**9.93**	**10.20**	**12.73**
PLS	-13.68	-8.16	-9.21	-10.04	-6.47	-5.98
Neural Network Median	**1.13**	**2.64**	**1.49**	0.22	-0.28	**2.48**

Swiss Finance Institute

Swiss Finance Institute (SFI) is the national center for fundamental research, doctoral training, knowledge exchange, and continuing education in the fields of banking and finance. SFI's mission is to grow knowledge capital for the Swiss financial marketplace. Created in 2006 as a public–private partnership, SFI is a common initiative of the Swiss finance industry, leading Swiss universities, and the Swiss Confederation.

swiss:finance:institute

c/o University of Geneva, Bd. Du Pont d'Arve 42, CH-1211 Geneva 4
T +41 22 379 84 71, rps@sfi.ch, www.sfi.ch