INNOVATION, ENTREPRENEURSHIP AND MANAGEMENT SERIES

Machine Learning for Asset Management

New Developments and Financial Applications

Edited by
Emmanuel Jurczenko

ISTE

WILEY

Machine Learning for Asset Management

Machine Learning for Asset Management

New Developments and Financial Applications

Edited by

Emmanuel Jurczenko

iSTE

WILEY

First published 2020 in Great Britain and the United States by ISTE Ltd and John Wiley & Sons, Inc.

ISTE Ltd
27-37 St George's Road
London SW19 4EU
UK

www.iste.co.uk

John Wiley & Sons, Inc.
111 River Street
Hoboken, NJ 07030
USA

www.wiley.com

Library of Congress Control Number: 2019955407

British Library Cataloguing-in-Publication Data
A CIP record for this book is available from the British Library
ISBN 978-1-78630-544-2

Contents

Foreword

The development of machine learning offers enormous opportunities to our industry, but it also presents some challenges: as machines continue to evolve, how much control should investment managers relinquish to technology to deliver the best outcomes for investors?

At Unigestion, we believe in collaborative intelligence and in the strength of humans and machines working together. We empower our investment teams with leading-edge technology to gain a deeper understanding of financial markets, in order to get better investment outcomes for our clients.

We strongly believe that machine learning can help active managers differentiate themselves from passive ones. There is huge potential for asset managers to use machine learning to support their investment decision-making, especially if backed up by human experience. Thanks to their ability to process much more complex patterns with better forecasting power, modern machine learning algorithms outperform traditional linear regression.

Machine learning is very good at finding statistical patterns through a mass of numbers, but those patterns are merely correlations amongst vast reams of data, rather than causative truths. As with any data-driven method, the data quality has a huge impact on the usefulness of the model output.

The principle of ‘garbage in, garbage out’ is also valid in this new quantitative world. For this reason, we believe investment managers must give an economic meaning to machine learning algorithms.

This groundbreaking book on *Machine Learning for Asset Management* represents a refreshing collaborative effort between sophisticated investment practitioners and researchers, to present practical application of machine learning methodologies. As one can see from the different chapters, machine learning can be applied to different parts of the investment process, from stock picking to tactical allocation, alpha signal enhancement or trading.

As a result, this comprehensive volume is a powerful tool to help practitioners keep abreast of developments in this fast-changing field, and to implement machine-learning methods into their investment value chain.

The future of asset management will likely involve a synthesis of human and artificial intelligence that harnesses the power of both. However, we need to clearly define the sharing of control between machine and manager. Whereas computers excel in responding to well-formulated questions with clear objectives, humans remain key in asking the right questions and interpreting the results.

Fiona FRICK
Group CEO, Unigestion

Acknowledgments

Firstly, I would like to thank the editorial team from ISTE. I would also like to express my gratitude to all the authors of the book for their contribution and trust. Finally, I would like to give special and warm thanks to Jérôme Teiletche, Head of Cross-Asset solutions at Unigestion, who introduced me to the subtleties of machine learning techniques in asset management.

1

Time-series and Cross-sectional Stock Return Forecasting: New Machine Learning Methods

This chapter extends the machine learning methods developed in Han *et al.* (2019) for forecasting cross-sectional stock returns to a time-series context. The methods use the elastic net to refine the simple combination return forecast from Rapach *et al.* (2010). In a time-series application focused on forecasting the US market excess return using a large number of potential predictors, we find that the elastic net refinement substantively improves the simple combination forecast, thereby providing one of the best market excess return forecasts to date. We also discuss the cross-sectional return forecasts developed in Han *et al.* (2019), highlighting how machine learning methods can be used to improve combination forecasts in both the time-series and cross-sectional dimensions. Overall, because many important questions in finance are related to time-series or cross-sectional return forecasts, the machine learning methods discussed in this chapter should provide valuable tools to researchers and practitioners alike.

1.1. Introduction

Researchers in finance increasingly rely on machine learning techniques to analyze Big Data. The initial application of the *least absolute shrinkage and selection operator* (Tibshirani 1996, LASSO) – one of the most popular machine learning techniques – in finance appears to be Rapach *et al.* (2013), who analyze lead-lag relationships among monthly international equity

Chapter written by David E. RAPACH and Guofu ZHOU.

Machine Learning for Asset Management: New Developments and Financial Applications, First Edition. Edited by Emmanuel Jurczenko.

returns in a high-dimensional setting. More recently, Gu *et al.* (2019) employ a comprehensive set of machine learning tools, including the LASSO, to analyze the time-series predictability of monthly individual stock returns, while Chinco *et al.* (2019) use the LASSO to predict individual stock returns one minute ahead. Freyberger *et al.* (2019) apply a nonparametric version of the LASSO to accommodate nonlinear relationships between numerous firm characteristics and cross-sectional stock returns. Kozak *et al.* (2019) use the LASSO in a Bayesian context to model the stochastic discount factor based on a large number of firm characteristics. Incorporating insights from Bates and Granger (1969); Diebold and Shin (2019), Han *et al.* (2019) propose procedures for forecasting cross-sectional returns using the information in more than 100 firm characteristics[1].

In this chapter, we show how the approach of Han *et al.* (2019), originally designed for forecasting cross-sectional stock returns, can be modified for time-series forecasting of the market excess return. A voluminous literature investigates market excess return predictability based on a wide variety of predictor variables[2]. In the presence of a large number of potential predictor variables, conventional forecasting methods are susceptible to in-sample overfitting, which often translates into poor out-of-sample performance. Rapach *et al.* (2010) employ forecast combination (Bates and Granger 1969) to incorporate the information in a large number of predictor variables in a manner that guards against overfitting. They find that a simple combination forecast – the average of univariate predictive regression forecasts based on the individual predictor variables – substantially improves out-of-sample forecasts of the US market excess return. Extending the methods of Han *et al.* (2019) to a time-series context along the lines of Diebold and Shin (2019), we describe how the *elastic net* (Zou and Hastie 2005), a well-known variant of the LASSO, can be used to refine the simple combination forecast, resulting in what we call the *combination elastic net* forecast. Intuitively, as explained by Han *et al.* (2019), the elastic net refinement allows us to more efficiently use the information in the predictor variables by selecting the most relevant predictors to include in the combination forecast. In an empirical application, we show that the combination elastic net approach indeed improves the

1 We focus on numerical data in this chapter. For textual analysis see, for example, Tetlock (2007); Loughran and McDonald (2011); Ke *et al.* (2019).

2 See Rapach and Zhou (2013) for a survey of the literature.

accuracy of US market excess return forecasts and provides substantive economic value to a mean-variance investor. Overall, our combination elastic net forecast appears to be among the best market excess return forecasts to date.

The rest of the chapter is organized as follows. Section 1.2 describes the construction of market excess return forecasts, including the combination elastic net forecast. Section 1.3 reports results for an empirical application centered on forecasting the US market excess return, using a variety of predictor variables from the literature. Section 1.4 outlines the construction of the cross-sectional return forecasts proposed by Han *et al*. (2019). Section 1.5 concludes this chapter.

1.2. Time-series return forecasts

1.2.1. *Predictive regression*

Stock market excess return predictability is typically analyzed in the context of a univariate predictive regression model:

$$r_t = \alpha + \beta x_{j,t-1} + \varepsilon_t, \tag{1.1}$$

where r_t is the period-t return on a broad stock market index in excess of the risk-free return, $x_{j,t}$ is the predictor variable, and ε_t is a zero-mean disturbance term. It is straightforward to use equation [1.1] to generate an out-of-sample forecast of r_{t+1} based on $x_{j,t}$ and data available through period t:

$$\hat{r}^{(j)}_{t+1|t} = \hat{\alpha}^{(j)}_{1:t} + \hat{\beta}^{(j)}_{1:t} x_{j,t}, \tag{1.2}$$

where $\hat{\alpha}^{(j)}_{1:t}$ and $\hat{\beta}^{(j)}_{1:t}$ are the ordinary least squares (OLS) estimates of α and β, respectively, in equation [1.1] based on data available from the start of the sample through t (i.e. the period of forecast formation).

Because there are a plethora of plausible predictor variables, it is advisable to aggregate information when forecasting the market excess return. The most obvious approach for incorporating information from multiple predictor variables is to specify a multiple predictive regression model:

$$r_t = \alpha + \sum_{j=1}^{J} \beta_j x_{j,t-1} + \varepsilon_t. \qquad [1.3]$$

It is again straightforward to use equation [1.3] to generate an out-of-sample forecast of r_{t+1} based on $x_{j,t}$ for $j = 1, \ldots, J$ and data available through t:

$$\hat{r}_{t+1|t}^{\text{OLS}} = \hat{\alpha}_{1:t}^{\text{OLS}} + \sum_{j=1}^{J} \hat{\beta}_{j,1:t}^{\text{OLS}} x_{j,t}, \qquad [1.4]$$

where $\hat{\alpha}_{1:t}^{\text{OLS}}$ and $\hat{\beta}_{j,1:t}^{\text{OLS}}$ are the OLS estimates of α and β_j, respectively, for $j = 1, \ldots, J$ in equation [1.3] based on data available through t.

Although the out-of-sample market excess return forecasts in equation [1.2] and [1.4] are easy to obtain, Goyal and Welch (2008) find that such forecasts, based on numerous popular predictor variables from the literature, fail to outperform the naive prevailing mean benchmark forecast on a consistent basis over time (as judged by the out-of-sample R^2 statistic, which we define below). The prevailing mean forecast ignores information in any predictor variable; it is simply the historical average excess return based on data available through t:

$$\hat{r}_{t+1|t}^{\text{PM}} = \frac{1}{t} \sum_{s=1}^{t} r_s. \qquad [1.5]$$

The prevailing mean forecast corresponds to the following simple data-generating process for the market excess return:

$$r_t = \alpha + \varepsilon_t, \qquad [1.6]$$

namely, the constant expected excess return (or random walk with drift) model. The Goyal and Welch (2008) findings pose important challenges for out-of-sample return predictability, as they indicate that exploiting the information in popular predictor variables via conventional regression methods does not improve forecast accuracy.

1.2.2. *Forecast combination*

The study of Goyal and Welch (2008) was influential in stimulating thinking about how to better use the information in predictor variables to forecast the market excess return. With respect to the univariate predictive regression forecast in equation [1.2], Rapach *et al*. (2010) argue that it is risky to rely on a single predictor variable, due to factors such as investor learning and structural change. Building on the seminal work of Bates and Granger (1969), Rapach *et al*. (2010) recommend forecast combination as a strategy for incorporating information from a variety of predictor variables. Forecast combination reduces forecast "risk" by diversifying across individual forecasts, similarly to diversifying across assets to reduce portfolio risk (Timmermann 2006). Specifically, Rapach *et al*. (2010) consider a combination forecast that takes the form of a simple average of the univariate predictive regression forecasts, based on $x_{j,t}$ for $j = 1, \ldots, J$ in equation [1.2]:

$$\hat{r}^{\mathrm{C}}_{t+1|t} = \frac{1}{J}\sum_{j=1}^{J}\hat{r}^{(j)}_{t+1|t}. \qquad [1.7]$$

They show that, in contrast to the conventional univariate and multiple predictive regression forecasts in equations [1.2] and [1.4], respectively, the combination forecast in equation [1.7] is able to deliver out-of-sample accuracy gains relative to the prevailing mean forecast, on a much more consistent basis over time.

How is it that – unlike the conventional multiple predictive regression forecast in equation [1.4], which also includes information from $x_{j,t}$ for $j = 1, \ldots, J$ – the combination forecast in equation [1.7] is able to improve out-of-sample performance? Rapach *et al*. (2010) point out that forecast combination is effectively a strong shrinkage estimator. They show that the combination forecast in equation [1.7] makes two adjustments to the conventional multiple predictive regression forecast in equation [1.4]: first, it replaces the OLS multiple regression coefficient estimates with their univariate counterparts, which reduces the role of multi-collinearity in producing imprecise parameter estimates; second, the combination forecast shrinks the univariate slope coefficients by the factor $1/J$, thereby shrinking the forecast to the prevailing mean benchmark.

The usefulness of shrinkage for improving out-of-sample market excess return forecasts stems from a delicate balance required for stock return forecasting. On the one hand, we want to incorporate information from a wide variety of potentially relevant predictor variables, especially since we do not want to neglect relevant information and cannot know *a priori* which predictors are the most relevant. On the other hand, incorporating information from numerous predictors via the multiple prediction regression forecast in equation [1.4] is inadvisable. Equation [1.4] is based on conventional estimation of the multiple predictive regression model in equation [1.3], which is susceptible to overfitting. Conventional OLS estimation maximizes the explanatory ability of the model over the estimation sample, which often leads to poor out-of-sample performance. Overfitting concerns are exacerbated as the number of explanatory variables increases and the signal-to-noise ratio in the data decreases. We encounter both of these challenges when forecasting stock returns: there are numerous plausible predictor variables, and the predictable component in returns is inherently limited. The combination forecast in equation [1.7] apparently provides an effective shrinkage strategy for incorporating information from numerous plausible predictor variables in a manner that avoids overfitting.

1.2.3. *Elastic net*

Machine learning techniques also provide a means for implementing shrinkage. Indeed, the popular LASSO estimator is a penalized regression approach that is explicitly designed to prevent overfitting via shrinkage. To compute a forecast based on the multiple predictive regression model in equation [1.3], instead of the OLS objective function, we estimate the coefficients using the LASSO objective function:

$$\underset{\alpha,\beta_1,\ldots,\beta_J \in \mathbb{R}}{\arg\min} \left[\frac{1}{2t} \sum_{s=1}^{t} \left(r_s - \alpha - \sum_{j=1}^{J} \beta_j x_{j,s-1} \right)^2 + \lambda \sum_{j=1}^{J} |\beta_j| \right], \quad [1.8]$$

where $\lambda \geq 0$ is a regularization parameter that controls the degree of shrinkage[3]. The first component of the LASSO objective function is the

3 Following standard practice, the predictor variables are standardized to have zero mean and unit variance before entering equation [1.8]. The final parameter estimates reflect the original scales of the predictor variables.

familiar sum of squared fitted residuals, so that the LASSO and OLS estimators coincide when $\lambda = 0$. The regularization parameter λ shrinks the coefficients towards zero. Unlike rigde regression (Hoerl and Kennard 1970), which relies on an ℓ_2 penalty term, the LASSO employs an ℓ_1 penalty, so that it permits shrinkage to exactly zero (for sufficiently large λ). Shrinkage to zero means that the LASSO also performs variable selection, which facilitates the interpretation of the fitted model.

To implement LASSO estimation, it is necessary to choose the value for λ. The most popular approach is K-fold cross-validation. However, the selection of the number of folds K and construction of the folds are largely arbitrary. The Hurvich and Tsai (1989) corrected version of the Akaike information criterion (Akaike 1973, AIC) provides an alternative to K-fold cross-validation for choosing λ. The corrected AIC is simpler to use in that it does not require arbitrary choices for the number and type of folds. Furthermore, Flynn *et al.* (2013) show that the corrected AIC has good asymptotic and finite-sample properties for choosing λ.

Zou and Hastie (2005) propose the elastic net (ENet) as a refinement to the LASSO that includes both ℓ_1 and ℓ_2 components in the penalty term. The ENet estimator is defined by the following objective function:

$$\underset{\alpha,\beta_1,\ldots,\beta_J\in\mathbb{R}}{\arg\min}\left[\frac{1}{2t}\sum_{s=1}^{t}\left(r_s-\alpha-\sum_{j=1}^{J}\beta_j x_{j,s-1}\right)^2+\lambda P_\delta\left(\beta_1,\ldots,\beta_J\right)\right], \quad [1.9]$$

where

$$P_\delta\left(\beta_1,\ldots,\beta_J\right)=0.5(1-\delta)\sum_{j=1}^{J}\beta_j^2+\delta\sum_{j=1}^{J}|\beta_j| \quad [1.10]$$

and $0 \leq \delta \leq 1$ is a parameter for blending the ℓ_1 and ℓ_2 components. A potential drawback of the LASSO is that it tends to somewhat arbitrarily select one predictor from a group of highly correlated predictors. In contrast, using $\delta = 0.5$ in equation [1.9] results in a stronger tendency to select the highly correlated predictors as a group (Hastie and Qian 2016). The corrected AIC can again be used to choose λ in equation [1.9].

A market excess return forecast based on ENet estimation of the multiple predictive regression model in equation [1.3] is given by:

$$\hat{r}_{t+1|t}^{\mathrm{ENet}} = \hat{\alpha}_{1:t}^{\mathrm{ENet}} + \sum_{j=1}^{J} \beta_{j,1:t}^{\mathrm{ENet}} x_{j,t}, \qquad [1.11]$$

where $\hat{\alpha}_{1:t}^{\mathrm{ENet}}$ and $\hat{\beta}_{j,1:t}^{\mathrm{ENet}}$ are the ENet estimates of α and β_j, respectively, for $j = 1, \ldots, J$ in equation [1.3]. Intuitively, we rely on the shrinkage properties of the ENet to generate a market excess return forecast that incorporates information from a potentially large number of predictor variables in a manner that guards against overfitting. Whether the ENet is an effective shrinkage strategy for forecasting the market excess return is ultimately an empirical issue. We investigate this issue in our empirical application in section 1.3[4].

1.2.4. *Combination elastic net*

Incorporating insights from Diebold and Shin (2019), we can also use machine learning techniques to refine the combination forecast in equation [1.7]. A potential drawback to equation [1.7] is that it may "overshrink" the forecast to the prevailing mean, thereby neglecting substantive relevant information in the predictor variables. In an effort to improve the combination forecast by exploiting more of the relevant information in the predictor variables (while still avoiding overfitting), we consider the following Granger and Ramanathan (1984) regression:

$$r_t = \eta + \sum_{j=1}^{J} \theta_j \hat{r}_{t|t-1}^{(j)} + \varepsilon_t, \qquad [1.12]$$

which we estimate via the elastic net to select the most relevant univariate forecasts to include in the combination forecast[5]. Specifically, to construct the combination elastic net (C-ENet) forecast, we first need to define an initial

4 We focus on the results for the ENet in our empirical application in section 1.3, although the results are qualitatively similar for the LASSO.

5 Again, the results are qualitatively similar in our empirical application in section 1.3 if we use the LASSO to estimate equation [1.12].

in-sample estimation period and corresponding holdout out-of-sample period; let t_1 denote the size of the initial in-sample period. We then proceed in three steps:

Step 1 For each predictor variable, we compute recursive univariate predictive regression forecasts based on equation [1.2] over the holdout out-of-sample period:

$$\hat{r}^{(j)}_{s|s-1} = \hat{\alpha}^{(j)}_{1:s-1} + \hat{\beta}^{(j)}_{1:s-1} x_{j,s-1}, \qquad [1.13]$$

for $s = t_1 + 1, \ldots, t$ and $j = 1, \ldots, J$.

Step 2 We estimate the Granger and Ramanathan (1984) regression in equation [1.12] via the ENet over the holdout out-of-sample period:

$$r_s = \eta + \sum_{j=1}^{J} \theta_j \hat{r}^{(j)}_{s|s-1} + \varepsilon_s, \qquad [1.14]$$

for $s = t_1 + 1, \ldots, t$. Let $\mathcal{J}_t \subseteq \{1, \ldots, J\}$ denote the index set of individual univariate predictive regression forecasts selected by the ENet in equation [1.14]. When estimating equation [1.14], we impose the restriction that $\theta_j \geq 0$ for $j = 1, \ldots, J$. This imposes the economically reasonable requirement that a univariate market excess return forecast be positively related to the realized excess return in order to be selected by the ENet in equation [1.14].

Step 3 We compute the C-ENet forecast as:

$$\hat{r}^{\text{C-ENet}}_{t+1|t} = \frac{1}{|\mathcal{J}_t|} \sum_{j \in \mathcal{J}_t} \hat{r}^{(j)}_{t+1|t}, \qquad [1.15]$$

where $|\mathcal{J}_t|$ is the cardinality of $\mathcal{J}_t$ and $\hat{r}^{(j)}_{t+1|t}$ is given by equation [1.2] for $j = 1, \ldots, J$.

The usefulness of the C-ENet approach for capturing the relevant information in numerous predictor variables, in a manner that guards against overfitting, is again ultimately an empirical issue. In our empirical application in section 1.3, we find that the C-ENet approach is indeed an effective strategy for forecasting the market excess return[6].

6 Observe that all the forecasts that we construct only use data available through t to forecast r_{t+1}, so that the forecasts do not entail “look-ahead” bias.

1.3. Empirical application

1.3.1. *Data*

We investigate the performance of the strategies discussed in section 1.2 for forecasting the monthly S&P 500 excess return. Using data available from Amit Goyal's website[7], we measure the excess return as the CRSP value-weighted S&P 500 return in excess of the risk-free return (based on the Treasury bill rate).

We consider 12 plausible predictor variables, which are illustrative of popular predictors used by academics and practitioners alike:

– *Log dividend-price ratio* (*DP*). Log of the 12-month moving sum of S&P 500 dividends minus the log of the S&P 500 price index.

– *Log earnings-price ratio* (*EP*). Log of the 12-month moving sum of S&P 500 earnings minus the log of the S&P 500 price index.

– *Volatility* (*VOL*). We follow Mele (2007) in measuring the annualized volatility for month t as $\sqrt{\frac{\pi}{2}}\sqrt{12}\hat{\sigma}_t$, where $\hat{\sigma}_t = \frac{1}{12}\sum_{s=1}^{12}|r_{t-(s-1)}|$.

– *Treasury bill yield* (*BILL*). Three-month Treasury bill yield minus the 12-month moving average of the three-month Treasury bill yield.

– *Treasury bond yield* (*BOND*). Ten-year Treasury bond yield minus the 12-month moving average of the ten-year Treasury bond yield.

– *Term spread* (*TERM*). Difference in yields on a ten-year treasury bond and a three-month treasury bill.

– *Credit spread* (*CREDIT*). Difference in yields on a AAA-rated corporate bond and a ten-year treasury bond.

– *Inflation* (*PPIG*). Producer price index (PPI) inflation rate.

– *Industrial production growth* (*IPG*). Growth rate of industrial production.

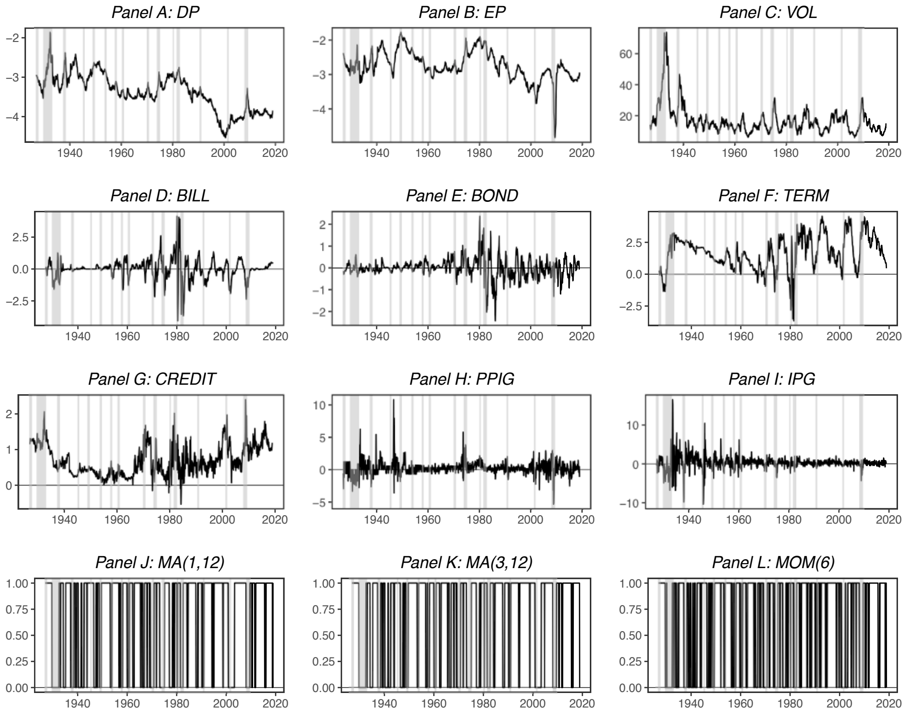
– *MA(1,12) technical signal* [*MA(1,12)*]. An indicator variable that takes a value of one (zero) if the S&P 500 price index is greater than or equal to (less than) the 12-month moving average of the S&P 500 price index.

– *MA(3,12) technical signal* [*MA(3,12)*]. An indicator variable that takes a value of one (zero) if the three-month moving average of the S&P 500 price

7 http://www.hec.unil.ch/agoyal/.

index is greater than or equal to (less than) the 12-month moving average of the S&P 500 price index.

– *Momentum technical signal* [*MOM(6)*]. An indicator variable that takes a value of one (zero) if the S&P 500 price index is greater than or equal to (less than) its value six months ago.



The figure depicts 12 predictor variables for 1927:01 to 2018:12. The predictor variable definitions are provided in section 1.3.1. Vertical bars delineate business-cycle recessions as dated by the National Bureau of Economic Research.

Figure 1.1. *Predictor variables*

The data used to construct the predictor variables are from Amit Goyal's website and the Federal Reserve Bank of St. Louis's Federal Reserve Economic Data (FRED)[8]. We account for the one-month publication lag in *PPIG* and *IPG*. We follow Neely *et al.* (2014) in defining indicator variables to include information from technical signals. Figure 1.1 portrays the 12

8 https://fred.stlouisfed.org/.

predictor variables for the 1927:01 to 2018:12 sample period. Visually, the predictors represent a variety of information sources.

1.3.2. *Forecasts*

We reserve the first two decades of the sample (1927:01 to 1946:12) as the initial in-sample estimation period. This provides us with an adequate number of observations to reasonably reliably estimate the predictive regression coefficients in equations [1.1] and [1.3]. The initial holdout out-of-sample period for computing the C-ENet forecast covers 1947:01 to 1956:12, so that we evaluate the out-of-sample forecasts for 1957:01 to 2018:12. The out-of-sample forecast evaluation period covers more than six decades, which allows us to analyze return predictability under a variety of economic conditions.

Figure 1.2 depicts the recursive slope coefficient estimates used to compute the predictive regression forecasts in equations [1.2], [1.4] and [1.11]. The black line in each panel delineates the OLS slope coefficient estimates for the multiple predictive regression model in equation [1.3]. The recursive estimates point to problems with the conventional OLS estimates of the multiple predictive regression model slope coefficients – the estimates often have the "wrong" sign (e.g. DP and $BILL$) and reach extreme values, which are manifestations of in-sample overfitting. Overfitting is not surprising when we rely on conventional methods to estimate a relatively high-dimensional predictive regression model in a noisy environment.

The blue line in each panel of Figure 1.2 delineates the recursive OLS slope coefficient estimates for the univariate predictive regression model in equation [1.1]. Compared to the recursive slope coefficient estimates for the multiple predictive regression model, the univariate estimates are generally much more stable. This reflects the increase in estimation precision afforded by the mitigation of multi-collinearity. Of course, the univariate estimates are potentially biased (due to omitted variable bias). However, in light of the bias-efficiency trade-off, the increase in estimation precision can outweigh the cost of the bias for the purpose of out-of-sample forecasting.