Time-varying predictability of stock return with a large number of predictors

Zhendong Sun*

This version: December 22, 2021

Abstract

We investigate the predictability of stock return with the time-varying parameters model including a large number of predictors (large-tvp model). We apply the dynamic variable selection (DVS) prior to handling the high-dimensional dataset, and propose a sequential method to efficiently estimate this model. We find the large-tvp model estimated by the proposed methodology outperforms other methods such as LASSO, PCA, and random forests, and it performs especially better during the financial crisis of 2007-2008.

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

1 Introduction

Predictability of stock return is one of the most interesting topics in empirical finance. Over the past several decades, much evidence has been found that the stock predictability could be changing over time (e.g., Dangl and Halling (2012), Farmer et al. (2019)). By incorporating time-varying coefficients in the predictive regression, the stock return becomes more predictable both in the practice of in-sample and out-of-sample forecasting.

While the previous research shows the time-varying coefficients regression does improve the accuracy of prediction, most of them only focus on a small-scale set of predictors. In the past decades, an increasing number of studies indicate that superior forecasts can be achieved by including more predictors (e.g., Gu et al. (2020), Rapach and Zhou (2020)). Hence, it is natural to ask whether we can further improve the stock return prediction by integrating time-varying coefficient regression with a high-dimensional set of predictors.

In this paper, we apply the large time-varying parameter model (large-tvp) into the practice of out-of-sample prediction of monthly S&P 500 excess return. To handle a large number of predictors, we use the dynamic variable selection (DVS) prior of Koop and Korobilis (2020). With DVS prior, we are able to dynamically select a small subset from all predictors to forecast the stock return while allowing the coefficients of selected predictors to evolve over time.

The contribution of this paper is twofold. First, we develop a new methodology to estimate the large-tvp model with dynamic variable selection prior. While much previous research proposes different methods to estimate this type of model, most of them focus on the practice of in-sample fit (e.g., Bitto and Frühwirth-Schnatter (2019), Chan et al. (2020)). However, those methods are not efficient for out-of-sample prediction since they require estimating the whole path of all the parameters whenever the new information is realized, which makes the computational cost huge especially for the case with a high-dimensional dataset. To overcome the bottleneck, we propose a new method based on the particle filter (see Amir-Ahmadi and Sun (2021)) and variational Bayes (see Koop and Korobilis (2020)). With the proposed method, we are able to sequentially estimate the parameters at each time with the newly realized observations, which is computationally efficient while yielding accurate forecasts.

The second contribution of this paper is on the empirical side. Namely, the large-tvp model achieves better prediction for stock return relative to the historical mean as well as other high-dimensional models with fixed parameters such as LASSO. This model also outperforms some machine learning methods such as random forests which are popular in empirical asset pricing in recent years. Besides the overall performance, the advantage of the large-tvp model becomes larger during the financial crisis of 2007-2008 when the stock market suffered from huge losses.

The outperformance of the large-tvp model does not only appear from the statistical perspective but also turns into the economic gain. Specifically, the optimal portfolio based on the forecasts by the large-tvp model yields a higher excess return than the benchmark and all the other competing models. Besides that, it suffers from a smaller drawdown during the financial crisis of 2007-2008.

The rest of the paper is organized as follows. Section 2 introduces the large-tvp model and the proposed method to estimate it. Section 3 introduces the dataset used in the empirical application and the setup of all the competing models. Section 4 reports the empirical results. Section 5 is the conclusion.

2 Methodology

This section briefly introduces the predictive regression with time-varying parameters. We outline the model specification as well as the efficient methodology to estimate the parameters in the model.

2.1 Predictive regression with time-varying parameters

Following a convention in much empirical literature, we use a linear predictive regression model to describe the relation between stock return and a set of predictors. To outline the time-varying parameters, we incorporate the regression into a state-space model, in which the vector of coefficients is treated as unobservable state variables following the random walk process.

Specifically, we estimate the model of form:

$$r_{t+1} = x_t'\beta_t + exp(h_t)\epsilon_t \qquad \epsilon_t \sim N(0,1) \tag{1}$$
$$\beta_t = \beta_{t-1} + \eta_t \qquad \eta_t \sim N(0,Q_t) \tag{2}$$
$$h_t = h_{t-1} + \xi_t \qquad \xi_t \sim N(0,W_t) \tag{3}$$

where r_{t+1} denotes the log excess market return and x_t is an $1 \times K$ vector of predictors. Equation (1) is called measurement equation which connects the stock return to the predictors through unobserved coefficients, and equation (2) is the state equation that describes the dynamic of unobserved state variables - the time-varying coefficients in this case. Besides assuming the coefficients changing over time, we also allow the stochastic volatility by modeling the natural log of variance as a random walk process in equation (3).

In Bayesian framework, we target the posterior distribution of all the unknown parameters

$$p(\theta_{1:t}, \psi_{1:t} | y_{1:t}, \theta_0, \psi_0) \tag{4}$$

where $\theta_{1:t} = \{\beta_{1:t}, h_{1:t}\}$, $\psi_{1:t} = \{Q_{1:t}, W_{1:t}\}$ and $y_{1:t}$ are the observations including stock return and predictors.

2.2 Sequential estimation

To estimate the posterior distribution of interest, the conventional method is based on the combination of Markov Chaim Monte Carlo (MCMC) and Kalman Filter (e.g., Koop and Korobilis (2010); Del Negro and Primiceri (2015)).

While the conventional method works well in many applications, Amir-Ahmadi and Sun (2021) propose a sequential method that is especially efficient for out-of-sample prediction[1].

This sequential method for time-varying parameter regression is built on the well-known particle filter. The idea of particle filter is that we first find an arbitrary proposal distribution $q(\theta_{1:t}, \psi_{1:t} | y_{1:t}, \theta_0, \psi_0)$ which we can directly sample from. Then we obtain many draws from this proposal distribution and assign a weight to each of them based on certain criteria. With a number

[1] Bognanni and Zito (2020) propose a similar method for time-varying parameter model, but they only allow the volatility changing over time while keeping the coefficients fixed.

of weighted draws (or called particles), we will get an approximation of the posterior distribution.

More specifically, we can rewrite the posterior described in equation (4) as below:

$$
\begin{aligned}
& p(\theta_{1:t}, \psi_{1:t} | y_{1:t}, \theta_0, \psi_0) && (5) \\
= & \underbrace{\frac{p(\theta_{1:t}, \psi_{1:t} | y_{1:t}, \theta_0, \psi_0)}{q(\theta_{1:t}, \psi_{1:t} | y_{1:t}, \theta_0, \psi_0)}}_{I} \times q(\theta_{1:t}, \psi_{1:t} | y_{1:t}, \theta_0, \psi_0) && (6)
\end{aligned}
$$

where term I is the so-called *importance weight*, which evaluates the importance of each draw from the proposal distribution $q(\cdot)$. To simplify the computation of term I, we can decompose it into a sequential form:

$$
\begin{aligned}
& \frac{p(\theta_{1:t}, \psi_{1:t} | y_{1:t}, \theta_0, \psi_0)}{q(\theta_{1:t}, \psi_{1:t} | y_{1:t}, \theta_0, \psi_0)} \\
= & \underbrace{\frac{p(y_t | y_{1:t-1}, \theta_t) p(\theta_t | \theta_{t-1}, \psi_t)}{q(\theta_t | \theta_{t-1}, \psi_t, y_t)}}_{\omega} \\
\times & \underbrace{\frac{p(\theta_{1:t-1}, \psi_{1:t-1} | y_{1:t-1}, \theta_0, \psi_0)}{q(\theta_{1:t-1}, \psi_{1:t-1} | y_{1:t-1}, \theta_0, \psi_0)}}_{W}
\end{aligned} \tag{7}
$$

where W is the importance weight in the last period, and ω is called *incremental weight*. In practice, we only need to calculate the incremental weight ω at each t because the previous importance weight W has already been computed and it will not be affected by new information.

In the time-varying parameter model, the posterior distribution of coefficients $\beta_{1:t}$ can be written analytically. Based on this feature, we are able to integrate out the state vector $\beta_{1:t}$ from the particle, which helps improve the numerical efficiency of the method. Specifically, we can decompose the posterior (4) as below:

$$
p(\beta_{1:t}, h_{1:t}, \psi_{1:t} | y_{1:t}) = \underbrace{p(\beta_{1:t} | h_{1:t}, \psi_{1:t}, y_{1:t})}_{I} \underbrace{p(h_{1:t}, \psi_{1:t} | y_{1:t})}_{II} \tag{8}
$$

For the term I in equation (8), there exists a closed form conditional on all the other parameters:

$$\beta_t \sim N(m_{t|t}, P_{t|t}) \tag{9}$$

where the mean $m_{t|t}$ and variance $P_{t|t}$ can be computed directly through a Kalman update step.

The term II in equation (8) includes the stochastic volatility parts as well as all the hyperparameters $\psi_{1:t}$, which can be estimated by the particle filter. The only difference compared to the original particle filter is that we cannot evaluate the incremental weight directly because the likelihood function $p(y_t|\theta_t, y_{1:t-1})$ is not available. However, we can approximate the function by the mean and variance of the coefficients. Specifically, the likelihood function shall be rewritten as follow:

$$\begin{aligned} p(y_t|h_t, y_{1:t-1}) &= \int p(y_t|\beta_t, h_t, y_{1:t-1})p(\beta_t|h_t, y_{1:t-1})db_t \\ &\approx p(y_t|m_{t|t-1}, P_{t|t-1}, h_t, y_{1:t-1}) \end{aligned} \tag{10}$$

where $m_{t|t-1}$ and $P_{t|t-1}$ are the mean and variance of coefficient b_t by the Kalman prediction, which is a function of the past data and the remaining parameters of the model. Notice that we use the Kalman prediction instead of the Kalman update of the coefficients because the current observation y_t is not available at this step.

2.3 Dynamic variable selection

One of the key steps in the sequential estimation method is to compute the posterior distribution of coefficient $\beta_{1:t}$. In a small-scale model, we are able to get an optimal solution by Kalman Filter. However, it is well-known that the Kalman Filter is inaccurate if the dimension of $\beta_{1:t}$ is large. A solution to this issue is to select a subset of the whole predictors by incorporating shrinkage prior on the coefficients $\beta_{1:t}$ (see Koop and Korobilis (2020)).

In this paper, we apply the dynamic variable selection prior (DVS) into the time-varying parameter model with a high-dimensional dataset. With DVS prior, we are able to select informative variables without updating the estimation of parameters in the past periods. While it may not be optimal in the practice of in-sample fit, it is much more efficient for out-of-sample forecasting in which we need to re-estimate the parameters whenever new observations are realized.

Following Koop and Korobilis (2020), we specify the DVS prior as below:

$$\beta_{k,t}|\gamma_{k,t},\tau_{k,t} \sim (1-\gamma_{k,t})N(0, c\times\tau_{k,t}^2) + \gamma_{k,t}N(0,\tau_{k,t}^2) \tag{11}$$
$$\gamma_{k,t}|pi_t \sim Bernoulli(\pi_{0,t}) \tag{12}$$
$$\tau_{k,t}^{-2} \sim Gamma(g_0, h_0) \tag{13}$$

where $k = 1, 2, \ldots, K$, and c, g_0 and h_0 are pre-determined hyperparameters for the priors. In general, we require a large $\tau_{k,t}$ while the c should be small. By doing so, the prior concentrates at zero when $\gamma_{k,t} = 0$, otherwise, it becomes uninformative since the variance is large.

To estimate the state-space model of equation (1) to (3) with the DVS prior, we need to combine two prior distribution described before:

$$\beta_t|\beta_{t-1}, Q_t \sim N(\beta_{t-1}, Q_t) \tag{14}$$
$$\beta_t|\gamma_{k,t},\tau_{k,t} \sim N(0, V_t) \tag{15}$$

where $V_t = diag(v_{1,t}, v_{2,t}, \ldots, v_{K,t})$ and $v_{j,t} = (1-\gamma_{k,t})N(0, c\times\tau_{k,t}^2)+\gamma_{k,t}N(0,\tau_{k,t}^2)$ for $j = 1, \ldots, K$. Equation (14) is equivalent to the state equation (2). Following Koop and Korobilis (2020), we can combine these two priors by rewriting the state equation (2) as below:

$$\beta_t = \tilde{F}_t\beta_{t-1} + \tilde{\eta}_t \qquad \eta_t \sim N(0, \tilde{Q}_t) \tag{16}$$

where $\tilde{Q}_t = Q_t^{-1} + V_t^{-1}$, and $\tilde{F}_t = \tilde{Q}_t + Q_t^{-1}$. Therefore, our target now is to estimate the state-space model with measurement equation (1) and state equation (16). With this revised state model, the posterior distribution of interest will be slightly different from equation (8). Specifically, the new posterior distribution can be written as below:

$$p(\beta_{1:t}, h_{1:t}, \eta_{1:t}, \psi_{1:t}|y_{1:t}) = \underbrace{p(\beta_{1:t}, \eta_{1:t}|h_{1:t}, \psi_{1:t}, y_{1:t})}_{I}\underbrace{p(h_{1:t}, \psi_{1:t}|y_{1:t})}_{II} \tag{17}$$

where $\eta_{1:t} = \{Q_{1:t}, V_{1:t}\}$ and $\psi_{1:t} = \{W_{1:t}\}$.

While the term I in equation (17) can be estimated by particle filter introduced in subsection (2.2), the closed form of term II no longer exists after incorporating the dynamic variable selection prior. Instead, we aim to find a good approximation of the term II so that we are still able to achieve an accurate estimation of the whole posterior distribution.

2.3.1 Variational Bayes

Following Koop and Korobilis (2020), we use the variational Bayes (VB) method to approximate the term I in equation (17). The VB method provides a suboptimal approximation for the posterior distribution of interest based on optimization. The main advantage of VB is that we can quickly fit many models with a large number of inputs (see Blei et al. (2017)). Hence, it is feasible to approximate the posterior of coefficients in each particle even though the number of particles will be large.

The main idea behind VB is to define an approximate densities $q(\cdot)$ that belongs to a family $\mathscr{F}$ of distributions with closed form, and minimizes the distance between the approximate densities and the exact posterior of interest, where the distance is measured by Kullback-Leibler divergence:

$$\begin{aligned} KL(q\|p) &= \int q(z)\log\left\{\frac{q(z)}{p(z|y)}\right\}\mathrm{d}z \\ &= \mathbb{E}_q[\log q(z)] - \mathbb{E}_q[\log p(z|y)] \\ &= \mathbb{E}_q[\log q(z)] - \mathbb{E}_q[\log p(z,y)] + \mathbb{E}_q[\log p(y)] \end{aligned} \tag{18}$$

Specifically, we aim to solve the optimization problem as below:

$$q^{\star}(\cdot) = \underset{q(\cdot)\in\mathscr{F}}{\arg\min} KL(q\|p) \tag{19}$$

In practice, it is always difficult to compute KL directly. However, we can optimize an alternative objective function called evidence lower bound (ELBO) defined as below:

$$ELBO(q) = \mathbb{E}_q[\log p(z,y)] - \mathbb{E}_q[\log q(z)] \tag{20}$$

In the large-tvp model, we try to find $q^{\star}(\beta_{1:t}, \eta_{1:t}, |h_{1:t}, \psi_{1:t} y_{1:t})$ that minimize the $KL(q||p)$ (or maximize the corresponding ELBO). For simplicity, we decompose the approximate density by the mean filed factorization:

$$q(\beta_{1:t}, \eta_{1:t}, |h_{1:t}, \psi_{1:t} y_{1:t}) = q(\beta_{1:t}|h_{1:t}, \psi_{1:t}, y_{1:t})q(\eta_{1:t}|h_{1:t}, \psi_{1:t}, y_{1:t}) \tag{21}$$

and it can be shown that the optimal choices for $q(\beta_{1:t}|\cdot)$ and $q(\eta_{1:t}|\cdot)$ are:

$$q(\beta_{1:t}|\cdot) = exp\{E_{q(\eta_{1:t}|\cdot)}[\log p(\beta_{1:t}|\eta_{1:t}, h_{1:t}, \psi_{1:t}, y_{1:t})]\} \tag{22}$$

$$q(\eta_{1:t}|\cdot) = exp\{E_{q(\beta_{1:t}|\cdot)}[\log p(\eta_{1:t}|\beta_{1:t}, h_{1:t}, \psi_{1:t}, y_{1:t})]\} \tag{23}$$

Equation (22) denotes the expectation of the posterior of $\beta_{1:t}$ over the approximated $q(\eta_{1:t}|\cdot)$ and similar for the equation (23) with swap the order of each variable. Hence, the $q(\beta_{1:t}|\cdot)$ is a function of $q(\eta_{1:t}|\cdot)$, and vice-versa, so that we can approximate these two densities iteratively until the value of objective reaches to a stable level. Algorithm (1) outline the VB method for $\hat{p}(\beta_{1:t}, \eta_{1:t}|h_{1:t}, \psi_{1:t}y_{1:t})$.

Algorithm 1: Variational Bayes for $\hat{p}(\beta_{1:t}, \eta_{1:t}|h_{1:t}, \psi_{1:t}y_{1:t})$

1. Initialize all the parameters, and let $n = 1$
2. For i = 1 to t
3. While $\{E_{q^{(n)}(\cdot)}[p(y_i|\beta_i^{(n)}, h_i)]$ - $E_{q^{(n-1)}(\cdot)}[p(y_i|\beta_i^{(n-1)}, h_i)]\} \geq$ c
4. Compute $\tilde{Q}_t^{(n)} = inv[inv(Q_t^{(n-1)}) + inv(V_t^{(n-1)})]$
5. Compute $\tilde{F}_t^{(n)} = \tilde{Q}_t^{(n)} inv(Q_t^{(n-1)})$
6. Compute $m_i^{(n)}$ and $P_i^{(n)}$ through Kalman Filter.
7. Compute $V_t^{(n)}$ through the DVS prior
8. n = n+1
9. end
10. end

2.4 Estimate the model

In this paper, we aim to estimate the large time-varying parameter (large-tvp) model specified as equation (1) and (16). To estimate the model, we propose a new method based on the variational Bayes and particle filter (VB-PF). Algorithm (2) outline the proposed algorithm.

Algorithm 2: VB-PF for large-tvp model

1. Initialize all the parameters, and let $n = 1$
2. For t = 1 to T
3. For n = 1 to N
4. Sample h_t from $N(h_{t-1}^{(n)}, W_{t-1}^{(n)})$
5. Update hyperparamet $W_t^{(n)}$ (IG distribution)
6. Update mean and variance of β_t, m_t^n and P_t^n
   based on the step (3) to (9) in algorithm (1)
7. Compute the incremental weight by equation (10)
8. end
9. end

3 Study design

This section describes the data used in the paper, and the alternative methods for comparison.

3.1 Data description

We evaluate the performance of methodology discussed in section 2 using monthly excess stock return, which is computed by the CPSP value-weighted S&P 500 return minus the risk-free return based on the Treasury bill rate[2].

We consider 34 predictors including the most widely-used variables used by previous research. The data ranges from July 1969 to December 2019, to make sure that all the data series have realized observation.

[2]The data is available from Amit Goyal's website.

Table 1: Data description

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
FF	Finance	Fred Data	Federal Fund Rate
SMB	Finance	French's website	Average return on the three small portfolios minus the average return on the three big portfolios
HML	Finance	French's website	Average return on the two value portfolios minus the average return on the two growth portfolio
RMW	Finance	French's website	Average return on the two robust operating profitability portfolios minus the average return on the two weak operating profitability portfolios
CMA	Finance	French's website	Average return on the two conservative investment portfolios minus the average return on the two aggressive investment portfolios
Mom	Finance	French's website	Average return on the two high prior return portfolios minus the average return on the two low prior return portfolios
PAYEMS	Macro	Fred Data	Percentage change 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	Percentage change of real personal income;
INDPRO	Macro	Fred Data	Percentage change of industrial production;
TCU	Macro	Fred Data	Capitial Utilization;
CPIAUCSL	Macro	Fred Data	Percentage change of CPI for all commodities;
CPILFESL	Macro	Fred Data	Percentage change of CPI for core commodities;
CPIUFDSL	Macro	Fred Data	Percentage change of CPI for food;
CPIENGSL	Macro	Fred Data	Percentage change of CPI for energy;
PCE	Macro	Fred Data	Percentage change of Personal Consumer Expenditure (PCE)
PCEPILFE	Macro	Fred Data	Percentage change of Personal Consumption Expenditures (PCE) Excluding Food and Energy
WPSFD49207	Macro	Fred Data	Percentage change of PPI final good
WPSID61	Macro	Fred Data	Percentage change of PPI intermediate good
WPSID62	Macro	Fred Data	Percentage change of PPI raw good

Table (1) gives a brief description about the included predictors. We normalized all the data by the sample mean and variance of each predictor up to the period we make prediction.

To avoid the forward-looking bias, we take one-month lag of all the financial variables and two months lag of the macroeconomic variables. In particular, when we predict the stock return r_{t+1}, we will use the predictors x_t from financial sector and x_{t-1} from macroeconomic sector.

3.2 Alternative methods

We compare the proposed method with some alternatives used in the previous research, which can be divided into two categories.

The first category focuses on the high-dimensional set of predictors while assuming the predictive model is linear with fixed parameters. This category includes PCA, LASSO, and optimal combination (OC).

The second category allows non-linearity in the model while being able to deal with a large-scale dataset. This category includes tree-based models such as Random forest and gradient boosted regression tree (GBRT).

Table (2) describes all the methods and corresponding setup for comparison.

Table 2: Alternative models

Category I	
Model	**Description**
PCA - 200	Use PCA over a 200-month rolling window
PCA - expanding	Use PCA over an expanding window
LASSO - 200	Use LASSO over a 200-month rolling window
LASSO - expanding	Use LASSO over an expanding window
OC - 200	Use optimal combination (OC) over a 200-month rolling window
OC - expanding	Use optimal combination (OC) over an expanding window
Category II	
Model	**Description**
Random Forest	Random forecast with 500 trees in the ensemble
GBDT	Gradient boosted regression trees (GBRT) with 500 trees in the ensemble

4 Empirical results

This section reports the out-of-sample predictability with time-varying parameter model including high-dimensional dataset of predictors.

4.1 Statistical Evaluation

To evaluate the statistical accuracy of forecasts, we report the mean squared forecast error (MSFE). Furthermore, we compute the out-of-sample R^2_{OOS} statistic:

$$R^2_{OOS} = 1 - \frac{\sum_{t=\bar{t}}^{T}(r_t - \hat{r}_t)^2}{\sum_{t=\bar{t}}^{T}(r_t - \bar{r}_t)^2} \tag{24}$$

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 predictable part in the monthly stock return. In practice, a small value of R^2_{OOS} statistic can generate significant degree of excess return[3]
.

Table (3) summarized the statistical evaluation for different models. We distinguish four forecast periods: 1980+, 1990+, 2000+ and 2010+. The R^2_{OOS} statistic is generally small (or negative) for all the models, which implies the fact that the predictability of stock return is very limited. However, it is still possible to extract signals from a large number of predictors, which

[3] Campbell and Thompson (2008) argue that a monthly R^2_{OOS} statistic of 0.5% indicates an economically significant signal. See also Rapach and Zhou (2013).

Table 3: Statistical Evaluation

Model	Forecast period: 1980+		Forecast period: 1990+		Forecast period: 2000+		Forecast period: 2010+	
	MSFE	R-squared	MSFE	R-squared	MSFE	R-squared	MSFE	R-squared
PCA_200	4.301	-1.14%	4.134	-2.22%	4.193	0.48%	3.685	-4.44%
PCA_expanding	4.337	-2.86%	4.183	-4.66%	4.277	-3.54%	3.619	-0.72%
LASSO_200	4.435	-7.51%	4.253	-0.13%	4.386	-8.89%	3.692	-0.76%
LASSO_expanding	4.325	-2.23%	4.099	-0.52%	4.199	0.19%	3.596	0.58%
OC_200	4.285	-0.55%	4.123	-1.70%	4.254	-2.44%	3.619	-0.72%
OC_expanding	4.286	-0.44%	4.118	-1.49%	4.217	-0.69%	3.597	0.48%
Random Forest	4.392	-5.47%	4.239	-7.54%	4.324	-5.86%	3.646	-2.23%
GBRT	4.745	-23.17%	4.574	-25.16%	4.658	-22.58%	3.9	-16.98%
Large-tvp (VB-PF)	4.263	0.64%	4.077	0.55%	4.169	1.60%	3.599	0.34%

helps to obtain a sizable R^2. In particular, the PCA with a 200-month rolling window yields an R^2 of 0.48% in the forecast period 2000+. The LASSO with expanding window produces an R^2 of 0.19% in the period 2000+, and 0.58% in the period 2000+. The optimal combination (OC) with expanding window yields an R^2 of 0.48% in the period 2000+. The large time-varying parameter predictive model (Large-tvp) estimated by the proposed VB-PF method produces an R^2 of 0.64% in the period 1980+, 0.55% in the period 1990+, 1.60% in the period 2000+ and 0.34% in the period 2010+.

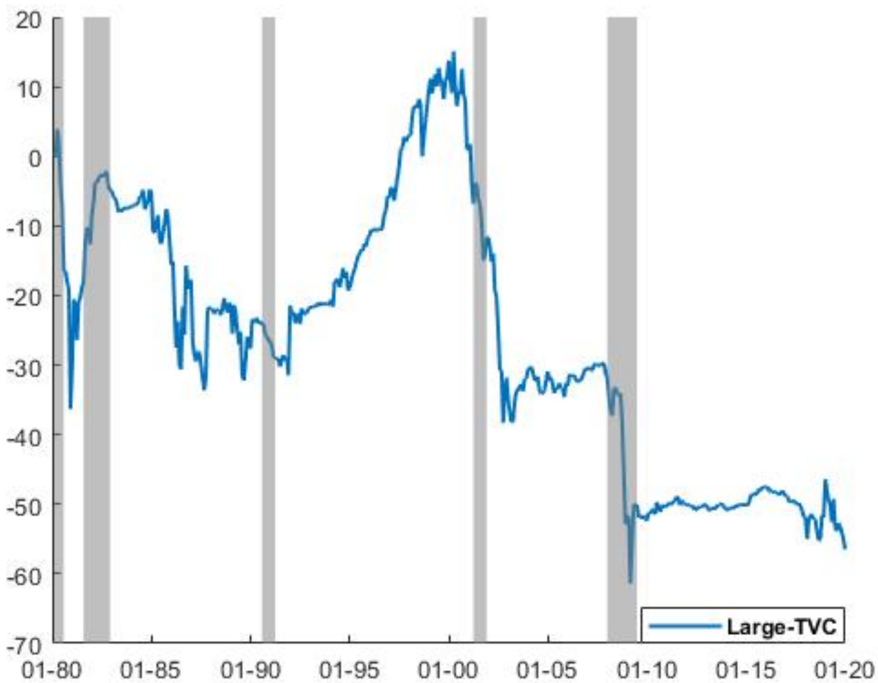
Among all the methods, the large-tvp model is the only model which yields a sizable R^2 in all periods. The best performance of large-tvp model is achieved in the period 2000+, which includes the financial crisis of 2007-2008. This implies that the large-tvp tends to perform better especially during the period with higher uncertainty (See e.g., Amir-Ahmadi and Sun (2021)).

Figure (1) shows the cumulative differences in forecast error between large-tvp model and benchmark, which illustrates the performance of the model during different periods. The cumulative differences in forecast error experienced significant decrease during the financial crisis of 2007-2008, which implies the advantage of large-tvp model over benchmark get larger during this period. It also echos the result we find in the table (3).

4.2 Economic Evaluation

So far we have shown that the large-tvp model improve the prediction for stock return statistically. In addition to that, we also want to discuss whether the statistical predictability can be turned into economic gains for an investor in an asset allocation context (e.g., Campbell and Thompson (2008), Rapach and Zhou (2020)). In particular, we consider an investor

Figure 1: Cumulative differences in forecast error



with a single-period horizon and the mean-variance preferences:

$$\arg\max_{w_{t+1|t}} w_{t+1}\hat{r}_{t+1} - 0.5\gamma w_{t+1}^2 \hat{\sigma}_{t+1}^2 \tag{25}$$

where γ is the coefficient of relative risk aversion, ω_{t+1} is the allocation to equities, $\hat{r}_{t+1}$ is the predicted stock return and $\hat{\sigma}_{t+1}$ is the estimated variance of stock return. The solution to equation (25) can be written as:

$$w_{t+1}^* = \left(\frac{1}{\gamma}\right)\left(\frac{\hat{r}_{t+1}}{\hat{\sigma}_{t+1}^2}\right) \tag{26}$$

In practice, we assume $\gamma = 5$, and estimate $\hat{\sigma}_{t+1}$ by sample variance over a 240-month rolling window. We restrict the weight of equitiesω_{t+1} from 0 to 1, which makes sure the allocation is always feasible in reality.

Table (4) reports the log of the cumulative excess return of the optimal portfolio created based on the forecasts by different models. When the portfolio is based on the historical mean of excess return, it yields a log of the cumulative excess return of 2.34 (6.02% annually) with an annualized volatility of 11.87% (the Sharp ratio is 0.51) in the period 1980+. With the portfolio based on the prediction by the large-tvp model, we obtain a log of the cumulative excess return of 2.66 (6.88% annually) with an annualized volatility of 11% (the Sharp ratio is 0.61), which is better than the benchmark portfolio.

Figure (2) shows the log of the cumulative excess return and corresponding equity weight. An important feature of the large-tvp model observed from figure (2) is that the allocation weight of equity changes more frequently than the portfolio built on the historical mean, and it can be even zero in a certain period. In reality, investors will reallocate the portfolio with the performance of the market. In particular, they expect to hold even zero risky assets when the market performs extremely badly. With the large-tvp model, the investor tends to decrease the holding on risky assets during a recession, and the weight on equity can get as low as zero in the financial crisis of 2007-2008, which helps the portfolio suffer smaller drawdowns than the portfolio based on the benchmark.

4.3 Analysis of time-varying predictors

After having the overall outperformance of large-tvp model both statistically and economically, we discuss the properties of predictability in more

Figure 2: Analysis of optimal portfolio

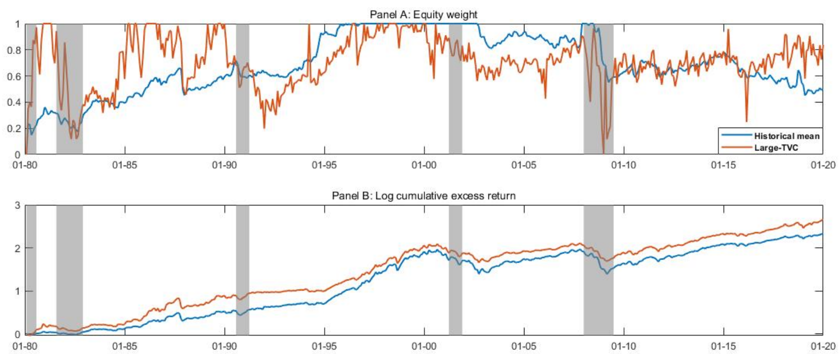


Table 4: Log of cumulative return

Model	1980+	1990+	2000+	2010+
Historical_mean	2.34	1.85	0.47	0.72
PCA_200	2.52	1.74	0.93	0.80
LASSO_expanding	2.46	1.82	0.75	0.89
OC_expanding	2.15	1.32	0.48	0.80
Large-tvp (VB-PF)	2.66	1.83	0.65	0.80

details.

The first important feature of predictability of stock return with a high-dimensional dataset of predictors is that only a smaller number of predictors will be included in the predictive model to get a good out-of-sample forecast. This results from the fact that the model is easily over-fitted with more input variables especially for the variable of interest with a low signal-to-noise ratio, such as the stock return data. We define the sparsity of the model at time t by the number of selected predictors:

$$Sparsity(t) = \sum_{k=1}^{K} \mathbb{I}_{k,t}(|\beta_{k,t}| > c) \tag{27}$$

where $\mathbb{I}$ is an indicator function which will take the value of 1 when the absolute value of corresponding coefficient $|\beta_{k,t}|$ is larger than the threshold c.

Besides the number of predictors selected in each period, people might also wonder how long each selected predictor will last. We measure this by duration of selected predictor defined as below:

$$Duration(k) = \sum_{t=\underline{t}}^{\bar{t}} \mathbb{I}_{k,t}(|\beta_{k,t}| > c) \tag{28}$$

where $\underline{t}$ and $\bar{t}$ are the beginning and the end of the evaluation sample.

Table (5) reports the average number of selected predictors at each time over the whole evaluation sample (1980+), and the average, minimum, and maximum duration of all the selected predictors. For all the models, only a relatively small number of predictors are included in the predictive models. Compared with the model with fixed parameters, the large-tvp model includes much fewer predictors on average at each time, and the average duration of each selected predictor is shorter. This might not be surprising because the existence of time-varying parameters might partly include the information contained in other predictors, which shrink more predictors to zero. We will explore this feature in future research.

A potential drawback of allowing time-varying parameters is that the model might be overfitted in the presence of sharp change in the variable of interest. In the high-dimensional scenario, we might worry that the number of selected predictors might increase sharply to get a better model fit during certain periods. Figure (3) shows the sparsity of each model in different

Figure 3: Number of selected predictors

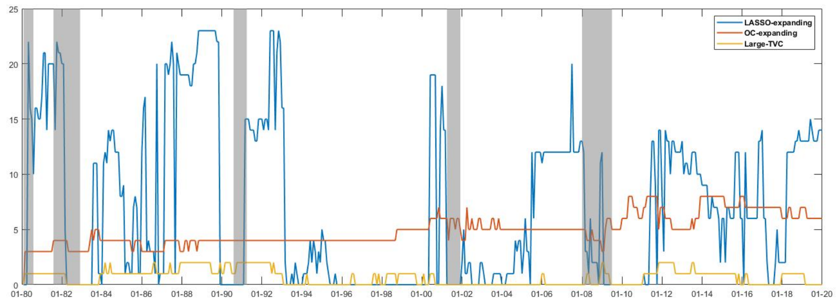


Table 5: Analysis of sparsity and duration

Model	# Selected predictors	Average duration	Min duration	Max duration
LASSO_200	5.94	83.71	5	176
LASSO_expanding	7.04	105.41	2	255
OC_200	3.58	107.125	1	284
OC_expanding	4.89	180.38	6	479
Large-tvp (VB-PF)	0.64	34.44	2	112

periods. For the large-tvp model, the sparsity actually evolves slowly even during the period with high uncertainty such as the financial crisis of 2007-2008. This illustrates that the large-tvp model is stable over time.

So far we have discussed the sparsity of the model with many predictors. We next move on to discuss the properties of each selected predictor in more detail. Much previous research shows that there exists short-horizon return predictability in the US stock market, that is, with certain predictors, the excess return is predictable in a short period while it cannot be predicted in long periods (e.g, Farmer et al. (2019)). Since the large-tvp model allows each predictor included and excluded over time, it is possible to capture the short-term predictability of stock return.

To measure the short-term predictability, we first introduce the concept "pocket" which means a continuous period that the coefficient of a specific predictor is non-zero. We define the pocket indicator $\mathbb{I}_{k,t}(|\beta_{k,t}| > c)$ which is the same as in the equation (27). The number of pocket $N_p(k)$ is defined as the number of times that the pocket indicator shift from zero to one for the predictor k, which can be computed as:

Table 6: Analysis of pocket

Predictors	# Pockets	Average duration	Min duration	Max duration
DFY	10	7.5	1	24
AWHMAN	6	3.83	1	11
SVAR	4	16.75	1	47
NTIS	2	56	2	112
CPIENGSL	2	10.5	2	21
RPI	2	1.5	1	2
SMB	1	3	3	3
DFR	1	3	3	3
FF	1	2	2	2

$$N_p(k) = \sum_{t=\underline{t}}^{\bar{t}} (1 - \mathbb{I}_{k,t})\mathbb{I}_{k,t+1} \tag{29}$$

and the duration of each pocket is given by

$$Dur_p(k) = t_1 - t_0 + 1 \tag{30}$$

where t_1 and t_0 are the start and the end date of certain pocket.

Table (6) reports the analysis of pockets of all the selected predictors in the large-tvp model. In general, the pockets of each predictor last shortly and the number of pockets is small. The largest number of pockets is 10 for the predictor of DFY, and the longest pockets is detected for the predictor of NTIS which lasts 112 month.

5 Conclusion

In this paper, we apply the time-varying parameter model with a large number of predictors (large-tvp) into the practice of out-of-sample prediction for the monthly stock return measured by the S&P 500 index. We incorporate the dynamic variable selection (DVS) prior to deal with the high-dimensional dataset, and propose a sequential method to efficiently estimate the model of interest. We find the large-tvp model outperforms other competing models such as LASSO, PCA, and random forest both statistically and economically.

References

Amir-Ahmadi, P., C. Matthes, and M.-C. Wang (2018). Choosing prior hyperparameters: with applications to time-varying parameter models. *Journal of Business & Economic Statistics*, 1–13.

Amir-Ahmadi, P. and Z. Sun (2021). Forecasting with time-varying parameter vars: An efficient sequential monte carlo approach. *none*.

Bitto, A. and S. Frühwirth-Schnatter (2019). Achieving shrinkage in a time-varying parameter model framework. *Journal of Econometrics 210*(1), 75–97.

Blei, D. M., A. Kucukelbir, and J. D. McAuliffe (2017). Variational inference: A review for statisticians. *Journal of the American statistical Association 112*(518), 859–877.

Bognanni, M. and J. Zito (2020). Sequential bayesian inference for vector autoregressions with stochastic volatility. *Journal of Economic Dynamics and Control*, 103851.

Bryzgalova, S., J. Huang, and C. Julliard (2019). Bayesian solutions for the factor zoo: We just ran two quadrillion models. *Available at SSRN 3481736*.

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

Chan, J. C., E. Eisenstat, and R. W. Strachan (2020). Reducing the state space dimension in a large tvp-var. *Journal of Econometrics 218*(1), 105–118.

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

Creal, D. (2012). A survey of sequential monte carlo methods for economics and finance. *Econometric reviews 31*(3), 245–296.

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

Del Negro, M. and G. E. Primiceri (2015). Time varying structural vector autoregressions and monetary policy: a corrigendum. *The review of economic studies 82*(4), 1342–1345.

Dong, X., Y. Li, D. Rapach, and G. Zhou (2021). Anomalies and the expected market return. *Baruch College Zicklin School of Business Research Paper* (2020-02), 02.

Farmer, L., L. Schmidt, and A. Timmermann (2019). Pockets of predictability. *Available at SSRN 3152386*.

Gu, S., B. Kelly, and D. Xiu (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies 33*(5), 2223–2273.

Herbst, E. P. and F. Schorfheide (2015). *Bayesian estimation of DSGE models*. Princeton University Press.

Huber, F., G. Koop, and L. Onorante (2020). Inducing sparsity and shrinkage in time-varying parameter models. *Journal of Business & Economic Statistics*, 1–15.

Koop, G. and D. Korobilis (2010). *Bayesian multivariate time series methods for empirical macroeconomics*. Now Publishers Inc.

Koop, G. and D. Korobilis (2013). Large time-varying parameter vars. *Journal of Econometrics 177*(2), 185–198.

Koop, G. and D. Korobilis (2020). Bayesian dynamic variable selection in high dimensions. *Available at SSRN 3246472*.

Pitt, M. K. and N. Shephard (1999). Filtering via simulation: Auxiliary particle filters. *Journal of the American statistical association 94*(446), 590–599.

Primiceri, G. E. (2005). Time varying structural vector autoregressions and monetary policy. *The Review of Economic Studies 72*(3), 821–852.

Rapach, D. and G. Zhou (2013). Forecasting stock returns. In *Handbook of economic forecasting*, Volume 2, pp. 328–383. Elsevier.

Rapach, D. E. and G. Zhou (2020). Time-series and cross-sectional stock return forecasting: New machine learning methods. *Machine Learning for Asset Management: New Developments and Financial Applications*, 1–33.

Safikhani, A. and A. Shojaie (2020). Joint structural break detection and parameter estimation in high-dimensional nonstationary var models. *Journal of the American Statistical Association*, 1–14.