Project report: time-varying predictability of stock return with high-dimensional data*

Zhendong Sun†

This version: August 29, 2022

1 Introduction

In this project, we discuss the predictability of stock return, which is one of the most attractive and difficult topics in finance. The challenge of predicting stock return is generally threefold: (a) extremely low signal-to-noise ratio, (b) time-varying validness of predictors, and (c) a large number of available predictors. To illustrate these, we use the figure of correlation coefficients between the return of S&P 500 index and predictors. Later we will give more details about the data description.

As shown in Figure (1.a), the (absolute) correlation coefficients between the predictors and stock return are generally below 0.15, which implies the predictors are not very informative for the stock return[1]. This property makes the task more difficult when the number of candidate predictors is large because we need to pick up some useful information from many uninformative variables.

Figure (1.b) illustrates another challenge for predicting stock, the time-varying validness of predictors. In the figure, the color denotes the value of correlation, the y-axis is the name of predictors, and the x-axis is the year. The figure shows a clear pattern that the correlations between most predictors and stock return are changing over years, which implies the validness of predictors is changing. In the practice, we could improve the predictive performance of stock return if we can capture the variation.

To tackle the challenges of predicting stock return, we propose a method that can deal with low signal-to-noise ratio and time-varying validness in a high-dimensional scenario. The idea of the method is built on the time-varying parameters (TVP) model that has been popular in recent years in empirical macroeconomics and finance. We incorporate the TVP model with a model combination method that can better handle the noisy data as the stock return. We name the proposed model as the dynamic selection and combination (DSC) model because it can dynamically select and combine predictors at each time. To apply the proposed DSC model efficiently, we develop a particle filter within variational Bayes (PF-VB) method. We will outline the methodology in the next section.

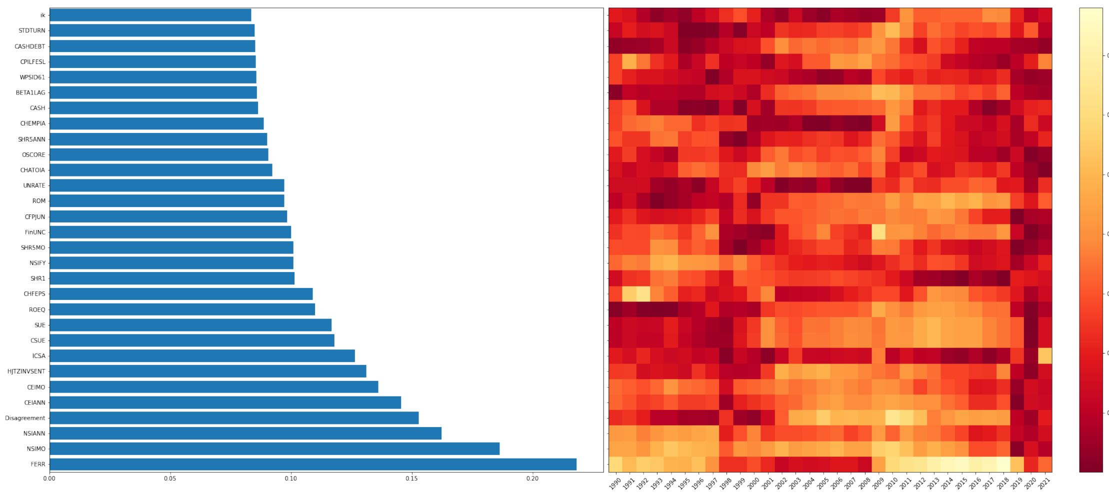
With the proposed model, we investigate the predictability of return of S&P 500 index with over 100 predictors that are used in previous studies. We find the proposed method yield significantly better performance statistically and economically compared with the standard linear model as well as some machine learning methods. We also discuss the rationale behind the outperformance, including the pattern of variable selection and the relation to the real economy.

*This is a summarized report of my paper, Time-varying predictability of stock return with high-dimensional data. The full paper and replication code are available here.

†University of Illinois at Urbana-Champaign, zs6@illinois.edu

[1]In this figure, we only show 30 predictors with the highest correlation coefficients due to the limitation of space, which implies the correlation of most predictors in the whole dataset are smaller than 0.1.

Figure 1: Correlation coefficients between predictors and stock return



(a) Correlation between predictors and stock return (b) Correlation over time

In the section of the report, we will outline the methodology of the paper. Then we will analyze the data with the proposed method and competitors. In the end, we will wrap up this report with a short conclusion section.

2 Methodology

The proposed method is fundamentally a forecast combination model which contains two stages. In the first stage, we generate the forecasts by each predictor through linear predictive regression, which extracts the correlated component for the stock return from each predictor. After that, we include all the forecasts into a time-varying parameters linear model. We incorporate the dynamic variable selection prior (DVS) to deal with the high-dimensional data.

Following the convention in much empirical literature, we start with a linear predictive regression model to describe the relation between stock return and a set of predictors, which can be written as

$$r_{t+1} = \alpha_i + \theta_i x_{i,t} + \epsilon_{t+1} \tag{1}$$

where r_{t+1} is the log return on a stock index minus the risk-free interest rate, and $x_{i,t}$ is the predictor of interest. Based on the equation, we can easily obtain an out-of-sample forecast of r_t:

$$\hat{r}_{i,t+1} = \hat{\alpha}_i + \hat{\theta}_i x_{i,t} \tag{2}$$

where $\hat{\alpha}_i$ and $\hat{\theta}_i$ are the estimators based on the observations available up to period t, and $\hat{r}_{i,t+1}$ is the forecast by $x_{i,t}$.

To better utilize the information from multiple predictors, previous studies advise combining the forecasts by each predictor. Namely, we can consider the forecast combination as a linear regression model:

$$r_{t+1} = \Theta \hat{R}_{t+1} + e_t \tag{3}$$

where $\hat{R}_{t+1}$ is a vector of forecasts generated by equation (2), Θ is the vector of corresponding combination weight, and we assume $e_t \sim N(0, H)$.

Because we aim to investigate the time-varying predictability of stock return, it is necessary to specify the model with dynamic coefficients (or weights) rather than fix it. In particular, we specify the model with time-varying combination weights as:

$$
\begin{aligned}
r_{t+1} &= \Theta_t \hat{R}_{t+1} + e_t && (4)\\
\Theta_t &= \Theta_{t-1} + \eta_t && (5)\\
H_t &= H_{t-1} + \xi_t && (6)
\end{aligned}
$$

where we assume $e_t \sim N(0, H_t)$, $\eta_t \sim N(0, Q_t)$ and $\xi_t \sim N(0, W_t)$. Moreover, we restrict $\forall \theta_{i,t} > 0$, where $\theta_{i,t}$ denotes the ith element of vector, because it has an economic meaning to restrict the forecasts by each individual predictors to be positively related to the realized excess market return.

The linear predictive model with a large number of variables always performs poorly in the practice of out-of-sample prediction due to the overfitting issue, so we need to incorporate a procedure of dimensionality reduction so that we can find a smaller counterpart of the original large model. In this project, we incorporate the dynamic variable selection prior (DVS) into the proposed dynamic combination model described in equation:

$$
\begin{aligned}
\theta_{i,t} | \gamma_{i,t}, \tau_{i,t} &\sim (1 - \gamma_{i,t}) N(0, \nu \times \tau_{i,t}^2) + \gamma_{i,t} N(0, \tau_{i,t}^2) && (7)\\
\gamma_{i,t} | \pi_t &\sim Bernoulli(\pi_{0,t}) && (8)\\
\tau_{i,t}^{-2} &\sim Gamma(g_0, h_0) && (9)
\end{aligned}
$$

where $i = 1, 2, \ldots, I$, and ν, g_0 and h_0 are predetermined hyperparameters for the priors. In practice, we setup a large value of $\tau_{i,t}$ while the ν is small. By doing so, the prior concentrates at zero when $\gamma_{i,t} = 0$, otherwise, it becomes uninformative because of the large variance.

To estimate the state-space model of equation (4) to (6) with the DVS prior, we need to combine two prior distributions described as below:

$$
\begin{aligned}
\Theta_t | \Theta_{t-1}, Q_t &\sim N(\Theta_{t-1}, Q_t) && (10)\\
\Theta_t | V_t &\sim N(0, V_t) && (11)
\end{aligned}
$$

where $V_t = diag(v_{1,t}, v_{2,t}, \ldots, v_{I,t})$, and $v_{i,t} = (1 - \gamma_{i,t}) N(0, \nu \times \tau_{i,t}^2) + \gamma_{i,t} N(0, \tau_{i,t}^2)$ for $i = 1, \ldots, I$. Equation (10) is equivalent to the state equation (5). We can combine these two priors by rewriting the state equation (5):

$$
\Theta_t = \tilde{F}_t \Theta_{t-1} + \tilde{\eta}_t \qquad (12)
$$

where $\tilde{\eta}_t \sim N(0, \tilde{Q}_t)$, $\tilde{Q}_t = (Q_t^{-1} + V_t^{-1})^{-1}$, and $\tilde{F}_t = \tilde{Q}_t \times Q_t^{-1}$.

To estimate the model efficiently, we propose a particle filter within variational Bayes (PF-VB) method. The algorithm is outlined in appendix, and the detail of the methods and replication code can be found here.

3 Empirical results

In this project, we evaluate the performance of methodology by using the monthly excess stock return in the US with 143 predictors ranging from January 1926 to December 2019. These predictors have been investigated in previous research. Table (A1) in appendix gives a brief description and data source of all the included predictors.

We compare the proposed DSC model with alternatives from the previous research, and we divide those methods into three categories. The first category focuses on the high-dimensional set of predictors

Table 1: **Alternative models**

Model	Description
Category I	
Model	**Description**
PCA	Use principal component analysis (PCA) over an expanding window
ENet	Use elastic net (ENet) over an expanding window
CENet	Use combination Enet (CEnet) over an expanding window
Category II	
Model	**Description**
RF	Random forest (RF)
NN1	Neural network (NN) with one layer of 32 neurons
NN2	Neural network (NN) with two layers, one with 32 and one with 16 neurons
Category III	
Model	**Description**
DSC-ORIG	Use origin estimated Θ without normalization to generate prediction
DSC-NORM	Normalize Θ to sum to unity, then generate prediction
DSC-EQ	Equalize the non-zero elements of Θ, normalize them to sum to unity, and then generate prediction

while assuming the predictive model is linear with fixed parameters. This category includes PCA, ENet, and combination ENet (CENet). The methods from the second category can capture the nonlinearity in a data-rich environment. This category includes Random forest (RF) and neural network (NN). The last categroy include the proposed DSC model with different normalization weights. Table (1) describes all the methods and corresponding setup for comparison.

To evaluate the statistical accuracy of forecasts, we report the out-of-sample R^2_{OOS} computed by:

$$R^2_{OOS} = 1 - \frac{\sum_{t=\bar{t}}^{T}(r_t - \hat{r}_t)^2}{\sum_{t=\bar{t}}^{T}(r_t - \bar{r}_t)^2} \tag{13}$$

where $\hat{r}_t$ is the forecasts by the model of interest and $\bar{r}_t$ denotes the historical mean of stock return. A positive R^2_{OOS} statistic implies the model of interest is able to extract the predictable part in the monthly stock return. In practice, a small value of R^2_{OOS} statistic can generate significant degree of excess return.

Figure 2: Statistical Evaluation

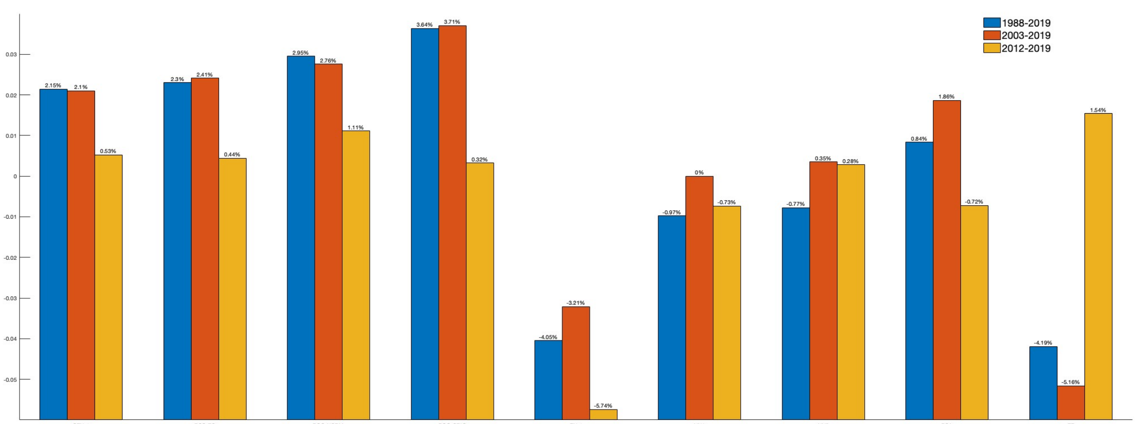


Figure (2) summarizes the statistical evaluation for different models. The R^2_{OOS} statistic is generally small (or negative) for all the models, which implies that the predictable part of a stock return is very limited. However, it is still possible to extract signals from a large number of predictors that help to obtain a sizable R^2_{OOS}. For instance, the CENet model yields an R^2_{OOS} of 2.23% during 1988–2019. The PCA model produces an R^2_{OOS} of 0.84% during 1988–2019. The RF yields an R^2_{OOS} of 0.95% during 2012–2019.

Among all the methods, the proposed DSC model is the only model that yields a sizable R^2_{OOS} in all periods, and it also beats alternatives during most periods. The DSC model generates an overall R^2_{OOS} of 3.23% by the DSC-ORIG, 2.94% by the DSC-NORM, and 2.67% by the DSC-EQ. Compared with all other models, the CENet achieves the most similar performance with the out-of-sample R^2 of 2.23% relative to the DSC model, which implies the gains in performance could mainly result from allowing the time variations in combination weights.

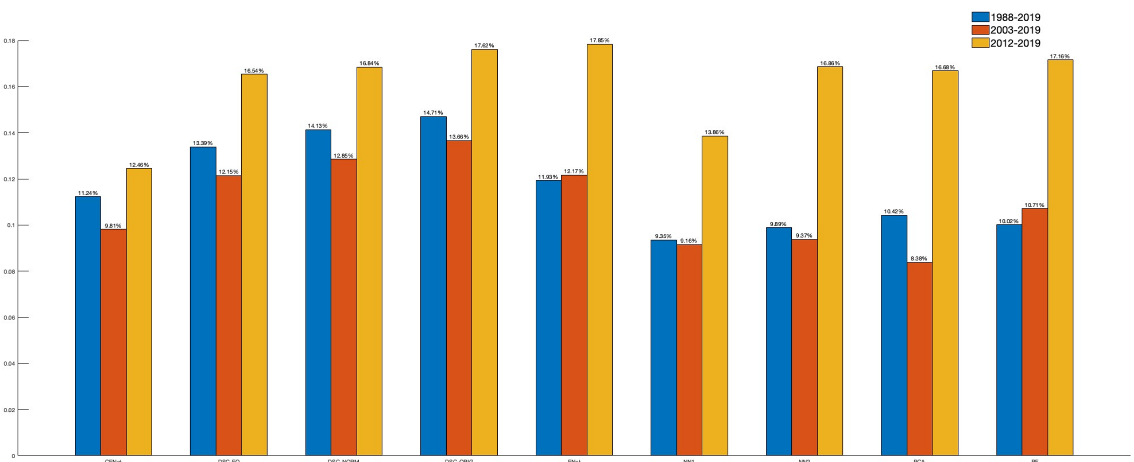
Besides the statistical predictability, we also want to discuss whether we can get economic gains for an investor in an asset allocation context. In particular, we consider the allocation to equities by an investor with a single-period horizon and the mean-variance preferences:

$$w^*_{t+1} = \left(\frac{1}{\gamma}\right)\left(\frac{\hat{r}_{t+1}}{\hat{\sigma}^2_{t+1}}\right) \tag{14}$$

In practice, we assume $\gamma = 3$ and estimate $\hat{\sigma}_{t+1}$ by sample variance over a 60-month rolling window. We restrict the weight of equities ω_{t+1} from -1 to 2, which ensures the allocation is feasible in reality.

Figure (3) reports the annualized market excess return of the optimal portfolio based on the forecasts by different models. Unlike the statistical evaluation, most of the models yield a sizable return that reflects the overall prosperity of the U.S. stock market over past decades. Among all the methods, the DSC model with different normalized weights can generate the highest annualized excess return of 17.12% by the DSC-ORIG, 16.99% by the DSC-NORM, and 16.47% by the DSC-EQ during all periods, which leads to a large profit in a long period. The best performance is achieved by the DSC-NORM model during 2012–2019 and yields an annualized return of 20.65%. Similar to the result by statistical evaluation, the CENet achieves the closest performance to the DSC model with an annualized return of 16.02%.

Figure 3: Economic Evaluation

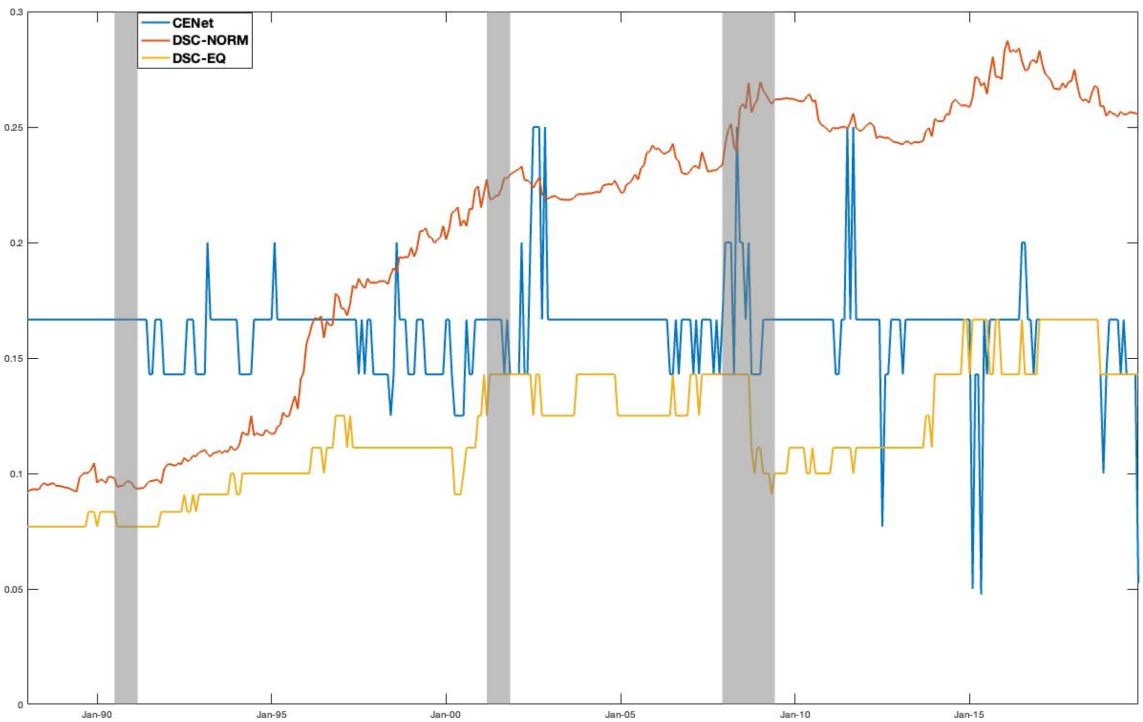


After reviewing the overall performance of the DSC model, we will discuss the properties of predictability in more detail.

Because the main ability of the DSC model is to dynamically select and combine predictors, we may ask whether the outperformance of the DSC model results from its capability of capturing the time-varying validness of individual predictors. To answer the question, we discuss the relationships among the variables selected and combined by the DSC model, the overall performance of the model, and the performance of each individual predictor in the following subsections.

The first metric is the *concentration* of the model, which measures the distribution of the combination weights at each time. We define the *concentration* of the model at time t as

Figure 4: Concentration of the model



$$C_t = \sum_{i=1}^{I} \theta_{i,t}^2 \tag{15}$$

where $\theta_{i,t}$ is the combination weights assigned to predictor i at time t. The value of C ranges from 0 to 1, and a higher value of C_t implies that most of the combination weights are assigned to a few predictors; that is, the model is more concentrated.

Figure (4) shows the concentration of the CENet, DSC-NORM, and DSC-EQ. Among these models, the CENet keeps a stable level of concentration with some occasional spikes that could result from noise in the data. Both DSC-NORM and DSC-EQ have an increasing concentration over time, while the trend of DSC-EQ is less significant because the model will smooth the change by equalizing the weights of all the selected predictors. Compared with the relatively constant concentration of the CENet, the increasing trend in concentration could be more reasonable based on the evidence from previous research because there are many fewer predictors that could generate positive out-of-sample R^2 after 1988 compared with the period of 1947–2002.

In addition to the concentration of models, the variations in combination weights could also affect the forecasting performance. In this part, we only focus on the variation of the DSC-NORM model because the other methods will induce extra variation by equalizing the weights of selected predictors.

To measure the variation of the model, we define the metric V_t as:

$$V_t = \sum_{i=1}^{I} (\theta_{i,t} - \theta_{i,t-1})^2 \tag{16}$$

where $\theta_{i,t}$ is the combination weights assigned to predictor i at time t. A higher value of V_t implies more volatile combination weights at time t.

Figure 5: Variation in the combination weights

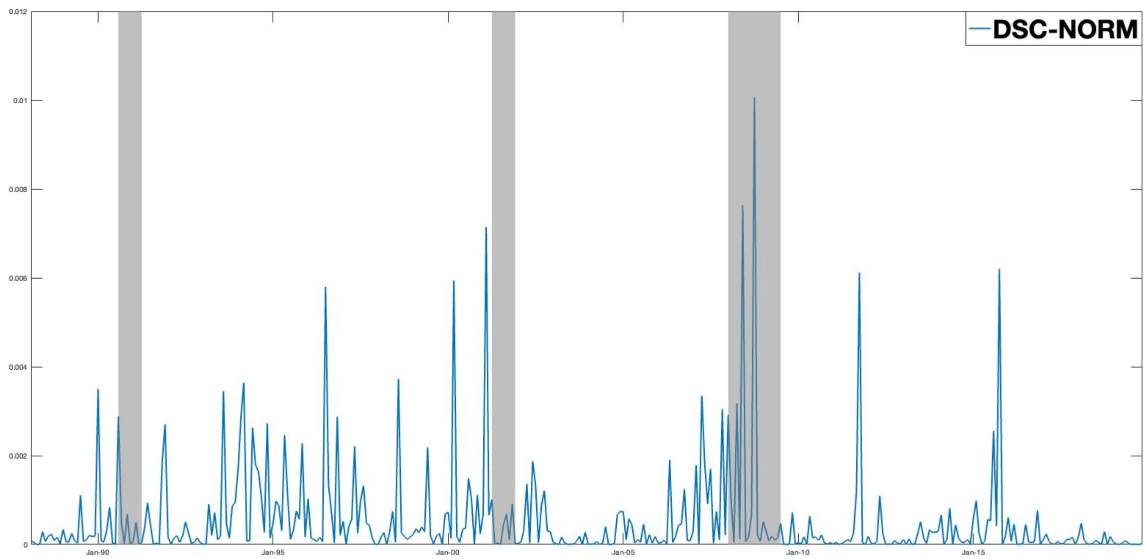


Figure (5) shows the variations of the DSC-NORM model over time. At first glance, we observe some spikes around the early 2000s when the dot-com bubble bursts and during the financial crisis of 2007–2009. This implies that the DSC model tends to adjust the combination weights more during uncertain periods.

So far, we have discussed the general patterns of how the DSC model selects and combines forecasts by different predictors. Next, we will pay more attention to how each predictor is selected in the model. In particular, we focus on the posterior inclusion probability (PIP) of individual predictors, which measures the probability of the predictor included in the model.

Intuitively, when the predictor is more valid, it is more likely to be included in the model. To measure this relationship, we compute the correlation coefficients between the PIP and the out-of-sample R^2 of every predictor at each time[2].

Table 2: **Posterior inclusion probability of individual predictors**

Predictors	Average PIP	coefficients	p-value
RSUP	37.36%	0.58	0.000
NSIFY	37.35%	0.39	0.000
NSIANN	26.98%	0.44	0.000
SHR1	25.30%	0.45	0.000
CHEMPIA	25.02%	0.51	0.000
CEIMO	18.81%	0.38	0.000
cay	16.74%	0.29	0.000
CSUE	11.61%	0.68	0.000
SUE	10.47%	0.70	0.000

Table (2) reports the correlation coefficients between PIP and the performance of individual predictors. From the second column of the table, we observe that the PIP of all variables is positively related to the out-of-sample R^2 of every predictor. Because the DSC model tends to include the more valid predictor with a higher probability, this relationship could be beneficial to the overall performance of the model.

[2]Because of space limitations, we only report 9 variables whose average PIP ranges from 0.1 to 0.9 over time, which implies the variables are not included (excluded) during the whole evaluation period.

4 Conclusion

In this project, we investigate the time-varying predictability of stock returns in a high-dimensional scenario. To overcome the technical difficulties of estimating a large set of coefficients dynamically, we propose the DSC model and the estimating strategy PF-VB. With the proposed DSC model, we achieve better out-of-sample performance for predicting the S&P500 return evaluated by both statistical and economic measures. We also find that predictability is related to the pattern of variable selection and combination.

A Appendix

A.1 Algorithm

Algorithm (1) outlines the VB method for $\hat{p}(\Theta_{1:t}, V_{1:t}|H_{1:t}, \Psi_{1:t}r_{1:t}, \hat{R}_{1:t})$.

Algorithm 1: VB for $\hat{p}(\Theta_{1:t}, V_{1:t}|H_{1:t}, \Psi_{1:t}, r_{1:t}, \hat{R}_{1:t})$

1. Initialize all the parameters and let $n = 1$
2. For t = 1 to T
3. While $\{E_{q^{(n)}(\cdot)}[p(r_t|\Theta_t^{(n)}, h_t)]$ - $E_{q^{(n-1)}(\cdot)}[p(r_t|\Theta_t^{(n-1)}, h_t)]\} \geq$ c,
   and $n < 100$
4. Compute $\tilde{Q}_t^{(n)} = inv[inv(Q_t) + inv(\hat{V}_t^{(n-1)})]$
5. Compute $\tilde{F}_t^{(n)} = \tilde{Q}_t^{(n)} inv(Q_t)$
6. Compute $m_i^{(n)}$ and $P_i^{(n)}$ through a Kalman filter.
7. Compute $\hat{V}_t^{(n)}$ with the DVS prior
8. n = n+1
9. end
10. end

With the VB method, we obtain the approximation for the posterior of combination weights that follows the normal distribution:

$$\Theta_t \sim N(m_{t|t}, P_{t|t}) \tag{17}$$

As mentioned in Section (2), we want to impose the non-negative restriction on each element of $\Theta_{1:t}$ because the forecasts by individual predictors shall be positively related to the realized excess market return. To impose the restriction, we use the idea of a restricted Kalman filter that projects the estimated states obtained by a naive Kalman filter onto a restricted space.

Specifically, we solve the following optimization problem with the given restriction:

$$\begin{aligned} \tilde{m}_{t|t} &= \text{argmin}_m (m - m_{t|t})^T (m - m_{t|t}) \\ \text{s.t.} &\qquad Am < a \end{aligned} \tag{18}$$

where $m_{t|t}$ is the mean of states computed by unrestricted Kalman update in Algorithm (1). This can be considered as a quadratic programming problem and be solved efficiently.

Algorithm (2) outlines the VB-PF for the DSC model:

Algorithm 2: PF-VB for the DSC model

1. Initialize all the parameters and let $n = 1$
2. For t = 1 to T (length of sample)
3. For n = 1 to N (number of particles)
4. Sample H_t from $N(H_{t-1}^{(n)}, W_{t-1}^{(n)})$
5. Update hyperparamet $\Psi_t^{(n)}$ (IG distribution)
6. Update mean and variance of Θ_t, $m_{t|t}^{(n)}$ and $P_{t|t}^{(n)}$ based on Steps (3) to (9) in Algorithm (1)
7. Obtain restricted mean $\tilde{m}_{t|t}^{(n)}$ by solving optimization problem (18)
8. Compute the incremental weight
9. end
10. end

A.2 Data description

Table A1: Full list of predictors

Data series	Category	Resource	Description
D/P	Finance	Goyal's website	Log of dividends on the S&P 500 index minus the log of prices
D/Y	Finance	Goyal's website	Log of dividends on the S&P 500 index minus the log of one-month-lagged prices
E/P	Finance	Goyal's website	Log of earnings minus the log of prices
BM	Finance	Goyal's website	Book value at the end of previous year dividedy by the end of month market value
SVAR	Finance	Goyal's website	Sum of squared daily return on the S&P 500
DFY	Finance	Goyal's website	BAA-rated corporate bond yields minus AAA-rated corporate bond
DFR	Finance	Goyal's website	Returns on long-term corporate bonds minus returns on long-term government bonds
NTIS	Finance	Goyal's website	12-month moving sums of net issues by NYSE listed stocks divided by the total market capitalization of NYSE stocks
LTR	Finance	Goyal's website	Return of log-term US Treasury bond
LTY	Finance	Goyal's website	Yield of log-term US Treasury bond
TBL	Finance	Goyal's website	3-month treasury-bill rate
Agg_Liq	Finance	Robert Stambaugh's website	Levels of aggregated liquidity
Innov_Liq	Finance	Robert Stambaugh's website	Innovations in aggregated liquidity
Fin_UNC	Finance	Sydney Ludvigson's website	A measure of financial uncertainty
Macro_UNC	Finance	Sydney Ludvigson's website	A macroeconomic of financial uncertainty
Real_UNC	Finance	Sydney Ludvigson's website	A real economic of financial uncertainty
BW_INV_SENT	Finance	Dashan Huang's website	Investment sentiment index by Barker and Wurgler(2006)
HJTZ_INV_SENT	Finance	Dashan Huang's website	Investment sentiment index by Huang, Jiang, Tu and Zhou (2015)
Disagrement	Finance	Dashan Huang's website	Disagreement index by Huang, Li and Wang (2021)
FF	Finance	Fred Data	Federal Fund Rate
M1	Finance	Fred Data	M1
ADS_index	Macro	Fed of Philadelphia	Aruoba-Diebold-Scotti Business Conditions Index
PAYEMS	Macro	Fred Data	Log diff of total nonfarm payroll
UNRATE	Macro	Fred Data	Unemployment rate
ICSA	Macro	Fred Data	Initial Claim
UEMPMEAN	Macro	Fred Data	Average weeks unemployments;
AWHMAN	Macro	Fred Data	Average weekly hour
RPI	Macro	Fred Data	Log difference of real personal income
INDPRO	Macro	Fred Data	Log difference of industrial production
TCU	Macro	Fred Data	Capitial Utilization
CPIAUCSL	Macro	Fred Data	Log difference of CPI for all commodities
CPILFESL	Macro	Fred Data	Log difference of CPI for core commodities
CPIUFDSL	Macro	Fred Data	Log difference of CPI for food;
CPIENGSL	Macro	Fred Data	Log difference of CPI for energy;
PCE	Macro	Fred Data	Log difference of Personal Consumper Expenditure (PCE)

Continued on next page

Table A1 – continued from previous page

Data series	Category	Resource	Description
PCEPILFE	Macro	Fred Data	Log difference of Personal Consumption Expenditures (PCE) Excluding Food and Energy
WPSFD49207	Macro	Fred Data	Log difference of PPI final good
WPSID61	Macro	Fred Data	Log difference of PPI intermediate good
WPSID62	Macro	Fred Data	Log difference of PPI raw good
CAY	Macro	Goyal's website	Consumption, wealth, income ratio
IK	Macro	Goyal's website	Investment-capital ratio
GDP_GROWTH	Macro	Fred Data	Log difference of Real GDP
GCEC_GROWTH	Macro	Fred Data	Log difference of Government Consumption Expenditures and Gross Investment
ABSACC	Anomaly Portfolio	Dave Rapach's website	Absolute value of accruals
ACC	Anomaly Portfolio	Dave Rapach's website	Accruals
AGE	Anomaly Portfolio	Dave Rapach's website	Firm age
AGR	Anomaly Portfolio	Dave Rapach's website	Asset growth
BETA1	Anomaly Portfolio	Dave Rapach's website	Short-term beta
BETA1LAG	Anomaly Portfolio	Dave Rapach's website	Short-term smoothed beta
BETA3	Anomaly Portfolio	Dave Rapach's website	Long-term beta
BETA3LAG	Anomaly Portfolio	Dave Rapach's website	Long-term smoothed beta
BMP	Anomaly Portfolio	Dave Rapach's website	Book to market
CASH	Anomaly Portfolio	Dave Rapach's website	Cash to assets
CASHDEBT	Anomaly Portfolio	Dave Rapach's website	Cash flow to debt
CASHPR	Anomaly Portfolio	Dave Rapach's website	Cash productivity
CEIANN	Anomaly Portfolio	Dave Rapach's website	Composite equity issuance, annual rebalancing
CEIFY	Anomaly Portfolio	Dave Rapach's website	Composite equity issuance, fiscal-year rebalancing
CEIMO	Anomaly Portfolio	Dave Rapach's website	Composite equity issuance, monthly rebalancing
CFPIA	Anomaly Portfolio	Dave Rapach's website	Industry-adjusted cash flow to price
CFPJUN	Anomaly Portfolio	Dave Rapach's website	Cash flow to price
CHATOIA	Anomaly Portfolio	Dave Rapach's website	Industry-adjusted change in asset turnover
CHEMPIA	Anomaly Portfolio	Dave Rapach's website	Industry-adjusted % change in employees
CHFEPS	Anomaly Portfolio	Dave Rapach's website	Change in forecasted earnings per share
CHINV	Anomaly Portfolio	Dave Rapach's website	Change in inventories
CHPMIA	Anomaly Portfolio	Dave Rapach's website	Industry-adjusted change in profit margin
CHTAX	Anomaly Portfolio	Dave Rapach's website	% change in tax expense
CINVEST	Anomaly Portfolio	Dave Rapach's website	Capital investment
CSUE	Anomaly Portfolio	Dave Rapach's website	Composite earnings surprise
CURRAT	Anomaly Portfolio	Dave Rapach's website	Current ratio
DEPR	Anomaly Portfolio	Dave Rapach's website	Depreciation to gross PP&E
DOLVOL	Anomaly Portfolio	Dave Rapach's website	Dollar trading volume
DROAQ	Anomaly Portfolio	Dave Rapach's website	Change in quarterly return on assets
DROEQ	Anomaly Portfolio	Dave Rapach's website	Change in quarterly return on equity
EAR	Anomaly Portfolio	Dave Rapach's website	Earnings announcement return
EGR	Anomaly Portfolio	Dave Rapach's website	Book equity growth
EP	Anomaly Portfolio	Dave Rapach's website	Earnings to price
FERR	Anomaly Portfolio	Dave Rapach's website	Earnings forecast error
FP	Anomaly Portfolio	Dave Rapach's website	Failure probability
GPA	Anomaly Portfolio	Dave Rapach's website	Gross profitability to assets
GRLTD	Anomaly Portfolio	Dave Rapach's website	Growth in long-term debt
GRLTNOA	Anomaly Portfolio	Dave Rapach's website	Growth in long-term net operating assets
HERF	Anomaly Portfolio	Dave Rapach's website	Industry sales concentration
HIRE	Anomaly Portfolio	Dave Rapach's website	% change in employees
IDIOVOL	Anomaly Portfolio	Dave Rapach's website	Idiosyncratic return volatility
ILLIQ	Anomaly Portfolio	Dave Rapach's website	Illiquidity
INDMOM12M	Anomaly Portfolio	Dave Rapach's website	Twelve-month industry momentum
INDMOM1M	Anomaly Portfolio	Dave Rapach's website	One-month industry momentum
MAXRET	Anomaly Portfolio	Dave Rapach's website	Maximum daily return
MOM12M	Anomaly Portfolio	Dave Rapach's website	Twelve-month momentum
MOM1M	Anomaly Portfolio	Dave Rapach's website	One-month momentum
MOM36M	Anomaly Portfolio	Dave Rapach's website	36-month momentum
MOM6M	Anomaly Portfolio	Dave Rapach's website	Six-month momentum
MS	Anomaly Portfolio	Dave Rapach's website	Growth score
MVE	Anomaly Portfolio	Dave Rapach's website	Market equity
NINCR	Anomaly Portfolio	Dave Rapach's website	Number of quarters with consecutive earnings increase
NOA	Anomaly Portfolio	Dave Rapach's website	Net operating assets
NSIANN	Anomaly Portfolio	Dave Rapach's website	Net share issuance, annual rebalancing
NSIFY	Anomaly Portfolio	Dave Rapach's website	Net share issuance, fiscal-year rebalancing
NSIMO	Anomaly Portfolio	Dave Rapach's website	Net share issuance, monthly rebalancing
OL	Anomaly Portfolio	Dave Rapach's website	Operating leverage
OPERPROF	Anomaly Portfolio	Dave Rapach's website	Operating profitability
ORGCAP	Anomaly Portfolio	Dave Rapach's website	Organization capital to assets
OSCORE	Anomaly Portfolio	Dave Rapach's website	O-score
PCHCURRAT	Anomaly Portfolio	Dave Rapach's website	% change in current ratio
PCHDEPR	Anomaly Portfolio	Dave Rapach's website	% change in depreciation to gross PP&E
PCHGMSALE	Anomaly Portfolio	Dave Rapach's website	% change in gross margin minus % change in sales
PCHQUICK	Anomaly Portfolio	Dave Rapach's website	% change in quick ratio
PCHSALEINV	Anomaly Portfolio	Dave Rapach's website	% change in sales minus % change in inventories
PCHSALEINVT	Anomaly Portfolio	Dave Rapach's website	% change in sales to inventories
PCHSALEREC	Anomaly Portfolio	Dave Rapach's website	% change in sales minus % change in accounts receivable
PCHSALESGM	Anomaly Portfolio	Dave Rapach's website	% change in sales minus % change in SG&A
PS	Anomaly Portfolio	Dave Rapach's website	Fundamental score
QUICK	Anomaly Portfolio	Dave Rapach's website	Quick ratio
RD	Anomaly Portfolio	Dave Rapach's website	R&D expense to market
RETVOL	Anomaly Portfolio	Dave Rapach's website	Return volatility
ROA	Anomaly Portfolio	Dave Rapach's website	Return on assets
ROAQ	Anomaly Portfolio	Dave Rapach's website	Quarterly return on assets
ROAVOL	Anomaly Portfolio	Dave Rapach's website	Volatility of return on assets
ROE	Anomaly Portfolio	Dave Rapach's website	Return on equity
ROEQ	Anomaly Portfolio	Dave Rapach's website	Quarterly return on equity
ROIC	Anomaly Portfolio	Dave Rapach's website	Return on invested capital
ROM	Anomaly Portfolio	Dave Rapach's website	Return on market equity
RSUP	Anomaly Portfolio	Dave Rapach's website	Revenue surprise
SALECASH	Anomaly Portfolio	Dave Rapach's website	Sales to cash
SALEINV	Anomaly Portfolio	Dave Rapach's website	Sales to inventories
SALEREC	Anomaly Portfolio	Dave Rapach's website	Sales to accounts receivable
SGR	Anomaly Portfolio	Dave Rapach's website	Sales growth
SHR1	Anomaly Portfolio	Dave Rapach's website	Short-term share issuance
SHR5ANN	Anomaly Portfolio	Dave Rapach's website	Long-term share issuance, annual rebalancing
SHR5MO	Anomaly Portfolio	Dave Rapach's website	Long-term share issuance, monthly rebalancing
SP	Anomaly Portfolio	Dave Rapach's website	Sales to price

Continued on next page

Table A1 – continued from previous page

Data series	Category	Resource	Description
SPI	Anomaly Portfolio	Dave Rapach's website	Special items
STDACC	Anomaly Portfolio	Dave Rapach's website	Standard deviation of accruals
STDCF	Anomaly Portfolio	Dave Rapach's website	Standard deviation of cash flows
STDTURN	Anomaly Portfolio	Dave Rapach's website	Standard deviation of turnover
SUE	Anomaly Portfolio	Dave Rapach's website	Standardized earnings surprise
TANG	Anomaly Portfolio	Dave Rapach's website	Tangibility
TIBI	Anomaly Portfolio	Dave Rapach's website	Taxable income to book income
TURN	Anomaly Portfolio	Dave Rapach's website	Total turnover
TURN3	Anomaly Portfolio	Dave Rapach's website	Average turnover, three months
TURNL	Anomaly Portfolio	Dave Rapach's website	Lagged total turnover
ZEROAVG	Anomaly Portfolio	Dave Rapach's website	Average number of turnover-adjusted zero trading volume
ZEROTOT	Anomaly Portfolio	Dave Rapach's website	Total number of turnover-adjusted zero daily volume