Predicting Stock Returns Using Firm Characteristics: A Bayesian Model Averaging Approach

Shan Chen *

November 26, 2018

Abstract

We apply the Bayesian model averaging approach to the cross-section of stock returns with totally 94 firm characteristics. The Stochastics Search Variable Selection approach yields approximately 20 firm characteristics providing reliable forcasting of average returns in the sense that the coefficients are statistically significant over time. The set of significant firm characteristics would be larger if we further consider the probability that a variable should be included, which is estimated based on the BMA framework, in the both pre-2003 and post-2003 sample. The SSVS approach also generates economically significant out-of-sample performances in different sample periods.

JEL Classification : C11, G11, G14

*Shan Chen (shachen2-c@my.cityu.edu.hk), Department of Economics and Finance, City University of Hong Kong.

1 Introduction

A comphrehensive study of Green, Hand and Zhang (2017) examines the return predictability of 94 firm characteristics on US stock returns using Fama and MacBeth (1973) regressions, in respond to the challenge of 'factor zoo' posed by Cochrane (2011). They find that only 12 firm characteristics provide independent and reliable information for 1-month ahead stock return in those non-microcap stocks after adjusted for the data-snooping biases (Benjamini and Yekutieli, 2001; Harvey, Liu and Zhu, 2016). Also, a structure break of return predictability is detected in 2003: the predictive ability of many firm characteristics vanishes substantially after 2003, even though they are statistically significant in the multivariate predictive regressions before 2003.

However, as argued by Han, He, Rapach and Zhou (2018), the study of Green, Hand and Zhang (2017) relies on ordinary/weighted least squares to estimate cross-sectional multivariate regressions simultaneously including 94 firm characteristics, and the estimation of high-dimensional predictor space may suffer the problem of overfitting. Han, He, Rapach and Zhou (2018) emphasize the importance of out-of-sample performance in evaluating predictive power and propose the mean combination approach (Rapach, Strauss and Zhou, 2010), which is essentially a shrinkage forecast to mitigate the overfitting problem.

In this paper, we use an alternative method to tackle the high-dimension predictive problem. Specifically, we apply a Bayesian Model Averaging (BMA) approach estimated by the Stochastic Search Variable Selection (SSVS) developed by George and McCulloch (1993). Rather directly conduct a variable selection procedure, BMA computes the probability associated with different model (hence different combinations of all variables) and averages over different models weighted by the model likelihood to make forecasting. This procedure incorporates the model uncertainty into the estimation and also gives the probability that the corresponding variable should be included, which offers another measure in evaluating the importance of the variable in the predictability problem.

We show two primary results. Firstly, the SSVS estimation procedure yields a larger set of firm characteristics that forecast 1-month ahead average return reliably over time, up to 19 firm characteristics, in terms of the estimators of predictive coefficients, compared to 12 of Green, Hand and Zhang (2017). However, if we further consider the probability of firm characteristics being selected, the number of important determinants would be much larger. Also, the probabilities of firm characteristics selected are stable in different sample periods, both pre-2003 and post-2003. Secondly, the hedge portfolio constructed based on the forecasted returns of SSVS generates sizable average returns in the out-of-sample periods, and the returns are stable at both pre-2003 and post-2003 periods.

This paper is contributed to understanding how the average stock return is related to firm characteristics cross-sectionally. The success of (Fama and French, 1992, 1993) three-factor model, augmented the original CAPM by two additional factors: size (SMB) and the book-to-market ratio (HML), inspires many studies to explore the relevant firm characteristics to average stock returns. For example, profitability (Novy-Marx, 2013) and investment (Cooper, Gulen and Schill, 2008), are tow prominent firm characteristics in recent studies, which are also included into the asset pricing models as two new factors, e.g. Fama and French (2015), Hou, Xue and Zhang (2015) and Stambaugh and Yuan (2017). We do not focus on a type of firm characteristics but study multiple firm characteristics simultaneously, inspired by Lewellen (2015). We evaluate the predictive ability of all firm characteristics by applying BMA method, which the probability that a firm characteristics is selected estimated by SSVS may serve as an important indicator.

This work is also related to the growing literature that study how to summarize the information from a large number of firm characteristics, or 'shrinking' the characteristics space into a parsimonious while important factor model in the sense of capturing cross-section of stock returns adequately. Lettau and Pelger (2018) and Kelly, Pruitt and Su (2018) propose modeling approaches with Instrumental/Generalized Principal Component Analysis to model cross-sectional stock returns. Light, Maslov and Rytchkov (2017) and Freyberger, Neuhierl and Weber (2017) address the statistical challenges of high dimensions of predictors using Partial Least Squares (PLS) and adaptive group LASSO. Kozak, Nagel and Santosh (2017) use shrinkage estimation to select a subset of characteristics portfolios with good out-of-sample predictive power of average returns. The BMA method does not formally 'shrink' the variable space but assigns the probabilities of variable being selected to evaluate the reliability of predictors and could possibly improve the out-of-sample forecasting.

2 Firm Characteristics

To ensure the results comparable, we use the same 94 firm characteristics dataset as Green, Hand and Zhang (2017) and also Han, He, Rapach and Zhou (2018), from January 1980 to December 2017.[1] The common stocks on the NYSE, AMEX and NASDAQ that have nonmissing market capitalization at the end of the previous month on CRSP and nonmissing common equity value on Compustat annual are included. The monthly return of firm i at month t is matching to firm characteristics available at the end of month $t-1$. For annual firm characteristics, we assume they are available at the end of month $t-1$ if firm's

[1]We thank Jeremiah Green to share the SAS code to construct those firm characteristics from the data of CRSP, Compustat and I/B/E/S.

fiscal year-end is at least 6 months before month $t-1$; for quarterly firm characteristics, they are available at the end of month $t-1$ if firm's fiscal quarter end is at least 4 months before month $t-1$. The I/B/E/S and CRSP data are aligned in calendar time using the I/B/E/S statistical period and the CRSP end date.

Following Green, Hand and Zhang (2017), we winsorize firm characteristics (except those dummy and categorical variables) at the 1st and 99th percentiles and also standardize them by subtracting cross-sectional mean and dividing by standard deviation at each month. Given only small fraction of stocks (less than 4% of the full sample) have non-missing records for all 94 characteristics, we replace the missing values with the standardized mean value of zero. For I/B/E/S-based firm characteristics[2], we again follow GHZ to use them only starting in January 1989, due to sufficient data coverage of I/B/E/S. Table 1 lists all 94 firm characteristics acronyms and related definitions.

3 Bayesian Model Averaging Estimation

Consider a standard linear predictive multivariate regression model:

$$\boldsymbol{r_t} = \boldsymbol{X_{t-1}}\boldsymbol{\beta_t} + \boldsymbol{\epsilon_t} \tag{1}$$

where $\boldsymbol{r_t}$ is a n-by-1 vector of n stock returns at time t, $\boldsymbol{X_{t-1}}$ is a n-by-K+1 matrix of $K+1$ associated predictors (including a constant term) for n stocks at time $t-1$, $\boldsymbol{\beta_t}$ is a K+1-by-1 vector of predictive coefficients and $\boldsymbol{\epsilon_t}$ is a n-by-1 vector of error term.

In the cross-sectional stock return predictability problem, the number of predictors could be very large. For example, in Green, Hand and Zhang (2017), the cross-sectional multivariate regression simultaneously include 94 firm characteristics as predictive variables. While this analysis is common in the finance literature, ordinary least squares estimation of high dimensional model suffers the well-known problem of overfitting. A popular trend in the literature is developing new methods to shrink the variable space by different methods: machine learning (Freyberger, Neuhierl and Weber, 2017; Han, He, Rapach and Zhou, 2018), partial least squares (Light, Maslov and Rytchkov, 2017), and principal component analysis (Kelly, Pruitt and Su, 2018; Lettau and Pelger, 2018).

A Bayesian approach offers an alternative solution for this problem. Bayesian Model Averaging (BMA) takes a weighted average of estimates or forecast from all models, weights given by the model likelihood. Specifically, suppose there are total S models, denoted as

[2]There are seven firm characteristics that employ I/B/E/S data: chnanalyst, chfeps, disp, fgr5yr, nanalyst, sfe, and sue.

$M_1, M_2, ..., M_S$, and $\boldsymbol{\beta_t}$ is the parameters to be estimated, then the probability rules imply:

$$p(\boldsymbol{\beta_t}|\boldsymbol{r_t}) = \sum_{i=1}^{S} p(\boldsymbol{\beta_t}|\boldsymbol{r_t}, M_i)p(M_i|\boldsymbol{r_t}) \tag{2}$$

where $p(\boldsymbol{\beta_t}|\boldsymbol{r_t}, M_i)$ is the posterior probability of $\boldsymbol{\beta_t}$ under M_i and $p(M_i|\boldsymbol{r_t})$ is the posterior model probability of M_i. Therefore, rather carry out a typical model selection to find out an optimal model, BMA incorporates the model uncertainty into the estimation procedure by averaging over all combinations of predictors, and hence different models, to make the forecasting inference.

Given a large number of predictors, the total number of all possible models combined by those predictors could be extremely huge: if there are K predictors, the possible combinations and thus the number of models would be 2^K, which will make estimate every model be impossible due to the heavy computation demand, even with natural conjugate prior and analytical forms for posterior density. Some popular algorithms, including Markov Chain Monte Carlo Model Composition (MC^3) and Occam's window (see Madigan and Raftery (1994), Madigan and York (1995), Raftery, Madigan and Hoeting (1997)), are developed to reduce the computation task, but still, if K is large enough (in our application, 94 firm characteristics), it is likely that some predictors could never be drawn if the number of draws is not sufficiently large, which requires an extremely strong computational power.

Another popular method is to turn this model space problem into an estimation problem by using hierarchical priors, such as Stochastic Search Variable Selection (SSVS) developed by George and McCulloch (1993) (see also Bańbura, Giannone and Reichlin (2010), Korobilis (2008)). The key idea of SSVS is to propose a hierarchical shrinkage prior of $\boldsymbol{\beta_t}$ with the form:

$$\beta_{t,i}|\gamma_i \sim (1-\gamma_i)N(0, \tau_{0,i}^2) + \gamma_i N(0, \tau_{1,i}^2), \text{ for } i = 1, 2, ..., K \tag{3}$$

Normally, $\tau_{0,i}$ is small while $\tau_{1,i}$ is large and γ_i is either 0 or 1. Therefore, when $\gamma_i = 0$, the tight prior shrinks the coefficient to be near zero; when $\gamma_i = 1$, the prior is non-informative and the estimation of coefficient is data-driven. Loosely speaking, γ determines whether a variable to be selected ($\gamma = 1$) or omitted ($\gamma = 0$). γ_i is treated as a unknown parameter and need to be estimated. With the posterior density of γ, BMA is essential to averages over restricted ($\gamma = 0$) and unrestricted ($\gamma = 1$) models. In the following, we formalize the Bayesian setting:

3.1 Likelihood Function

With the notation of equation (1), we assume that[3]:

$$\boldsymbol{\epsilon} \sim N(\mathbf{0}, h^{-1} I_n) \tag{4}$$

h is the error precision frequently appeared in the Bayesian regression and equal to σ^{-2}, where σ^2 is the variance of each element from $\boldsymbol{\epsilon}$ (assuming homoscedasticity). I_n is a n-by-n identity matrix. Then the likelihood function $p(\boldsymbol{r}|\boldsymbol{\beta}, h)$ is a normal density function.

3.2 Prior

We use standard independent Normal-Gamma Priors:

$$p(\boldsymbol{\beta}, h) = p(\boldsymbol{\beta})p(h), \boldsymbol{\beta} \sim N(\underset{\sim}{\boldsymbol{\beta}}, \underset{\sim}{V}), h \sim G(\underset{\sim}{s}^{-2}, \underset{\sim}{\nu}) \tag{5}$$

where $\underset{\sim}{\boldsymbol{\beta}} = \mathbf{0}$ to shrink coefficients to $\mathbf{0}$, $\underset{\sim}{V} = DD$, D is diagonal matrix with elements:

$$d_i = \begin{cases} \tau_{0,i}, \gamma_i = 0 \\ \tau_{1,i}, \gamma_i = 1 \end{cases} \quad \text{for } i = 1, ..., K \tag{6}$$

As discussed above, $\gamma_i \in \{0, 1\}$ indicating whether variable $\boldsymbol{x_i}$, the i-th column of $\boldsymbol{X}$ is included and $\tau_{0,i}^2/\tau_{1,i}^2$ is the small/large prior variance for each coefficient β_i.

For the distribution of h, $\underset{\sim}{s}^{-2}$ and $\underset{\sim}{\nu}$ are the prior mean and degrees of freedom of h, respectively[4]. We set $\underset{\sim}{s}^2 = 0.01$ and $\underset{\sim}{\nu} = 0$, i.e. relatively non-informative priors.

So far, conditional on γ, all priors are standard setting, leading to standard independent Normal-Gamma posteriors for both $\boldsymbol{\beta}$ and h. For the prior of γ, George and McCulloch (1993) suggest to specify a Bernoulli distribution for each γ_i:

$$Prob(\gamma_i = 0) = 1 - \underset{\sim}{q_i}, \ Prob(\gamma_i = 1) = \underset{\sim}{q_i} \tag{7}$$

A non-informative prior choice is $\underset{\sim}{q_i} = 0.5$, implying that each predictor is equally likely to be included as excluded.

[3] For simplicity, we drop subscript t in the following section.

[4] A common definition of Gamma distribution is characterized by a shape parameter k and a scale parameter θ, that is $h \sim G(k, \theta)$. In our application, it is convenient to define the Gamma distribution by its expectation μ and the degree of freedom ν as it could explicitly give the statistics we interest in. By defining $k = \frac{\nu}{2}$ and $\theta = \frac{2\mu}{\nu}$, one could easily transform two different parameterizations for a same Gamma distribution.

The remaining question is how to specify the prior covariance $\underline{V}$, or $\tau_{0,i}^2$ and $\tau_{1,i}^2$. Rather subjectively select two variances without having confident prior information, George and McCulloch (1993) suggest to use 'default semi-automatic approach' that chooses $\tau_{0,i}^2$ and $\tau_{1,i}^2$ based on the initial estimation procedure. In particular, we first perform OLS procedure from regression on equation (1) with all predictors and get the standard error $\hat{\sigma}_i$ of each OLS coefficient $\hat{\beta}_i$. Let $\tau_{0,i}$ and $\tau_{1,i}$ be proportional to $\hat{\sigma}_i$:

$$\tau_{0,i} = \frac{1}{c} \times \hat{\sigma}_i \text{ and } \tau_{1,i} = c \times \hat{\sigma}_i \tag{8}$$

for constant c, e.g. $c = 10$ in our application. This finishes all priors specification.

3.3 Posterior

With priors mentioned above, we have the following posterior densities:

The posterior distribution of γ_i is Bernoulli:

$$Prob(\gamma_i = 0|\boldsymbol{r}, \gamma) = 1 - \bar{q}_i,\ Prob(\gamma_i = 1|\boldsymbol{r}, \gamma) = \bar{q}_i \tag{9}$$

where

$$\bar{q}_i = \frac{\frac{1}{\tau_{1,i}} exp(-\frac{\beta_i^2}{2\tau_{1,i}^2})\underline{q}_i}{\frac{1}{\tau_{1,i}} exp(-\frac{\beta_i^2}{2\tau_{1,i}^2})\underline{q}_i + \frac{1}{\tau_{0,i}} exp(-\frac{\beta_i^2}{2\tau_{0,i}^2})(1 - \underline{q}_i)} \tag{10}$$

Under γ, we have standard results for $\boldsymbol{\beta}$ and h:

$$\begin{aligned} \boldsymbol{\beta}|\boldsymbol{r}, h, \gamma &\sim N(\bar{\boldsymbol{\beta}}, \bar{V}) \\ h|\boldsymbol{r}, \boldsymbol{\beta}, \gamma &\sim G(\bar{s}^2, \bar{\nu}) \end{aligned} \tag{11}$$

where

$$\bar{\boldsymbol{\beta}} = \bar{V}(\underline{V}^{-1}\underline{\boldsymbol{\beta}} + h\boldsymbol{X}\boldsymbol{r}) \tag{12}$$

$$\bar{V} = (\underline{V} + h\boldsymbol{X}'\boldsymbol{X})^{-1} \tag{13}$$

$$\bar{s}^2 = [(\boldsymbol{r} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{r} - \boldsymbol{X}\boldsymbol{\beta}) + \underline{s}^2\underline{\nu}]/\bar{\nu} \tag{14}$$

$$\bar{\nu} = \underline{\nu} + n \tag{15}$$

3.4 Valued-Weighted Estimation

Following Green, Hand and Zhang (2017) and Han, He, Rapach and Zhou (2018), we also apply three strategies to alleviate the problem from those small market value stocks: ordinary least squares (OLS) with all available stocks, OLS without microcap stocks (defined as those stocks of market capitalization below 20th percentile of NYSE-only market capitalization), and weighted least squares (WLS) weighted by stock's market capitalization at month $t-1$. OLS-based strategies are straightforward in our Bayesian procedure, either by including all available stocks or excluding all microcap stocks in the estimation, and WLS could also be applied by transforming the data matrix. Denoting the n-by-n weighting matrix $\boldsymbol{W_t}$, where n is the number of stocks available in month t, each diagonal element is $w_{i,t}$, the weight for stock i corresponding to its market capitalization at month $t-1$ and 0 otherwise[5]. Since $\boldsymbol{W_t}$ is a positive definite symmetric matrix, we could define a transform matrix P_t by $P_t P_t' = \boldsymbol{W_t}^{-1}$ and do the following transformation on equation (1):

$$\boldsymbol{r_t}^* = P_t \boldsymbol{r_t}, \boldsymbol{X_{t-1}}^* = P_t \boldsymbol{X_{t-1}}, \boldsymbol{\epsilon_t}^* = P_t \boldsymbol{\epsilon_t} \tag{16}$$

Therefore, we could do estimation similar to WLS under the same procedure after the data transformation.

3.5 Computation

Given all posteriors, we apply the most common Markov Chain Monte Carlo algorithm, Gibbs sampler to sample γ, $\boldsymbol{\beta}$ and h for 100,000 draws with 10,000 discarded as the burn-in sample. We begin the draws with initial values of $\gamma = \mathbf{1}$ and $h = 0.01^{-2}$. The threshold of the probability of γ is 0.5 and therefore

$$\gamma_i = \begin{cases} 1, & Prob(\gamma_i = 1|\boldsymbol{r}) > 0.5 \\ 0, & \text{otherwise} \end{cases} \tag{17}$$

We then calculate the estimators of $Prob(\gamma|\boldsymbol{r})$, $\boldsymbol{\beta}$ and h by Monte Carlo integration. Specifically, let total draws be S (including S_0 burn-in sample), and θ be the parameter we interest in, the estimator of θ, $\hat{\theta}$, is:

$$\hat{\theta} = \frac{1}{(S - S_0)} \sum_{s=S_0+1}^{S} \theta^{(s)} \tag{18}$$

[5] We scale the weights by $w_{i,t} = n \times \frac{m_{i,t-1}}{\sum_{n}^{i=1} m_{i,t-1}}$ so that $\sum_{i=1}^{n} w_{i,t} = n$, where $m_{i,t-1}$ is the market capitalization of stock i at month $t-1$.

where $\theta^{(s)}$ denotes the s-th draw from the sampling. We mainly interest in the posterior mean and standard deviation of $Prob(\gamma|\boldsymbol{r})$, each element of which is the probability that the corresponding predictor is included, and $\boldsymbol{\beta}$, the coefficient vector.

The data frequency is monthly. At each month t, we perform the Bayesian procedure above with the return vector $\boldsymbol{r_t}$ against the characteristics matrix $\boldsymbol{X_{t-1}}$ cross-sectionally. Therefore, at each time t, we have the estimator of $\boldsymbol{\beta_t}$ and $Prob(\gamma_t|\boldsymbol{r_t})$ by their posterior means based on the available information upon time t.

4 Empirical Results

4.1 Fama-MacBeth Tests

We follow the Fama and MacBeth (1973) regression procedure in the sample period from January 1980 to December 2017 but the cross-sectional regression at the first stage is estimated based on the SSVS procedure at each month t. To evaluate the predictive power of firm characteristics on average returns, we simultaneously include all 94 firm characteristics in the cross-sectional estimation. Since the microcap stocks only take up about 3% of the total market capitalization of NYSE/AMEX/NASDAQ universe (Fama and French (2008), Hou, Mo, Xue and Zhang (2018)), we address the possible bias from overweighting microcap stocks by applying three weighting strategies discussed in Section 3.4, which is related to the WLS, OLS excluding microcap and OLS including all stocks in Green, Hand and Zhang (2017) and Han, He, Rapach and Zhou (2018), and also denoted three strategies with the same notations.

With the estimator of $\boldsymbol{\beta_t}$ at each month t, we take the time-series average and compute the standard error adjusted by Newey-West standard error of 12 month lags. Following Harvey, Liu and Zhu (2016), we use the t-statistics cutoff of 3.

Table 2 presents the main results of evaluating all 94 firm characteristics simultaneously using the full sample period from January 1980 to December 2017. The estimated coefficient with adjusted $|t\text{-stat}| \geq 3$ is shown with a bold font. In Column (2) estimated using WLS, 12 firm characteristics present significant predictive power in the whole sample period (Dispersion in forecasted EPS, Earnings-to-price ratio, Amihud's measurement of illiquidity, Industry momentum, 1-month momentum, Market capitalization, Number of analysts covering stock, Number of quarters of earnings increase, %Change in sales minus %change in receivables, R&D over market capitalization, Share turnover, Zero trading days). If applying OLS excluding microcap estimation, 11 firm characteristics provide reliable information on 1-month ahead return (Asset growth, Change in forecasted EPS, Change in number of analysts, Earnings announcement return, 1-month momentum, Financial statement score, Number of quarters of earnings increase, R&D over market capitalization, Return volatility, Volatility of liquidity measured by share turnover, Share turnover). Further incorporating microcap stocks in OLS estimation increase the number of significant firm characteristics to 27.

Green, Hand and Zhang (2017) reports that there is a break in return predictability around 2003 and find the predictability falls sharply after 2003, the post-2003 sample. We perform the same procedure using the post-2003 sample period from January 2004 to December 2017 and the results are presented in Table 3. Given the t-stat cutoff of 3, only 4 (Earnings-to-price ratio, Amihud's measurement of Illiquidity, Market capitalization, Number of analysts covering stock) and 2 (Change in number of analysts, Earnings announcement return) firm characteristics show persistent predictive power over the post-2003 sample period via WLS and OLS excluding microcap, respectively. Even though OLS including all stocks increases the number of significant coefficients to 12, it drops by more than 50% compared to the full sample period.

Taken the results in Table 2 and Table 3 together, we find similar results to Green, Hand and Zhang (2017) findings by employing standard Fama and MacBeth (1973) regressions, while the set of significant firm characteristics is slightly different. Pooling WLS and OLS excluding microcap results yields a set of 19 firm characteristics in the full sample period and that of 6 firm characteristics in the post-2003 sample period out of 94 firm characteristics. In other words, there are only a few of firm characteristics independently provide reliable information when predicting average stock return, in terms of the significance of coefficient over time. Also, in the post-2003 period, there even less firm characteristics show the reliable predictive power of equity return.

4.2 Probability of Firm Characteristics Selected

The discussion above is purely based on the time-series average of the estimator of $\boldsymbol{\beta_t}$, in our case, the posterior mean of $\boldsymbol{\beta_t}$. However, in Bayesian model averaging, we also estimate the $Prob(\gamma_t|\boldsymbol{r_t})$, the probability that firm characteristic should be included in the model, and hence γ_t has directly implication for evaluating the importance of the corresponding firm characteristics in the predictive regressions. Similar to the analysis in Section 4.1, we also take the time-series average of $Prob(\gamma_t|\boldsymbol{r_t})$ estimated from the cross-sectional regression at each month t from January 1980 to December 2017.

Table 4 presents the time-series average of $Prob(\gamma|\boldsymbol{r_t})$ for three weighting strategies and for different sample periods. For the full sample period, most of selection probabilities are above 40% and only two below 30% in WLS estimation, similar to the findings in Han, He, Rapach and Zhou (2018), indicating that the data favors a larger number of firm characteristics to be included into the predictive model compared to the results when only consider the significance of $\boldsymbol{\beta_t}$ over the sample period. Column (5) and (8) report the probability of OLS excluding microcap and OLS estimation, repectively and the results are similar while the probabilities are slightly smaller than those of WLS estimation, most of them above 30% Hence, it suggests that most of firm characteristics are relevant to stock return along the time, when measuring by the probability of being selected, $Prob(\gamma_t|\boldsymbol{r_t})$.

Another important observation is that for most firm characteristics, the probabilities of selected are quite stable before and after 2003, regardless of different weighting strategies. We again split the sample into two parts: pre-2003 period, from Juanaury 1980 to December 2002; post-2003 period, from January 2004 to December 2017 and calculate the time-series average of the probabilities. The difference between the pre-2003 probability and the post-2003 probability is fairly small for the vast majority of firm characteristics, within 5% in most cases. Thus, the small difference in probabilities from two periods indicates that the result is not mainly driven by the pre-2003 sample and most firm characteristics are equally likely to be included in two sub-sample periods.

In consistent with the evidences from Han, He, Rapach and Zhou (2018), based on the probability of firm characteristics selected, there are a sizable number of firm characteristics affects stock return over time while the most relevant firm characteristics to stock return may also be varied substantially in different point in time.

4.3 Hedge Portfolio

Following Lewellen (2015), Green, Hand and Zhang (2017) and Han, He, Rapach and Zhou (2018), we also analyze the out-of-sample hedge portfolio performance to further examine the firm characteristics predictive power in terms of economics benefits. Specifically, at the end of each month t, we sort stocks into deciles according to the predicted return using the information available upon at the end of month t, and then form the hedge portfolio by buying the decile with highest forecasted return and selling the decile with lowest forecasted return. Similar to Green, Hand and Zhang (2017), the deciles are determined by using breakpoints of NYSE-only stock /All-but-microcap stock/All available stock forecasted return for WLS/OLS excluding microcap/OLS cases respectively. The decile portfolios are calculated by value weighted for WLS and by equal weighted for OLS excluding microcap/OLS.

The results of Table 2 and Table 4 indicates that although there are many firm characteristics relevant to 1-month ahead average return, the set of important firm characteristics is varying substantially over time. Therefore, if we generate the predicted return of $t+1$ only relying on the coefficient $\boldsymbol{\beta_t}$ estimated by using the return $\boldsymbol{r_t}$ and the firm characteristics $\boldsymbol{X_{t-1}}$, the forecasting performance could be weak due to the time-varying effect. This problem is also common for evaluating cross-sectional predictive regressions over time and a common but ad-hoc method in the literature is to use the rolling average of coefficients available at time t to serve as the estimator of $\boldsymbol{\beta_{t+1}}$, from Fama and MacBeth (1973).

In our estimation procedure, SSVS not only estimates the $\boldsymbol{\beta_t}$, but also gives $Prob(\gamma_t|\boldsymbol{r_t})$, which describes the relative importance of the corresponding predictor and may possibly serve as an indicator of the reliability of that variable in different time points. Rather directly calculate the forecasted return of month $t+1$ of each stock using $\boldsymbol{\beta_t}$ multiplied by the observed characteristics available at the end of month t, we incorporate the $Prob(\gamma|\boldsymbol{r_t})$ to scale the $\boldsymbol{\beta_t}$ to form the predicted returns. Specifically, we compute the rolling average of $Prob(\gamma|\boldsymbol{r_t})$ using 120-month window and then multiply $\boldsymbol{\beta_t}$ by its corresponding probability (linked by the same firm characteristics). The forecasted return of each stock is simply the modified $\boldsymbol{\beta_t}$ times the firm characteristics in month t. For comparison, we also restrict the sample period from January 1990 to December 2017.

Table 5 presents the portfolio performances computed by different methods. GHZ refers to the method of 120-month rolling average coefficient from Fama and MacBeth (1973) regressions in Green, Hand and Zhang (2017), Mean and LASSO refer to the mean combination and LASSO approach in Han, He, Rapach and Zhou (2018). SSVS is the approach used in the paper. We report the average return, volatility and the t-statistics adjusted by Newey-West standard error of 12 month lags. Again, we also split the sample into the pre-2003 and post-2003 subsamples.

For the WLS cases from column (2) to (4), GHZ method has the highest t-statistics of 3.63 in the full sample period but the performance is mainly driven by the pre-2003 period since the hedge portfolio only yields an average return of 0.004% after 2003. Both methods in Han, He, Rapach and Zhou (2018) generates not only sizable average return with moderate t-statistics, but also show predictability in the post-2003 sample. The portfolio of SSVS method has a stable performance and moderate t-statistics in different sample periods, but the average return is relatively low compared to other methods.

Table 5 column (5) to (7) present the results for the OLS excluding microcap case. GHZ, Mean and LASSO have the similar story as in WLS case. GHZ forecast works fine in the full sample but the performance is biased toward the pre-2003 sample and fail to generate remarkable average return after 2003; Mean and LASSO perform more stable in different sample periods with significant t-statistics except LASSO in the post-2003 period at 10% level. SSVS also preserves the stable performances in both pre-2003 and post-2003 periods and the average returns are remarkably high, 2.38% in the full sample and 1.38% in the post-2003 sample. For the OLS case, all approaches maintain predictability in all sample periods, but GHZ again suffers a sharp decline in the post-2003 sample while other methods are relatively stable over time.

5 Conclusions

We apply a Bayesian model averaging method to study the return predictability of 94 firm characteristics cross-sectionally. A nice feature of BMA estimation is that it also estimates the probability that the corresponding variable is included or not. Given the common practice in cross-sectional predictive problem is taking time-series average of the coefficient at each time point, the probability is a natural tool to evaluate the significance of variable over time, supplement to the coefficient estimator. Our results show that combining both estimated coefficients and probabilities could improve the forecasted return and also the out-of-sample performance.

Consistent with the findings of Han, He, Rapach and Zhou (2018), there are larger number of firm characteristics matter at each time point from 1980 to 2017 but this set of firm characteristics may also vary over time. An interesting question following these results may be that how the predictive power of individual firm characteristics varies and hence how the set of significant firm characteristics varies over time. An additional factor that captures this time-varying feature may also be linked to the expected return cross-sectionally. The probability of firm characteristics selected could be informative for such tasks.

Table 1: Firm Characteristics Acronyms

This table provides the acronyms and related firm characteristics given by Green, Hand and Zhang (2017). The Appendix in Green, Hand and Zhang (2017) offers detailed definitions for all 94 firm characteristics.

Acronym	Firm Characteristics	Acronym	Firm Characteristics
absacc	Absolute accurals	mom1m	1-month momentum
acc	Working capital accruals	mom36m	36-month momentum
aeavol	Abnormal earnings announcement volume	ms	Financial statement score
age	# of years since first Compustat coverage	mve	Market capitalization
agr	Asset growth	mve_ia	Industry-adjusted market capitalization
baspread	Bid-ask spread	nanalyst	# of analysts covering stock
beta	Market beta	nincr	# of quarters of earnings increases
bm	Book-to-market ratio	operprof	Operating profitability
bm_ia	Industry-adjusted book-to-market ratio	orgcap	Organizational capital
cash	Cash holdings	pchcapx_ia	Industry-adjusted Δ% in capital expenses
cashdebt	Cash flow to debt	pchcurrat	Δ% in current ratio
cashpr	Cash productivity	pchdepr	Δ% in depreciation
cfp	Cash-flow-to-price ratio	pchgm_pchsale	Δ% in gross margin - Δ% in sales
cfp_ia	Industry-adjusted cahs-flow-to-price ratio	pchsale_pchinvt	Δ% in sale - Δ% in inventory
chatoia	Industry-adjusted Δ in asset turnover	pchsale_pchrect	Δ% in sales - Δ% in A/R
chcsho	Δ in shares outstanding	pchsale_pchxsga	Δ% change in sales - Δ% in SG&A
chempia	Industry-adjusted change in employees	pchsaleinv	Δ% sales-to-inventory
chfeps	Δ in forecasted EPS	pctacc	Percent accruals
chinv	Δ in inventory	pricedelay	Price delay
chmom	Δ in 6-month momentum	ps	Financial statements score
chnanalyst	Δ in number of analysts	rd	R&D increase
chpmia	Industry-adjusted Δ in profit margin	rd_mve	R&D to market capitalization
chtx	Δ in tax expense	rd_sale	R&D to sales
cinvest	Corporate investment	realestate	Real estate holdings
convind	Convertible debt indicator	retvol	Return volatility
currat	Current ratio	roaq	Return on assets
depr	Depreication / PP&E	roavol	Earnings volatility
disp	Dispersion in forecasted EPS	roeq	Return on equity
divi	Dividend initiation	roic	Return on invested capital
divo	Dividend omission	rsup	Revenue surprise
dy	Dividend-to-price	salecash	Sales-to-cash
ear	Earnings announcement return	saleinv	Sales-to-inventory
egr	Growth in common shareholder equity	salerec	Sales-to-receivables
ep	Earnings-to-price	secured	Secured debt
fgr5yr	Forecasted growth in 5-year EPS	securedind	Secured debt indicator
gma	Gross profitability	sfe	Scaled earnings forecast
grCAPX	Growth in capital expenditures	sgr	Sales growth
grltnoa	Growth in long-term net operating assets	sin	Sin stocks
herf	Industry sales concentration	sp	Sales-to-price
hire	employee growth rate	std_dolvol	Volatility of liquidity ($ trading volume)
idiovol	Idiosyncratic return volatility	std_turn	Volatility of liquidity (share turnover)
ill	Illiquidity	stdcf	Cash flow volatility
indmom	Industry momentum	sue	Unexpected quarterly earnings
invest	Capital expenditure	tang	Debt capacity / firm tangibility
IPO	New equity issue	tb	Tax income-to-book income
lev	Leverage	turn	Share turnover
mom12m	12-month momentum	zerotrade	Zero trading days

Table 2: Fama-MacBeth Regressions of Monthly Returns on 94 Firm Characteristics

This table reports the results from the monthly Fama-MacBeth regressions of cross-sectional returns in month t against 94 firm characteristics in month $t-1$ simultaneously. At each month, we estimate the coefficients associated with 94 firm characteristics via the SSVS approach by three weighting strategies: (1) All stocks, WLS: estimation weighted by the market value of stock in month $t-1$; (2) All-but-microcap, OLS: using all-but-microcap stocks; (3)All stocks, OLS: using all available stocks. The characteristics are winsorized at 1% and 99% and standardized to have a zero mean and unit standard deviation. Missing characteristics are set to be zero. The data sample is from January 1980 to December 2017, covering all common stocks listed in NYSE, AMEX and NASDAQ. The coefficients are the time-series mean of the monthly estimated coefficients×100 and the t-statistics are adjusted by Newey-West standard error of 12 month lags.

Characteristics	All stocks, WLS		All-but-microcap, OLS		All stocks, OLS	
	Coefficient	t-stat	Coefficient	t-stat	Coefficient	t-stat
absacc	-0.05	-0.69	-0.04	-2.13	-0.03	-1.70
acc	-0.00	-0.04	-0.02	-1.14	-0.01	-0.53
aeavol	0.09	1.28	-0.01	-0.72	0.01	1.33
age	0.23	1.69	-0.01	-0.58	0.03	1.61
agr	-0.31	-2.32	**-0.05**	**-3.06**	**-0.08**	**-4.93**
baspread	0.21	1.82	-0.06	-0.45	0.13	1.85
beta	-0.05	-0.50	0.01	0.12	0.05	0.83
bm	0.10	1.13	0.05	1.76	**0.11**	**4.31**
bm_ia	-0.41	-1.55	-0.04	-1.03	-0.02	-0.82
cash	0.32	2.78	0.13	2.63	**0.16**	**4.34**
cashdebt	-0.05	-0.78	0.05	1.35	-0.02	-1.07
cashpr	0.29	1.20	-0.01	-1.76	-0.01	-0.89
cfp	0.09	0.79	-0.05	-0.88	**0.08**	**4.16**
cfp_ia	0.38	1.39	0.06	1.92	0.03	1.79
chatoia	0.06	0.67	0.03	2.43	0.02	2.20
chcsho	-0.17	-1.82	-0.00	-0.39	-0.03	-2.63
chempia	-0.02	-0.12	0.00	0.25	-0.01	-0.74
chfeps	0.09	1.12	**0.07**	**3.41**	**0.07**	**3.79**
chinv	-0.10	-1.20	-0.02	-1.10	-0.04	-2.66
chmom	-0.22	-2.41	-0.04	-1.36	-0.00	-0.00
chnanalyst	-0.44	-1.97	**-0.03**	**-4.14**	-0.03	-2.65
chpmia	0.04	0.39	0.03	1.62	0.02	1.30
chtx	0.07	1.16	0.00	0.11	**0.05**	**4.64**
cinvest	0.06	0.87	-0.06	-2.22	-0.01	-0.83
convind	-0.65	-2.51	-0.00	-0.02	**-0.08**	**-3.40**
currat	-0.05	-0.78	0.01	0.66	0.01	0.50
depr	0.05	0.70	0.02	0.77	0.05	2.84
disp	**-0.10**	**-3.08**	-0.03	-1.59	-0.04	-2.94
divi	-0.30	-0.99	-0.08	-1.50	-0.05	-1.41
divo	0.35	0.85	-0.00	-0.04	0.05	1.02
dy	-0.01	-0.08	-0.03	-1.16	-0.01	-0.58
ear	0.04	0.65	**0.06**	**6.10**	**0.07**	**6.91**
egr	0.10	1.24	-0.03	-2.20	-0.02	-1.49
ep	**0.35**	**3.16**	-0.03	-0.54	**0.10**	**3.30**
fgr5yr	0.04	0.51	0.02	0.37	0.01	0.36
gma	0.13	1.09	0.04	1.30	0.05	2.00
grcapx	-0.06	-0.47	-0.03	-2.64	-0.02	-2.36
grltnoa	0.04	0.32	-0.02	-1.23	-0.02	-1.40
herf	-0.07	-0.96	-0.01	-0.44	-0.02	-1.45
hire	-0.09	-0.65	0.00	0.02	-0.02	-0.76

(Continued)

Characteristics	All stocks, WLS		All-but-microcap, OLS		All stocks, OLS	
	Coefficient	t-stat	Coefficient	t-stat	Coefficient	t-stat
idiovol	-0.10	-0.67	-0.14	-1.88	-0.14	-2.47
ill	**0.38**	**5.31**	-0.72	-1.12	**0.31**	**7.04**
indmom	**0.65**	**5.67**	0.09	2.47	**0.28**	**5.81**
invest	-0.09	-0.82	-0.04	-1.71	-0.05	-2.76
IPO	-0.90	-2.82	0.05	0.50	**-0.21**	**-3.21**
lev	0.01	0.07	0.03	0.69	-0.01	-0.13
mom12m	-0.08	-0.88	0.15	1.86	0.10	1.55
mom1m	**-1.18**	**-6.98**	**-0.41**	**-5.81**	**-0.69**	**-8.38**
mom36m	0.04	0.41	-0.03	-1.36	-0.03	-1.23
ms	0.09	1.12	**0.05**	**4.07**	**0.06**	**3.88**
mve	**-3.45**	**-6.12**	-0.22	-2.96	**-0.41**	**-5.27**
mve_ia	-0.22	-0.75	0.00	0.12	0.02	1.34
nanalyst	**2.92**	**4.24**	0.02	0.78	**0.13**	**3.55**
nincr	**0.30**	**3.46**	**0.04**	**4.29**	**0.06**	**6.31**
operprof	-0.06	-0.78	0.02	2.10	0.01	0.60
orgcap	-0.10	-1.02	0.01	0.37	0.02	0.63
pchcapx_ia	0.10	1.14	0.02	1.54	0.03	1.54
pchcurrat	0.00	0.01	-0.03	-2.12	0.01	0.61
pchdepr	-0.12	-1.51	-0.00	-0.16	-0.01	-1.18
pchgm_pchsale	-0.08	-1.03	0.02	1.69	0.01	0.71
pchsale_pchinvt	-0.10	-0.87	0.00	0.42	0.01	1.49
pchsale_pchrect	**0.23**	**3.34**	-0.01	-0.72	0.02	2.71
pchsale_pchxsga	-0.18	-1.63	-0.00	-0.17	-0.01	-0.74
pchsaleinv	-0.05	-0.62	0.01	0.97	0.01	0.65
pctacc	-0.15	-1.47	0.00	0.25	-0.02	-2.64
pricedelay	-0.04	-0.44	0.01	0.46	-0.00	-0.50
ps	-0.08	-0.96	0.02	1.18	**0.03**	**3.74**
rd	0.07	0.30	0.08	2.81	**0.14**	**3.76**
rd_mve	**0.41**	**3.33**	**0.23**	**3.40**	**0.30**	**6.19**
rd_sale	0.24	2.12	0.04	0.54	0.05	1.66
realestate	-0.11	-0.69	0.02	0.88	0.03	1.27
retvol	-0.03	-0.22	**-0.31**	**-4.15**	**-0.30**	**-6.18**
roaq	-0.07	-0.55	0.03	1.07	0.08	2.85
roavol	-0.08	-0.83	0.03	1.16	-0.01	-0.36
roeq	0.16	1.44	0.05	2.73	**0.06**	**3.19**
roic	-0.07	-0.69	0.04	1.02	0.00	0.16
rsup	0.01	0.21	0.04	1.56	0.04	1.99
salecash	0.10	0.43	0.01	2.00	-0.00	-0.43
saleinv	0.17	2.56	0.01	1.07	0.02	2.13
salerec	-0.02	-0.31	0.03	1.37	0.02	1.30
secured	0.05	0.45	-0.11	-1.30	-0.02	-0.72
securedind	-0.00	-0.02	0.07	0.95	0.03	0.56
sfe	-0.01	-0.06	-0.14	-0.65	0.08	1.50
sgr	-0.06	-0.41	-0.02	-0.80	-0.04	-2.28
sin	2.51	1.82	0.29	2.95	**0.34**	**3.14**
sp	-0.18	-1.41	0.05	1.10	0.02	0.54
std_dolvol	-0.28	-2.27	-0.02	-0.82	-0.07	-2.14
std_turn	0.40	2.30	**0.11**	**4.66**	**0.19**	**4.74**
stdcf	0.04	0.52	-0.01	-0.40	-0.02	-1.15
sue	0.15	1.85	0.08	2.42	**0.11**	**5.66**
tang	-0.08	-0.70	0.03	1.29	0.02	1.28
tb	0.09	1.32	0.01	0.57	0.02	2.33
turn	**-1.08**	**-4.87**	**-0.16**	**-5.13**	**-0.32**	**-8.11**
zerotrade	**-0.59**	**-4.38**	-0.07	-1.23	**-0.18**	**-5.47**

Table 3: Fama-MacBeth Regressions of Monthly Returns on 94 Firm Characteristics in Post-2003 Sample

This table reports the results from the monthly Fama-MacBeth regressions of cross-sectional returns in month t against 94 firm characteristics in month $t-1$ simultaneously. At each month, we estimate the coefficients associated with 94 firm charactersitics via the SSVS approach by three weighting strategies: (1) All stocks, WLS: estimation weighted by the market value of stock in month $t-1$; (2) All-but-microcap, OLS: using all-but-microcap stocks; (3)All stocks, OLS: using all available stocks. The characteristics are winsorized at 1% and 99% and standardized to have a zero mean and unit standard deviation. Missing charracteristics are set to be zero. The data sample is from January 2004 to December 2017, covering all common stocks listed in NYSE, AMEX and NASDAQ. The coefficients are the time-series mean of the monthly estimated coefficients×100 and the t-statistics are adjusted by Newey-West standard error of 12 month lags.

Characteristics	All stocks, WLS		All-but-microcap, OLS		All stocks, OLS	
	Coefficient	t-stat	Coefficient	t-stat	Coefficient	t-stat
absacc	-0.09	-0.95	-0.05	-1.75	-0.02	-0.89
acc	-0.10	-0.64	0.03	1.27	-0.02	-0.97
aeavol	-0.07	-0.76	-0.01	-0.46	0.01	0.81
age	0.17	1.20	-0.01	-0.35	0.01	0.51
agr	0.09	0.39	-0.06	-1.89	-0.06	-2.24
baspread	0.08	0.35	-0.11	-0.52	-0.12	-1.40
beta	-0.19	-1.27	-0.04	-0.37	0.01	0.10
bm	0.16	1.23	-0.05	-1.46	0.04	1.42
bm_ia	0.10	0.44	0.03	1.19	0.01	0.24
cash	0.13	0.74	0.02	0.39	0.13	2.58
cashdebt	-0.07	-0.62	0.00	0.05	-0.01	-0.52
cashpr	-0.05	-0.32	-0.02	-1.19	0.00	0.08
cfp	0.12	0.74	-0.04	-0.82	**0.09**	**3.26**
cfp_ia	-0.35	-1.71	0.05	1.33	0.04	1.19
chatoia	-0.00	-0.03	-0.01	-0.34	0.01	0.45
chcsho	-0.36	-1.83	0.00	0.33	-0.02	-0.77
chempia	-0.29	-1.03	0.01	0.48	-0.02	-0.94
chfeps	0.04	0.29	0.04	1.35	0.06	2.04
chinv	-0.10	-0.93	-0.04	-1.41	-0.01	-0.47
chmom	-0.28	-2.14	0.00	0.04	-0.03	-0.72
chnanalyst	-0.84	-2.00	**-0.03**	**-4.44**	**-0.05**	**-3.63**
chpmia	0.08	0.63	0.03	0.88	0.02	0.85
chtx	0.08	0.87	-0.01	-0.55	**0.03**	**3.11**
cinvest	-0.06	-0.49	-0.06	-1.01	-0.02	-0.78
convind	-0.34	-0.80	0.04	0.81	-0.08	-1.70
currat	-0.13	-1.41	-0.00	-0.20	-0.00	-0.12
depr	0.10	0.93	-0.01	-0.62	0.02	1.10
disp	-0.12	-2.83	-0.03	-1.35	-0.03	-1.66
divi	-0.39	-0.94	-0.04	-0.91	-0.03	-0.61
divo	0.03	0.06	0.00	0.02	-0.05	-0.59
dy	-0.12	-1.17	-0.02	-0.61	-0.01	-0.29
ear	-0.03	-0.29	**0.06**	**4.27**	**0.07**	**5.19**
egr	0.21	1.62	-0.00	-0.14	0.02	1.08
ep	**0.44**	**3.14**	-0.09	-0.83	0.11	2.07
fgr5yr	-0.06	-0.53	0.00	0.10	0.00	0.11
gma	-0.09	-0.58	0.07	1.82	0.03	1.32
grcapx	-0.16	-1.84	-0.03	-1.22	-0.02	-1.68
grltnoa	-0.06	-0.36	0.00	0.06	0.01	0.34
herf	-0.05	-0.43	-0.03	-2.08	-0.01	-0.68
hire	0.13	0.59	0.01	0.57	0.02	0.66

(Continued)

Characteristics	All stocks, WLS		All-but-microcap, OLS		All stocks, OLS	
	Coefficient	*t*-stat	Coefficient	*t*-stat	Coefficient	*t*-stat
idiovol	0.08	0.43	-0.07	-0.65	-0.03	-0.48
ill	**0.38**	**4.23**	0.04	0.03	**0.14**	**5.75**
indmom	0.52	2.43	0.05	1.34	**0.16**	**3.76**
invest	0.08	0.47	-0.01	-0.26	-0.00	-0.07
IPO	-0.97	-1.57	0.29	1.41	-0.11	-0.80
lev	-0.34	-1.70	-0.13	-1.40	-0.13	-1.29
mom12m	-0.16	-1.43	-0.08	-0.50	-0.08	-0.59
mom1m	-0.41	-1.76	-0.16	-1.66	-0.24	-2.45
mom36m	-0.01	-0.04	-0.00	-0.14	-0.01	-0.37
ms	-0.12	-1.00	0.03	1.63	0.03	1.10
mve	**-1.84**	**-4.98**	-0.12	-0.97	-0.22	-2.15
mve_ia	-0.01	-0.02	-0.00	-0.29	-0.01	-0.41
nanalyst	**1.82**	**3.37**	-0.03	-1.00	0.07	1.41
nincr	0.12	1.23	0.02	1.76	**0.04**	**3.08**
operprof	-0.09	-0.55	0.01	0.80	0.01	0.90
orgcap	-0.04	-0.25	-0.05	-1.13	-0.07	-1.13
pchcapx_ia	0.14	0.95	0.03	0.70	0.08	1.39
pchcurrat	-0.05	-0.49	-0.01	-0.49	0.02	1.68
pchdepr	-0.10	-0.91	-0.01	-0.41	0.00	0.06
pchgm_pchsale	-0.24	-1.72	0.02	0.94	-0.01	-0.39
pchsale_pchinvt	-0.12	-0.66	-0.00	-0.37	0.01	0.89
pchsale_pchrect	0.38	2.91	-0.02	-0.96	0.03	1.85
pchsale_pchxsga	0.01	0.07	0.02	0.90	-0.00	-0.14
pchsaleinv	-0.23	-1.72	-0.02	-0.65	-0.01	-1.82
pctacc	-0.10	-1.08	-0.02	-1.77	-0.00	-0.44
pricedelay	-0.14	-1.68	0.02	0.38	-0.01	-0.37
ps	-0.14	-1.09	-0.03	-1.06	0.01	1.21
rd	0.06	0.19	0.05	1.02	0.07	2.00
rd_mve	0.42	1.94	0.19	2.01	**0.26**	**4.49**
rd_sale	0.12	1.02	0.07	0.55	0.03	0.58
realestate	-0.09	-0.57	0.01	0.18	0.03	1.76
retvol	-0.42	-1.85	0.01	0.09	**-0.24**	**-3.43**
roaq	0.04	0.27	-0.01	-0.27	0.10	2.87
roavol	-0.07	-0.45	0.01	0.30	-0.01	-0.48
roeq	0.05	0.36	-0.00	-0.14	-0.02	-1.10
roic	-0.21	-1.28	-0.07	-0.98	-0.10	-2.43
rsup	-0.01	-0.08	0.04	1.04	0.02	0.63
salecash	-0.15	-1.70	0.02	1.64	0.00	0.43
saleinv	0.15	1.17	0.02	1.24	0.01	0.63
salerec	0.07	0.51	0.04	0.62	0.01	0.52
secured	-0.13	-0.90	-0.02	-1.20	-0.01	-0.78
securedind	0.14	0.52	-0.01	-0.27	0.04	0.84
sfe	0.11	0.70	-0.08	-0.18	**0.28**	**3.37**
sgr	-0.25	-1.01	0.00	0.10	-0.07	-2.37
sin	0.39	1.14	0.41	2.53	0.34	1.98
sp	0.09	0.49	0.08	1.78	0.10	1.83
std_dolvol	-0.23	-1.47	-0.03	-0.97	-0.06	-1.09
std_turn	-0.33	-1.05	0.00	0.11	0.02	0.24
stdcf	0.10	0.67	-0.04	-0.74	-0.04	-1.41
sue	0.16	1.42	0.13	2.27	**0.14**	**5.65**
tang	-0.36	-2.33	0.06	1.86	-0.02	-0.59
tb	-0.02	-0.22	0.01	1.05	0.01	1.50
turn	-0.36	-0.81	-0.10	-2.10	**-0.20**	**-3.57**
zerotrade	-0.06	-0.81	0.10	0.63	-0.04	-1.53

Table 4: Stochastic Search Variable Selection Frequencies

This table reports the stochastics search variable selection frequncies for 94 firm characteristics in cross-sectional regressions. The frequency is calculated based on the time-serise average of $Prob(\gamma_t|\boldsymbol{r_t})$ esimated via SSVS at each month t. We peform three weighting schemes: (1) All stocks, WLS: estimation weighted by the market value of stock in month $t-1$; (2) All-but-microcap, OLS: using all-but-microcap stocks; (3)All stocks, OLS: using all available stocks. The full sample is from January 1980 to December 2017, the pre-2003 sample is from January 1980 to December 2002 and the post-2003 sample is from January 2004 to December 2017.

Characteristics	All stocks, WLS			All-but-microcap, OLS			All stocks, OLS		
	Full	Pre2003	Post 2003	Full	Pre2003	Post 2003	Full	Pre2003	Post 2003
absacc	53 %	51 %	55 %	20 %	20 %	19 %	22 %	23 %	20 %
acc	58 %	57 %	59 %	19 %	19 %	20 %	21 %	20 %	22 %
aeavol	45 %	39 %	56 %	20 %	19 %	23 %	19 %	15 %	25 %
age	55 %	53 %	58 %	22 %	23 %	21 %	23 %	24 %	21 %
agr	51 %	50 %	52 %	21 %	19 %	23 %	21 %	20 %	24 %
baspread	63 %	63 %	63 %	39 %	32 %	51 %	47 %	48 %	45 %
beta	61 %	62 %	60 %	59 %	60 %	57 %	57 %	58 %	56 %
bm	59 %	58 %	60 %	27 %	25 %	29 %	34 %	35 %	31 %
bm_ia	45 %	41 %	51 %	25 %	23 %	28 %	20 %	17 %	26 %
cash	48 %	41 %	58 %	34 %	31 %	39 %	30 %	28 %	34 %
cashdebt	49 %	48 %	52 %	25 %	24 %	27 %	26 %	26 %	27 %
cashpr	51 %	50 %	52 %	19 %	19 %	18 %	21 %	23 %	18 %
cfp	58 %	55 %	61 %	24 %	24 %	23 %	28 %	27 %	30 %
cfp_ia	47 %	44 %	51 %	26 %	24 %	29 %	24 %	23 %	26 %
chatoia	55 %	51 %	61 %	20 %	19 %	22 %	20 %	18 %	22 %
chcsho	55 %	51 %	60 %	17 %	18 %	17 %	21 %	19 %	25 %
chempia	52 %	50 %	53 %	22 %	21 %	23 %	18 %	17 %	20 %
chfeps	33 %	25 %	40 %	26 %	28 %	24 %	25 %	25 %	26 %
chinv	56 %	53 %	58 %	23 %	21 %	27 %	21 %	21 %	20 %
chmom	56 %	55 %	56 %	39 %	38 %	43 %	34 %	33 %	36 %
chnanalyst	32 %	27 %	37 %	23 %	23 %	24 %	18 %	18 %	18 %
chpmia	52 %	51 %	52 %	30 %	32 %	29 %	28 %	28 %	29 %
chtx	48 %	40 %	59 %	25 %	27 %	22 %	22 %	22 %	22 %
cinvest	51 %	47 %	56 %	23 %	20 %	26 %	23 %	20 %	27 %
convind	55 %	54 %	56 %	21 %	18 %	24 %	20 %	18 %	22 %
currat	50 %	51 %	50 %	21 %	22 %	21 %	20 %	21 %	19 %
depr	57 %	59 %	55 %	24 %	24 %	24 %	26 %	27 %	24 %
disp	20 %	19 %	21 %	27 %	27 %	27 %	20 %	20 %	20 %
divi	45 %	42 %	49 %	19 %	21 %	15 %	17 %	18 %	17 %
divo	46 %	45 %	45 %	20 %	22 %	18 %	19 %	18 %	20 %
dy	56 %	56 %	58 %	28 %	30 %	25 %	28 %	29 %	26 %
ear	50 %	45 %	58 %	24 %	24 %	25 %	23 %	23 %	23 %
egr	51 %	50 %	52 %	20 %	19 %	22 %	20 %	19 %	21 %
ep	63 %	63 %	62 %	28 %	25 %	33 %	34 %	34 %	34 %
fgr5yr	26 %	28 %	24 %	35 %	42 %	27 %	31 %	37 %	24 %
gma	50 %	47 %	55 %	27 %	26 %	29 %	25 %	25 %	25 %
grcapx	48 %	49 %	45 %	21 %	19 %	23 %	19 %	19 %	20 %
grltnoa	60 %	58 %	63 %	19 %	19 %	19 %	20 %	21 %	19 %
herf	54 %	53 %	54 %	24 %	25 %	24 %	22 %	22 %	21 %
hire	53 %	51 %	58 %	20 %	19 %	21 %	19 %	18 %	21 %
idiovol	61 %	62 %	59 %	36 %	37 %	35 %	42 %	45 %	38 %
ill	64 %	62 %	66 %	20 %	21 %	10 %	39 %	45 %	30 %

(Continued)

Characteristics	All stocks, WLS			All-but-microcap, OLS			All stocks, OLS		
	Full	Pre2003	Post 2003	Full	Pre2003	Post 2003	Full	Pre2003	Post 2003
indmom	60 %	59 %	61 %	45 %	47 %	43 %	47 %	48 %	48 %
invest	55 %	52 %	59 %	23 %	21 %	26 %	22 %	20 %	24 %
IPO	53 %	57 %	48 %	26 %	27 %	25 %	25 %	26 %	24 %
lev	52 %	47 %	60 %	42 %	43 %	40 %	38 %	35 %	43 %
mom12m	56 %	53 %	61 %	56 %	56 %	56 %	51 %	50 %	53 %
mom1m	75 %	76 %	74 %	55 %	56 %	54 %	63 %	70 %	54 %
mom36m	48 %	49 %	46 %	32 %	31 %	33 %	28 %	28 %	30 %
ms	46 %	41 %	52 %	20 %	21 %	20 %	20 %	21 %	20 %
mve	66 %	67 %	64 %	39 %	37 %	42 %	43 %	45 %	39 %
mve_ia	51 %	48 %	57 %	21 %	21 %	21 %	24 %	25 %	23 %
nanalyst	45 %	48 %	42 %	28 %	28 %	27 %	31 %	34 %	27 %
nincr	42 %	36 %	51 %	23 %	25 %	19 %	19 %	20 %	17 %
operprof	49 %	48 %	50 %	19 %	19 %	19 %	20 %	21 %	19 %
orgcap	54 %	52 %	59 %	31 %	33 %	28 %	27 %	27 %	27 %
pchcapx_ia	51 %	50 %	52 %	27 %	21 %	38 %	24 %	20 %	33 %
pchcurrat	54 %	52 %	57 %	20 %	19 %	20 %	22 %	22 %	22 %
pchdepr	58 %	61 %	53 %	21 %	20 %	23 %	22 %	23 %	22 %
pchgm_pchsale	53 %	52 %	54 %	21 %	20 %	22 %	22 %	21 %	24 %
pchsale_pchinvt	52 %	51 %	55 %	14 %	15 %	14 %	17 %	17 %	17 %
pchsale_pchrect	53 %	52 %	56 %	21 %	21 %	22 %	22 %	22 %	23 %
pchsale_pchxsga	55 %	54 %	56 %	22 %	22 %	21 %	22 %	22 %	23 %
pchsaleinv	49 %	46 %	53 %	14 %	14 %	15 %	16 %	17 %	14 %
pctacc	41 %	39 %	46 %	18 %	17 %	20 %	17 %	17 %	16 %
pricedelay	56 %	55 %	58 %	20 %	17 %	22 %	20 %	21 %	20 %
ps	52 %	49 %	56 %	20 %	21 %	19 %	18 %	19 %	17 %
rd	52 %	51 %	52 %	21 %	22 %	20 %	22 %	23 %	20 %
rd_mve	59 %	61 %	58 %	34 %	33 %	36 %	35 %	34 %	37 %
rd_sale	50 %	52 %	47 %	32 %	29 %	36 %	31 %	31 %	31 %
realestate	59 %	57 %	61 %	26 %	26 %	27 %	25 %	24 %	25 %
retvol	63 %	61 %	67 %	29 %	32 %	24 %	40 %	44 %	34 %
roaq	56 %	50 %	65 %	27 %	27 %	28 %	29 %	26 %	32 %
roavol	51 %	45 %	58 %	26 %	25 %	27 %	26 %	24 %	28 %
roeq	57 %	54 %	64 %	23 %	23 %	22 %	26 %	27 %	25 %
roic	56 %	58 %	53 %	28 %	24 %	34 %	32 %	32 %	34 %
rsup	59 %	55 %	66 %	27 %	24 %	31 %	29 %	27 %	33 %
salecash	51 %	53 %	48 %	17 %	15 %	20 %	18 %	18 %	17 %
saleinv	46 %	49 %	42 %	17 %	18 %	17 %	16 %	16 %	16 %
salerec	50 %	49 %	51 %	33 %	32 %	34 %	26 %	24 %	30 %
secured	52 %	52 %	53 %	22 %	21 %	23 %	21 %	22 %	21 %
securedind	56 %	57 %	56 %	20 %	20 %	21 %	22 %	21 %	24 %
sfe	47 %	39 %	56 %	33 %	28 %	36 %	35 %	25 %	46 %
sgr	53 %	55 %	51 %	22 %	21 %	25 %	20 %	19 %	22 %
sin	28 %	26 %	29 %	23 %	23 %	24 %	19 %	19 %	19 %
sp	57 %	54 %	62 %	32 %	32 %	30 %	31 %	29 %	34 %
std_dolvol	52 %	51 %	55 %	24 %	28 %	18 %	34 %	30 %	41 %
std_turn	58 %	54 %	65 %	25 %	28 %	22 %	34 %	35 %	34 %
stdcf	43 %	37 %	54 %	24 %	22 %	27 %	22 %	19 %	27 %
sue	66 %	61 %	69 %	26 %	24 %	28 %	34 %	33 %	35 %
tang	54 %	52 %	58 %	25 %	25 %	26 %	23 %	25 %	21 %
tb	44 %	42 %	46 %	20 %	21 %	19 %	17 %	18 %	16 %
turn	62 %	60 %	65 %	38 %	40 %	37 %	47 %	46 %	48 %
zerotrade	58 %	59 %	56 %	17 %	19 %	12 %	34 %	38 %	24 %

Table 5: Hedge Portfolio Performance

The table reports summary statistics for spread portfolios formed from the out-of-sample forecasted returns by different forecasting methods. At the end of each month t, stocks are sorted into deciles accroding to the predicted return using the information available upon at the end of month t, and the hedge portfolio is formed by buying the decile with highest forecasted return and selling the decile with lowest forecasted return. The declies are determined by using breakpoints of NYSE-only stock /All-but-microcap sotck/All available stock forecasted return for WLS/OLS excluding microcap/OLS caces respecitively. The decile portfolios are calculated by value weighted for WLS and by equal weighted for OLS excluding microcap/OLS. GHZ refers to Green, Hand and Zhang (2017), Mean and LASSO refers to Han, He, Rapach and Zhou (2018), SSVS is the approach in this paper. The full sample period is from January 1990 to December 2017.

	WLS			OLS Excl. Microcap			OLS		
Method	Mean	Volatility	t-stat.	Mean	Volatility	t-stat.	Mean	Volatility	t-stat.
Panel A: Full out-of-sample Period (1990:01-2017:12)									
GHZ	0.98%	4.95%	3.63	1.24%	5.13%	4.43	2.95%	4.14%	13.07
Mean	0.99%	7.88%	2.29	1.60%	10.27%	2.86	1.78%	8.03%	4.07
LASSO	1.16%	7.20%	2.95	1.38%	8.73%	2.90	1.93%	7.15%	4.95
SSVS	0.34%	4.99%	1.24	2.38%	8.34%	5.22	2.72%	10.11%	4.93
Panel B: Pre 2003 out-of-sample Period (1990:01-2002:12)									
GHZ	2.03%	5.79%	4.39	2.59%	6.19%	5.23	4.45%	4.39%	12.68
Mean	1.21%	8.61%	1.74	2.30%	12.95%	2.22	2.41%	9.55%	3.15
LASSO	1.45%	8.08%	2.24	2.24%	11.06%	2.53	2.46%	8.46%	3.63
SSVS	0.22%	6.34%	0.43	3.35%	10.52%	3.98	3.77%	12.88%	3.66
Panel C: Post 2003 out-of-sample Period (2004:01-2017:12)									
GHZ	0.004%	3.86%	0.01	0.11%	3.58%	0.39	1.68%	3.46%	6.31
Mean	0.78%	7.20%	1.14	0.93%	7.17%	1.68	0.99%	6.33%	2.03
LASSO	0.93%	6.42%	1.87	0.60%	5.95%	1.30	1.28%	5.66%	2.93
SSVS	0.39%	3.48%	1.43	1.38%	5.58%	3.20	1.46%	6.46%	2.93

References

Bańbura, M., D. Giannone, and L. Reichlin, 2010, "Large Bayesian vector auto regressions," *Journal of Applied Econometrics*, 25 (1), 71–92.

Benjamini, Y. and D. Yekutieli, 2001, "The control of the false discovery rate in multiple testing under dependency," *Annals of statistics*, 1165–1188.

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

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

Fama, E. F. and K. R. French, 1992, "The Cross-Section of Expected Stock Returns," *Journal of Finance*, 47 (2), 427–465.

Fama, E. F. and K. R. French, 1993, "Common Risk Factors in the Returns on Stocks and Bonds," *Journal of Financial Economics*, 33 (1), 3–56.

Fama, E. F. and K. R. French, 2008, "Dissecting Anomalies," *Journal of Finance*, 63 (4), 1653–1678.

Fama, E. F. and K. R. French, 2015, "A Five-Factor Asset Pricing Model," *Journal of Financial Economics*, 116 (1), 1–22.

Fama, E. F. and J. D. MacBeth, 1973, "Risk, Return, and Equilibrium: Empirical Tests," *Journal of Political Economy*, 81, 607–636.

Freyberger, J., A. Neuhierl, and M. Weber, 2017, "Dissecting characteristics nonparametrically," working paper, National Bureau of Economic Research.

George, E. I. and R. E. McCulloch, 1993, "Variable selection via Gibbs sampling," *Journal of the American Statistical Association*, 88 (423), 881–889.

Geweke, J., 2005, *Contemporary Bayesian econometrics and statistics*, 537. John Wiley & Sons.

Green, J., J. R. Hand, and X. F. Zhang, 2017, "The characteristics that provide independent information about average us monthly stock returns," *Review of Financial Studies*, 30 (12), 4389–4436.

Han, Y., A. He, D. Rapach, and G. Zhou, 2018, "How Many Firm Characteristics Drive Us Stock Returns?" working paper, University of North Carolina at Charlotte, Emory University, Saint Loius University, and Washington University in St. Loius.

Harvey, C. R., Y. Liu, and H. Zhu, 2016, "... And the Cross-Section of Expected Returns," *Review of Financial Studies*, 29, 5–68.

Hou, K., H. Mo, C. Xue, and L. Zhang, 2018, "Which Factors?" *Review of Finance*, Forthcoming.

Hou, K., C. Xue, and L. Zhang, 2015, "Digesting Anomalies: An Investment Approach," *Review of Financial Studies*, 28 (3), 650–705.

Kelly, B., S. Pruitt, and Y. Su, 2018, "Characteristics are covariances: A unified model of risk and return," working paper, National Bureau of Economic Research.

Koop, G., 2003, *Bayesian Econometrics*. John Wiley and Sons UK.

Korobilis, D., 2008, "Forecasting in vector autoregressions with many predictors," in *Bayesian Econometrics*. Emerald Group Publishing Limited, 403–431.

Kozak, S., S. Nagel, and S. Santosh, 2017, "Shrinking the cross section," working paper, National Bureau of Economic Research.

Lettau, M. and M. Pelger, 2018, "Estimating latent asset-pricing factors," working paper, National Bureau of Economic Research.

Lewellen, J., 2015, "The Cross-section of Expected Stock Returns," *Critical Finance Review*, 4 (1), 1–44.

Light, N., D. Maslov, and O. Rytchkov, 2017, "Aggregation of information about the cross section of stock returns: A latent variable approach," *The Review of Financial Studies*, 30 (4), 1339–1381.

Madigan, D. and A. E. Raftery, 1994, "Model selection and accounting for model uncertainty in graphical models using Occam's window," *Journal of the American Statistical Association*, 89 (428), 1535–1546.

Madigan, D. and J. York, 1995, "Bayesian graphical models for discrete data," *International Statistical Review*, 215–232.

Newey, W. K. and K. D. West, 1987, "Hypothesis testing with efficient method of moments estimation," *International Economic Review*, 777–787.

Novy-Marx, R., 2013, "The Other Side of Value: The Gross Profitability Premium," *Journal of Financial Economics*, 108 (1), 1–28.

Raftery, A. E., D. Madigan, and J. A. Hoeting, 1997, "Bayesian model averaging for linear regression models," *Journal of the American Statistical Association*, 92 (437), 179–191.

Rapach, D. E., J. K. Strauss, and G. Zhou, 2010, "Out-of-sample equity premium prediction: Combination forecasts and links to the real economy," *Review of Financial Studies*, 23 (2), 821–862.

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